COMPARATIVE ANALYSIS OF COUNTING METHODS USING DOCUMENTED RANKED PREFERENCE ELECTIONS

BRANDON CALIA, JULIA CONIGLIARI, GILLIAN MCGUIRE, AND JEROMY SIVEK

Abstract. This study uses ballots from real-world ranked preference elections to compute the frequency of certain pathologies. We focus primarily on the frequency of disagreement between different counting methods, which is to say the frequency that the choice of method itself decides an election. We notice and quantify that disagreements between methods become more likely as an election becomes closer to a tie. We also compute the frequency of spoiler candidates for each method. And we extract some information about voter behavior that we hope will inform further simulation of ranked preference elections. We use data from 521 single-seat and 27 multi-seat political and large-organization elections and 77 sets of ballots from Major League Baseball awards.

Keywords: Voting Theory, Social Choice Theory, Ranked Preference Voting, Spatial Voting Model, Condorcet, Baldwin, Borda, Instant Runoff

1. Introduction

In ranked choice voting, voters rank a set of candidates in order of preference. Given a set of such 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 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] who proved that certain simple and valuable properties cannot all be possessed by the same counting method. Arrow's theorem has been widely discussed and further clarity comes from Amartiya Sen [14] and others.

In practice, most ranked preference elections use a counting method called Instant Runoff (IR). In IR, whichever candidate has the fewest first place votes is *eliminated* from every ballot, and the candidates below the eliminated candidate are elevated each by one position. Then the ballots are re-counted and the process of elimination continues until one candidate has a majority of first place votes and becomes the winner.

In this study we also consider three well-known alternative counting methods ascribed to Baldwin, Borda, and Condorcet. We give brief descriptions of those methods here. More background information on these methods is widely available.

In Borda's method, each ballot position is traditionally worth a number of points equal to the number of *lower* positions on the ballot. If there are N ballot positions, then 1st place is worth $N-1$ points, 2nd place is worth $N-2$ points, and on until last place on a given ballot is worth 0 points. The candidate with the greatest number of points is the winner of Borda's method. In fact, a scoring system is considered a Borda method if the gaps between positions have equal value. In the traditional scoring described above, each position is worth one more point than the position below it. Because candidate number can exceed ballot size, we decided to let last place be worth one point and therefore 1st place was worth N points throughout this paper.

Baldwin's method uses eliminations like IR. But the elimination criterion is that the candidate with the fewest Borda points at a given stage is eliminated at that stage. Whichever candidate is left after every other candidate is eliminated in this way is deemed the Baldwin winner.

Date: October 17, 2023.

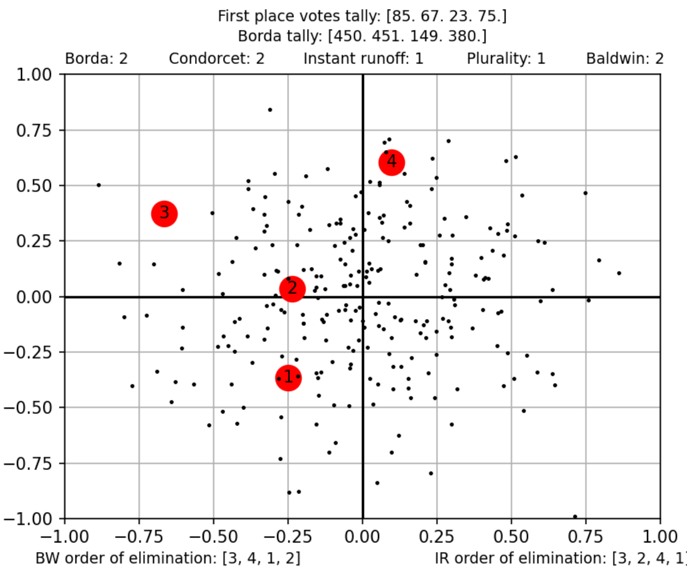
Condorcet's method focuses on hypothetical pairwise matchups between candidates on the ballot. Candidate C is said to win the pairwise matchup with candidate X if the number of ballots where C is ranked higher than X is greater than the number of ballots where X is ranked higher than C. If candidate C wins its matchups with every other candidate on the ballot, then candidate C is the Condorcet winner. One important note is that not every ranked preference election has a Condorcet winner although there were Condorcet winners in 99.0% of the real-world elections we processed.

We also compute hypothetical Plurality winners of ranked preference elections. We do this by simply assigning each voter's plurality vote to the candidate that they ranked first, while understanding that voters might choose a different candidate if they knew they would only get one vote.

Our focus here is on the cases where the methods disagree. We processed the ballots from 625 real-world elections to identify the number of disagreements and compute some other parameters, including the frequency that voters will choose to rank only one candidate or completely fill their ballots. Our results are shared in section 2.

One particularly famous pathology in elections is the so-called spoiler effect. Sometimes, a non-viable candidate (a *spoiler*) will siphon votes (or Borda points, etc.) from one candidate more than others and thereby determine the result of the election. In this paper, we check for such spoiler candidates in 444 historical elections with single winners and relatively large voter numbers. Our notion of spoiler is weaker than Arrow's famous Independence of Irrelevant Alternatives requirement.

This study is further motivated by the attempt to create maximally accurate simulations of ranked preference elections. Evidence suggests that a mixed spatial model provides for accurate simulation of real-world election dynamics [13], [8], [3]. Given a spatial simulation like the one pictured below, one can compute the *utility score* that each voter has for each candidate. This would create a complete ranking of candidates for each voter, but leaves open the question of how complete the voter's ballot will be.



This example illustrates a simulated election (with parameters established by this study) and includes one of the most common method disagreements seen in real elections. The black dots represent voters, and the numbered circles represent candidates. IR and Plurality disagree with the rest of the counting methods on this election. 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 matchup, however, 2 beats 1, and seems to benefit from a more central location in that matchup.

Beyond understanding the frequency of disagreements and spoilers, we eventually hope to use such simulations (improved by information from real elections) to point toward "true values" of disagreement frequencies and to classify the underlying causes of method disagreements.

2. Disagreement Frequency and Voter Behavior

Ranked preference elections have been held around the world, the data for some of which are available online [10], [7]. A program was written to take in the data from these ballot databases and determine the winner of the election using the counting methods considered in this study. In fact, there were a nontrivial number of elections where a different candidate would have been elected if the counting method changed.

Before discussing general findings, we consider some specific examples to see how the counting methods can disagree in real elections. It is interesting to consider that a simple change in counting method would lead to a different winner in these cases.

2.1. **2007 American Psychological Association Presidential Election.**

In 2007, James H. Bray was elected the American Psychological Association's next president in a ranked preference election with 13318 voters and IR counting. In fact, Bray was also a Condorcet winner and the plurality winner. Borda and Baldwin, however, would have selected a different winner (whose name is not specified on the ballot data). Below are a few metrics regarding the election. Bray is the candidate in the second position, whose statistics are in bold:

Borda tally: [23469, **39424**, 39518, 32605, 30500]

First place votes tally: [1396, **4126**, 3929, 1838, 2029]

Second place votes tally: [1212, **2442**, 2626, 2805, 2116]

Ballot appearance tally: [9302, **10710**, 10914, 9936, 9990]

What follows is Bray's pairwise comparison against all other candidates, reported in portion of ballots where Bray is preferred. Ballots containing neither of the two candidates being compared are not counted.

[.7184, **1**, .5099, .5934, .6316]

As noted above, the methods tend to disagree more frequently when we see near-tie conditions in the Borda tally. Indeed, Bray and the candidate in the 3rd position (who we will call candidate C) have uncommonly close Borda scores (within 0.24% of Bray's score). Bray has the most first place votes, but candidate C receive a similar number. Candidate C receives more second place votes. Furthermore, candidate C appeared on over 200 more ballots than Bray. If we consider being on the ballot a sign of approval, candidate C would have thus won an approval method election. Similarly, the pairwise comparison between Bray and candidate C is extremely close. Bray is only preferred by 51% of voters who put either Bray or candidate C on a ballot.

There is another quirk in this example that illustrates a theoretical nuance. Assuming complete ballots, it can be proven that the Baldwin method will always pick a Condorcet winner if such exists. This property (choosing a Condorcet winner if one exists) is frequently called "Condorcet consistency". We note with this election that Baldwin's method loses Condorcet consistency when ballots can be incomplete (as they must be allowed to be for practical purposes in most cases).

2.2. **2009 Burlington Mayoral Election.**

Another dramatic instance of ranked preference counting methods disagreeing is the 2009 Burlington, Vermont Mayoral election. 8,890 voters submitted ballots with the following candidates running:

1. Bob Kiss	2. Andy Montroll	3. James Simpson
4. Dan Smith	5. Kurt Wright	6. Write-In

Burlington used instant runoff to decide the winner, which selected Bos Kiss, the incumbent. Kurt Wright, however, was the plurality winner. Moreover, Andy Montroll was a Condorcet winner who would have also been selected by both Borda and Baldwin. Here are the metrics for this election as computed from the data found on PrefLib [7]:

Borda tally: [23405, 26162, 6805, 21998, 23046, 665]

First place votes tally: [2586, 2063, 35, 1306, 2953, 37]

Second place votes tally: [1404, 2637, 307, 2107, 995, 49]

Ballot appearance tally: [6185, 6706, 3391, 6094, 6090, 243]

Pairwise comparison matrix for Kiss, Montroll, and Wright:
Values show the number of voters preferring the candidate in the row over the candidate in the column

Candidate	Kiss	Montroll	Wright
Kiss	-	3477	4314
Montroll	4067	-	4597
Wright	4067	3668	-

We see that the Condorcet winner was not selected. Montroll was preferred by 56% to 44% over Wright and 54% to 46% over Kiss, who was elected. Montroll also appeared on the most ballots, and had the highest Borda score.

Wright, the plurality winner, was in fact a Condorcet loser among the top 3 candidates. He loses pairwise comparisons to both Kiss and Montroll. Wright would have the *worst* performance among the top three under IR (an election using reversed ballots). Interestingly, if Wright - clearly not a natural winner - had not been in this election, Montroll would have won instead of Kiss, making Wright a spoiler. This speaks to IR's failure to satisfy "Independence of Irrelevant Alternatives" (a condition in Arrow's famous theorem).

Another intriguing aberration of this election is that it exhibits "non-monotonicity". The monotonicity criterion for voting methods is satisfied if a winning candidate cannot be made to lose by giving them more support, or vice versa. If 753 voters preferring Wright over Kiss switched their support to Kiss over Wright, Kiss would have lost the election. Wright would be eliminated before Montroll instead of the other way around, letting Montroll beat Kiss in the final round [4].

Amid dissatisfaction with this result, Burlington citizens voted to repeal instant runoff voting in 2010. Subsequently they chose to reinstate it for city council elections in 2021.

2.3. **Summary of Processed Elections.** We processed 452 single seat and 27 multi-seat political elections. Beyond the above examples, we accessed databases of elections from Berkley (CA), Dublin (Ireland), Glasgow (Scotland), Meath (Ireland), Minneapolis (MN), New York (NY), Oakland (CA), Pierce (WA), San Francisco (CA), San Leandro (CA), Takoma Park (WA), UK's Labor Party, various Vermont precincts, and a few other places. Some of these databases were slightly modified (error correction done carefully and affecting a tiny portion of ballots) from those found at "PrefLib: A Library of Preferences" [7] and the Cast Vote Records at Harvard's Dataverse collected by Deb Otis [10].

For all seats we checked for Plurality, Borda, Instant-Runoff, and Baldwin winners. For single seat elections, we also computed the Condorcet winner. We found that 88.1% of the winners in single seat elections were the same no matter the method used. We found that 88.8% of those winning seats in multi-seat elections would have done so for every method checked. The tables below have data for the frequency that pairs of methods elected the same candidates. For pairs involving Condorcet's method, we report the portion of agreement over the 99.6% of cases where a Condorcet winner exists.

Single Seat Political	Baldwin	Borda	Condorcet	IR	Plurality
Baldwin	-	.969	.942	.934	.889
Borda	.969	-	.956	.951	.914
Condorcet	.942	.956	-	.984	.927
Instant Runoff	.934	.951	.984	-	.931
Plurality	.889	.914	.927	.931	-

The apparent closeness of these numbers obscures some trends that we saw while examining certain subsets of the data. First, the Plurality method does tend to disagree with other methods much more frequently than the more advanced methods disagree with each other. We also see some clustering of method choices (on certain data subsets) whereby IR agrees more with Condorcet and Plurality than the other methods considered here while Borda and Baldwin agree more frequently with each other. These sorts of trends are more evident when elections are closer to "ties" in a sense that is clarified below.

We also processed ballots from 27 multi-seat political elections. These elections were held in Aspen (CO), Dublin (Ireland), Glasgow (Scotland), and Minneapolis (MN). There were 98 total seats decided in these 27 elections.

For a multi-seat election with N seats to be assigned, we used the following modified counting methods. For Plurality, we ranked candidates by number of first-place votes and assigned seats to the highest N candidates in this ranking. Similarly, the modified Borda method assigned seats to candidates with the top N Borda scores.

The runoff methods are equally natural to generalize. For Baldwin and IR, we simply eliminated candidates (using Borda scores and first places votes respectively, as per the original methods) until there were N winners remaining.

We observed the following frequencies of agreement between the methods. Here, we consider methods to agree if they select the same candidate even if a different order. For example, in 2009 Minneapolis held an election for 3 seats to the Parks and Recreation Board using ranked-preference ballots. If instant-runoff had been used, then candidates C, D, and H (in our naming) would have won seats. If Baldwin had been used, candidates D, E, and H would have won seats. So we consider IR and Baldwin to agree on 2 of the 3 seats in this election and sum these numbers over all elections in the database. These elections all come from [7].

	Plrlty-Borda	Borda-IR	Bldwn-Borda	Plrlty-IR	Bldwn-Plrlty	Bldwn-IR
Multi Seat	.898	.918	1	.959	.898	.918

2.4. **Frequency of Ballot Blanks.** Voters in ranked preference elections are typically given the option of submitting incomplete ballots. Over a diverse set of 430 collections of such ballots from single-seat elections [10], we measured voters' decisions to submit complete or semi-complete ballots. We ignored those ballots that were left totally blank, suspecting that voters went to the polls only to vote in other elections that were being held that day. We hope that this data provides the reader with insight into voter behavior and we plan to use this data to inform the behavior of simulated voters in other experiments.

The most common behavior we observed was the submission of a ballot with only one candidate selected. We measured that 41.8% of voters across these 430 elections submitted only a first preference. The inter-election standard deviation of the portion of voters ranking only a single candidate was 21.3%.

Voters choosing to rank only one candidate may be accustomed to plurality elections and simply chose not to modify their approach to voting. We also note here that this may help explain why the frequencies of agreement all exceeded our expectations. If all voters submit only one candidate, then every method will agree. And - without currently being able to prove a rigorous theorem to this effect - we conjecture that the probability of method disagreement will increase as the portion of such ballots decreases. Further discussion of this can be found in section 4.

We observed that 26.4% of voters submitted complete ballots. The data-set included elections where the number of candidates was greater than the number of ballot positions and elections where these numbers were the same. The inter-election standard deviation of this measure was 25.2%. And we observed that voters, on average, completed 59.0% of the available slots on their ballots with an inter-election standard deviation of 14.9%.

2.5. **MLB Award Ballots.** The Baseball Writer's Association of America (BBWAA) votes for eight total awards across two leagues using ranked preference voting [9]. Different awards use different modifications of the Borda system. League MVP award ballots have 10 positions with point values of 14 for first place, nine for second, eight for third, seven for fourth, and so on down to one for tenth. Other awards have fewer ballot positions and use smaller modifications of Borda scoring.

In the 77 sets of BBWAA ballots we processed, there were 72 cases where all of this study's methods agreed with each other and also agreed with the modified scoring system used by BBWAA. This 93.5% total agreement rate exceeds the 85.4% rate of such in the single-winner elections that we processed. There were only two elections without a Condorcet winner. Further, in two of the five aberrant cases, all methods agree except for plurality where there is a tie.

We ascribe this significantly higher rate of total agreement primarily to the fact that voters names are published, which pressures individual voters to agree with a consensus choice when one emerges.

One particularly interesting case of disagreement is the 2016 AL Cy Young Award. Rick Porcello won the award and also would have won a Borda count or an Instant Runoff election with these ballots.

But Justin Verlander had the most first-place votes by a significant six-vote margin (out of 30 votes). With just this information, one might expect Verlander to win the other methods in a landslide. But every voter who didn't put Verlander in first place had Porcello higher than Verlander on their ballot. So Porcello won the pairwise matchup 16-14 and therefore wins if Condorcet's method is used. Because he appears on every ballot and is a Condorcet winner, Porcello is also the Baldwin winner of this election.

Also working against Verlander in 2016 was that two voters left him off of their five-candidate ballot completely, whereas Porcello appeared on every ballot. And of the 16 ballots where Porcello appeared ahead of Verlander, only seven of them had Verlander at the rank directly behind Porcello. By contrast, when Verlander appeared higher on the ballot, Porcello was in the next possible slot ten of the possible 14 times.

We speculate that this may be an instance of strategic voting. The voters who pushed Verlander down or off their ballots had a significant impact on this election, and they may have done this intentionally to improve Porcello's chances with the knowledge that Verlander was a legitimate contender. Scanning the literature, we find no evidence to support this conjecture.

Regardless, we believe that this election shows how Baldwin, Borda, Condorcet, and IR methods tend to select less-polarizing choices that do fairly well on a greater number of ballots.

2.6. **Tideman Trade Union Data.** A collection of ballots due to Tideman [11] contain the results of trade union elections and similar. The organizations themselves were granted anonymity and the data was hosted by the Center for Range Voting [15]. We note that these elections had smaller voter number than many political elections and possessed other quirks. We were not able to process every file in that repository and found the following frequencies of agreement. There was a Condorcet winner in 66 of the 69 processed elections and the Condorcet-agreement frequencies below are taken over the cases where the Condorcet winner exists.

69 Elections	Baldwin	Borda	Condorcet	IR	Plurality
Baldwin	-	.913	.864	.812	.739
Borda	.913	-	.894	.841	.768
Condorcet	.864	.894	-	.985	.818
Instant Runoff	.812	.841	.985	-	.797
Plurality	.739	.768	.818	.797	-

Merging these with the political elections above, we have the following frequencies of agreement over 521 single-winner ranked-choice elections. In 85.4% of these elections, all methods agreed on the winner. There were Condorcet winners in 99.0% of these elections.

521 Elections	Baldwin	Borda	Condorcet	IR	Plurality
Baldwin	-	.962	.932	.917	.869
Borda	.962	-	.948	.937	.894
Condorcet	.932	.948	-	.984	.913
Instant Runoff	.917	.937	.984	-	.914
Plurality	.869	.894	.913	.914	-

Note that the few dozen elections introduced here contained a higher density of cases where the methods disagreed. We suspect that the bigger data-set contains more cases where there was a *consensus* winner and the elections were not close enough for methodological details to determine the result. For example, the bigger data-set includes the 2020 Democratic Presidential primary in Hawaii, which Joe Biden won by a large margin after Bernie Sanders (the last of his rivals remaining in the race) withdrew before voters finished casting ballots. There were other noncompetitive elections including 2022's election for House District 38, 2017's election for Minneapolis's 13th Ward Representative to City Council, and others. There is further discussion of method disagreement as a function of the closeness of the election in section 4.

3. Spoilers in Ranked Preference Elections

The **spoiler effect** is one of the well-known drawbacks of electoral systems. If similar candidates enter an election and split the votes of their shared constituency, it can help elect the least-favored alternative for that constituency. Famous spoilers in presidential politics include Ross Perot and Ralph Nader. And Abraham Lincoln may have been elected president with the help of a spoiler [12].

Mitigating the spoiler effect is one attractive aspect of ranked-preference voting methods [5]. Using the ballots from 444 political and large organizational elections ([10], [7]) we measured the number of candidates who spoiled elections.

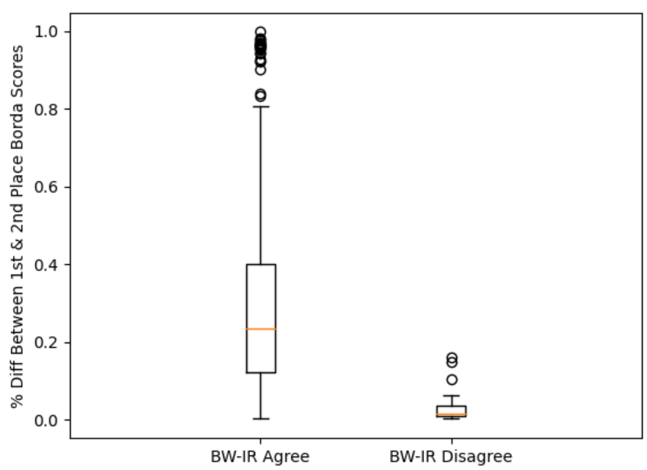
Here we define a **spoiler** to be a candidate who does not win the election, but whose removal from all ballots changes the winner of the election. Naturally, spoilers will occur less for runoff methods that are designed to eliminate irrelevant alternatives during vote counting. However, by our somewhat broad definition, spoilers are still possible for IR and Baldwin's method (even though we observed 0 Baldwin spoilers).

444 Elections	Plurality Spoilers	Borda Spoilers	IR Spoilers	Baldwin Spoilers
Number/ Frequency	29/ 0.0653	8/ 0.0180	5/ 0.0113	0/ 0.000

4. Method Disagreements and Close Elections

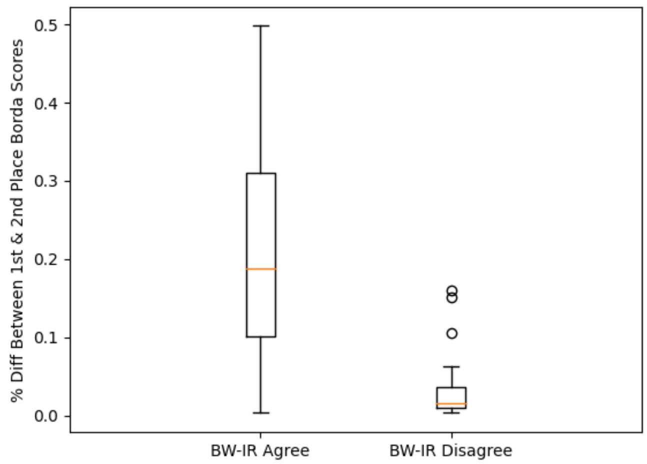
We observe that methods are more likely to disagree when the elections are approximately *ties* in some underlying sense. Here we use the relative closeness of the top two candidates' Borda scores as a measure of a given election's closeness.

Below is a box plot of the Borda gaps for elections where Baldwin's method agreed with IR next to the same for cases where those methods disagreed. These numbers come from 430 elections found at [10] and there were 25 Baldwin-IR disagreements in this data.



Here we see that the methods can agree under any conditions, but that disagreement is much more likely when the elections are close. We see that the disagreeing cases are also among the very closest elections with only a few outliers.

Considering only those cases where the second highest Borda score is at least half the size of the highest Borda score, we include all disagreements while excluding those elections with only one truly viable candidate. The box plot of those cases is below.



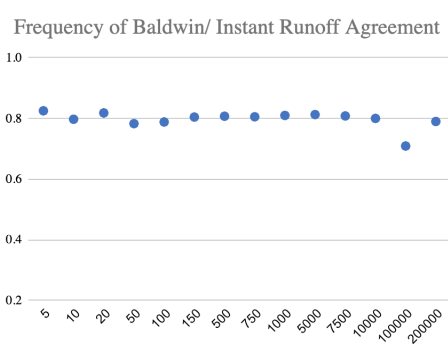
4.1. **Uniform Ballot Simulations.** There are interesting and accurate ways to simulate voter behavior. Much literature emphasizes the accuracy of mixed spatial models ([13], [8], [3]) as illustrated by our simulation in section 1. There is also a well-known graph theoretic voter model which focuses on the way voters influence each other before votes are cast [6].

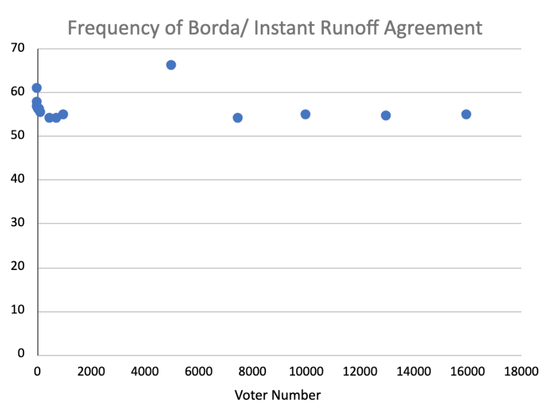
One naive way to model voter behavior is to simply assign preferences to voters using a uniform distribution. We have found that this method does not match real world elections. But it does provide us with insights into counting methods.

When N simulated candidates are distributed randomly with uniformly equal probabilities of landing in each spot on each voter's ballot, the election is likely to be close to an N-way tie. As previously noted, when the top candidates are equally well positioned, the frequency of method disagreement increases. And we conjecture that the same holds for other pathologies.

We believe that some pairs of methods are fundamentally more susceptible to disagreement. And using uniform simulations - despite their politically unrealistic nature - can help clarify this in two ways. First, because they provide so many near ties, the data-sets will be denser in disagreements. And second, because uniform simulations are totally free of implementation choices (spatial models like the one illustrated in section 1 require careful calibration and non-trivial choices about voter distribution) they can provide a more sterile and homogeneous environment for attempting to identify the aspects of the counting methods themselves that are leading to disagreements.

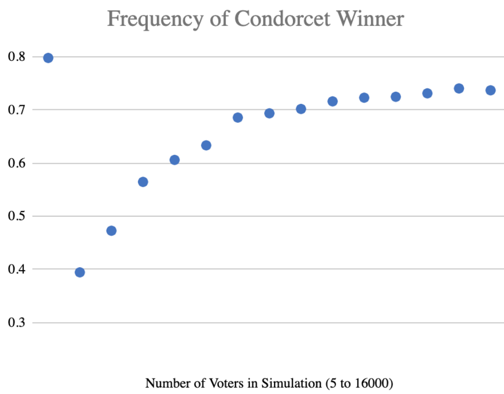
Below we have graphs showing the frequency of agreement between IR and the Borda and Baldwin methods. Each dot represents a sample of over 1100 trials. Every simulation had five candidates. Other pairwise comparisons produced similar looking graphs, with Plurality disagreeing most frequently with other methods.

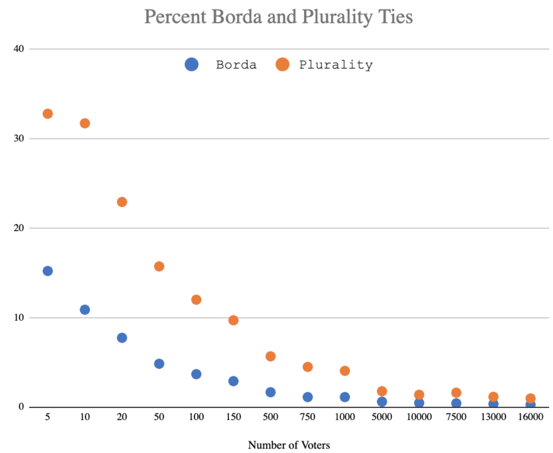




One interesting observation about this data is that one detects horizontal asymptotes in the graphs above. We have attempted simple combinatorial proofs that these asymptotes must exist, but have not yet succeeded.

We also found that while the probability of a tie decreases in these simulations, we are of course staying close to a tie. And we found apparently asymptotic behavior in the probability of a Condorcet winner. This quantity - like others observed here - showed a much higher frequency of pathology than the real-world data.





5. Concluding Remarks

The different data presented here tell a consistent story. Counting methods which consider relative rankings tend to agree more with each other than those that see only first place votes. Beyond this, Baldwin and Borda tend to agree with each other most frequently (natural, perhaps, considering Baldwin's use of Borda points) while IR and Condorcet were frequently in agreement in our data. And all uses of ranked preference ballots allow voters to express opinions that the common plurality method does not.

And the main natural questions about methods disagreeing in real elections and realistic simulations remain. 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 note that while disagreements are somewhat infrequent, they still occur often enough to effect the life of a community using ranked preference elections regularly. As illustrated in the APA and Burlington examples above, when these elections are close to ties, methodological quirks can be the tie-breaker. This may strike voters as less fair even than the spoiler effect which reflects a shortcoming of plurality elections but can be ascribed directly to voters' actions, as noted in the Burlington case specifically.

There are other well-understood reasons why communities may stick with plurality voting beyond the inertia of the status quo. Plurality has the simplest ballot and the most easily understood formula for choosing a winner. Complexity can trigger suspicion, particularly if the counting is invisible. There is also the fact that plurality entrenches the two-party system, a tendency known widely as Duverger's Law [2]. And powerful parties (for example, Democrats and Republicans in the United States) may work against changes that threaten the stable duopoly after it emerges.

As a matter of personal opinion, all authors would like to see one of the above methods used over plurality for elections involving our communities because of the way they facilitate more meaningful and nuanced expression by voters. And we consider the less frequent occurrence of spoilers to indicate the value of the advanced methods.

References

[1] Kenneth J Arrow. *Social choice and individual values*, volume 12. Yale university press, New Haven, CT, 2012.

[2] Maurice Duverger. *Political parties : their organization and activity in the modern state*. Methuen, London, UK, 1964.
[3] James M Enelow and Melvin J Hinich. *The spatial theory of voting: An introduction*. Cambridge University Press, Cambridge, UK, 1984.
[4] A. Gierzynski, W. Hamilton, and W. D. Smith. Burlington vermont 2009 irv mayor election. `https://www.rangevoting.org/Burlington.html`, 2009. Retrieved June, 2022.
[5] R. Hutchinson. Defining the spoiler effect. `https://fairvote.org/defining-the-spoiler-effect/`, 2023.
[6] Thomas M Liggett. Stochastic models of interacting systems. *The Annals of Probability*, 25(1):1–29, 1997.
[7] N. Mattei, T. Walsh, and R. Simon. Preflib: A library for preferences. `https://www.preflib.org/`, 2023.
[8] Samuel Merrill III. Discriminating between the directional and proximity spatial models of electoral competition. *Electoral Studies*, 14(3):273–287, 1995.
[9] Baseball Writer's Association of America. Award index. `https://bbwaa.com/awards/`, 2022.
[10] Deb Otis. Single winner ranked choice voting CVRs. `https://doi.org/10.7910/DVN/AMK8PJ`, 2022.
[11] Florenz Plassmann and Nicolaus Tideman. How frequently do different voting rules encounter voting paradoxes in three-candidate elections? *Social Choice and Welfare*, 42:31–75, 2014.
[12] William Poundstone. *Gaming the vote: Why elections aren't fair (and what we can do about it)*. Macmillan, New York, NY, 2008.
[13] George Rabinowitz and Stuart Elaine Macdonald. A directional theory of issue voting. *American political science review*, 83(1):93–121, 1989.
[14] Amartya Sen. *Collective choice and social welfare*. Harvard University Press, Cambridge, MA, 2018.
[15] W. Smith and N. Tideman. An election collection. `https://www.rangevoting.org/TidemanData.html`. Retrieved June, 2022.