Public Choice (2014) 161:1–9
DOI 10.1007/s11127-013-0118-2

Frequency of monotonicity failure under Instant Runoff Voting: estimates based on a spatial model of elections

Joseph T. Ornstein · Robert Z. Norman

Received: 5 October 2012 / Accepted: 7 September 2013 / Published online: 17 October 2013

Abstract It has long been recognized that Instant Runoff Voting (IRV) suffers from a defect known as nonmonotonicity, wherein increasing support for a candidate among a subset of voters may adversely affect that candidate's election outcome. The expected frequency of this type of behavior, however, remains an open and important question, and limited access to detailed election data makes it difficult to resolve empirically. In this paper, we develop a spatial model of voting behavior to approach the question theoretically. We conclude that monotonicity failures in three-candidate IRV elections may be much more prevalent than widely presumed (results suggest a lower bound estimate of 15 % for competitive elections). In light of these results, those seeking to implement a fairer multi-candidate election system should be wary of adopting IRV.

Keywords Voting theory · Instant Runoff Voting · Agent-based modeling · Monotonicity

JEL Classification D72

1 Introduction

Instant Runoff Voting (IRV) is a ranked-ballot voting system that selects a single winner by successively eliminating the candidate with the fewest first place votes until a single candidate receives a majority. IRV, like many systems that employ a successive-elimination procedure, violates the monotonicity criterion (Smith 1973), meaning there exist some conditions under which increasing the support for a candidate (without changing the voters' rankings for any of the other candidates) would be harmful to that candidate.

J.T. Ornstein (✉)
Department of Political Science, University of Michigan, 5700 Haven Hall, 505 South State Street, Ann Arbor, MI 48109-1045, USA
e-mail: ornstein@umich.edu

R.Z. Norman
Department of Mathematics, Dartmouth College, 340 Kemeny Hall, Hanover, NH 03755, USA

Fishburn and Brams (1983) describe a paradox that may result from this defect, which they term the "more-is-less paradox". This paradox, which we will refer to as a *monotonicity failure*, is characterized as a situation in which the IRV winner would lose if ranked higher by some subset of voters. The converse, an election in which the IRV loser would win if ranked lower by some subset of voters, is also a type of monotonicity failure. We term the former paradox an *upward* monotonicity failure, and the latter a *downward* monotonicity failure. In this paper, we will focus exclusively on the prevalence of upward monotonicity failures under IRV, leaving downward monotonicity failure as a topic for future research.

How often could we reasonably expect to observe this behavior in real-world IRV elections? One study (Allard 1996) suggests that, if the United Kingdom adopted IRV for its general elections, this paradox would occur only about once per century. If this were true, then the monotonicity criterion would be cause for little practical concern. However, more recent theoretical work on the topic (Lepelley et al. 1996; Norman 2006, 2012; Miller 2012) finds that the proportion of possible IRV elections that exhibit a monotonicity paradox is non-trivial.

Here we build on this body of work by estimating IRV's monotonicity failure rate in the context of a spatial model of voter behavior. This method allows us to avoid the assumption that all electoral outcomes are equiprobable, and to explore the conditions under which IRV may be more or less vulnerable to monotonicity failure.

2 The 2009 Burlington mayoral election: a case study

Before introducing the model, we present a case study from the 2009 mayoral election in Burlington, VT, which illustrates the key features of an upward monotonicity failure. Table 1 reports the ballot results for the three candidates remaining before the final elimination—the Republican (R), Democrat (D), and Progressive (P).

The Republican candidate had the most first-place votes, with 3,297. The Democrat had 2,554 first-place votes, and the Progressive incumbent had 2,982. Although the Democrat was the Condorcet winner (a majority of voters preferred him in all two way contests), he received the fewest first-place votes and so was eliminated. After the Democrat's votes were redistributed to the other two candidates, the Progressive won the election 4,314 to 4,067.

It was a very competitive election, as the winner's margin of victory was 247 votes (2.8 % of the electorate). A small shift in support from the Progressive to the Republican would have resulted in a Republican victory. But what if Burlington's electorate had been composed of even more Progressive voters? If we shift the Progressive candidate up in 750 rankings, we can construct the following election profile (Table 2, changes in bold).

Table 1 Ballot results from the 2009 Burlington, VT mayoral election (Laatu and Smith 2009)

Ranking	1513	495	1289	1332	767	455	2043	371	568
1st	R	R	R	D	D	D	P	P	P
2nd	D	P		P	R		D	R	
3rd	P	D		R	P		R	D	

Each column represents a unique rank-order ballot cast by voters, and the top row denotes the number of voters who submitted that ballot. For instance, the first column denotes that 1513 voters ranked the Republican first, the Democrat second, and the Progressive third. In the Burlington election, voters were permitted to submit incomplete rankings (a "truncated ballot"). Though this form of IRV is also susceptible to monotonicity failure, for simplicity of analysis we do not include it in the model.

Table 2 Hypothetical ballot results for Burlington election

Ranking	1513	**195**	**839**	1332	767	455	2043	**1121**	568
1st	R	**R**	**R**	D	D	D	P	**P**	P
2nd	D	**P**		P	R		D	**R**	
3rd	P	**D**		R	P		R	**D**	

As a result of this mass shift in voting, the first-place vote totals would now be: R—2,547, D—2,554, and P—3,732. If the Progressive won the election in which those 750 voters supported the Republican instead of him, then a monotonic voting system would award him this election as well. However, in this hypothetical election, the Republican is eliminated instead of the Democrat, and after redistributing the Republican's votes, the Progressive loses to the Democrat 3,927 to 4,067.

The Burlington election offers a compelling illustration of monotonicity failure's practical importance, but such detailed IRV ballot data are rare. Therefore, in order to estimate the frequency of its occurrence, we rely in this paper on a spatial model of elections. In the following section, we describe the conditions under which a three-candidate election will exhibit an upward monotonicity failure, which we use for the analysis of the model.[1]

3 Necessary and sufficient conditions for monotonicity failure

The outcome of any election conducted using Instant Runoff Voting can be represented by a vector P, termed an election *profile*. Each element in P is a non-negative integer denoting the number of voters who cast a particular rank-ordered ballot. The sum of all elements in P is V, the number of voters. For three-candidate elections, we denote the candidates A, B, and C, where A is the IRV winner and C is the candidate with the fewest first-place votes. Therefore the elements in P are $\langle a_1, a_2, b_1, b_2, c_1, c_2\rangle$ as in Table 3.

An election profile P exhibits an upward monotonicity failure if there exists a profile P' that is identical to P except that candidate A is ranked higher by a subset of voters, but candidate A is not the IRV winner. Formally, there must exist non-negative integers λ_1 and λ_2 such that we can construct an election profile

$$P' = \langle a_1 + \lambda_1 + \lambda_2, a_2, b_1 - \lambda_1, b_2 - \lambda_2, c_1, c_2\rangle$$

in which candidate A is not a majority winner, candidate B is eliminated, and candidate C is the IRV winner. According to our definitions from Sect. 1, profile P' will exhibit a *downward monotonicity failure*. There are two conditions which, jointly, are necessary and sufficient for the existence of P', shown in (1) and (2) below.

The first is that P must be a *competitive* election, defined as a profile in which[2]

$$c > \frac{V+2}{4} \tag{1}$$

[1] Our analysis here closely parallels work by Lepelley et al. (1996) and Miller (2012), but differs in its construction of condition (1). In our analysis of the spatial model (Sect. 4), we ignore cases where two candidates tie for fewest first-place votes, so we include a stronger version of condition (1) than in these previous papers.

[2] For convenience, we define $c = c_1 + c_2$.

Table 3 A three-candidate election profile

Ranking	a_1	a_2	b_1	b_2	c_1	c_2
1st choice	A	A	B	B	C	C
2nd choice	B	C	A	C	A	B
3rd choice	C	B	C	A	B	A

The second condition necessary for P to exhibit an upward monotonicity failure is that candidate C must be preferred over A by a majority of voters. This implies that, absent a tie, either candidate C is the Condorcet winner, or that there is no Condorcet winner (the election exhibits a majority cyclic triple). This condition can be expressed by:

$$c + b_2 > \frac{V}{2} \tag{2}$$

The necessity of condition (1) follows from the requirement that candidate A cannot receive a majority of first-place votes or tie under P'. Formally

$$(b_1 - \lambda_1) + (b_2 - \lambda_2) + c > \frac{V}{2} \tag{3}$$

Because we've also specified that candidate B is eliminated under P', we can combine

$$(b_1 - \lambda_1) + (b_2 - \lambda_2) \le c - 1 \tag{4}$$

with (3), yielding equation (1).

The necessity of condition (2) follows from the requirement that candidate C wins under P' with $c + b_2 - \lambda_2$ votes. Since λ_2 is non-negative, it is necessary that P satisfy (2).

To prove the sufficiency of conditions (1) and (2), we will show that when P satisfies (1) and (2), there must exist an election profile P' in which candidate B receives fewer first-place votes than C, and candidate C receives a majority following B's elimination. Formally, it will suffice to show that under these conditions there exist non-negative integers λ_1 and λ_2 that satisfy (4) and

$$c + b_2 - \lambda_2 > \frac{V}{2} \tag{5}$$

Let $\lambda_1 = b_1$ and λ_2 be the largest integer that satisfies (5). It follows by definition that λ_1 is non-negative and from (2) that λ_2 is non-negative. It follows by contradiction that these values for λ_1 and λ_2 will also satisfy (4), because when $(b_2 - \lambda_2) \ge c$, (1) implies $c + b_2 - \lambda_2 > \frac{V+2}{2}$, so λ_2 is not the largest integer satisfying (5). This completes the proof.

In the model presented in the following section, we use (1) and (2) to assess the monotonicity failure rate of a set of simulated elections.

4 The model

The model we develop here is a two-dimensional spatial model, a method used widely to analyze behavior in elections and performance among voting systems (Downs 1957; Merrill 1988; Kenny and Lotfinia 2005). Such modeling has also been used to inform disputes about the merits of IRV in particular (Fraenkel and Grofman 2004; Horowitz 2004).

The positions of candidates and voters are represented by points on a two-dimensional issue space, and each voter's preference ranking is constructed by taking the reverse order of the Euclidean distance to each candidate within the space.[3] Formally, let n represent the number of issues ($n = 2$ for this paper), let x_{ji} represent the ideal point for the jth voter's ith issue, and let y_i represent candidate y's position on the ith issue. Voters rank each candidate y in increasing order of utility, given by

$$u_j(y) = \left(\sum_{i=1}^{n}(x_{ji} - y_i)^2\right)^{-1/2} \tag{6}$$

Voters' ideal points are randomly seeded across the issue space using one of four stylized distributions. These distributions are not drawn from data, but are constructed to mirror features from plausible real-world electorates. Such non-empirical distributions allow us to gauge the model's behavior over a wider range of scenarios than empirical data alone would permit.

In the base case of the model, the ideal points on each axis are drawn from a Gaussian distribution with mean 0 and standard deviation 0.25. In the "polarized" case, ideal points are drawn at random from one of two bivariate Gaussian distributions, one centered at point $(-0.25, -0.25)$ and the other centered at $(0.25, 0.25)$, each with standard deviation 0.1. This configuration mimics an election in which voters are split into two polarized camps. In Sect. 5, we present the model results from two instantiations of this distribution: one where voters are equally likely to be in either camp (Balanced Polarized), and one where voters are 1.5 times more likely to position themselves in one camp than the other (Unbalanced Polarized).

The final, "multiparty", distribution randomly draws half of the voters' ideal points from the polarized distributions described above, one-quarter of the voters from distributions centered at $(0.25, -0.25)$ and $(-0.25, 0.25)$, and one-quarter of the voters from a distribution centered at $(0.0, 0.0)$. Each distribution has standard deviation 0.1. This configuration represents an election in which there are three major camps and two smaller camps. An illustration of these three scenarios is in Fig. 1.

We model candidates as *boundedly rational adaptive agents*, following (Kollman et al. 1992; De Marchi 1999; Laver 2005). Each candidate in the model attempts to maximize the number of first-place votes received, but does not know his or her optimal location given the locations of other candidates, and receives information only through periodic polling. All three candidates begin the election at random positions in the voter distribution,[4] and each period they determine how many first-place votes they would receive if the election were held immediately. Candidates then change their positions on one or both issues by a fixed increment if that adjustment would result in a higher first-place vote count. The size of this fixed increment is 0.01, sufficiently small such that agents will not "overstep" an advantageous position. The positions of voters do not change during the course of the simulation.

Since outcomes are stochastic, we model 5,000 elections for each type of voter distribution (varying by the number of periods that candidates may move during each election, L). Each election is parameterized with 1,001 voters, though outcomes are robust to halving or doubling this value, or setting V even or odd. The simulation code is available from the authors on request.

[3] The qualitative results presented in this paper are robust to alternate specifications of utility, including city block distance and squared Euclidean distance.

[4] In the polarized case, we ensure that there is at least one candidate in each "camp", and two in the larger camp.

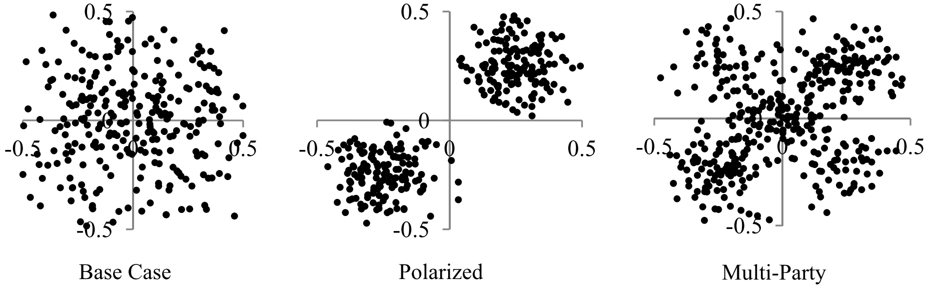

Fig. 1 Examples of the three voter distributions utilized in the model. The balanced and unbalanced polarized distributions are visually indistinguishable, and so only one case is illustrated here

5 Results

The results from the model suggest that upward monotonicity failures are likely to occur with significant frequency under IRV, and that this frequency increases with competitiveness. Depending on the type of voter distribution and length of the simulation, the simulated elections exhibit monotonicity failure in anywhere from 0.7 % to 51 % of all cases, and between 15 % and 51 % of competitive elections (excluding ties, which account for approximately 1.1 % of simulated elections).

Monotonicity failure rates for each voter distribution appear to stabilize at higher values of L (Fig. 2). The Unbalanced Polarized distribution exhibits the highest rate of monotonicity failure (approximately 50 % at $L > 40$), the Base Case and Balanced Polarized distributions exhibit monotonicity failures in 9 % to 12 % of simulated elections, and the Multiparty distribution exhibits the most infrequent monotonicity failures (a lower bound of 0.7 %).

The simulation length parameter (L) appears to have a varied effect on the rate of monotonicity failure. As L increases, the Base Case and Polarized distributions exhibit competitive elections as defined by (1) more frequently, which in turn results in a higher overall monotonicity failure rate. By contrast, the proportion of competitive elections tends to decrease in the Multiparty distribution as L increases, as illustrated by Fig. 3.

In the base case and polarized distributions, candidates will tend to settle on a local equilibrium given enough time. In the base case, candidates position themselves near the yolk of the distribution (McKelvey 1986) centered on $(0, 0)$. This central positioning increases the chance of a three-way competitive election. Candidates in the two polarized distributions tend to locate near the yolks of their respective camps. This invariably leads to a competitive election in the Unbalanced Polarized case, where the two candidates in the larger camp each take 30 % of the vote, but rarely results in a competitive election in the Balanced Polarized case, where two candidates must split roughly 50 % of the vote between them. On average, 50 % of competitive elections in either polarized distribution result in the elimination of the Condorcet winner, and therefore exhibit a monotonicity failure (Fig. 4).

By contrast, candidates in the Multiparty distribution never settle into a local equilibrium near the yolk, regardless of simulation length. This is likely because candidates at the center of the distribution who have captured the vote of one of the smaller peripheral camps have an incentive to compete for one of the larger peripheral camps instead. This sets off a race between two candidates to gain the support of the periphery, destabilizing the equilibrium at the yolk, where elections are three-way competitive. As indicated by Figs. 3 and 4, very few elections conducted with the Multiparty voter distribution are competitive, but those

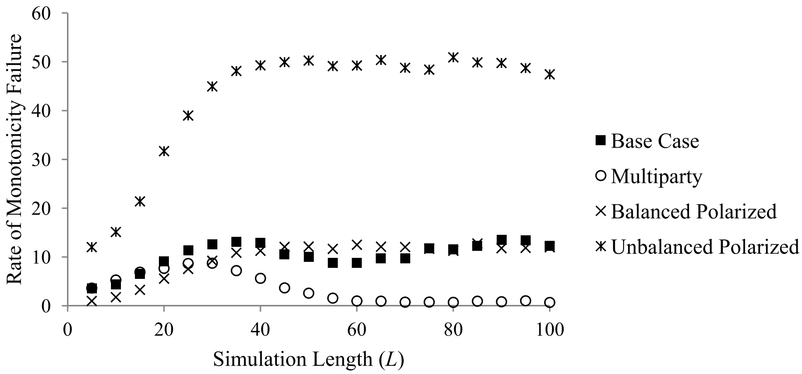

Fig. 2 Nonmonotonic rate (%) by simulation length (L). Values derived from 5,000 simulated elections for each value of L (ties excluded)

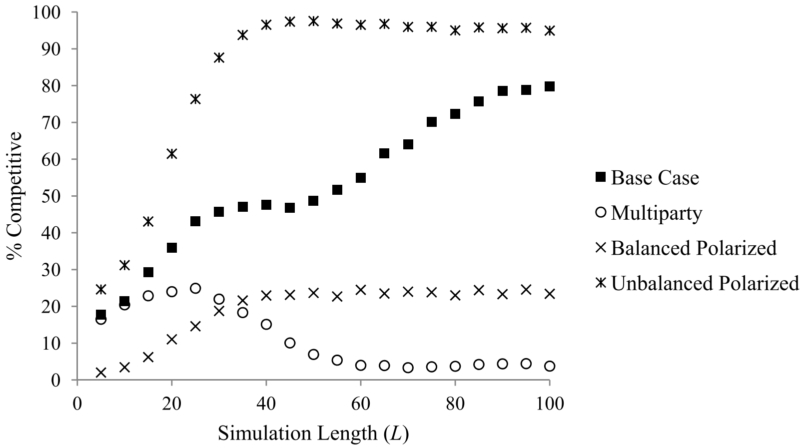

Fig. 3 Percentage of elections that are competitive as a function of L. Values derived from 5,000 simulated elections for each value of L (ties excluded)

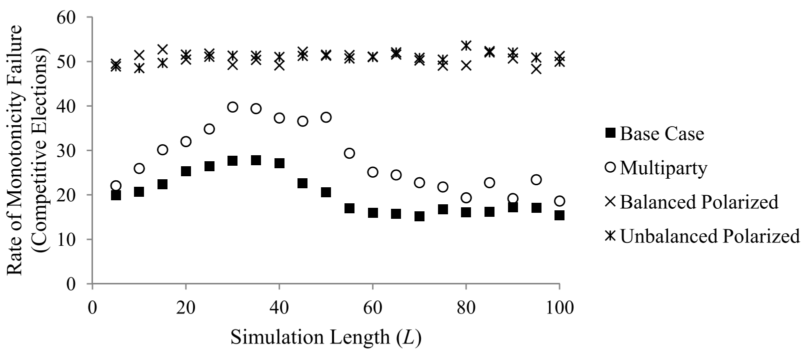

Fig. 4 Nonmonotonic rate (%) of competitive elections by simulation length (L). Values derived from 5,000 simulated elections for each value of L (ties excluded)

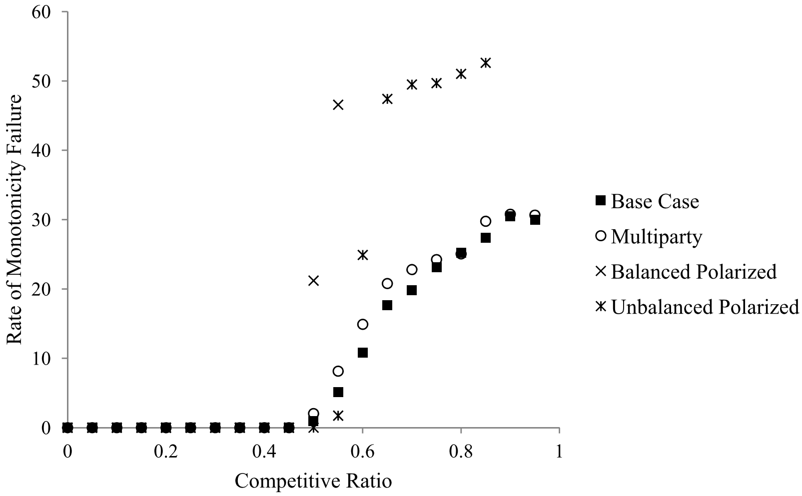

Fig. 5 Nonmonotonic rate (%) of elections by competitive ratio. This chart is derived from 500,000 runs of the simulation for each distribution. Each *point* illustrates the rate of monotonicity failure for elections with competitive ratio between x and $(x + 0.05)$

that are competitive exhibit monotonicity failure roughly 20 % of the time. Indeed, competitive simulated elections exhibited montonicity failure at least 15 % of the time for all parameterizations (Fig. 4).

In addition to reporting overall monotonicity failure rates, we investigate whether an election's degree of competitiveness has an effect on its monotonicity failure rate. To do so, we construct 500,000 simulated elections for each voter distribution and plot the rate of monotonicity failure as a function of an election's *competitive ratio* (defined as the ratio of first-place votes received by candidate C to first-place votes received by candidate A; elections with higher ratios are more competitive). Figure 5 illustrates how the rate of monotonicity failure increases with competitive ratio for all four voter distributions.

Finally, it is notable that very few of the model's simulated profiles exhibited majority cyclic triples. Of elections run with the Base Case distribution, only 0.4 % exhibited a majority cyclic triple, 0.05 % in the Multiparty distribution, and 0.01 % for each of the Polarized distributions. Since monotonicity failures can occur only when the election profile exhibits a majority cyclic triple or when IRV fails to elect the Condorcet winner, this result indicates that the monotonicity failures simulated here occur primarily due to IRV's Condorcet inefficiency in competitive elections.

6 Conclusion

We have demonstrated here in a spatial model of voter behavior that upward monotonicity failures arise in a non-trivial percentage of simulated elections. The lower bound estimate of 15 % in competitive elections represents a testable prediction of the model, and suggests that three-way competitive races will exhibit unacceptably frequent monotonicity failures under IRV. We also find that the rate of monotonicity failure increases with an election's degree of competitiveness, a finding that holds true for all of the distributions studied. We restrict our attention in this paper to the three-candidate case for largely pragmatic reasons; the closed-form method for determining which profiles exhibit monotonicity failure (Sect. 3) greatly reduces the computational complexity of our model. The general case with more than three candidates is a promising topic for future research.

Of course, upward monotonicity failure is not the only major defect of IRV, and future work will need to examine the frequency of other paradoxes to which IRV is subject. Perhaps the only definitive way these questions can be resolved is by examining a broad body of data from real IRV elections. Such a body does not yet exist, though it is telling that out of only two IRV elections in Burlington, VT, there has already been one recorded instance of nonmonotonicity. If widespread use of Instant Runoff Voting continues, then we can expect to see many more.

Acknowledgements The authors thank Ross A. Hammond and the anonymous reviewers for insightful comments and suggestions. Any errors remain the responsibility of the authors.

References

Allard, C. (1996). Estimating the probability of monotonicity failure in a U.K. general election. *Voting Matters*, *5*, 1–7.

De Marchi, S. (1999). Adaptive models and electoral instability. *Journal of Theoretical Politics*, *11*(3), 393–419.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper.

Fishburn, P., & Brams, S. (1983). Paradoxes of preferential voting. *Mathematics Magazine*, *56*(4), 207–214.

Fraenkel, J., & Grofman, B. (2004). A neo-Downsian model of the alternative vote as a mechanism for mitigating ethnic conflict in plural societies. *Public Choice*, *121*, 487–506.

Horowitz, D. L. (2004). The alternative vote and interethnic moderation: a reply to Fraenkel and Grofman. *Public Choice*, *121*, 507–516.

Kenny, L. W., & Lotfinia, B. (2005). Evidence on the importance of spatial voting models in presidential nominations and elections. *Public Choice*, *123*(3–4), 439–462.

Kollman, K., Miller, J., & Page, S. (1992). Adaptive parties in spatial elections. *American Political Science Review*, *86*, 929–937.

Laatu, J., & Smith, W. (2009). *The rank order votes in the 2009 Burlington mayoral election*. http://rangevoting.org/JLburl09.txt.

Laver, M. (2005). Policy and the dynamics of political competition. *American Political Science Review*, *99*(2), 263–281.

Lepelley, D., Chantreuil, F., & Berg, S. (1996). The likelihood of monotonicity paradoxes in run-off elections. *Mathematical Social Sciences*, *31*, 133–146.

McKelvey, R. (1986). Covering, dominance, and institution-free properties of social choice. *American Journal of Political Science*, *30*(2), 283–314.

Miller, N. R. (2012). Monotonicity failure in IRV elections with three candidates. In *World meeting of the public choice societies, 2012*.

Merrill, S. (1988). *Making multicandidate elections more democratic*. Princeton: Princeton University Press.

Norman, R. Z. (2006). Some properties of STV in close and polarized elections. In *2006 meeting of the society for social choice and welfare*.

Norman, R. Z. (2012). Frequency and severity of some STV paradoxes. In *World meeting of the public choice societies, 2012*.

Smith, J. H. (1973). Aggregation of preferences with variable electorate. *Econometrica*, *41*, 1027–1041.