Susceptibility to strategic voting: a comparison of plurality and instant-runoff elections[1]

Andrew C. Eggers[†]

Tobias Nowacki[†]

Abstract

Advocates of the instant-runoff voting system (IRV) often argue that it is less susceptible to strategic voting than plurality. Is this true? How should we define and measure a voting system's susceptibility to strategic voting? Previous research in this area is unsatisfying, partly because it ignores the uncertainty voters face when they vote; we introduce a better approach. We find that, when beliefs are precise and other voters are expected to vote sincerely, more voters would benefit from voting strategically in IRV than in plurality (contrary to what advocates suggest). The anticipated benefit for these voters is small, however, and for the average voter the benefit of taking strategy into account is many times larger in plurality than IRV – especially when beliefs are imprecise and/or voters expect other voters to behave strategically. The methods we introduce can be used to study other properties of voting systems when voters are strategic.

[1]This version: December 1, 2021. Thanks to Steve Callander, Gary Cox, Justin Grimmer, Matias Iaryczower, Paul Klemperer, Dimitri Landa, Gregory Martin, David Myatt, Karine van der Straeten, and seminar audiences at Trinity College Dublin, the Toulouse School of Economics, the Democracy and Polarization Lab at Stanford University, and the ECPR Standing Group on Strategic Interactions for valuable feedback. Thanks to Mats Ahrenshop, Tak Huen Chau, and Tom Robinson for research assistance. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. Earlier versions were circulated starting June 2019 under the title "Comparing strategic voting incentives in plurality and instant-runoff elections".

[†]University of Chicago. email: `aeggers@uchicago.edu`

[†]Stanford University. email: `tnowacki@stanford.edu`

In recent years, the instant-runoff voting system (IRV) has been a popular reform proposal in several countries. In IRV (also known by several other terms, including ranked-choice voting and the alternative vote[2]), voters rank the candidates and the winner is determined by successively eliminating less-popular candidates. Large-scale referendums to replace plurality voting with IRV failed in the UK in 2011, British Columbia in 2018, and Massachusetts in 2020, but succeeded in San Francisco in 2002, Maine in 2016, and New York City in 2019. According to its advocates, IRV chooses winners with broader support, discourages negative campaigning, and minimizes strategic voting, among other benefits.[3]

Although many aspects of IRV elections have now been studied (e.g. Farrell and McAllister, 2006; Fraenkel and Grofman, 2006; Horowitz, 2006; John, Smith, and Zack, 2018; McDaniel, 2018), previous research leaves some important questions unanswered. Consider IRV's resistance to strategic voting, which is sometimes offered as its most important advantage over plurality.[4] When a candidate is eliminated in IRV, ballots ranking that candidate first are effectively transferred to the next-ranked candidate. This suggests that the incentive to abandon hopeless candidates is lower in IRV than in plurality. But IRV allows for other types of strategic voting, some rather perverse: for example, one might help elect one's preferred candidate by ranking one's *least*-preferred candidate first (e.g. Fishburn and Brams, 1983; Dummett, 1984). It remains an open question whether, for realistic assumptions about the circumstances voters face in real elections, the opportunities for strategic voting are (as advocates claim) more constrained in IRV than in plurality. Dummett (1984), for example, ventured that "a voter who has understood the workings of [IRV], and who has some information about the probable intentions of the others, will have nearly as much incentive to vote strategically." This conjecture has never been carefully tested, however, because (as we explain below) previous efforts fail to capture the uncertainty implied in the scenario Dummett imagines.

[2]IRV's other names include ranked-choice voting (RCV), the alternative vote (AV), preferential voting, single-transferable vote (STV), and the Hare system. We use instant-runoff voting because the term is widely used and more descriptive than "ranked-choice voting" or "the alternative vote".

[3]See e.g. FairVote's "Ranked Choice Voting 101" at `https://www.fairvote.org/rcv#rcvbenefits`, visited 1 August 2020.

[4]For example, the first argument offered in favor of IRV by UK Deputy Prime Minister Nick Clegg in parliamentary testimony in advance of the 2011 UK referendum was that IRV "stops people from voting tactically and second-guessing how everybody else will vote in their area." See "The Coalition Government's programme of political and constitutional reform: Oral and written evidence", 15 July 2010, HC 358-i, published 22 October 2010 (link).

In this paper, we seek to rigorously assess the common claim that IRV is less susceptible to strategic voting than plurality; to do so, we develop concepts and tools that should be useful for addressing other questions about voting systems and strategic voting. We define a voting system's susceptibility to strategic voting as the expected benefit to the voter from voting strategically (i.e. casting the expected-utility maximizing vote) instead of voting sincerely (i.e. simply reporting the sincere preference). To measure susceptibility to strategic voting in realistic scenarios, we start with preference data from 160 election surveys (a total of around 220,000 voters); we then infer reasonable beliefs these voters might hold about election outcomes in each voting system given others' preferences and a novel model of belief formation informed by empirical measures of forecasting uncertainty. This allows us to measure the expected utility of each possible ballot for each voter under a range of assumptions about the prevalence of strategic voting in the electorate; from this, we can compute how many voters would benefit from voting strategically and by how much, which together constitute our measure of susceptibility to strategic voting.

Our analysis shows that, for a wide range of assumptions about how beliefs are formed, IRV is less susceptible to strategic voting on average than plurality voting, although the opportunities to benefit from strategic voting in IRV may be more widespread than previously recognized. If we assume that voters have relatively precise beliefs about election outcomes and expect other voters to vote sincerely, the average gain in expected utility from voting strategically (rather than always sincerely) is about five times higher in plurality compared to IRV; that gap becomes wider when we assume that voters have less precise beliefs about election outcomes and/or expect other voters to vote strategically. We also find that Dummett and other IRV skeptics are correct in supposing that informed voters might find ample opportunities for strategic voting in IRV: when beliefs are precise and other voters are expected to vote sincerely, the proportion of voters who benefit from strategic voting is higher on average in IRV, and much higher for some preference profiles. But when beliefs become less precise, or when other voters are expected to vote more strategically, these opportunities for strategic voting tend to disappear. This main finding holds when we consider possible effects of the voting system on preferences or the number of parties competing. We also find that *outcomes* depend more on strategic voting in plurality than in IRV.

Although we are not the first to compare opportunities for strategic voting across voting systems (e.g. Chamberlin and Featherston, 1986; Saari, 1990; Bartholdi and Orlin, 1991; Green-Armytage, Tideman, and Cosman, 2016), our approach improves on previous efforts in two significant ways.[5] Most importantly, to our knowledge we are the first to assess susceptibility to strategic voting while taking into account the uncertainty that voters face at the time they decide how to vote. Starting with the Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975), previous research evaluating the possibility of strategic manipulation has assumed that voters know other voters' votes.[6] Ignoring uncertainty simplifies many aspects of the problem, but it may produce misleading conclusions about opportunities for strategic voting in actual elections: even if a voting system sometimes creates situations where a voter would *regret* a sincere vote (as Gibbard-Satterthwaite shows it must), it may very rarely create situations where a voter could *anticipate* (given realistic uncertainty) that a non-sincere vote would be optimal, even if she perfectly understands how the system operates.

Our second key innovation is to model voters' beliefs in a way that accounts for strategic behavior by other voters. Previous researchers have assessed opportunities for strategic manipulation assuming that everyone votes sincerely. Again, this simplification may mislead: the incentive to vote strategically may be stronger or weaker when voters take into account others' strategic behavior. To capture a range of possibilities about the prevalence of strategic voting in the electorate, we introduce a model of belief formation in which strategic voters with realistically uncertain beliefs respond myopically to a sequence of polls; we measure the incentive to vote strategically at each stage of this process, with the first iteration capturing the assumption that voters expect others to vote sincerely and later iterations converging on a strategic voting equilibrium. Our approach thus includes the assumptions typical of previous work (no uncertainty, sincere voting) as a special case but allows us to relax those assumptions to account for uncertainty and strategic voting by other voters. This approach reveals an important

[5]We also use a larger and more representative set of preference distributions, as explained below.

[6]As explained below, this statement applies to research in social choice theory and computational social choice that compares manipulability of voting systems. The study of strategic voting in economics and political science has always taken uncertainty seriously, but it has not sought to measure and compare susceptibility to strategic voting across voting systems.

qualitative difference between IRV and plurality: in plurality the incentive to vote strategically tends to *increase* as others vote more strategically (because trailing candidates become more hopeless), while in IRV the reverse is true.

Although this paper focuses on comparing susceptibility to strategic voting in IRV and plurality, we emphasize that the methods we develop are useful for other purposes. Our tools for measuring pivot probabilities in three- and four-candidate IRV elections make it possible to answer other questions about IRV (e.g. assessing the *ex ante* probability of monotonicity failure, studied from an *ex post* perspective by Ornstein and Norman (2014)) and can be extended to handle more candidates. Our iterative polling algorithm can be used to locate an equilibrium in any voting system given arbitrary preferences and offers a way to study other properties of voting systems (such as aggregate voter utility or the probability of electing a Condorcet winner) while relaxing the assumption that voters vote sincerely. Our entire framework for studying susceptibility to strategic voting can be extended to assess other voting systems. By necessity we do not explore these possibilities in this paper, but we hope to lay a foundation for future work on other properties and other voting systems.

1 Orientation

1.1 Why measure susceptibility to strategic voting?

Voting theorists have argued that we should prefer a voting system that is less susceptible to strategic voting (e.g. Saari, 1990; Tideman, 2018), meaning a system that is less likely to reward voters who manipulate the result by submitting a vote that differs from their sincere preference, i.e. a "misaligned vote" (Kawai and Watanabe, 2013). Two types of manipulability are potentially concerning.

First, one might be concerned about the *ex post* manipulability of a voting system, by which we mean the frequency with which voters might regret a sincere vote after the results are announced. It is natural and unavoidable for voters who favored unsuccessful candidates to feel disappointment with the outcome, but it seems desirable to avoid situations where voters regret having submitted a sincere ballot, as may have been the case for left-leaning Floridians who voted for Ralph Nader in the 2000 U.S. presidential election. Voters who recognize *ex post* a chance to improve the outcome with a misaligned vote may feel dismay about the vote they or others cast; they may also question the legitimacy of a result that could have been reversed by clever manipulation such as not voting (e.g. Fishburn and Brams, 1983) or submitting an incomplete ranking of candidates (Fishburn and Brams, 1984).

Second, one can also be concerned about the *ex ante* manipulability of a voting system, meaning the frequency with which voters could obtain a better expected election outcome by submitting a misaligned vote. A system that is more *ex ante* manipulable raises several concerns. To the extent that voters respond to the opportunity to manipulate by submitting misaligned votes, the aggregated ballot counts may diverge from the preference distribution in the electorate, making the election difficult to interpret (Satterthwaite, 1973; Riker, 1982; Green-Armytage, Tideman, and Cosman, 2016): did the voters who voted for candidate A really prefer that candidate, or did they vote for A for strategic reasons? Voters evidently derive expressive benefits from voting for candidates they sincerely support (Hamlin and Jennings, 2011; Pons and Tricaud, 2018), so a system that induces widespread misaligned voting also deprives many voters of these expressive benefits. Furthermore, there is evidence that some types of voters (e.g. poorer ones) are less able or inclined to vote strategically (Eggers and Vivyan, 2020), which suggests that a system that is more *ex ante* manipulable disadvantages these voters to a greater extent.[7] Finally, a voting system that is more *ex ante* manipulable creates stronger incentives for voters to pay attention to polls, for media outlets to conduct polls, and for parties to spread polling information (and misinformation); all of this effort might be better spent on other activities, such as scrutinizing candidates' track records or policy proposals (though see Dowding and Van Hees, 2008). Interpretability, expressive utility, fairness, and efficiency arguments thus all favor voting systems that are less *ex ante* manipulable.

[7]It is not known how heterogeneity in strategic behavior varies across voting systems, but it is a reasonable guess that voting systems that produce lower incentives to vote strategically also produce less heterogeneity in observed strategic behavior.

1.2 The limitations of previous research

Perhaps surprisingly, previous research on susceptibility to strategic voting focuses exclusively on *ex post* manipulability. The Gibbard-Satterthwaite Theorem states that any reasonable voting system is *ex post* manipulable in some circumstance; the manipulability literature (e.g. Chamberlin and Featherston, 1986; Saari, 1990; Favardin and Lepelley, 2006; Ornstein and Norman, 2014; Green-Armytage, Tideman, and Cosman, 2016; Tideman, 2018) quite reasonably builds on this result by measuring the proportion of likely election outcomes that are *ex post* manipulable. More precisely, both the manipulability literature and the social choice literature from which it emerges assume perfect information about the election result, which describes the situation after the election takes place but is plainly unrealistic as a description of the environment in which voters typically decide how to vote. The manipulability literature is also unsatisfying because it checks for opportunities to manipulate when everyone votes sincerely, which is unrealistic at least in plurality elections, where a substantial proportion of voters are believed to cast misaligned votes (e.g. Kawai and Watanabe, 2013).

The lack of research into the *ex ante* manipulability of voting systems would be unproblematic if *ex ante* and *ex post* approaches always produced the same conclusions, but they do not. For example, consider a plurality election in which candidates A, B, and C receive nearly equal support, with candidate A defeating candidate B by just one vote and C finishing slightly behind B. Clearly *ex post* manipulation is possible: for example, two supporters of candidate C who prefer B over A could elect B by switching their votes from C to B. But given a poll predicting the same near three-way tie, there is little *ex ante* reason for these voters to switch to B: each pair of candidates is approximately equally likely to be tied for first, so switching from C to B is just as likely to backfire (in the event of a B-C tie) or to lead to a wasted vote (in the event of an A-C tie) as it is to pay off (in the event of an A-B tie). More generally, any system that rewards a misaligned vote in one set of circumstances but punishes the same vote in another similar set of circumstances will look more manipulable *ex post* than *ex ante*, because the *ex ante* perspective considers both the benefits and risks of a misaligned vote.[8]

[8] A similar critique applies to research (mainly in computer science) assessing the complexity of manipulation in various voting systems (see Faliszewski and Procaccia, 2010, for a review). Bartholdi and Orlin (1991) showed that the problem of computing the optimal vote in IRV is NP-complete given perfect information and thus may

By contrast, the game theoretic literature in political science and economics treats strategic voting as an instance of choice under uncertainty (e.g. Myerson and Weber, 1993; Fey, 1997; Myerson, 2002; Myatt, 2007; Laslier, 2009; Bouton, 2013; Bouton and Gratton, 2015) and focuses on characterizing the equilibrium of voting games. To our knowledge, no previous research has applied this literature's view of strategic voting (which takes account of uncertainty and considers how voters' incentives depend on other voters' behavior) to the problem of comparing voting systems' susceptibility to strategic voting. That is our objective in this paper.

2 Susceptibility to strategic voting: a new approach

We seek to measure the extent to which different voting systems encourage strategic voting from an *ex ante* perspective: that is, given information voters might have before an election takes place, to what extent might voters expect to be rewarded for casting a misaligned vote? In this section we explain our approach, starting with defining strategic voting and susceptibility to strategic voting before moving to issues of operationalization.

2.1 Strategic voting as expected utility maximization

Following the rational choice tradition (e.g. McKelvey, 1972; Cox, 1997), we view strategic voting as an instance of choice under uncertainty. In that framework, voters have preferences over the candidates competing (which can be captured by a Von Neumann-Morgenstern utility function) and beliefs about the probability of various election outcomes (which can be captured by a probability density/mass function). *Voting strategically* means choosing the ballot that, given preferences and beliefs, yields the highest expected utility; by contrast, *voting sincerely* means simply choosing the ballot that most closely matches one's preferences. A *misaligned vote* is a vote that differs from the voter's sincere preference; strategic voting thus implies casting a misaligned vote in some circumstances, while sincere voting does not. Our terminology reflects that in Kawai and Watanabe (2013); readers should note, however, that others have used "strategic vote/voting" to mean what we refer to as "misaligned vote/voting."

be intractable as the number of voters and candidates increases. Conitzer, Sandholm, and Lang (2007) show that manipulation in IRV given perfect information is *not* hard (in the complexity sense) with a fixed number of candidates, but may be hard when we introduce uncertainty.

More formally, suppose an election takes place involving K candidates $c_1, c_2, \ldots, c_K$ in which voters may submit one of B distinct ballots. Voter i's utility if candidate c_j is elected is given by a utility function $u_i(c_j)$. Let b_i denote the ballot submitted by voter i, let $\mathbf{v}_{-i} = (v_1, v_2, \ldots, v_B)$ indicate the proportion of other voters casting each distinct ballot (where $\sum \mathbf{v}_{-i} = 1$), and let $w(b_i, \mathbf{v}_{-i}) \in \{c_1, c_2, \ldots, c_K\}$ denote the winning candidate when i submits b_i and others' votes are given by $\mathbf{v}_{-i}$. Finally, let $f(\mathbf{v}_{-i})$ denote beliefs about election outcomes, specified as a probability mass function defined over the sample space of $\mathbf{v}_{-i}$. Then i's expected utility from casting ballot b_i is

$$\overline{u}_i(b_i) \equiv \sum_{\mathbf{v}_{-i}} u_i\left(w\left(b_i, \mathbf{v}_{-i}\right)\right) f(\mathbf{v}_{-i}) \tag{1}$$

and voting strategically means choosing $b_i^* = \arg\max_{b_i} \overline{u}_i(b_i)$.

2.2 Susceptibility to strategic voting

We will say that a voting system is susceptible to strategic voting to the extent that it puts voters in circumstances where, given information available before the election, their expected utility from strategic voting is higher than their expected utility from sincere voting. Compared to other definitions based on the possibility of *ex post* manipulation, this definition speaks more directly to concerns that voters might submit misaligned ballots or exert effort determining whether they should do so.

Let $\delta_i \equiv \overline{u}_i(b_i^*) - \overline{u}_i(b_i^{\text{sincere}})$ denote the gain in expected utility voter i receives from voting strategically instead of sincerely (where b_i^* and b_i^{sincere} denote a strategic vote and sincere vote, respectively). This gain is zero if $b_i^* = b_i^{\text{sincere}}$ and positive otherwise. Our main estimand is $E\left[\delta_i\right]$, where expectations are taken over circumstances we might expect voters to face, i.e. combinations of preferences and beliefs that might be observed in a given system. We will also investigate the probability that a misaligned vote is optimal (the "prevalence" of strategic voting incentives) and the expected benefit of voting strategically conditional on a misaligned vote being optimal (the "magnitude" of strategic voting incentives), which can be seen as components of $E\left[\delta_i\right]$ and/or alternative measures of susceptibility to strategic voting.

2.3 Estimating susceptibility to strategic voting

Given this definition of susceptibility to strategic voting, we might operationalize and estimate it in various ways. Ideally we might run a large randomized control trial in which we randomly assign voting systems to a large number of polities and, after allowing time for voters and candidates to respond to their assigned voting system, use a survey to measure voters' preferences and beliefs and compare across systems. More realistically, we could conduct observational studies comparing strategic voting incentives across systems in actual use; in practice, however, it can be difficult to disentangle differences in susceptibility to strategic voting from differences in the polities where different systems are used, particularly for uncommon systems.[9] Susceptibility to strategic voting (and voters' responses to strategic voting incentives) could also be studied in the lab (e.g. Blais et al., 2016; Hix, Hortala-Vallve, and Riambau-Armet, 2017).

Rather than study actual elections (whether as part of an experiment or an observational study), we study hypothetical elections simulated using empirical preference data and a novel model of belief formation. These simulations require stronger assumptions than empirical alternatives, but they offer a practical way to study the properties of even obscure voting methods without intervening in real elections. The next few paragraphs explain our assumptions about preferences and beliefs; Appendix A explains how we compute pivot probabilities from beliefs and compute the expected utility of each ballot for each voter in our surveys.

2.3.1 Preferences from election surveys

In actual elections, voters' preferences over candidates/parties (denoted $u_i(c_j)$ above) reflect the positions parties adopt and the characteristics of party leaders, among other things; these features, as well as the number and type of parties that compete, are well known to be features of the electoral system. A "general equilibrium" comparison of voting systems would thus model the effect of the voting system on (at least) party entry and positioning. Our approach is more "partial equilibrium": we use party ratings from recent electoral surveys to capture typical party preferences in existing systems, and we hold fixed these preferences as we compare alternative

[9] Several studies compare strategic voting in PR and plurality systems (e.g. Bargsted and Kedar, 2009; Abramson et al., 2010). Blais (2004) and Dolez and Laurent (2010) study strategic voting in runoff elections and Farrell and McAllister (2006) discusses instances of apparent *ex post* manipulability in Australian IRV elections. These studies focus on voter behavior rather than voter incentives.

voting rules. Our analysis thus aims to capture how strategic voting incentives differ across systems when we hold fixed the distribution of preferences and thus (implicitly) the parties competing and their strategies. While we do not fully endogenize preferences in this paper, we do extend our baseline analysis by allowing four parties and conditioning on the existing electoral system, which lets us speculate about how second-order effects of the electoral system might influence our conclusions.

More specifically, we measure voter preferences using numerical ratings of parties from 160 national election surveys in 56 unique countries, collected through the Comparative Study of Election Systems (CSES) waves 1-4 (1996-2016).[10] In each survey, respondents are asked to rate each of the main parties on a 0 to 10 scale, where 0 means the respondent "strongly dislikes" that party and 10 means the respondent "strongly likes" that party. In our baseline analysis we retain these numerical ratings for the three largest parties in each survey (based on national vote share) and add a small amount of random noise (so that there is a unique sincere vote for every voter) to form 160 distinct preference distributions, one for each survey. (Green-Armytage (2014) and Eggers and Vivyan (2020) discuss the suitability of party ratings as utility measures.) The average survey has just under 1,400 respondents who rate at least these three parties, for a total of over 220,000 usable respondents across all the surveys. Because the countries in the survey differ widely in population, and some countries have more surveys in the dataset than others, when we combine results across CSES cases we weight by country population and the number of surveys the country contributes to the CSES,[11] thus characterizing incentives for the typical citizen across the countries in the CSES. In further analysis we consider preferences over the largest four parties and subset our results by the existing electoral system, so that we can compare e.g. IRV using preferences over the top four parties found in PR countries to plurality using preferences over the top three parties found in majoritarian countries.

[10]See `http://www.cses.org`. There are 162 election surveys in these four waves, but we exclude Belarus in 2008 and Lithuania in 1997 because they record preferences on only two parties.

[11]Specifically, we weight voter i in country j by $\frac{w_i N_j}{n_j}$, where w_i is the normalized survey weight assigned to respondent i (with $\sum w_i = 1$ in each poll), N_j is country j's population, and n_j is the number of surveys from country j in the dataset.

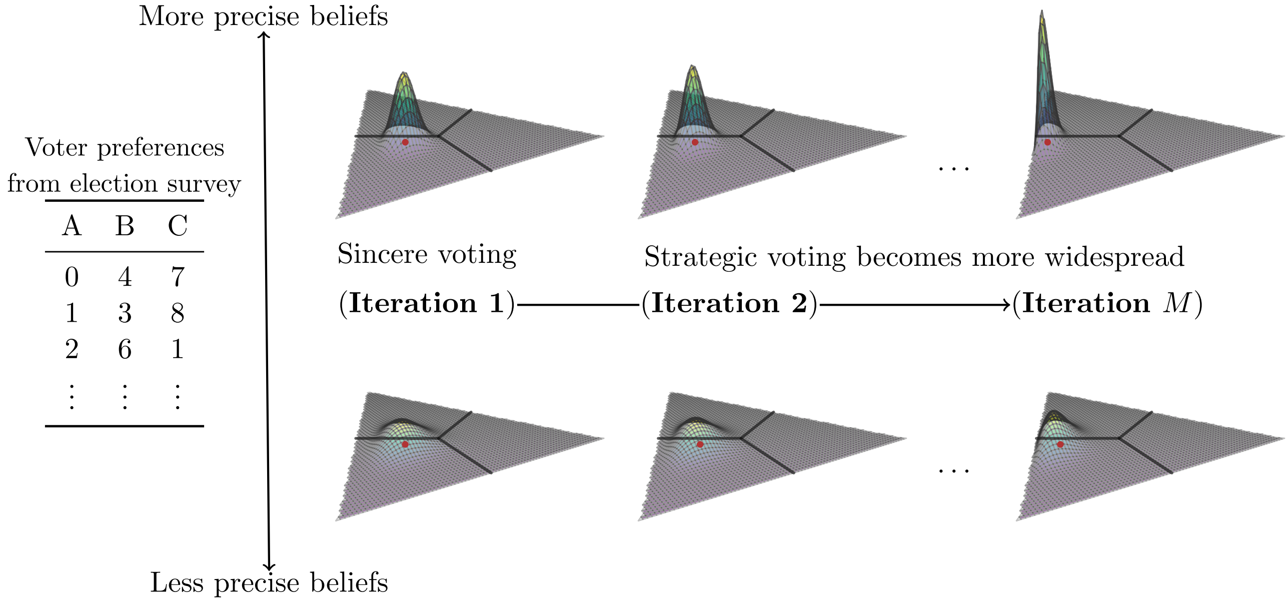


Figure 1: Diagram illustrating the iterative polling algorithm

2.3.2 Beliefs from an iterative polling algorithm

Given a distribution of preferences drawn from an election survey, what beliefs (i.e. distribution over possible election results) would it be reasonable to assume? As noted above, the standard approach in manipulability research is to assume perfect information and sincere voting, which in this case would imply assuming that voters in a given survey know that the election outcome will reflect the sincere distribution of preferences in that survey. A game theorist might instead focus on Nash equilibria given some uncertainty about voter actions or the distribution of types: for plurality, this might imply assuming that voters in a given survey expect almost everyone to vote for just two candidates.

Rather than focusing on beliefs at either the sincere result or an equilibrium result, we conduct our analysis for a range of beliefs between these extremes. Following Fisher and Myatt (2017), we model voters' common beliefs as a Dirichlet distribution, which can be described by two parameters: the location (or expected value) $\overline{\mathbf{v}} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_B)$,[12] with one element for each of B possible ballots, and a scalar precision parameter s. For each set of analysis we choose a value of s from a range informed by recent empirical work: Fisher and Myatt (2017) find that English voters' beliefs are characterized by $s = 10$, while Eggers, Rubenson,

[12] Beliefs relate to others' votes ($\mathbf{v}_{-i}$), but we henceforth omit the subscript to lighten notation.

and Loewen (2021) and Eggers and Vivyan (2020) find that forecasters in Canada and the UK have beliefs characterized by $s = 53$ and $s = 85$, respectively.[13] Each of the ternary plots in Figure 1 illustrates Dirichlet beliefs for a three-candidate plurality election. The red dot in each plot represents the location parameter $\overline{\mathbf{v}}$; the top row of diagrams illustrate more precise beliefs (higher s), while the bottom row of diagrams illustrate less precise beliefs (lower s).

For each election survey, we then trace out a sequence of expected results $\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots$ using a novel iterative polling algorithm as follows.[14] The first expected result $\overline{\mathbf{v}}_1$ is the sincere voting result, i.e. the proportion of respondents in the survey who would cast each ballot if voting sincerely. Each subsequent expected result $\overline{\mathbf{v}}_m$ is a weighted average of the previous result ($\overline{\mathbf{v}}_{m-1}$) and voters' best response to beliefs centered at the previous result (given belief precision s), with the weight on voters' best response given by a mixing parameter λ. Figure 1 illustrates this progression of beliefs from one iteration to the next given different assumptions about uncertainty. The sequence may converge on a fixed point, which can be considered a strategic voting equilibrium. We compute each system's susceptibility to strategic voting for beliefs centered at each step along the sequence, which allows readers to assess each voting system under a range of assumptions about the prevalence of strategic behavior.

The iterative polling algorithm can be interpreted in various ways. It can been be seen as a model of how (expected) election results change over a series of polls or elections, where we assume that voters are inattentive (in that only a fraction λ best-respond at each poll/election, with others sticking to their previous choice) and myopic (in that they do not consider how their action affects future polls/elections); in this view, strategic incentives we measure at lower or higher iterations indicate how incentives might evolve over a campaign or over several elections. (We suspect that observed election results tend to be found somewhere along the sequence produced by the algorithm; we leave for another paper the task of testing that conjecture and exploring the properties of the algorithm more fully.) It could also be seen as a model of levels

[13] As Fisher and Myatt (2017) point out, an observer with an uninformative Dirichlet prior over vote shares who observes a random sample of s voting intentions has Dirichlet posterior beliefs with precision s; thus s can be seen as the size of the poll that informs voter beliefs.

[14] Computer scientists have also written on "iterative voting", which (like our iterative polling algorithm) refers to a procedure in which agents first vote sincerely and then make myopic strategic adjustments (see Meir, 2018, for a review). The difference is that our agents adjust their voting intention based on imprecise beliefs centered at previous poll results, whereas their agents know exactly how others have voted and adjust their votes to achieve a (known) better result.

of rationality (Stahl and Wilson, 1994), where the kth iteration captures the beliefs of level-k voters who assume a distribution of level $1, \ldots, k-1$ voters governed by λ;[15] in that view, the strategic incentives we measure at lower or higher iterations are those that would be perceived by voters with lower or higher degrees of strategic sophistication. More simply, it could be seen as method of locating a strategic voting equilibrium, which is useful in systems (like IRV) where equilibrium has not yet been characterized and does not require subscribing to any particular theory of how voters respond to polls. This method of finding an equilibrium implicitly contains an equilibrium refinement: it selects an equilibrium that can be reached by a process of iterative best responses starting at the sincere result. Several authors have used the idea of iterative best responses to justify refinements to strategic voting equilibria (e.g. Palfrey and Rosenthal, 1991; Fey, 1997; Myatt, 2007);[16] we instead apply iterative best responses as an analytical tool with results subject to various interpretations.

3 Strategic voting in plurality and IRV elections

To this point we have described a general approach to measuring susceptibility to strategic voting. Before applying this approach to plurality and IRV elections, we briefly discuss the qualitative nature of strategic voting in each system.

To begin with, note that rather than considering all possible election outcomes, a strategic voter can focus on *pivot events* (Myerson and Weber, 1993), i.e. situations where a single vote can determine the winner.[17] That is, the ballot that maximizes expression 1 when we sum over all possible outcomes is the same as the ballot that maximizes a version of expression 1 where we sum over only pivot events. Similarly, the gain from strategic voting compared to sincere voting δ_i is the same whether we compute expected utility over all possible outcomes or only over pivot events.[18] To understand strategic voting in plurality and IRV we will therefore focus on pivot events.

[15] For example, the first iteration captures the beliefs of strategic voters who know the distribution of preferences and believe other voters are not strategic.

[16] These authors focus on expectationally stable equilibria as defined by Palfrey and Rosenthal (1991). Our algorithm will locate a globally expectationally stable equilibrium if there is one.

[17] In terms of the notation above, pivot events describe $\{\mathbf{v}_{-i} : w(\mathbf{v}_{-i}, b_i) \neq w(\mathbf{v}_{-i}, b_i')\}$ for some b_i, b_i'.

[18] We describe how to calculate the probability of any pivot event in IRV and plurality in Appendix A.2.

In a three-candidate plurality election, the relevant pivot events are the three possible ties for first.[19] A strategic voter whose preference over candidates is $A \succ B \succ C$ votes B if the probability of a B-C tie for first is sufficiently high relative to an A-B or A-C tie for first. For reasonable specifications of beliefs, a candidate expected to finish third or lower is less likely to be involved in a tie for first than a candidate expected to finish first or second (Fisher and Myatt, 2017); thus strategic voters tend to abandon trailing candidates, producing Duvergerian results.

Now consider a three-candidate IRV election: the candidate who receives the fewest first-place votes is eliminated, and the winner is the remaining candidate who is ranked higher on the majority of all ballots (including those that ranked the eliminated candidate first). In such an election there are twelve relevant pivot events to consider. There are three pairs of candidates who, having not been eliminated in the "first round", could be tied in the final tally, such that a single ballot could determine the outcome; each of these might be called a "second-round" pivot event. (For example, it could be that C receives the fewest first-place votes and is eliminated, and A is ranked higher than B on exactly half of all ballots.) Second-round pivot events never reward misaligned votes: if the winner will be candidate A or B, one cannot do better than submit a sincere ordering of those two candidates. Then there are nine "first-round" pivot events in which a pair of candidates ties for second in top rankings, with the identity of the ultimate winner depending on which one is eliminated. (For example, B and C could be tied for second in first-place votes, and A would win if C were eliminated but C would win if B were eliminated.) First-round pivot events do reward misaligned votes. At the pivot event just mentioned, for example, a voter with preference order $A \succ B \succ C$ or $A \succ C \succ B$ could elect A instead of C by ranking the candidates BAC;[20] a voter with preference order $B \succ C \succ A$ could elect C instead of A by ranking the candidates CBA. Note, however, that each of these tactics could backfire if another pivot event took place, including a second-round pivot event.

[19] We could further distinguish between the events where A wins one more vote than B, A and B win the same number of votes, and B wins one more vote than A; for simplicity, suppose that alphabetical order is used to determine the winner if two candidates win the same number of votes, and assume that it is equally likely that A and B have the same number of votes and A leads B by one vote. Then we can refer to a single A-B pivot event.

[20] These are illustrations of non-monotonicity, a characteristic of all runoff systems. The tactic is sometimes referred to as voting for a "pushover" (Bouton and Gratton, 2015) or a "turkey" (Cox, 1997).

While there are six ways to rank three candidates, a strategic voter in a three-candidate IRV election optimally chooses among only three possible ballots: for example, a voter with preference order $A \succ B \succ C$ chooses among ABC, BAC, and CAB. To see why, note that there are first-round pivot events that reward ranking one's second or third choice first (examples appear in the previous paragraph), but the ranking of the other two candidates could only matter in a second-round pivot event, which never rewards a misaligned vote.

We may shed additional light on strategic voting in IRV by comparing it to strategic voting in a conventional runoff election, which has been more extensively studied (e.g. Bouton, 2013; Ornstein and Norman, 2014; Bouton and Gratton, 2015). The two systems are similar in many respects related to strategic voting, including the possibility of helping one candidate win by ranking another candidate higher. There are two main differences. First, in conventional runoff elections a voter could vote for candidate B in the first round and then, after A and B qualify for the second round, switch to A; this might be attractive if, by voting for B, the voter could prevent C (a more dangerous opponent for A) from advancing. In IRV, by contrast, the same voter can help B advance by submitting a BAC ballot, but this effectively commits the voter to supporting B over A in the second round as well (where the tactic might backfire). Second, in conventional runoff elections voters have an incentive to help a leading candidate secure a majority in the first round; this leads to Duvergerian equilibria in Bouton (2013), with only two candidates winning first-round votes. In an IRV election, where it is not possible to lose support from one round to the next, this incentive is irrelevant. If candidate A is one vote short of a first-round majority, then an additional first-place vote could affect the outcome only if supporters of the third-place candidate *never* rank A second; but then A is also one vote short of a second-round majority, so the circumstance is subsumed in a second-round pivot event.

4 Analysis

4.1 Iterative polling algorithm

We begin by describing the results of the iterative polling algorithm, which provides the sequence of expected results that forms the basis of beliefs in our main analysis. When not otherwise

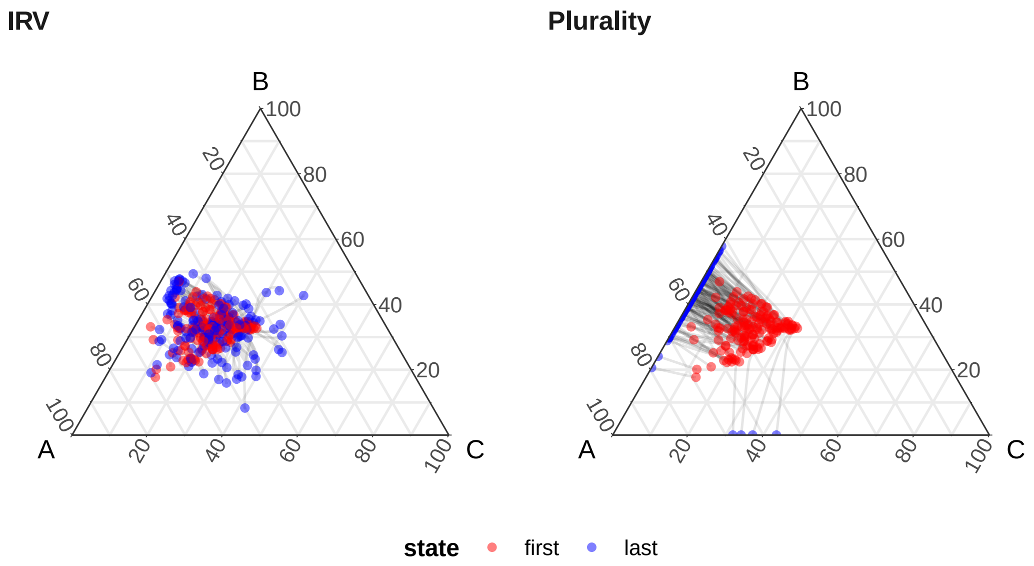


Figure 2: Evolution of ballot share vectors for all 160 CSES election surveys for both IRV (left) and plurality (right). Red dots indicate the first hypothetical poll result, blue dots indicate the 250th hypothetical poll result

specified we use $s = 85$ (the level of precision associated with UK election forecasters by Eggers and Vivyan (2020)) and mixing parameter $\lambda = .05$, but we present comparable results for other parameters in the appendix and our main results with other levels of precision in Figure 4. A key insight is that the algorithm's ballot shares converge towards a fixed point in both plurality and IRV; while this is well known and expected in plurality, the algorithm's results in IRV suggest the existence of strategic voting equilibria in IRV, which (to our knowledge) have not yet been documented. Appendix C contains results for other parameter values; we discuss these robustness checks at many points below.

Figure 2 uses a ternary diagram to represent the share of first-preference votes in IRV (left) and plurality (right) in the first hypothetical poll (red dots), i.e. the sincere profile, and the 250th hypothetical poll (blue dots); a gray line traces the intervening polls. In each CSES case we have labeled the parties such that A has the largest share of top rankings and B the second highest; the results of the first poll are therefore all in the lower left corner of the ternary diagram.

In plurality, the iterative polling algorithm traces a path directly from the sincere profile to a Duvergerian equilibrium in which two parties receive essentially all the votes. In almost all cases, the two parties receiving votes are the ones receiving the most sincere preferences (A and B). (The few exceptions were cases where B and C started off nearly tied in sincere preferences and a substantial proportion of voters abandoned B for A, such that B trailed C after a few iterations and subsequently lost all support.) In IRV, by contrast, the iterative polling algorithm in all cases converges on an "interior" point, i.e. one where all candidates receive some first-preference support. Appendix C confirms that convergence takes place in all CSES cases in both systems.

In plurality, the precision parameter s affects how quickly the algorithm converges to a Duvergerian equilibrium: higher belief precision makes a vote for the trailing candidate more obviously ineffective and thus speeds desertion of this candidate. (See Figures A.4 and A.5 in the Appendix.) In IRV, by contrast, s noticeably affects the location of the equilibrium (as shown by Figures A.14 and A.15). The reason for this difference is that, within the range of s we consider and for results in the neighborhood of the equilibria we find, the choice of s has a larger impact on relative pivot probabilities in IRV than in plurality. Near a Duvergerian equilibrium in plurality, a single pivot event (a tie between the two frontrunners) is much more likely than the others; s affects how much more likely, but it cannot reverse the order of pivot events, so the same result (roughly) is an equilibrium at a range of s. Near IRV equilibria, by contrast, several pivot probabilities are relevant, and the choice of s can affect not only the relative magnitudes of the pivot probabilities but also their rank ordering. Thus a vote share vector $\overline{\mathbf{v}}$ that is an IRV equilibrium at one value of s will not be an IRV equilibrium at another value of s.

The multiplicity of equilibria in plurality is well known, and can be illustrated with our algorithm: if we replace the starting profile with a result in which candidates B and C are clearly in the lead, for example, we always end up at the equilibrium where those two candidates win almost all votes. In IRV, by contrast, we find that (for a given value of s) the algorithm converges toward the same point regardless of the starting point, suggesting a unique equilibrium. Figure 3 illustrates this for four CSES cases. In each plot, each red dot indicates a (randomly chosen) starting point, each gray line traces the path of the algorithm, and each blue dot indicates the endpoint after 250 iterations. In each of these CSES cases all of the paths appear to lead to the same point.[21] The apparent uniqueness of equilibrium in IRV across 160 cases strongly suggests the uniqueness of equilibria in IRV more generally and deserves more study.

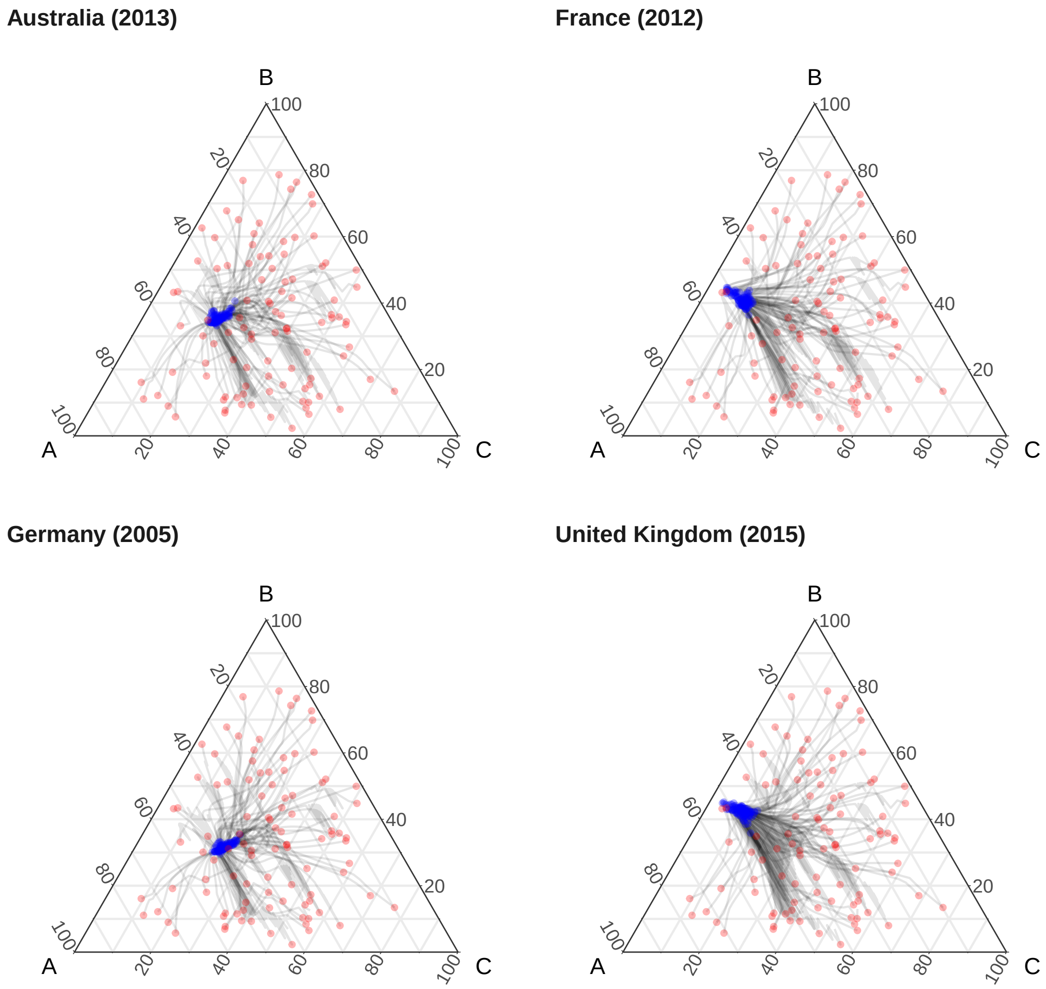


Figure 3: Path of first-preference shares along iterative polling algorithm for select cases in IRV from random starting points (in red) to 60th iteration (blue).

[21] Figure A.16 in the Appendix shows for all CSES cases that all paths from random starting points are still converging toward the fixed point we locate when starting from the sincere profile.

4.2 Susceptibility to strategic voting

Our iterative polling algorithm yields a sequence of expected results $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M$ given precision s and mixing parameter λ for each election survey in the CSES. At each stage of each sequence, we compute δ_i (the expected utility benefit from strategic voting vs. sincere voting) for each voter; averaging this across voters gives us a measure of each system's susceptibility to strategic voting.

In Figure 4 we summarize δ_i across our 220,000 CSES respondents in plurality and IRV in three different ways. The plots in the left column focus on $E[\delta_i]$ computed within CSES cases (thin lines) and across all CSES respondents (thick lines, weighted as described in footnote 11) separately for plurality (orange) and IRV (blue) at each of the first 60 iterations[22] of the polling algorithm (horizontal axis) and different values of the precision parameter s ($s = 10$ on top, $s = 55$ in middle, $s = 85$ at bottom). Note that δ_i is measured in the units of the CSES party ratings (where 0 is "strongly dislike" and 10 is "strongly like") multiplied by the assumed size of the electorate; an expected benefit of .4 in an electorate of 1 million, for example, indicates that the average voter expects to be .4/1,000,000 points (on the 0-10 scale) more pleased with the winner if she were to switch from sincere voting to strategic voting.

The clear conclusion is that IRV is less susceptible to strategic voting on average than plurality, in the sense that it creates smaller average incentives to vote strategically. At $s = 10$ (approximately the level of belief precision Fisher and Myatt (2017) ascribe to UK voters), the average incentive to vote strategically is low for both systems at beliefs close to the sincere profile (i.e. to the left of the diagram), but as voters respond strategically to polls the benefit of strategic voting in plurality increases while the benefit in IRV decreases further. At $s = 85$ the difference in expected benefit is marked even at the sincere profile, grows as voters respond strategically to polls over the first several iterations, and then remains flat with further iterations. More specifically, at the first iteration the expected benefit of strategic voting in plurality is around 5 times higher than in IRV for $s = 85$ and 20 times higher for $s = 10$; by the 60th iteration the ratio ranges from around 30 (for $s = 85$) to around 70 (for $s = 10$).

[22] There is less change after the first 60 iterations; the results for all 250 iterations at different values of s appear in Appendix D.

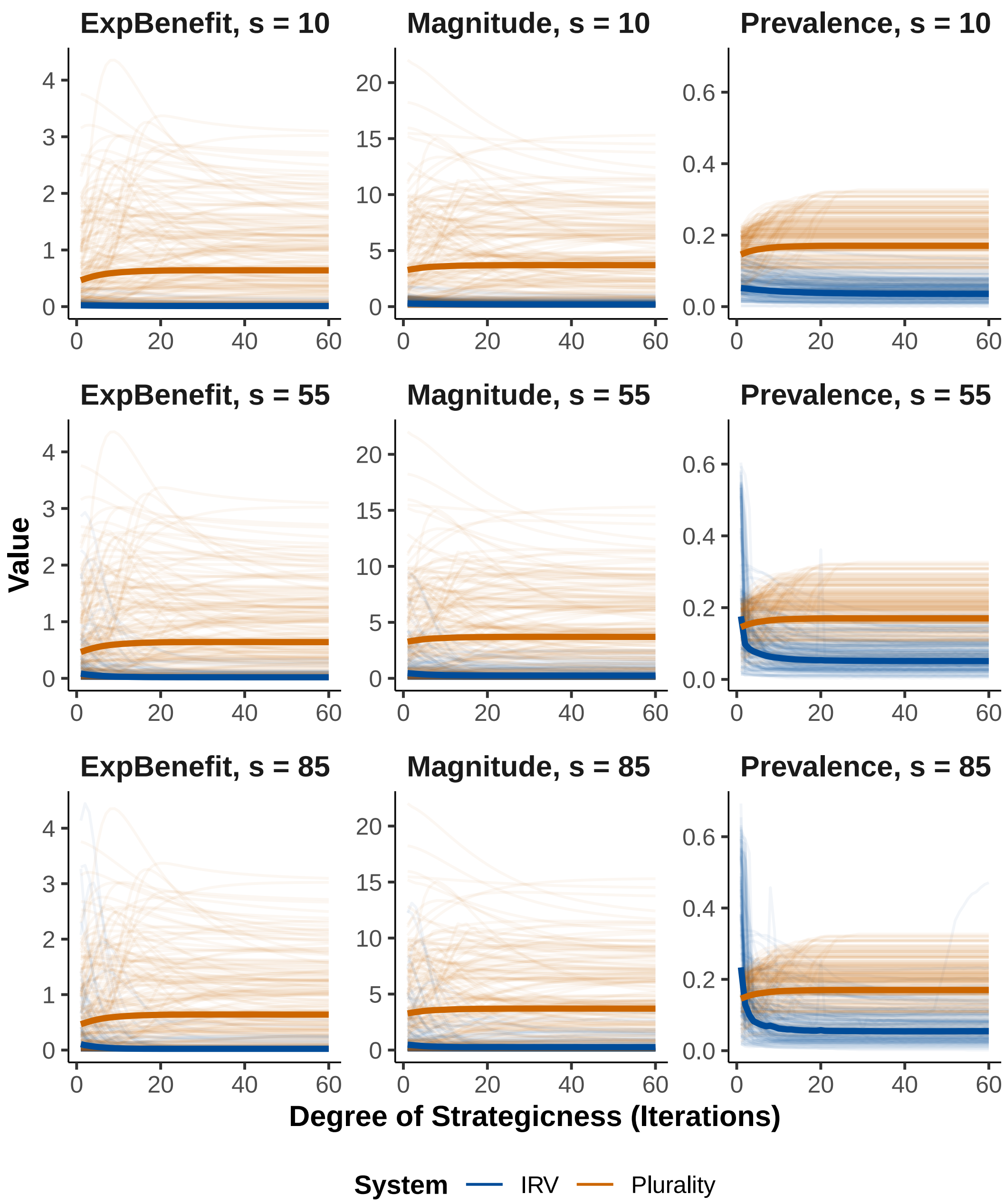


Figure 4: Expected benefit, magnitude, and prevalence of strategic voting at three values of s (belief precision)

Although the size of the difference depends on parameters and assumptions, the conclusion that IRV is less susceptible to strategic voting on average holds across parameters. Notably, across all iterations and choices of s, the expected benefit of strategic voting for plurality in the best case (iteration 1, $s = 10$) is substantially higher than the expected benefit of strategic voting for IRV in the worst case (iteration 1, $s = 85$). Furthermore, sensible departures from our assumptions about belief formation only exacerbate the average difference between plurality and IRV: putting some weight on other Duvergerian equilibria in plurality (i.e. those where voters abandon one of the more sincerely popular candidates) would only increase the share of voters who optimally cast a misaligned vote, making plurality appear even more susceptible to strategic voting.

To better understand these differences in susceptibility to strategic voting, we decompose our measure into the *magnitude* of the benefit (i.e. how much benefit is there for voters who would benefit from a misaligned vote?) and the *prevalence* of the benefit (i.e. what proportion of voters would benefit from a misaligned vote?). More formally,

$$\underbrace{E[\delta_i]}_{\substack{\text{Expected}\\ \text{benefit}}} = \underbrace{E[\delta_i \mid \delta_i > 0]}_{\text{Magnitude}} \times \underbrace{E[\mathbb{1}\{\delta_i > 0\}]}_{\text{Prevalence}}.$$

Thus the magnitude corresponds to the intensive margin of δ_i and the prevalence corresponds to the extensive margin of δ_i.

The plots in the second and third columns of Figure 4 show magnitude and prevalence across plurality and IRV for each level of belief precision. The plots indicate that magnitude and prevalence both play a role in producing the overall differences we observe: voters who expect to benefit from a misaligned vote do so by less on average in IRV (magnitude) and there are usually fewer voters who expect to benefit from a misaligned vote in IRV (prevalence). Note, however, that in IRV the prevalence near the sincere profile is fairly high: around one-fifth of voters optimally submit a misaligned vote at $s = 55$ and $s = 85$ in the first few iterations, which is higher than the equilibrium prevalence in plurality. (The magnitude in both cases is much lower in IRV than in plurality.) Thus when beliefs are precise and other voters are expected to vote insincerely, one is more likely to detect an opportunity to benefit from strategic voting in

IRV on average, even if the size of that opportunity is small.

Our analysis so far has suggested that IRV is less susceptible to strategic voting on average across all CSES respondents, but this leaves open the possibility that IRV might be more susceptible to strategic voting for some preference profiles. In Figure 5 we compare susceptibility to strategic voting in the two voting systems for each CSES survey at iteration 1 (top row) and iteration 60 (bottom row). At iteration 1, the expected benefit of strategic voting is lower in IRV for a large majority of cases, and the same is true for magnitude. For prevalence the picture is more mixed: in most cases no more than 1/5 of voters optimally cast a misaligned vote in either system, but there is a substantial minority of cases where many more voters would optimally cast a misaligned vote in IRV. (These are cases where many voters who favor a leading candidate optimally rank that candidate second in IRV.) At iteration 60, however, there are essentially no CSES cases where any measure of susceptibility to strategic voting is higher in IRV than in plurality. Figures A.23 and A.24 in the Appendix show a similar story for $s = 10$ and $s = 55$, respectively, except that the proportion of cases where IRV is more susceptible to strategic voting than plurality by any measure is lower at the first iteration when beliefs are less precise. Thus the conclusion from disaggregated analysis mirrors that from the aggregated analysis: more voters may benefit from strategic voting in IRV when beliefs are precise and others are expected to vote sincerely, but by other measures and in other circumstances IRV is less susceptible to strategic voting than plurality.

4.3 Feedback mechanisms and strategic voting in plurality and IRV

The results in Figure 4 suggest that strategic voting incentives in plurality and IRV depend differently on expectations about others' strategic behavior: strategic voting incentives are highest in IRV when voters expect others to vote sincerely but they are highest in plurality when voters expect others to vote strategically. Why is this the case, and what does it suggest about the likely prevalence of strategic voting in these systems?

The bandwagon logic of strategic voting in plurality is well-understood (e.g. Cox, 1997). If a given candidate is trailing in a poll, then this candidate will trail by even more when other voters respond to the poll; thus if my best naive response to the poll is to abandon a

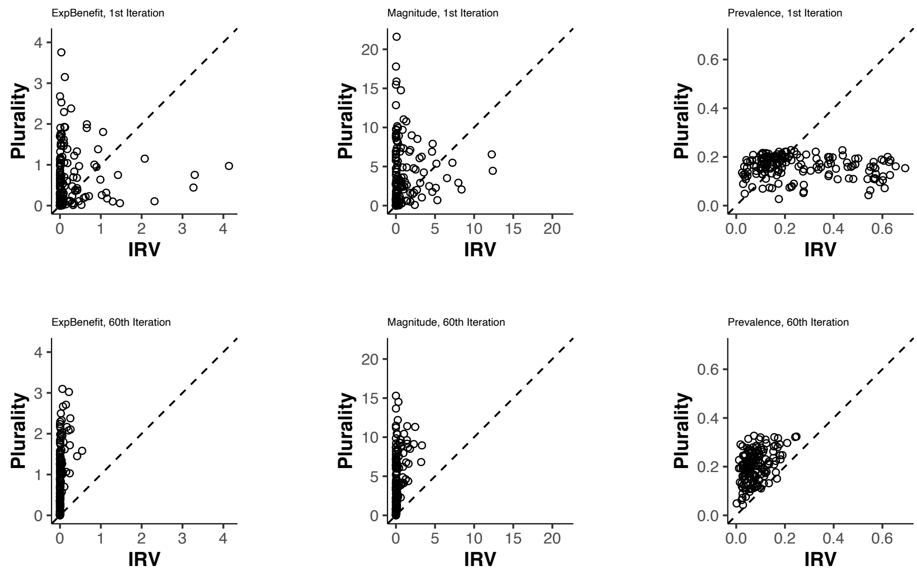


Figure 5: Expected benefit, magnitude, and prevalence of strategic voting in plurality (vertical axis) vs. IRV (horizontal axis), where each dot is one CSES survey ($s = 85, \lambda = 0.05$).

candidate in favor of my second choice, the incentive to do so is likely to be even larger when I take into consideration other voters' responses to the poll.[23] Strategic voting in plurality is thus characterized by positive feedback: strategic responses to a particular pattern of expected results (i.e., one candidate trailing the others) tend to exacerbate that pattern of expected results, inducing more of the same strategic response.[24]

Strategic voting in IRV, by contrast, has a stronger tendency toward negative feedback. To see why, consider a simple example involving three candidates (A, B, and C), with A expected to get the most top ratings; preferences are expected to be mostly single-peaked with B as the centrist, so that B is ranked second on most ballots ranking A or C first. Some of A's supporters may then strategically rank C first, reasoning that C would be an easier second-round opponent (a "turkey" or "pushover") than B would: if B were eliminated her support would be divided between A and C, likely allowing A to maintain her lead, whereas if C were eliminated most of her support would go to B, possibly pushing B ahead of A. But if some A

[23]If others' desertions widen the expected margin between the top candidates, the effect is ambiguous.

[24]Myatt (2007) emphasizes the role of negative feedback in a model where voters receive private signals about which candidate is stronger; feedback here refers to the degree to which voters' respond to their own private signals. When public signals are added the bandwagon logic once again dominates. Bouton, Castanheira, and Llorente-Saguer (2017) highlight negative feedback in a case where aggregate uncertainty produces a bimodal belief distribution.

supporters strategically rank C first, A's expected lead over the others narrows, which makes further such desertions less appealing; also, because these desertions increase the share of ballots that rank C first and A second, the ballots become less single-peaked and there is less advantage to A from facing C instead of B in the second round (conditional on a first-round tie between those two candidates). This example highlights the more general tendency toward negative feedback in IRV: strategic responses to a particular pattern of expected results (here, single-peaked preferences with a non-centrist candidate leading) tend to *neutralize* that pattern of expected results, discouraging more of the same strategic response.[25]

This tendency toward negative feedback provides a new perspective on Dummett (1984)'s comment, cited in the introduction, that a well-informed voter could have nearly as much incentive to cast a misaligned vote in IRV as in plurality. In light of our analysis, it is true that a strategic voting enthusiast who assumes she is the only strategic actor may often recognize chances to cast a misaligned vote in IRV. But our analysis also indicates that these opportunities likely remain limited to a small proportion of the electorate, because the incentive to strategically respond to a poll all but disappear when one perceives that others will do so as well.

4.4 Taking into account second-order impacts of the voting method

So far we have studied strategic voting incentives in plurality and IRV for a fixed common set of preference distributions. The voting system may, however, affect entry decisions, policy positions, and other choices made by political actors. In this section we carry out additional analysis to provide a tentative sense of how these second-order impacts may affect our conclusions. Throughout this section we focus on the case where $s = 85$ and $\lambda = 0.05$; other results appear in Appendix E.[26]

To investigate how our conclusions might change if we took into account effects of the voting system on preferences, Figure 6 shows susceptibility to strategic voting separately according to the existing electoral system in the country. Within each subset of countries (those with

[25] Because positive feedback is also possible in IRV (e.g. if B is expected to finish last in the first round and some of C's supporters strategically rank B first to avoid electing A, making it more likely that B and C tie for second), the tendency toward negative feedback is not an intrinsic property of the voting system; it also depends on preferences and beliefs.

[26] In the figures below, we report weighted averages across all cases belonging to the same group.

plurality/majoritarian elections, PR elections, or ranked systems as in Australia and Ireland) the overall pattern holds. Moreover, the overall pattern obtains when we compare plurality given preferences from plurality countries to IRV given preferences from PR or ranked countries, although the difference in expected benefit of strategic voting is smaller.

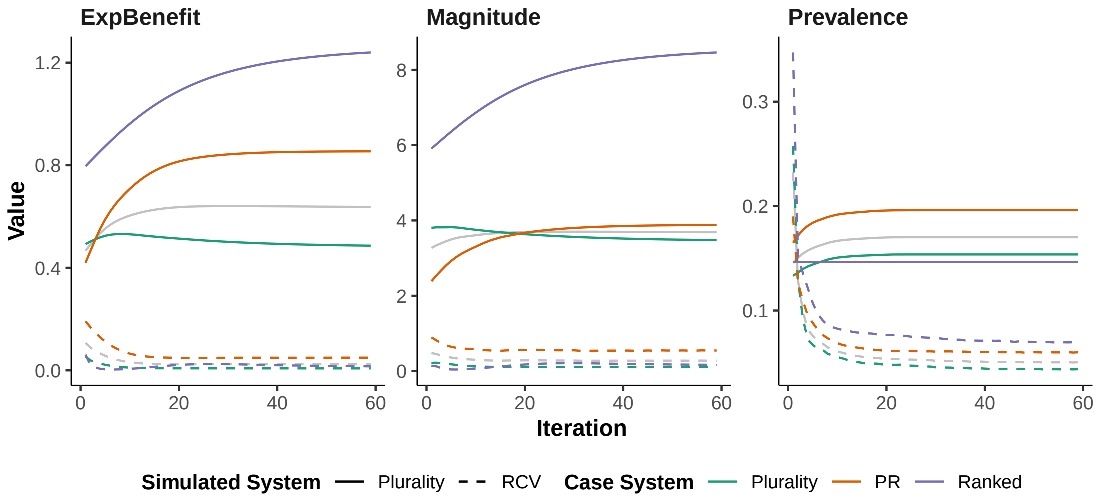


Figure 6: Comparison of expected benefit, magnitude and prevalence of strategic voting in plurality and IRV by cases' actual electoral system ($s = 85$, $\lambda = 0.05$). The grey line represents the (weighted) grand average across all cases.

To gain a sense of how entry might affect our conclusions, Figure 7 presents results for four parties and compares them with results in the three-party case.[27] The same pattern of results holds when we look at four parties instead of three, but (given that IRV tends to make entry more appealing) the more informative comparison may be four-party IRV vs. three-party plurality. We find that the expected benefit and magnitude measures are much higher in three-party plurality than four-party IRV, but average prevalence is slightly higher in four-party IRV. Thus voters in four-party IRV elections who consider strategy can expect to find more chances to benefit from a misaligned vote than in three-party plurality elections, but the expected benefit of voting strategically remains much lower.

Taken together, these results suggest that if we were to fully take into account the effect of the voting rule on candidate entry and positioning, we would draw the same qualitative conclusions as we do from our baseline analysis that holds these features fixed.

[27]The results are less smooth because we use Monte Carlo simulation to estimate four-party pivot probabilities.

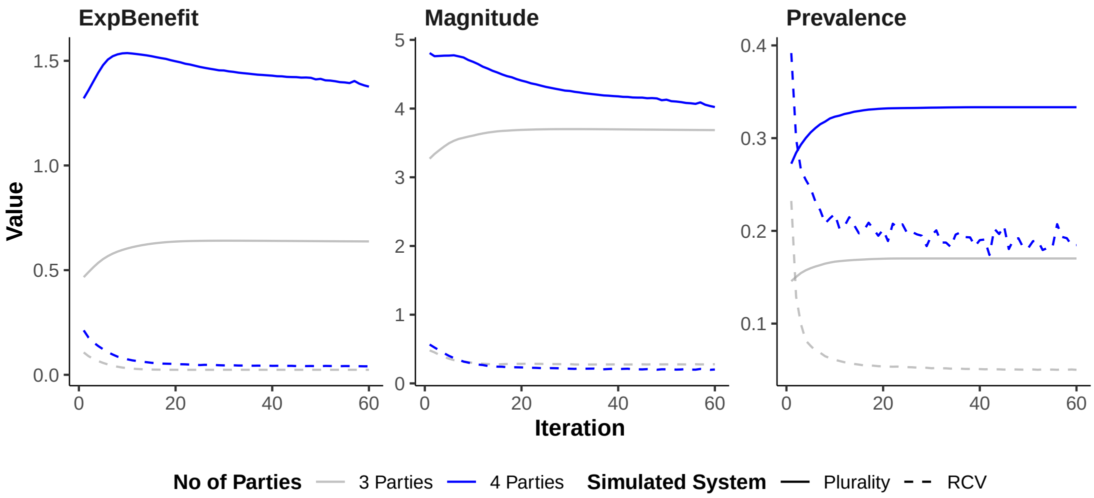


Figure 7: Comparison of expected benefit, magnitude and prevalence of strategic voting in plurality and IRV by number of simulated parties ($s = 85$, $\lambda = 0.05$).

4.5 Strategic voting and election outcomes

So far we have studied voting systems' susceptibility to strategic voting in terms of the incentives for voters to act strategically. Using the same tools we can also ask how strategic voting affects election outcomes and voter welfare. These questions deserve their own extended treatment. Here we offer some key results.

To analyze the impact of strategic voting on election outcomes, we first use Monte Carlo simulation to estimate each candidate's probability of winning at each iteration in each CSES case;[28] we then compare outcomes at the 1st iteration (no strategic voting) and the 60th iteration (widespread strategic voting).

As discussed above, strategic voting in plurality causes the initially trailing candidate's support to collapse; strategic voting in IRV has more complicated effects, with defections away from trailing candidates being counterbalanced in some cases with shifts toward those candidates and negative feedback playing an important role. Consistent with this, we find that strategic voting affects election outcomes more in plurality than in IRV on average. For each case we compute the Euclidean distance between the vector of victory probabilities at the 1st and 60th iteration; the mean (median) distance in plurality is .105 (.041) compared to .048 (.003) in IRV. Plurality is therefore also more susceptible to strategic voting in the sense that its outcomes

[28] We take 1M draws from the belief distribution for each iteration/case.

Table 1: Average Expected Utility and Pr(Condorcet Winner Wins) by Iteration and System

Average Expected Utility

Iteration/System	Plurality	IRV	Δ System
1	5.429	5.431	0.002 (0.301) [0.764]
60	5.436	5.423	-0.013 (-3.795) [0]
Δ Iteration	0.007 (1.287) [0.2]	-0.008 (-3.347) [0.001]	

Pr(Condorcet Winner Wins)

Iteration/System	Plurality	IRV	Δ System
1	0.825	0.826	0.002 (0.173) [0.863]
60	0.838	0.82	-0.018 (-2.085) [0.039]
Δ Iteration	0.013 (1.267) [0.207]	-0.006 (-1.012) [0.313]	

Note: Averages include case weights. Margins report weighted averaged of between-iteration (bottom row) and between-system (last column) differences, along with (t-statistic) and [p-value]

depend more on whether voters are strategic or not.

Turning to voter welfare, strategic voting seems to make voters on average slightly better off in plurality and slightly worse off in IRV, though the impacts are small and the conclusion depends on how we aggregate cases. Table 1 summarizes average expected utility – calculated by combining candidates' win probabilities with voters' ratings of the parties – and the average victory probability of the Condorcet Winner[29] grouped by simulated system (plurality vs. IRV) and the iteration (1st vs. 60). Across both measures, we find no difference across systems at the first iteration; as voters become more strategic, both measures improve in plurality but worsen in IRV, resulting in a statistically significant difference between systems at the 60th iteration. Strategic voting helps voters in plurality on average here because the Condorcet Winner/average utility maximizer is usually among the frontrunners, and strategic voting tends to help that candidate; the effects of strategic voting on IRV outcomes are more mixed.

[29] We refer here to the Condorcet Winner among survey respondents.

5 Conclusion

This paper has introduced a new approach to evaluating voting systems' susceptibility to strategic voting that addresses important shortcomings in previous work. Previous researchers have assessed the manipulability of voting systems by checking how often a voter or group of voters can benefit from a misaligned vote assuming perfect information and sincere voting by other voters. We focus instead on measuring strategic voting incentives as they might be perceived by sophisticated voters or elites in advance of an election, which requires allowing for uncertainty and relaxing the assumption that others vote sincerely. Although our method can be used to measure susceptibility to strategic voting in any electoral system, we focus on the contrast between IRV and plurality, as this has been a salient issue in recent electoral reform proposals in the U.S. and elsewhere. In hypothetical three-candidate elections based on preference data from 160 election surveys from 56 countries, we show that although opportunities for strategic voting may be more widespread in IRV when beliefs are precise and other voters vote sincerely, the average benefit of being strategic is much higher in plurality, especially when beliefs are imprecise and strategic voting widespread. We suggest that IRV is less susceptible to strategic voting on average partly because of negative feedback: typically, the more I expect others to respond strategically (and myopically) to a poll, the less opportunity for me to benefit from a misaligned vote. This contrasts with the well-known bandwagon effect in plurality elections, where the more other voters desert my preferred candidate the more I want to do so.

We anticipate that this paper may strike some theoretically-minded readers as too empirical (because it relies on preference data from a specific set of electoral surveys rather than studying voting system properties in a more abstract way) and some empirically-minded readers as too theoretical (because it studies hypothetical elections rather than real ones). To the charge of being too empirical, we respond that (as manipulability research going back to at least Chamberlin (1985) has recognized) one cannot characterize a voting method's susceptibility to strategic voting without specifying preferences; we have used party ratings from electoral surveys to approximate typical preference arrangements, but others are free to apply our methods using other assumptions about preferences. To the charge of being too theoretical, we emphasize the difficulty of empirically comparing the properties of voting systems that are used in disparate situations (or barely used at all); our approach allows us to vary the system while holding fixed important features, though of course we welcome complementary efforts to study these issues in the lab or with real elections.

We highlight three important caveats to our conclusions. First, susceptibility to strategic voting is only one of many criteria to consider in choosing a voting system; if we ignored other considerations we would choose a dictatorship. Second, we do not fully model the impact of the voting system on candidate entry and positioning (although we do attempt to show how these factors might affect our conclusions), which may leave us with an incomplete view of the systems' effects on strategic voting incentives as they are experienced by voters. Bringing these two points together, our measure of susceptibility to strategic voting should be considered along with measures of how the voting system affects candidate entry, candidate positioning, and other outcomes. Third, we have measured susceptibility to strategic voting assuming that voters have imprecise beliefs but otherwise perfectly understand the voting system and can compute the optimal vote given their beliefs; in practice, voters undoubtedly struggle to comprehend complex strategic possibilities and rely heavily on heuristics (Van der Straeten et al., 2010). Our analysis can therefore be seen as measuring hypothetical opportunities for strategic voting, which voters may discover only with the help of elite cues or voting advice applications.

References

Abramson, Paul R, John H Aldrich, André Blais, Matthew Diamond, Abraham Diskin, Indridi H Indridason, Daniel J Lee, and Renan Levine. 2010. "Comparing strategic voting under FPTP and PR." *Comparative Political Studies* 43(1): 61–90.

Bargsted, Matias A, and Orit Kedar. 2009. "Coalition-targeted Duvergerian voting: how expectations affect voter choice under proportional representation." *American Journal of Political Science* 53(2): 307–323.

Bartholdi, John J, and James B Orlin. 1991. "Single Transferable Vote Resists Strategic Voting." *Social Choice and Welfare* 8(4): 341–354.

Blais, André. 2004. "Strategic Voting in the 2002 French Presidential Election." In *The French Voter*. Springer pp. 93–109.

Blais, André, Jean-François Laslier, Karine Van der Straeten et al. 2016. *Voting Experiments*. Springer.

Bouton, Laurent. 2013. "A Theory of Strategic Voting in Runoff Elections." *American Economic Review* 103(4): 1248–88.

Bouton, Laurent, and Gabriele Gratton. 2015. "Majority Runoff Elections: Strategic Voting and Duverger's Hypothesis." *Theoretical Economics* 10(2): 283–314.

Bouton, Laurent, Micael Castanheira, and Aniol Llorente-Saguer. 2017. "Multicandidate Elections: Aggregate Uncertainty in the Laboratory." *Games and economic behavior* 101: 132–150.

Chamberlin, John R. 1985. "An Investigation into the Relative Manipulability of Four Voting Systems." *Behavioral Science* 30(4): 195–203.

Chamberlin, John R, and Fran Featherston. 1986. "Selecting a Voting System." *The Journal of Politics* 48(2): 347–369.

Conitzer, Vincent, Tuomas Sandholm, and Jérôme Lang. 2007. "When Are Elections with Few Candidates Hard to Manipulate?" *Journal of the ACM (JACM)* 54(3): 14.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge: Cambridge University Press.

Dolez, Bernard, and Annie Laurent. 2010. "Strategic Voting in a Semi-Presidential System with a Two-Ballot Electoral System. The 2007 French Legislative Election." *French Politics* 8(1): 1–20.

Dowding, Keith, and Martin Van Hees. 2008. "In Praise of Manipulation." *British Journal of Political Science* 38(January).

Dummett, Michael. 1984. *Voting Procedures*. Oxford University Press.

Eggers, Andrew C, and Nick Vivyan. 2020. "Who Votes More Strategically?" *American Political Science Review* .

Eggers, Andrew C, Daniel Rubenson, and Peter Loewen. 2021. "Who Votes More Strategically in Canada?" Forthcoming in *Journal of Politics*.

Faliszewski, Piotr, and Ariel D Procaccia. 2010. "AI's War on Manipulation: Are We Winning?" *AI Magazine* 31(4): 53–64.

Farrell, David M, and Ian McAllister. 2006. *The Australian Electoral System: Origins, Variations, and Consequences.* UNSW Press.

Favardin, Pierre, and Dominique Lepelley. 2006. "Some Further Results on the Manipulability of Social Choice Rules." *Social Choice and Welfare* 26(3): 485–509.

Fey, Mark. 1997. "Stability and coordination in Duverger's law: A formal model of preelection polls and strategic voting." *American Political science review* 91(1): 135–147.

Fishburn, Peter C, and Steven J Brams. 1983. "Paradoxes of Preferential Voting." *Mathematics Magazine* 56(4): 207–214.

Fishburn, Peter C, and Steven J Brams. 1984. "Manipulability of Voting by Sincere Truncation of Preferences." *Public Choice* 44(3): 397–410.

Fisher, Stephen D, and D Myatt. 2017. "Strategic Voting in Plurality Rule Elections." *Mimeo* .

Fraenkel, Jon, and Bernard Grofman. 2006. "Does the Alternative Vote Foster Moderation in Ethnically Divided Societies? The Case of Fiji." *Comparative Political Studies* 39(5): 623–651.

Frigyik, Bela A, Amol Kapila, and Maya R Gupta. 2010. Introduction to the Dirichlet Distribution and Related Processes. UWEE Technical Report UWEETR-2010-0006 Department of Electrical Engineering, University of Washington.

Gibbard, Allan. 1973. "Manipulation of Voting Schemes: A General Result." *Econometrica* 41(4): 587–601.

Goldberg, David. 1991. "What Every Computer Scientist Should Know about Floating-Point Arithmetic." *ACM Computing Surveys (CSUR)* 23(1): 5–48.

Green-Armytage, James. 2014. "Strategic Voting and Nomination." *Social Choice and Welfare* 42(1): 111–138.

Green-Armytage, James, T Nicolaus Tideman, and Rafael Cosman. 2016. "Statistical Evaluation of Voting Rules." *Social Choice and Welfare* 46(1): 183–212.

Hamlin, Alan, and Colin Jennings. 2011. "Expressive Political Behaviour: Foundations, Scope and Implications." *British Journal of Political Science* 41(3): 645–670.

Hix, Simon, Rafael Hortala-Vallve, and Guillem Riambau-Armet. 2017. "The Effects of District Magnitude on Voting Behavior." *The Journal of Politics* 79(1): 356–361.

Horowitz, Donald L. 2006. "Strategy Takes a Holiday: Fraenkel and Grofman on the Alternative Vote." *Comparative Political Studies* 39(5): 652–662.

John, Sarah, Haley Smith, and Elizabeth Zack. 2018. "The Alternative Vote: Do Changes in Single-Member Voting Systems Affect Descriptive Representation of Women and Minorities?" *Electoral Studies* 54: 90–102.

Kawai, Kei, and Yasutora Watanabe. 2013. "Inferring Strategic Voting." *American Economic Review* 103(2): 624–62.

Laslier, Jean-François. 2009. "The Leader Rule: A Model of Strategic Approval Voting in a Large Electorate." *Journal of Theoretical Politics* 21(1): 113–136.

McDaniel, Jason. 2018. "Does More Choice Lead to Reduced Racially Polarized Voting? Assessing the Impact of Ranked-Choice Voting in Mayoral Elections." *California Journal of Politics and Policy* 10(2).

McKelvey, Richard D. 1972. "A General Theory of the Calculus of Voting." *Mathematical applications in political science* .

Meir, Reshef. 2018. "Strategic Voting." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13(1): 1–167.

Myatt, David P. 2007. "On the Theory of Strategic Voting." *The Review of Economic Studies* 74(1): 255–281.

Myerson, Roger B. 2002. "Comparison of Scoring Rules in Poisson Voting Games." *Journal of Economic Theory* 103(1): 219–251.

Myerson, Roger B, and Robert J Weber. 1993. "A Theory of Voting Equilibria." *American Political Science Review* 87(1): 102–114.

Ornstein, Joseph T, and Robert Z Norman. 2014. "Frequency of Monotonicity Failure under Instant Runoff Voting: Estimates Based on a Spatial Model of Elections." *Public Choice* 161(1-2): 1–9.

Palfrey, T., and H. Rosenthal. 1991. "Testing game-theoretic models of free riding: new evidence on probability bias and learning." In *Laboratory Research in Political Economy*, ed. Thomas Palfrey. Ann Arbor, Michigan: University of Michican Press.

Pons, Vincent, and Clémence Tricaud. 2018. "Expressive Voting and Its Cost: Evidence from Runoffs with Two or Three Candidates." *Econometrica* 86(5): 1621–1649.

Riker, William H. 1982. *Liberalism against Populism.* San Francisco: WH Freeman.

Saari, Donald G. 1990. "Susceptibility to Manipulation." *Public Choice* 64(1): 21–41.

Satterthwaite, Mark A. 1973. "The Existence of a Strategy-Proof Voting Procedure: A Topic in Social Choice Theory." *Ph. D. Dissertation, University of Wisconsin* .

Satterthwaite, Mark Allen. 1975. "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions." *Journal of economic theory* 10(2): 187–217.

Stahl, II, Dale O, and Paul W Wilson. 1994. "Experimental Evidence on Players' Models of Other Players." *Journal of economic behavior & organization* 25(3): 309–327.

Tideman, Nicolaus. 2018. "How Should Votes Be Cast and Counted?" *The Oxford Handbook of Public Choice* 2: 1.

Van der Straeten, Karine, Jean-François Laslier, Nicolas Sauger, and André Blais. 2010. "Strategic, Sincere, and Heuristic Voting under Four Election Rules: An Experimental Study." *Social Choice and Welfare* 35(3): 435–472.

Appendix

Table of Contents

A Measuring the expected utility of strategic voting

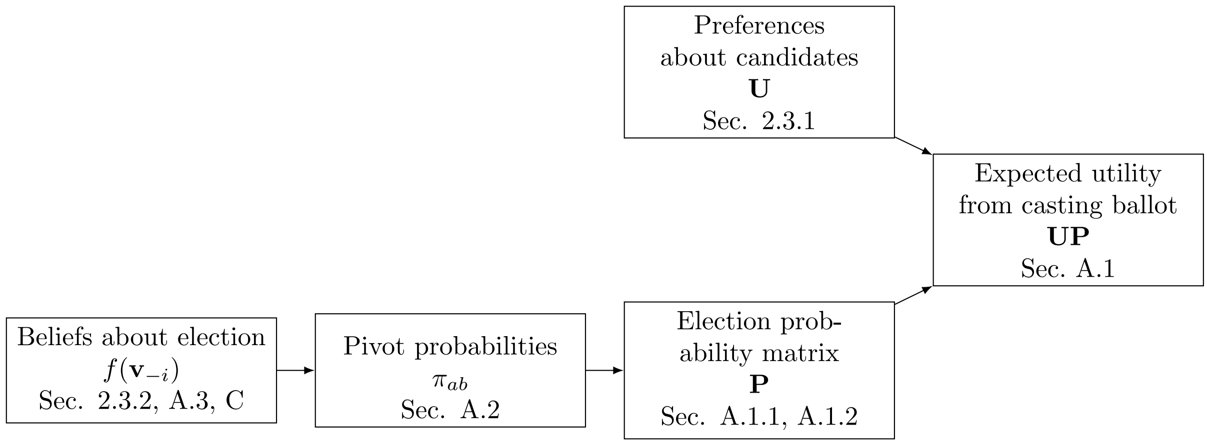


Figure A.1: Steps in calculating voters' expected utility from their preferences and beliefs. The entire process in the figure represents one iteration of our iterative polling algorithm; for the next iteration, beliefs are adjusted accordingly. The cross-references in each box point the reader towards more detail on each step.

In this section, we offer a more technical description of our new method of measuring voters' expected utility from voting strategically. Figure A.1 summarizes the main steps in the process of computing the expected utility of each ballot from preferences and beliefs; each block in the flowchart contains references to sections containing relevant detail for that step.

In what follows, we review the steps in this process roughly in reverse: we show how to compute the expected utility of each ballot in each voting system (assuming known preferences and pivot probabilities); we next show how to compute pivot probabilities given beliefs about likely election outcomes; finally, we describe how we model beliefs about likely election outcomes using in iterative polling algorithm.

A.1 Calculating the expected utility from each possible ballot

Suppose n voters participate in an election to choose a winner from a set of candidates denoted $\mathcal{C}$. We assume that these voters have Von Neumann-Morgenstern utility functions defined over the candidates, with $u_{i,c}$ denoting the utility of voter i from the election of candidate $c \in \mathcal{C}$. We can organize these utilities into a *utility matrix* $\mathbf{U}$ with one row per voter and one column per candidate; for example, given candidates $\{A, B, C\}$, $\mathbf{U}$ is

$$\mathbf{U} = \begin{bmatrix} u_{1A} & u_{1B} & u_{1C} \\ u_{2A} & u_{2B} & u_{2C} \\ \vdots & \vdots & \vdots \\ u_{nA} & u_{nB} & u_{nC} \end{bmatrix}.$$

We also assume that each voter is uncertain about how other voters will vote but all voters share a common belief about the probability of each possible election result, including the election results in which a single ballot could be decisive in various ways, i.e. *pivot events.* Let $\mathcal{B}$ be the set of all permissible ballots (i.e. distinct votes that can be cast) in the voting system. Let $p_{c,b}$ be the probability that candidate c is elected given one additional ballot $b \in \mathcal{B}$ is submitted, and organize these probabilities into an *election probability matrix* $\mathbf{P}$ with one row per candidate and one column per ballot (so its dimensions are $|\mathcal{C}| \times |\mathcal{B}|$). Then the expected utility of each voter from submitting each possible ballot is the *expected utility matrix* $\overline{\mathbf{U}} = \mathbf{UP}$ with n rows (one row per voter) and one column per ballot. From the expected utility matrix we can compute the optimal (strategic) ballot for each of our n voters as well as the difference in each voter's expected utility between casting the optimal ballot and casting a sincere ballot, which is a measure of the voter's strategic voting incentive. Studying strategic voting incentives in any voting system given voters' preferences and beliefs is thus essentially a problem of assembling the election probability matrix $\mathbf{P}$. We now show how to do this in three-candidate plurality and IRV elections given the probability of pivot events (i.e. *pivot probabilities*); later we show how to compute these pivot probabilities given beliefs.

A.1.1 The P matrix in plurality

In plurality, voters submit ballots naming one candidate, so the set of admissible ballots is the set of candidates. Given candidates $\{A, B, C\}$, the $\mathbf{P}$ matrix is

$$\mathbf{P} = \begin{bmatrix} p_{A,A} & p_{A,B} & p_{A,C} \\ p_{B,A} & p_{B,B} & p_{B,C} \\ p_{C,A} & p_{C,B} & p_{C,C} \end{bmatrix},$$

where e.g. $p_{B,A}$ indicates the probability B is elected when one votes for A. Let an election result be written as a vector $\mathbf{v} = (v_A, v_B, v_C)$, with e.g. v_A indicating the share of ballots naming candidate A. For each voter, this is implicitly the result among other voters ($\mathbf{v}_i$ in the main text), though we omit the subscript to lighten notation. We assume throughout that voters consider $\mathbf{v}$ to be a continuous random variable, with a common belief summarized by pdf $f(\mathbf{v})$; letting $\mathbf{v}$ be continuous simplifies the analysis by eliminating the possibility of ties. Given a total electorate of size N,[30] the probability that a single ballot could change the plurality winner from candidate j to candidate i is then

$$\pi_{ij} = \Pr\left(v_j - v_i \in \left(0, N^{-1}\right) \cap v_i > v_k\right). \tag{2}$$

Assuming each candidate is equally likely to finish just ahead of or just behind another candidate (so that $\pi_{ij} = \pi_{ji}$), the diagonal elements of $\mathbf{P}$ in plurality are

$$p_{i,i} = \pi_i + 2(\pi_{ij} + \pi_{ik}), \tag{3}$$

where i, j, and k are distinct candidates. That is, i wins (given an additional vote for i) if i would win in any case (which occurs with probability π_i), if i would finish either slightly behind or slightly ahead of j ($2\pi_{ij}$), or if i would finish either slightly behind or slightly ahead of k ($2\pi_{ik}$). The off-diagonal elements are

$$p_{j,i} = \pi_j + \pi_{jk} \tag{4}$$

where again i, j, and k are distinct candidates. That is, j wins (given an additional vote for i) if j would win in any case (π_j) or if j is slightly ahead of k (π_{jk}).

It will be convenient to work with a normalized version of $\mathbf{P}$ in which we set π_i to 0 for $i \in \{A, B, C\}$, thus ignoring results in which a single ballot could not determine the outcome. In that case $\mathbf{UP}$ produces a normalized (i.e. recentered) measure of expected utility that is sufficient for determining both the optimal ballot and the benefit of strategic voting.[31]

[30] Thus the n voters may be a sample of the larger electorate.
[31] Similarly, Myerson and Weber (1993) focus on the gain in expected utility relative to abstention.

Table 1: Pivot events in IRV

Label	Type	Description	Probability
$ij.2$	Second-round	i and j tie after k is eliminated in 1st round	$\pi_{ij.2} = \Pr\left(v_j + v_{kj} - \frac{1}{2} \in \left(0, N^{-1}\right) \cap v_k < v_i \cap v_k < v_j\right)$
$ij.ik$	First-round	i and j tie for 2nd in 1st round; only i would defeat k	$\pi_{ij.ik} = \Pr\left(v_j - v_i \in \left(0, N^{-1}\right) \cap v_j < v_k \cap v_k + v_{ik} > \frac{1}{2} \cap v_k + v_{jk} < \frac{1}{2}\right)$
$ij.kj$	First-round	i and j tie for 2nd in 1st round; only j would defeat k	$\pi_{ij.kj} = \Pr\left(v_j - v_i \in \left(0, N^{-1}\right) \cap v_j < v_k \cap v_k + v_{ik} < \frac{1}{2} \cap v_k + v_{jk} > \frac{1}{2}\right)$
$ij.ij$	First-round	i and j tie for 2nd in 1st round; both i and j would defeat k	$\pi_{ij.ij} = \Pr\left(v_j - v_i \in \left(0, N^{-1}\right) \cap v_j < v_k \cap v_k + v_{ik} < \frac{1}{2} \cap v_k + v_{jk} < \frac{1}{2}\right)$

Notes: In this table a "tie" indicates that one candidate finishes slightly ahead of the other, such that a single ballot could reverse the order of finish.

A.1.2 The P matrix in IRV

In an IRV election involving three candidates $\{A, B, C\}$, voters submit ballots ranking the candidates, so the admissible ballots are $\{AB, AC, BA, BC, CA, CB\}$ (where ij denotes a ballot that ranks candidate i first, j second, and (implicitly) k third). The $\mathbf{P}$ matrix then looks like

$$\mathbf{P} = \begin{bmatrix} p_{A,AB} & p_{A,AC} & p_{A,BA} & p_{A,BC} & p_{A,CA} & p_{A,CB} \\ p_{B,AB} & p_{B,AC} & p_{B,BA} & p_{B,BC} & p_{B,CA} & p_{B,CB} \\ p_{C,AB} & p_{C,AC} & p_{C,BA} & p_{C,BC} & p_{C,CA} & p_{C,CB} \end{bmatrix}$$

and an election result can be written as $\mathbf{v} = (v_{AB}, v_{AC}, v_{BA}, v_{BC}, v_{CA}, v_{CB})$.

A three-candidate IRV election can be considered to take place in two rounds: in the first round the candidate who receives the fewest first-place votes is eliminated; in the second round the winner is determined based on the ranking of the remaining two candidates on all ballots.[32] There are two classes of pivot events in IRV. In *second-round pivot events*, a single ballot determines who wins the second round. Let $ij.2$ denote the event that a single ballot ranking i above j could change the IRV winner from j to i in the second round, which (again assuming $\mathbf{v}$ is continuous) occurs when j is preferred to i on only slightly more than half of all ballots and k receives fewer top rankings than either i or j. The probability of this pivot event ($\pi_{ij.2}$) appears in the first row of Table 1. In *first-round pivot events* a single ballot determines the winner by determining who advances to the second round. If candidates i and j are essentially tied for second (in top rankings) in the first round, such that a single ballot determines which one advances, then there are three scenarios in which a single ballot could determine the winner: when either candidate (i or j) would defeat k in the second round (event $ij.ij$), when only i would defeat k in the second round (event $ij.ik$), and when only j would defeat k in the second round (event $ij.kj$). These events are described in Table 1, with the associated probability appearing in the final column.

To fill in the $\mathbf{P}$ matrix for IRV using pivot probabilities, we assume again that adjacent pivot events are equally likely: $\pi_{ij.2} = \pi_{ji.2}$ (i.e. each candidate is just as likely to trail as to lead another candidate in the second round) and $\pi_{ij.ik} = \pi_{ji.ki}$, $\pi_{ij.kj} = \pi_{ji.kj}$, and $\pi_{ij.ij} = \pi_{ji.ji}$ (i.e. each candidate is just as likely to trail as to lead another candidate for second in the first

[32]Descriptions of three-candidate IRV often note that the election ends in the first round if one candidate wins a majority of top rankings, but such a candidate would obviously win the second round so this step is superfluous.

round, for each possible way that the first round outcome could determine the winner). Then we have

$$p_{i,ij} = p_{i,ik} = \pi_i + 2(\pi_{ij.2} + \pi_{ik.2} + \pi_{ij.ij} + \pi_{ij.ik} + \pi_{ik.ik} + \pi_{ik.ij}) + \pi_{jk.ik} + \pi_{jk.ji},$$

meaning that i wins (given an additional ballot ranking i first) if i would win in any case (which occurs with probability π_i); if i would finish nearly tied with (i.e. just ahead of or just behind) j or k in the second round ($2(\pi_{ij.2} + \pi_{ik.2})$); if i would finish nearly tied with j or k for second in the first round and would win if it advanced ($2(\pi_{ij.ij} + \pi_{ij.ik} + \pi_{ik.ik} + \pi_{ik.ij})$); or if j and k would nearly tie for second in the first round, only one of them would lose to i in the second round, and that candidate is the one who advances ($\pi_{jk.ik} + \pi_{jk.ji}$). Similarly, we have

$$\begin{aligned} p_{i,jk} &= \pi_i + 2\pi_{jk.ik} + \pi_{ik.ik} + \pi_{ik.ij} \\ p_{i,ji} &= \pi_i + 2(\pi_{ik.2} + \pi_{jk.ik}) + \pi_{ik.ik} + \pi_{ik.ij}. \end{aligned}$$

The first expression states that i wins (given an additional ballot ranking j first and k second) if i would win in any case (which occurs with probability π_i); if j and k would nearly tie for second in the first round and only k would defeat i ($2\pi_{jk.ik}$), so that a ballot of ji ensures i's victory; or if i would finish the first round narrowly ahead of k for second place and would defeat j in the second round ($\pi_{ik.ik} + \pi_{ik.ij}$). The second expression states that i wins (given an additional ballot ranking j first and i second) in all the same situations plus when i would finish nearly tied with k in the second round ($2\pi_{ik.2}$). As explained above, in practice we set π_i to zero for $\{A, B, C\}$, which focuses on pivot events and produces a normalized measure of expected utility.

A.2 Computing pivot probabilities

Next, we show how to compute pivot probabilities in three-candidate plurality and IRV elections in order to be able to compute the election probability matrix $\mathbf{P}$.

In both systems, we assume that the distribution over election outcomes $f(\mathbf{v_{-i}})$ follows a Dirichlet distribution, which implies modeling the distribution of votes as a continuous random variable. One benefit of this assumption is that the probability of an exact tie between two candidates is zero, so we can avoid tedious complications about tie-breaking and making vs. breaking ties. The Dirichlet distribution's properties make it particularly convenient to work with, as we will see when we compute pivot probability for IRV elections.

A.2.1 Pivot Probabilities in Plurality

A single vote can move candidate j ahead of candidate i when $v_i - v_j \in (0, \frac{1}{N})$, where N is the total size of the electorate. Without loss of generality, the probability that a single vote can elect candidate 2 instead of 1 in a three-candidate race is

$$\pi_{12} = \Pr\left(v_1 - v_2 \in \left(0, \frac{1}{N}\right) \cap v_1 > v_3\right)$$

which given $f(\mathbf{v_{-i}})$ and an electorate of size N can be written as

$$\pi_{12} = \int_{\frac{1}{3}}^{\frac{1}{2}} \int_{v_1 - \frac{1}{N}}^{v_1} f(v_1, v_2, 1 - v_1 - v_2)\, dv_2\, dv_1.$$

If $f(\mathbf{v_{-i}})$ is smooth and the electorate N is large, then

$$\pi_{12} \approx \frac{1}{N} \int_{\frac{1}{3}}^{\frac{1}{2}} f(y, y, 1 - 2y)\, dy. \tag{5}$$

(Note that the same approximation would hold for the probability that a single vote can elect candidate 1 instead of 2, so we can collapse these events into a single pivot event π_{12}.) This approximation can be computed by numerical integration methods.[33]

A.2.2 Pivot Probabilities in IRV

To compute the probability of pivot events in IRV given Dirichlet beliefs, we make use of three well-known (Frigyik, Kapila, and Gupta, 2010) properties of the Dirichlet distribution:

Aggregation property: $(v_1, v_2, \ldots, v_i + v_j, \ldots v_B) \sim \mathrm{Dir}\big(\alpha_1, \alpha_2, \ldots, \alpha_i + \alpha_j, \ldots \alpha_B\big)$. (If two of the vote shares are added together to create a new, shorter vector of vote shares, the new vector of vote shares also follows a Dirichlet distribution, where the parameters corresponding to the summed-up vote shares are also summed up.)

Marginal distribution: $v_i \sim \mathrm{Beta}(\alpha_i, \sum_{-i} \alpha)$. (Unconditionally, any particular vote share follows a Beta distribution. This follows from the aggregation property and the observation that a Dirichlet distribution with two parameters is a Beta distribution.)

Conditional distribution: $(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_B \,|\, v_i) \sim (1 - v_i)\mathrm{Dir}(\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_B)$. (Conditional on i receiving share v_i, the remaining shares follow a rescaled Dirichlet distribution in which α_i is removed from the parameter vector.)

[33] Fisher and Myatt (2017) provide an analytical expression for relative pivot probabilities in three-candidate plurality contests given Dirichlet beliefs. Eggers and Vivyan (2020) validate a numerical approximation when there are more than three candidates.

We will use $f(\mathbf{v}; s\overline{\mathbf{v}})$ to indicate the Dirichlet density with parameters $s\overline{\mathbf{v}}$ evaluated at $\mathbf{v}$. Because the Beta density can be seen as a special case of the Dirichlet density, we will use $f(\cdot)$ for both. As in the main text, v_{ab} denotes the share of ballots ranking a first, b second, and (implicitly) c third; with v_{ac}, v_{ba} etc similar; v_a denotes the share of ballots listing a first, i.e. $v_a \equiv v_{ab} + v_{ac}$.

Probability of second-round pivot events: The probability of a trailing b by less than $\frac{1}{N}$ in the second round can be written

$$\Pr\left(v_c < v_a < \frac{1}{2} \cap v_c < v_b < \frac{1}{2} \cap v_a + v_{ca} - \frac{1}{2} \in \left(-\frac{1}{N}, 0\right)\right).$$

This can be factorized as

$$\Pr\left(v_a + v_{ca} - \frac{1}{2} \in \left(-\frac{1}{N}, 0\right)\right) \times \Pr\left(v_c < v_a \cap v_c < v_b \,\middle|\, v_a + v_{ca} - \frac{1}{2} \in \left(-\frac{1}{N}, 0\right)\right). \quad (6)$$

Using the aggregation property, the first term in expression 6 is

$$\int_{-\frac{1}{N}}^{0} \int_{0}^{\frac{1}{2}} f\left(y - x/4, \frac{1}{2} - y - x/4, \frac{1}{2} + x/2; s\overline{v}_a, s\overline{v}_{ca}, s(\overline{v}_b + \overline{v}_{cb})\right) dy\, dx$$

which is approximately

$$\frac{\sqrt{6}}{4}\frac{1}{N}\int_{0}^{\frac{1}{2}} f\left(y, \frac{1}{2} - y, \frac{1}{2}; s\overline{v}_a, s\overline{v}_{ca}, s(\overline{v}_b + \overline{v}_{cb})\right) dy.$$

(The approximation is exact if the density is flat in the immediate neighborhood of second-round ties between a and b.) To understand the leading $\frac{\sqrt{6}}{4}$ term: On the 3-dimensional unit simplex with vertices v_a, v_{ca}, and $1 - v_a - v_{ca}$, draw a line where $v_a + v_{ca} = \frac{1}{2}$; this locus characterizes pairwise ties between a and b and goes through the point $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. Now draw a line parallel to the first but shifted slightly so that it runs through the point $(\frac{1}{4} - \frac{1}{4N}, \frac{1}{4} - \frac{1}{4N}, \frac{1}{2} - \frac{1}{2N})$. Between the two lines is the narrow strip where b wins a pairwise contest between the two candidates but a single ballot could move a ahead of b. The width of this strip is $\sqrt{\left(\frac{1}{4N}\right)^2 + \left(\frac{1}{4N}\right)^2 + \left(\frac{1}{2N}\right)^2} = \frac{\sqrt{6}}{4N}$.

We now turn to the second term in expression 6. Given that $v_a = y$, $v_{ca} = \frac{1}{2} - y$, and $v_b + v_{cb} = \frac{1}{2}$, we note that $v_c < v_a$ implies $v_{cb} < 2y - \frac{1}{2}$ and $v_c < v_b$ implies $v_{cb} < \frac{y}{2}$; comparing the two conditions, note that the former binds when $y < \frac{1}{3}$ and the latter binds otherwise. Next, using all three properties of the Dirichlet noted above and given that $v_a + v_{ca} = \frac{1}{2}$,

$$(v_{cb} \mid v_a + v_{ca}) \sim \frac{1}{2}\text{Beta}\left(s\overline{v}_{cb}, s\overline{v}_b\right), \quad (7)$$

i.e. given that half the ballots list a first or list c first and a second, the proportion listing c first and b second (instead of b first) lies between 0 and 1/2; if we multiply the proportion by two, the result is distributed according to a Beta distribution with parameters $s\overline{v}_{cb}$ and $s\overline{v}_b$. Thus to find the probability that $v_{cb} < 2y - \frac{1}{2}$ (the binding constraint in the second term from expression 6 when $y < 1/3$), we integrate this distribution from 0 to $2y - \frac{1}{2}$; to find the probability that $v_{cb} < \frac{y}{2}$ (the binding constraint in the second term from expression 6 when $y > 1/3$), we integrate this distribution from 0 to $\frac{y}{2}$. Finally note that y (i.e. v_a) cannot be below 1/4; otherwise either a finishes last in first-preference votes or b receives more than half of first-preference votes. Combining all of this, we have

$$\begin{aligned}N\pi_{ab} &\approx \frac{\sqrt{6}}{4}\int_{\frac{1}{4}}^{\frac{1}{3}} f\left(y,\frac{1}{2}-y,\frac{1}{2};s\overline{v}_a,s\overline{v}_{ca},s(\overline{v}_b+\overline{v}_{cb})\right)\int_0^{2y-\frac{1}{2}} f\left(2z,1-2z;s\overline{v}_{cb},s\overline{v}_b\right)dz\,dy\,+\\ &\quad\int_{\frac{1}{3}}^{\frac{1}{2}} f\left(y,\frac{1}{2}-y,\frac{1}{2};s\overline{v}_a,s\overline{v}_{ca},s(\overline{v}_b+\overline{v}_{cb})\right)\int_0^{\frac{y}{2}} f\left(2z,1-2z;s\overline{v}_{cb},s\overline{v}_b\right)dz\,dy. \qquad (8)\end{aligned}$$

Note that the second and fourth densities are evaluated at $(v_{cb}=2z, v_b=1-2z)$ rather than $(v_{cb}=z, v_b=\frac{1}{2}-z)$ because of the $\frac{1}{2}$ in expression 7.

The analysis extends straightforwardly to the two other second-round pivot events by exchanging candidate labels.

Probability of first-round pivot events: First-round pivot event *ab.ab* takes place when *a* ties *b* for second place in first-preference votes and either candidate would win the election if the other were eliminated. Generally, the probability of *ab.ab* is

$$\Pr\Big(v_b-v_a\in\Big(-\frac{1}{2N},\frac{1}{2N}\Big)\cap v_b<v_c\cap v_a<v_c<\frac{1}{2}\cap v_a+v_{ba}>v_c+v_{bc}\cap v_b+v_{ab}>v_c+v_{ac}\Big), \quad (9)$$

which can be factorized as

$$\begin{aligned}&\Pr\Big(v_b-v_a\in\Big(-\frac{1}{2n},\frac{1}{2n}\Big)\cap v_b<v_c\cap v_a<v_c<\frac{1}{2}\Big)\times\\ &\Pr\Big(v_a+v_{ba}>v_c+v_{bc}\cap v_b+v_{ab}>v_c+v_{ac}\,\Big|\,v_b-v_a\in\Big(-\frac{1}{2N},\frac{1}{2N}\Big)\cap v_b<v_c\cap v_a<v_c<\frac{1}{2}\Big).\end{aligned}$$

Using the same approximation as above, the first line is approximately

$$\frac{1}{\sqrt{2}}\frac{1}{N}\int_{\frac{1}{4}}^{\frac{1}{3}} f\Big(z,z,1-2z;s\overline{v}_a,s\overline{v}_b,s\overline{v}_c\Big)dz.$$

Letting $v_a=v_b=z\in\left(\frac{1}{4},\frac{1}{3}\right)$, the second term becomes

$$\Pr\Big(v_{bc}<2z-\frac{1}{2}\cap v_{ac}<2z-\frac{1}{2}\,\Big|\,v_a=v_b=z\Big). \quad (10)$$

and again combining all three properties we have

$$\begin{aligned}(v_{bc}|v_a+v_c) &\sim z\text{Beta}\big(s\overline{v}_{bc},s\overline{v}_{ba}\big)\\ (v_{ac}|v_b+v_c) &\sim z\text{Beta}\big(s\overline{v}_{ac},s\overline{v}_{ab}\big).\end{aligned}$$

Putting together the above, we have

$$\begin{aligned}N\pi_{ab.ab} &\approx \frac{1}{\sqrt{2}}\int_{\frac{1}{4}}^{\frac{1}{3}} f\Big(z,z,1-2z;s\overline{v}_a,s\overline{v}_b,s\overline{v}_c\Big)\times\\ &\quad\int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z},\frac{z-x}{z};s\overline{v}_{bc},s\overline{v}_{ba}\right)dx\times\int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z},\frac{z-x}{z};s\overline{v}_{ac},s\overline{v}_{ab}\right)dx\,dz.(11)\end{aligned}$$

To get the probability of pivotal event *ab.ac* we reverse the last inequality in expression 9 (changing $v_b+v_{ab}>v_c+v_{ac}$ to $v_b+v_{ab}<v_c+v_{ac}$), which means changing the last term in expression 11

from $\int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z}, \frac{z-x}{z}; s\overline{v}_{ac}, s\overline{v}_{ab}\right) dx$ to $1 - \int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z}, \frac{z-x}{z}; s\overline{v}_{ac}, s\overline{v}_{ab}\right) dx$. The analysis extends straightforwardly to all other first-round pivot events by similarly reversing inequalities and/or exchanging candidate labels.

Checking consistency of numerical and simulation-based estimates: To check the validity of the numerical approach and compare the computational burden of the two approaches, we computed pivotal probabilities for 100 scenarios using the two approaches while varying the number of simulation draws. If our numerical approach is correct, the simulation results should converge on our numerical solutions as the number of simulations (and the computational burden of the simulation approach) increases. Below we show that this is the case.

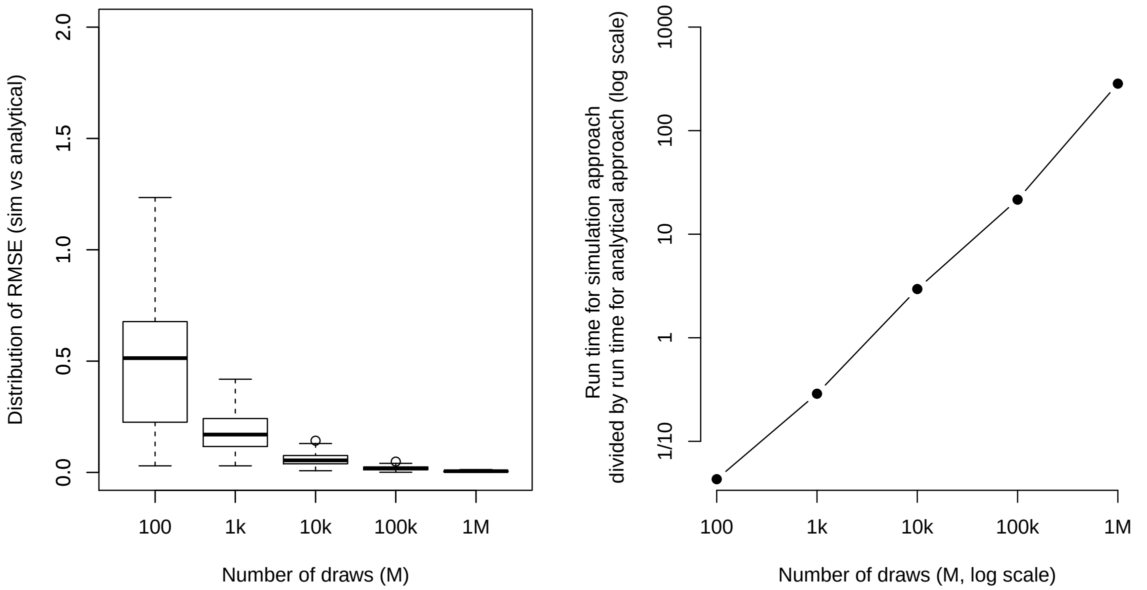
We begin by drawing J sets of Dirichlet parameter values at which we will calculate pivotal probabilities. Specifically, for scenario j we (1) draw a vector $\overline{\mathbf{v}}_j = (\overline{v}_{AB,j}, \overline{v}_{AC,j}, \overline{v}_{BA,j}, \overline{v}_{BC,j}, \overline{v}_{CA,j}, \overline{v}_{CB,j})$ from a Dirichlet distribution with parameters $(6, 4, 5, 5, 4, 6)$ and (2) draw s_j independently from a uniform distribution between 15 and 60. Together, $\overline{\mathbf{v}}_j$ and s_j define beliefs for scenario j. For each of these J scenarios there are 12 pivotal probabilities to compute. Let $\mathbf{T}$ denote the $J \times 12$ matrix of pivotal probabilities computed with our numerical approach, and let $\tilde{\mathbf{T}}_M$ denote the $J \times 12$ matrix of pivotal probabilities computed with our simulation method using M draws from the belief distribution. Our focus is on how the discrepancies between $\mathbf{T}$ and $\mathbf{T}_M$ vary with M. We summarize these discrepancies with two approaches.

First, for each M and for each of $J = 100$ we compute the root mean squared error (RMSE), or average discrepancy, between $\mathbf{T}$ and $\tilde{\mathbf{T}}_M$. That is, for a given M, we compute the RMSE for each row of $\mathbf{T}$ and $\tilde{\mathbf{T}}_M$. The left panel of Figure A.2 summarizes the distribution of these 100 RMSEs at each value of M. It shows that the distribution of RMSEs converges toward a point mass at zero as the number of draws from the belief distribution increases. As the simulation approach becomes more accurate, its computational burden also increases (as shown in the right panel): with M of 1 million, our machine takes over 250 times longer to compute the pivotal probabilities by simulation than by the analytical approach.[34]

Second, for each pivotal event we compute at each M the RMSE across the $J = 100$ scenarios between $\mathbf{T}$ and $\tilde{\mathbf{T}}_M$. That is, for a given M, we compute the RMSE for each column of $\mathbf{T}$ and $\tilde{\mathbf{T}}_M$. Figure A.3 summarizes how these RMSEs vary with M. It shows that the RMSE drops toward zero for all pivotal events as the number of number of draws from the belief distribution increases.

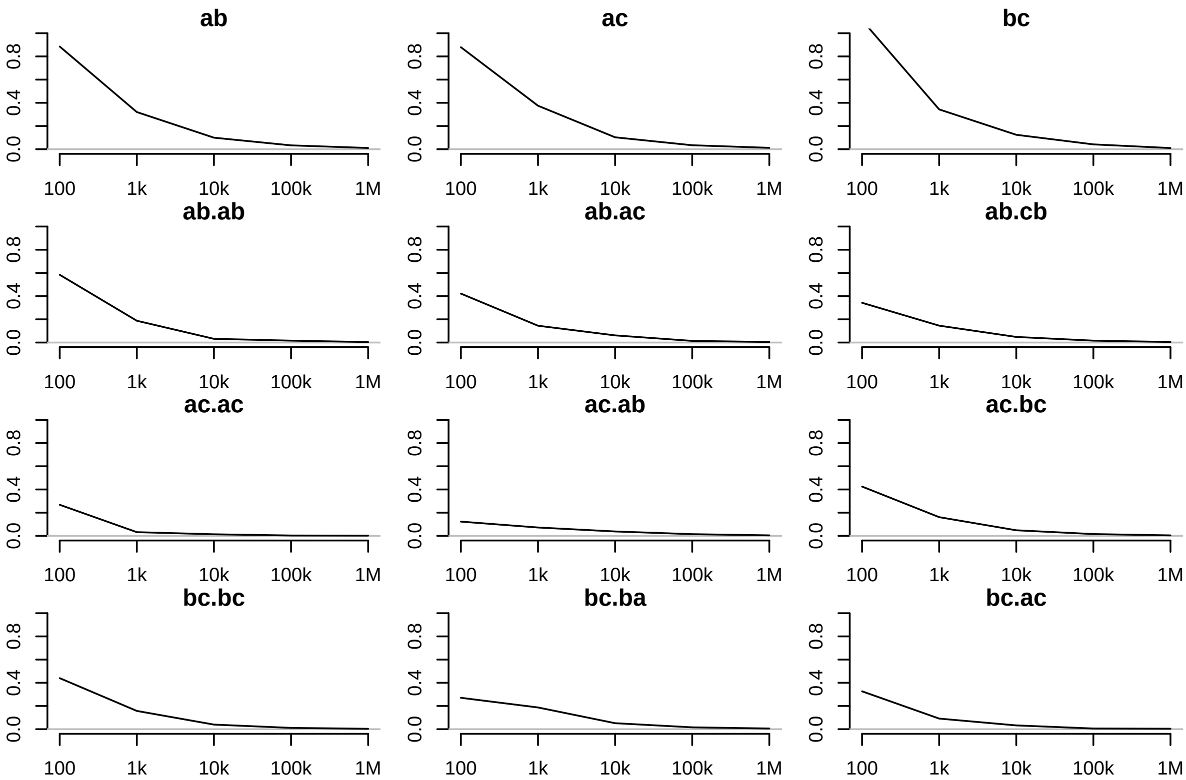
[34] Benchmarking performed on a 2017 MacBook Pro with 2.3 GHz processor and 16GB memory.

Figure A.2: Numerical/analytical approach agrees with simulations but is many times faster



Note: For each of 100 sets of belief parameters, we compute pivotal probabilities (1) analytically and (2) by simulation, with M draws from the belief distribution. We then calculate the RMSE across the 12 pivotal events between the analytical approach and the simulation approach for each of the 100 scenarios. The left figure shows, for each value of M (horizontal axis), that the distribution of the RMSEs across the 100 scenarios converges to a point mass at zero as the number of simulation draws increases. The right panel shows how the relative computational burden of the simulation approach increases as the number of simulation draws increases

Figure A.3: RMSE by pivotal event and number of draws in simulation



Note: For each of 100 sets of belief parameters, we compute pivotal probabilities (1) analytically and (2) by simulation, with M draws from the belief distribution. The figure shows, for each pivotal event, the average discrepancy (RMSE) between the two approaches as M increases.

A.3 Modeling beliefs about likely election outcomes

Finally, we provide greater detail about modeling beliefs about likely election outcomes. As stated in the main text, we rely on a Dirichlet distribution to describe voters' beliefs about the distribution of likely ballot shares. More formally, the distribution $f(\mathbf{v}_{-i})$ is defined as

$$f(\mathbf{v}_{-i}) = \mathrm{Dir}(\bar{\mathbf{v}} \times s) \tag{12}$$

where $\bar{\mathbf{v}}$ is the location parameter – describing the expected value of the distribution – and s is the scale parameter, scaling the uncertainty around that expectation. The choice of $\bar{\mathbf{v}}$ depends on our iterative polling algorithm. For the first iteration, we set the location parameter to correspond to the distribution of vote shares if everyone in the surveyed election had cast their ballot sincerely.

In each subsequent iteration m, the expected result at which the belief is centered is a weighted average of the previous expected result, $\bar{\mathbf{v}}_{\mathbf{m-1}}$, and voters' best response ($\bar{\mathbf{v}}^{\mathbf{BR}}$) to beliefs centered at $\bar{\mathbf{v}}_{m-1}$:

$$\mathbf{v_m} = \lambda \mathbf{v}^{BR}(\bar{\mathbf{v}}_{m-1}, s) + (1 - \lambda)\bar{\mathbf{v}}_{m-1}. \tag{13}$$

We run this algorithm for 250 iterations for each case. As the main text suggests, iterations at the beginning of the algorithm can be interpreted as assuming that most voters vote sincerely; while later iterations can be interpreted as assuming a highly strategic electorate.

B Selecting the optimal vote from expected utilities

In order to identify a voter's strategically optimal vote, we have to find the ballot choice that yields the highest expected utility. In technical terms, this is the row maximum of the matrix product of the utility matrix $\mathbf{U}$ and the pivotal probability matrix $\mathbf{P}$ (cf. Section 2 in the main paper). However, as some pivotal probabilities are of an extremely small magnitude (especially in IRV), we conventionally run into the 'floating point problem' when using computational approaches to identify the maximum (Goldberg, 1991). Essentially, finite memory means that computers can only store numbers with limited precision by approximating them to the closest defined floating point. As a consequence, tiny differences between numbers that lie in between two floating points are lost due to rounding. This problem affects the selection of row minima and maxima if the values are sufficiently small, and the difference between the two ballots under IRV with the same first preference depends on some very unlikely event. For example, for a voter with sincere preference ABC, voting ABC or ACB amounts to the same expected utility except for the case where A is eliminated in the first round and the voter is pivotal between B and C in the second round. If that event is sufficiently unlikely, the two expected utilities may appear to be the same and either could be chosen as the maximum, even though ABC must be strictly better for the voter (assuming the $B - C$ second-round tie has non-zero probability) and would be chosen if we had unlimited precision.

Ideally, we would increase the memory for each stored number, but this comes at a high computational cost. As a more feasible solution, we implement the following procedure to avoid selecting row maxima that are theoretically unjustified and only occur because of the floating point precision problem:

1. We order voters' preferences over ballots from most sincere to least sincere, assigning a higher weight w to more sincere ballots; for a CBA voter, for example the weights are $CBA = 6$, $CAB = 5$, $BCA = 4$, $BAC = 3$, $ACB = 2$, $ABC = 1$
2. We then add a small weight to voters' expected utility for each ballot, which is the above weight w multiplied by 1×10^{-10}, so e.g. we add 6×10^{-15} to the expected utility of a CBA ballot for a CBA-type voter.

This adjustment addresses the floating point problem. In some cases it may also override a very small but legitimate expected utility difference in favor of a sincere vote. Although we would prefer to address the floating point problem without introducing extra assumptions about voter behavior, nudging voters who are essentially indifferent between two ballots toward a more sincere vote arguably better captures what voters might do in practice. In our IRV simulations it appears that reducing the magnitude of the sincerity nudge allows for more oscillating patterns at a given λ but does not affect any of the main conclusions of the paper.

C Convergence of the iterative polling algorithm

In this section of the appendix we provide additional evidence that (a) the iterative polling algorithm converges at all in IRV; (b) it converges onto a seemingly unique equilibrium of ballot shares. In plurality, we can infer the equilibrium behavior from the vote share paths in Figure 2 alone; as all voters with a sincere preference for C have responded by voting strategically for A or B, the result is a (quasi-)Duvergerian equilibrium and everyone's best response is to continue voting as they did in response to the previous poll.[35]

In IRV, we cannot make the same inference as there is no general (analytical) characterization of strategic voting equilibria. However, we provide evidence that the algorithm converges onto a unique ballot share. We also provide evidence to suggest that this resulting equilibrium is robust to parameter choice (s, λ) as well as the 'starting point' of the equilibrium.

Notation. Let $\bar{\mathbf{v}}_{j,k}(s,\lambda)$ denote the (weighted) IRV ballot share vector (with six items) after the j^{th} iteration and in CSES case k. This vector is also the expected result upon which initial beliefs in the subsequent iteration $j+1$ are centered. For example, $\bar{\mathbf{v}}_{1,\text{AUS2013}}$ denotes the ballot shares at the end of the first iteration in the 2013 Australia case. This quantity, as defined by Equation 13, is a weighted average of voters' best response to initial beliefs in this iteration ($\mathbf{v}^{BR}_{j,k}(s,\lambda)$), and the initial expected result at the beginning of the iteration ($\bar{\mathbf{v}}_{j-1,k}(s,\lambda)$).

Next, let $d(\mathbf{m},\mathbf{n}) = \sqrt{(\mathbf{m}-\mathbf{n})(\mathbf{m}-\mathbf{n})}$ denote the Euclidean distance between two arbitrary vectors of the same length. We then define, $D_{j,k}$ (the quantity in Appendix C.1), as:

$$D_{j,k} = d(\bar{\mathbf{v}}^{BR}_{j,k}(s,\lambda),\ \bar{\mathbf{v}}_{(j-1,k)}(s,\lambda))$$

which is the Euclidean distance between the voters' best response to any given iteration in case k, and the ballot shares in the poll at the beginning of that iteration.[36]

Convergence onto a fixed point. In Appendix C.1, we report $D_{j,k}$ (the distance between voters' best response and the expected result at the beginning of the iteration), and show that the convergence behavior is robust to parameter choices.

Convergence onto an oscillating sequence. Appendix C.2 reports further context on the oscillating behavior under IRV and shows that when comparing the distance between an expected result and a lagged average of the algorithm output (smoothing any oscillation), the distance converges towards zero.

Convergence onto the same point across parameter values. In Appendix C.3, we provide evidence that the equilibria upon which the algorithm converges are robust to the parameter choice of λ.

Convergence onto the same point across starting points. Furthermore, Appendix C.4 suggests that the vote shares upon which the iterative polling algorithm converges in IRV may hold irrespective of the starting point. Together, these results characterize the behavior of the iterative polling algorithm under IRV and suggest that a general strategic voting equilibrium in IRV may exist.

[35]The same logic, merely with inverted party names, holds for the few cases where the eventual equilibrium pins A and C against each other.

[36]Alternatively, also denoted as the output of the algorithm in the iteration $j-1$.

C.1 Euclidean distances between best response and ballot share vector

We check the convergence of the polling algorithm for different parameter values of precision (s) and strategic responsiveness (λ). We present, for each iteration j and every case k, $D_{j,k}$, the distance between the ballot shares of best responses given certain parameter values, and the ballot shares in the poll at the beginning of the algorithm (using the same parameter values). We use a logarithmic scale to plot the distance, in order to highlight changes of a small magnitude when the ballot shares do not move much anymore after multiple iterations.

In IRV (left panels), we see that for a handful of cases, the distances decrease continuously and evenly. The remainder sees their distance drop until the 50th or 60th iteration before stagnating at very small values ($e^{-7} \approx 0.0009$). This behavior occurs because of the oscillations, whereby a small number of voters changes their strategic vote in a regular pattern, thus preventing the algorithm from reaching a 'true' fixed point. In a setting with low strategic responsiveness (smaller values of λ), the distance decreases more slowly as convergence is slower; in most cases, 250 iterations are not sufficient to reach full convergence. In contrast, with high strategic responsiveness, the algorithm settles into a pattern where the poll-to-poll distances are greater, since more voters are part of the 'oscillation'. Although a few CSES cases are sensitive to the choice of s, the broader convergence pattern and magnitude of oscillation distances (conditional on λ) appear robust.

In plurality, convergence occurs in a very regular and even fashion – there are no cases that get stuck in an oscillating pattern or stop converging towards zero. This corroborates theoretical knowledge about equilibria in plurality. However, the speed of convergence and variance between cases is sensitive to values of λ and s.

C.1.1 Medium strategic responsiveness ($\lambda = 0.05$)

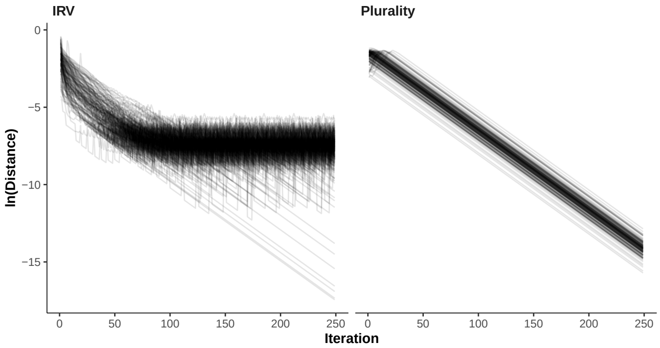


Figure A.4: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

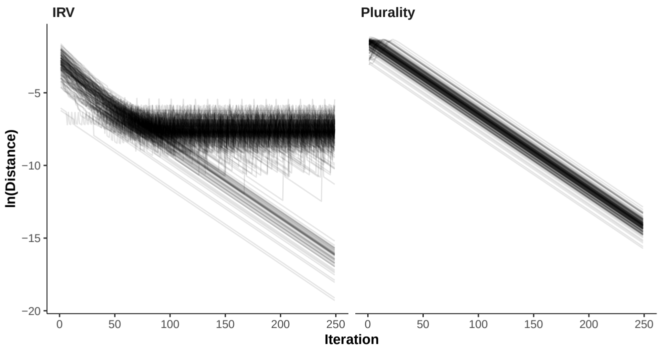


Figure A.5: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

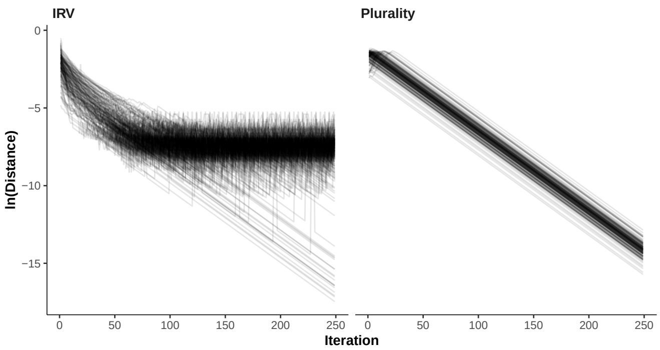


Figure A.6: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

C.1.2 Low strategic responsiveness ($\lambda = 0.01$)

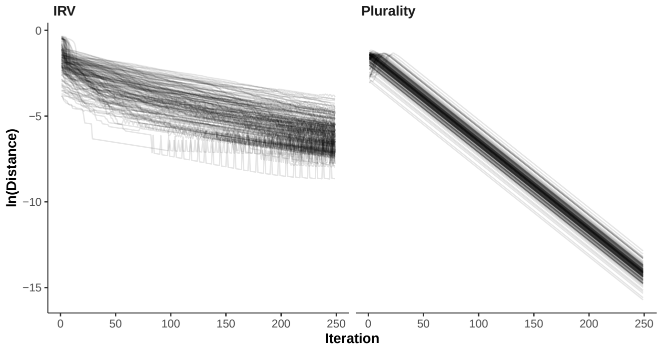


Figure A.7: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

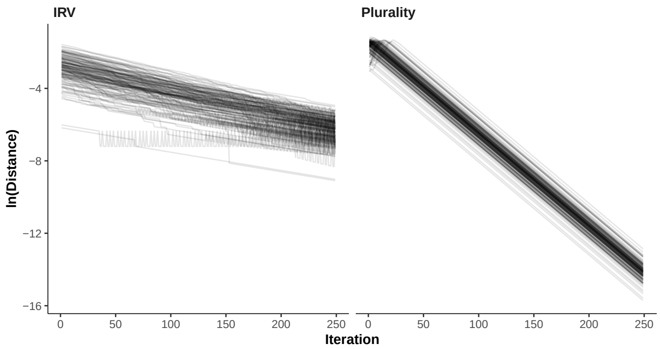


Figure A.8: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

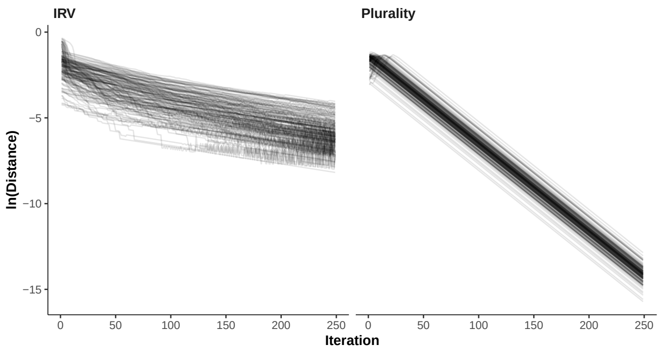


Figure A.9: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

C.1.3 High strategic responsiveness ($\lambda = 0.10$)

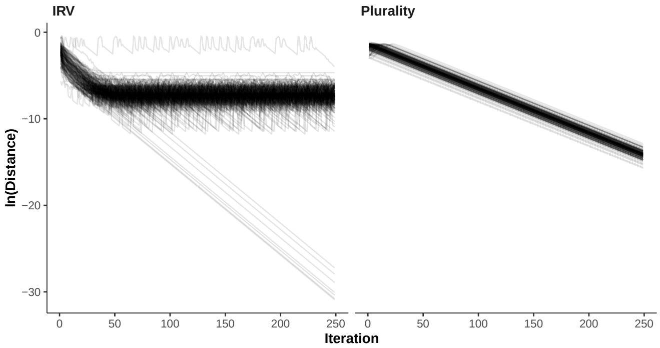


Figure A.10: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

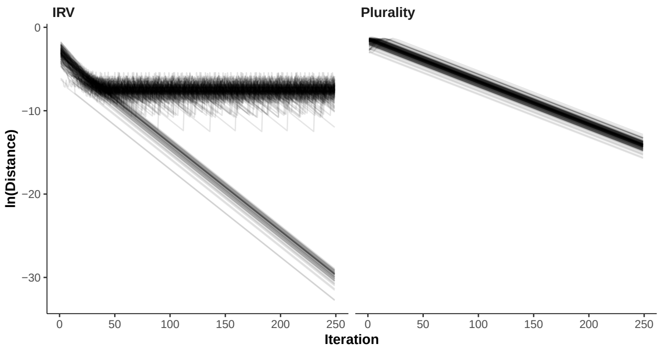


Figure A.11: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

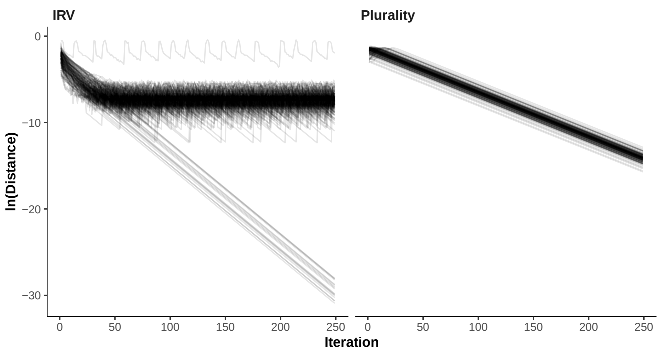


Figure A.12: Logged distance between poll result at the beginning of the iteration and voters' best response to it in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

C.2 Euclidean distances between best response and lagged ballot share vector

To provide further context on the oscillating behavior under IRV, we report the Euclidean distance (Figure A.13) between the resulting best response ballot shares after the poll at time j of the algorithm, and the average of poll vote shares between times $j-20$ and $j-10$:

$$D^{lag}_{j,k} = d(\bar{\mathbf{v}}^{BR}_{j,k}(s,\lambda), \bar{\mathbf{v}}_{((t-10,\ t-20),\ k)}(s,\lambda))$$

Here, too, the quantity of interest decreases as voters become more strategic; the majority of cases settles in a band between 0 and 0.01.[37] This behavior indicates that although there are changes from one iteration to another due to a small number of voters changing their optimal strategic response, the overall vote share does not move in great distances across multiple iterations. Occasional spikes occur when that pattern is disrupted and the vote share moves a larger distance before settling into a new oscillation again. Altogether, the examination of vote share distances along the iteration paths suggests that in IRV, the algorithm settles on either a direct fixed point, or an oscillating pattern where only a small number of voters changes their strategic response in a regular manner.

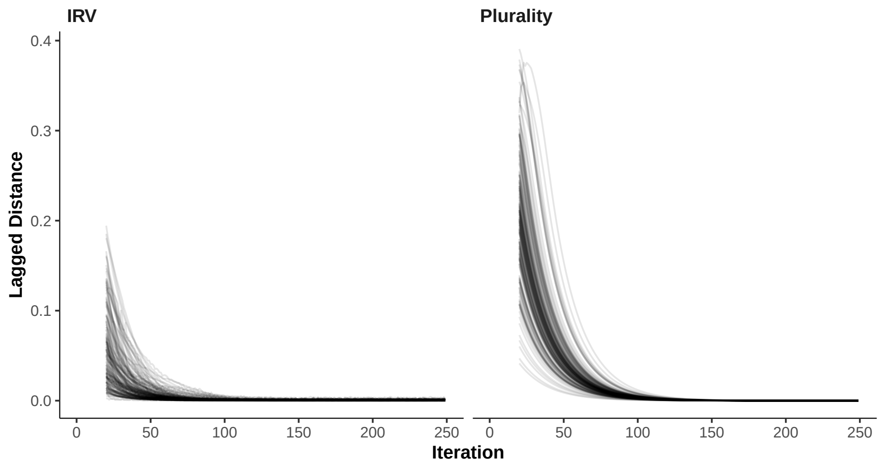


Figure A.13: Distance between the shares of voters' best responses after voters have been given a poll in IRV (with $\lambda = .05$ responding strategically), and the average of the respective vote shares 10 to 20 iterations ago.

[37]Note that for other parameter values [not shown], the range of this band will vary, but the general pattern holds.

C.3 Comparison of convergence paths relative to baseline case

We provide evidence that the equilibria upon which the algorithm converges are robust to the parameter choice of λ; we plot the distribution (across CSES cases) of distances between a jth poll with certain parameter values, and the resulting vote shares after the 250th poll in the baseline case ($s = 85$, $\lambda = 0.05$, Figure A.14), as well as after the 250th poll in the case with the same s, but holding $\lambda = 0.05$ (Figure A.15).

Formally, the quantities of interest are:

$$\begin{aligned} D_{j,k}^{base} &= d(\bar{\mathbf{v}}_{j,k}(s,\lambda), \bar{\mathbf{v}}_{(250,\ k)}(s = 85, \lambda = 0.05)) \\ D_{j,k}^{s-comp} &= d(\bar{\mathbf{v}}_{j,k}(s,\lambda), \bar{\mathbf{v}}_{(250,\ k)}(s, \lambda = 0.05)) \end{aligned}$$

The results show that although the algorithm converges on different ballot shares conditional on the choice of s – the densities of distances do not converge onto zero when compared to the baseline of $s = 85, \lambda = 0.05$ (Figure A.14), the equilibrium is robust to the choice of λ: when comparing distances across different values of λ, but holding s fixed (Figure A.15), we see differences in how quickly the algorithm converges (which is what λ determines by definition), but, ultimately, the distances converge towards zero.

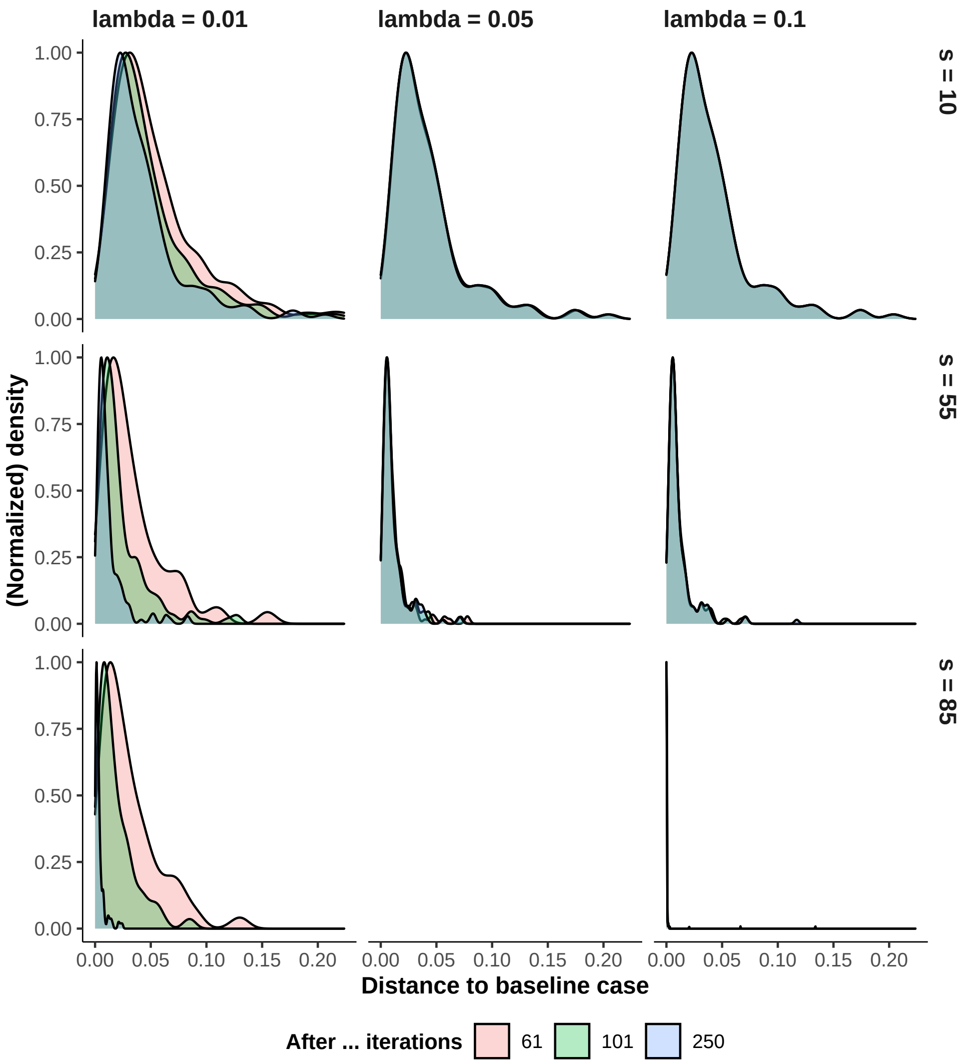


Figure A.14: Distribution of Euclidean distances across CSES cases between resulting vote shares in jth iteration under given parameter combination (information precision, s, and strategic responsiveness, λ) compared to 250th iteration in the baseline case ($s = 85$, $\lambda = 0.05$).

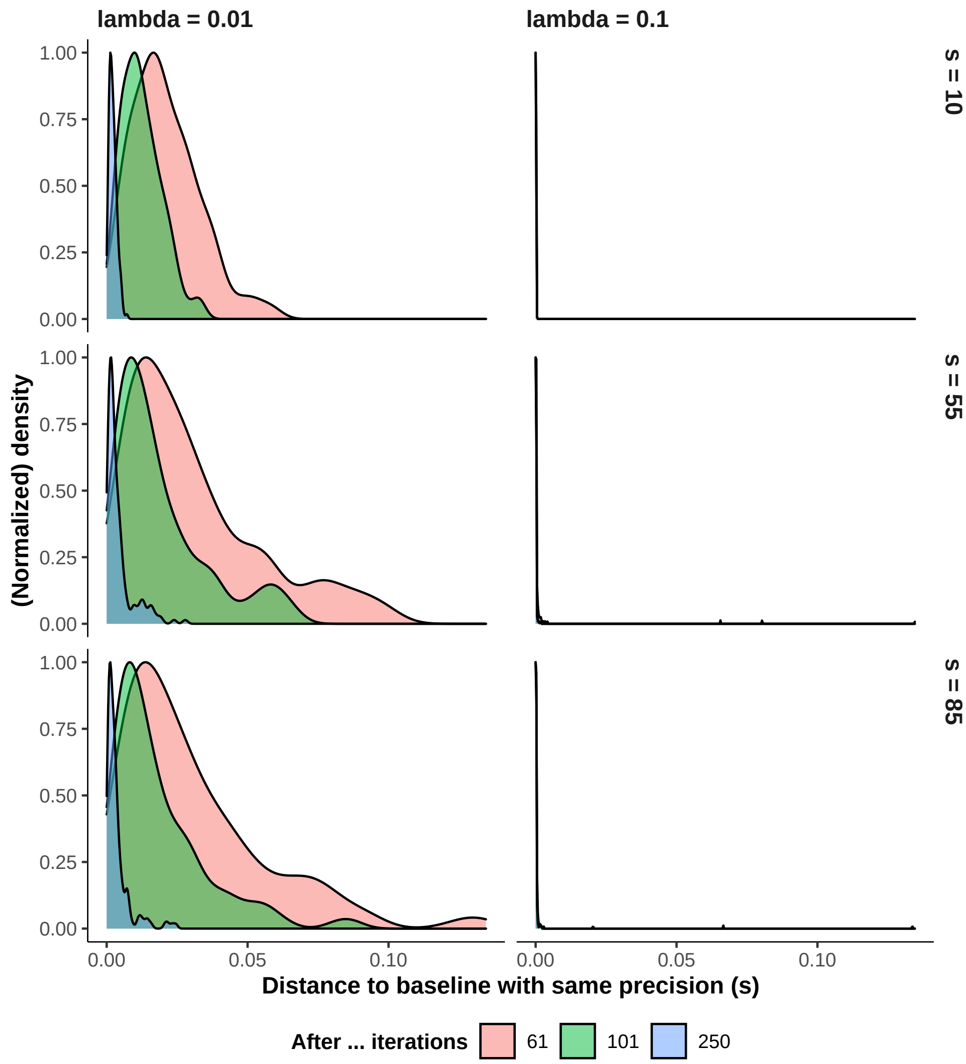


Figure A.15: Distribution of Euclidean distances across CSES cases between resulting vote shares in jth iteration under given parameter combination (information precision, s, and strategic responsiveness, λ) compared to 250th iteration in the case with same s but $\lambda = 0.05$.

C.4 Convergence under IRV with random starting points

In order to evaluate whether the CSES cases converge onto the same IRV strategic voting equilibrium irrespective of the initial belief about ballot shares, we draw 100 random ballot shares from a Dirichlet distribution with uniform density, and use these to initialize the polling algorithm. Let $q \in \{1, ..., 100\}$ denote the particular random draw. Formally, let $\tilde{\mathbf{v}}_{j,k}(s, \lambda, \tilde{\mathbf{v}}_0)$ denote the ballot share vector after j iterations for CSES case k, where the algorithm was initialized with the values s, λ, and a starting belief about ballot shares centered on $\mathbf{v}_0$. Then, the "random starting point distance to baseline case" refers to the distance between the ballot share vector after the jth iteration for case k and a random starting belief centered on $\mathbf{v}_q$, and the ballot share after the 250th iteration where the algorithm was initialized with baseline parameter values ($s = 85, \lambda = 0.05$), and the sincere ballot share profile for that case. Formally,

$$\tilde{d}_{j,k,q} = d(\tilde{\mathbf{v}}_{j,k}(s, \lambda, \bar{\mathbf{v}}_q), \mathbf{v}_{250,k}(s = 85, \lambda = 0.05, \mathbf{v}_{true}))$$

Figure A.16 summarises the distribution of distances between ballot shares starting at random points (with $s = 85, \lambda = 0.05$, i.e., baseline parameter values), and the 'converged' ballot share after 250 iterations starting at each case's sincere profile. Each point indicates the median, 90th or 99th quantile of the distribution of distances (y-axis) between the algorithm from random starting points and the converged ballot shares (after 250 iterations) coming from the sincere voting profile, for each case and after each iteration (x-axis). Formally, each point represents a summary statistic of all $\tilde{d}_{j,k,q}$ for each case k, and after each iteration j across all 100 random draws.

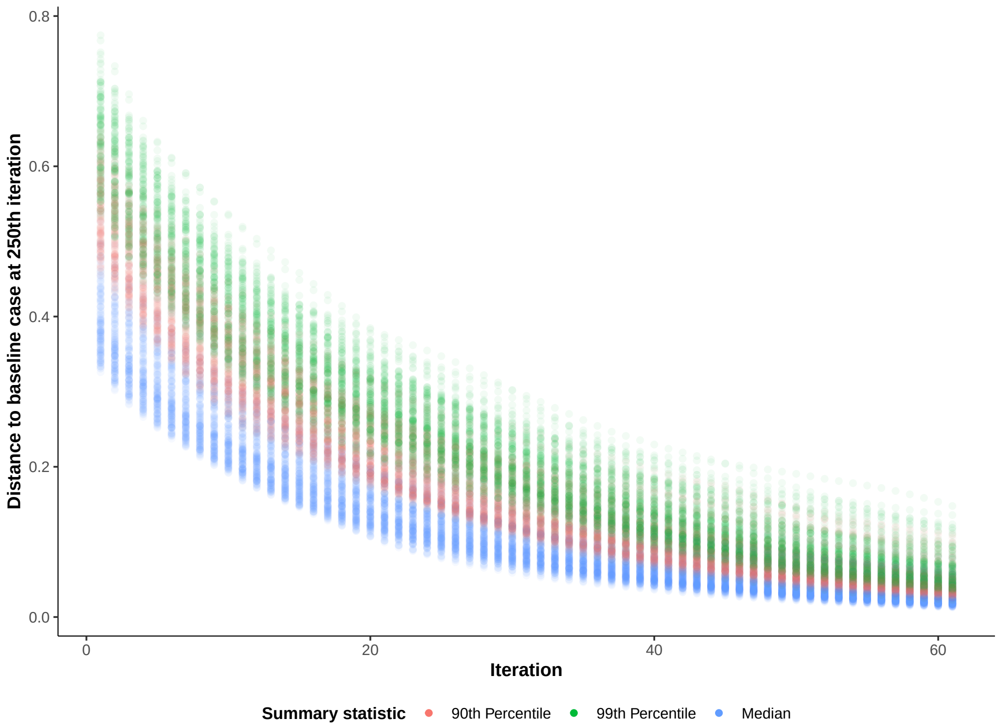


Figure A.16: Summary of distances between case-specific distributions of distances between ballot shares after iterations from random starting points, and the converged ballot shares in the baseline case

D Robustness of expected benefit results to number of iterations

In this section, we present results from Figure 4 extended to 250 (rather than 60) iterations. The results do not change substantially beyond 60 iterations.

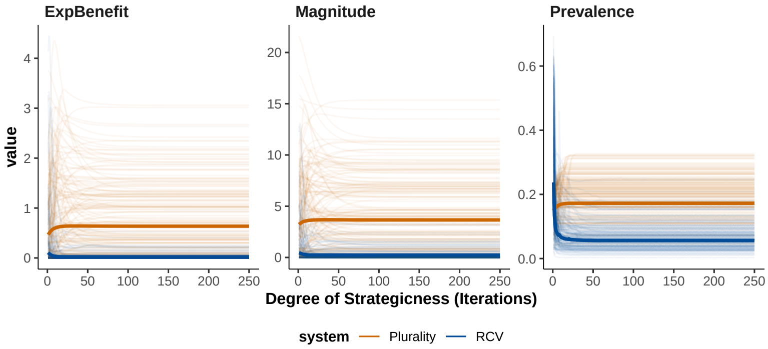


Figure A.17: Expected benefit, magnitude, and prevalence of strategic voting with high ($s = 85$) belief precision, and medium strategic responsiveness ($\lambda = 0.05$).

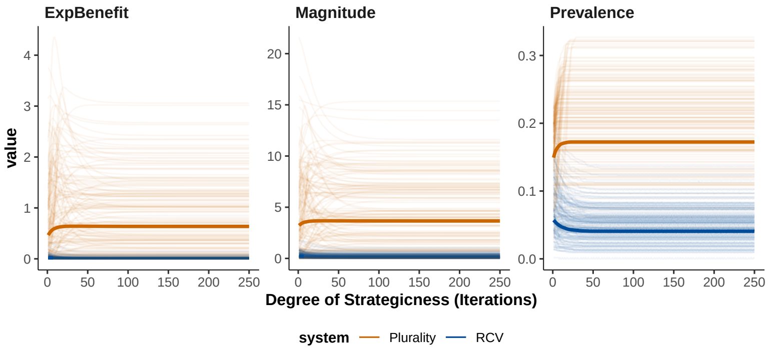


Figure A.18: Expected benefit, magnitude, and prevalence of strategic voting with low ($s = 10$) belief precision, and medium strategic responsiveness ($\lambda = 0.05$).

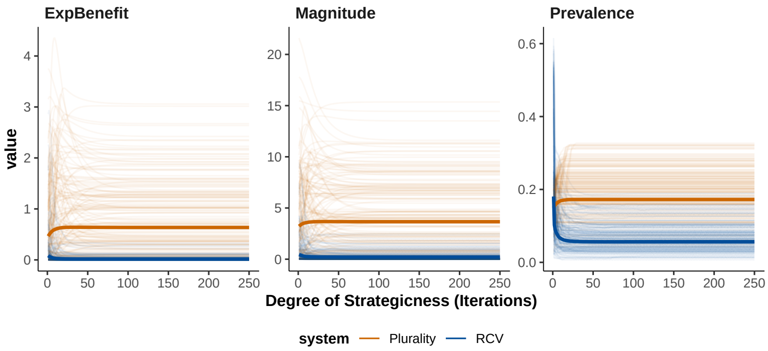


Figure A.19: Expected benefit, magnitude, and prevalence of strategic voting with low ($s = 10$) and medium ($s = 55$, right) belief precision, and medium strategic responsiveness ($\lambda = 0.05$).

E Robustness of Results Grouped By Electoral Systems

In this section we provide additional results akin to Figure 6, but for different parameter combinations.

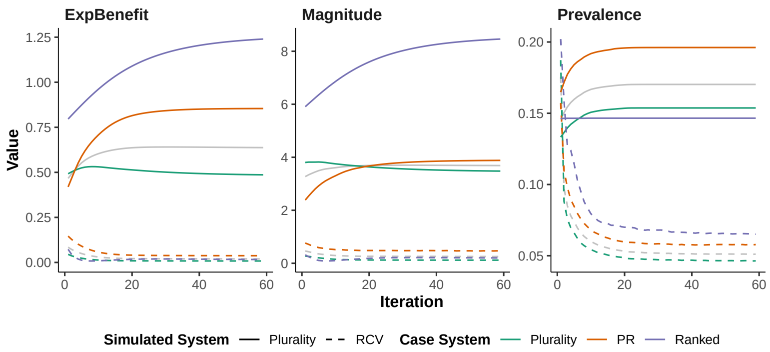


Figure A.20: Expected benefit, magnitude, and prevalence of strategic voting with medium ($s = 55$, right) belief precision, and medium strategic responsiveness ($\lambda = 0.05$). Grouped by electoral system used in cases.

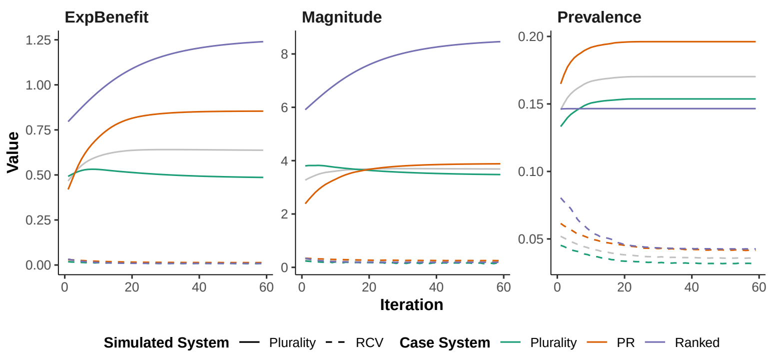


Figure A.21: Expected benefit, magnitude, and prevalence of strategic voting with medium ($s = 10$, right) belief precision, and medium strategic responsiveness ($\lambda = 0.05$). Grouped by electoral system used in cases.

F Additional Results on Voter Welfare

Below, we report the distribution of within-case differences of voter welfare metrics (average expected utility and Pr(Condorcet Winner wins)) when compared across both iterations and simulated systems.

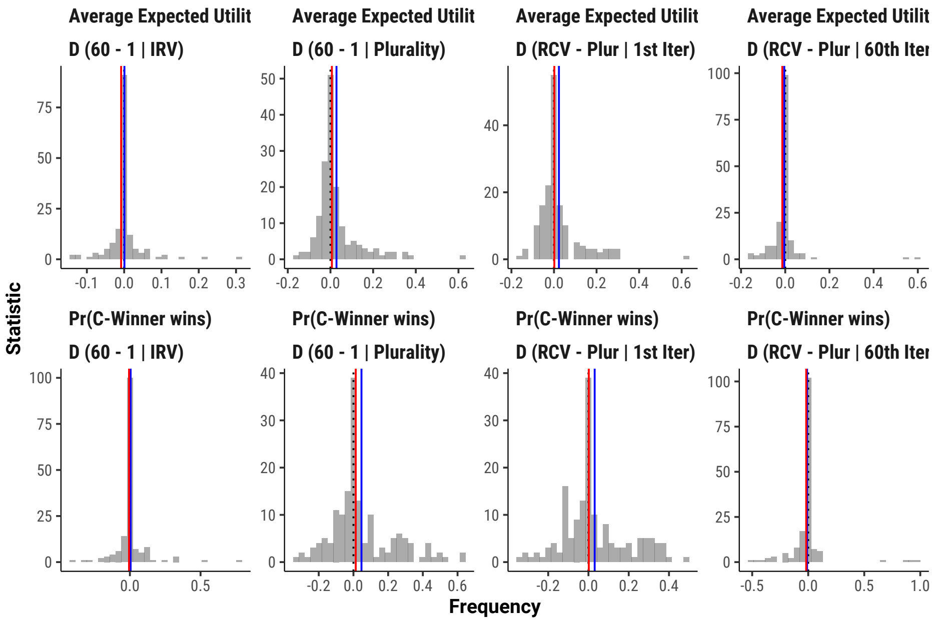


Figure A.22: Within-case differences of average expected utility and Pr(Condorcet Winner wins) when compared across iterations (holding system constant) and system (holding iterations constant). Blue lines indicate raw means across cases; red lines indicated weighted means across cases.

G Case-by-case comparisons

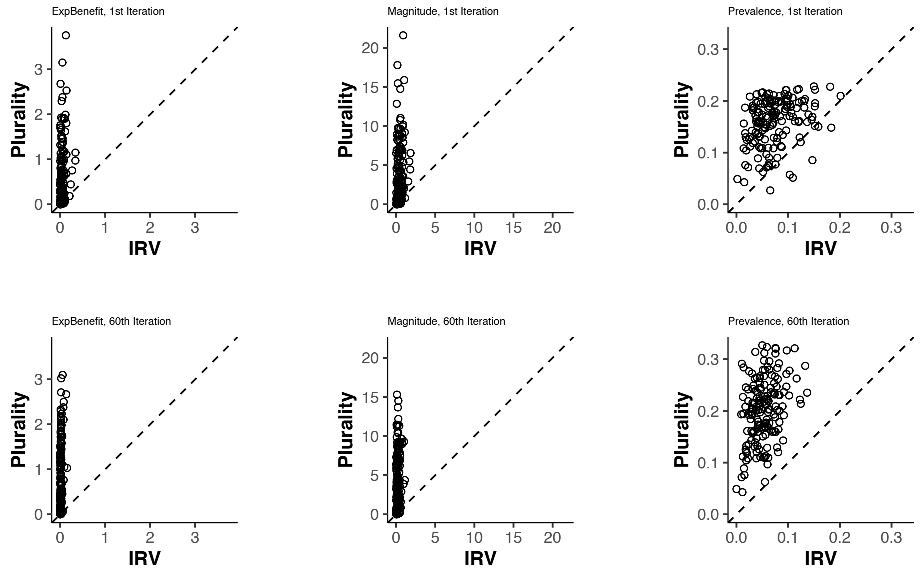


Figure A.23: Expected benefit, magnitude, and prevalence of strategic voting in plurality (vertical axis) vs. IRV (horizontal axis) with low ($s = 10$) belief precision and medium strategic responsiveness ($\lambda = 0.05$), where each dot is one CSES survey.

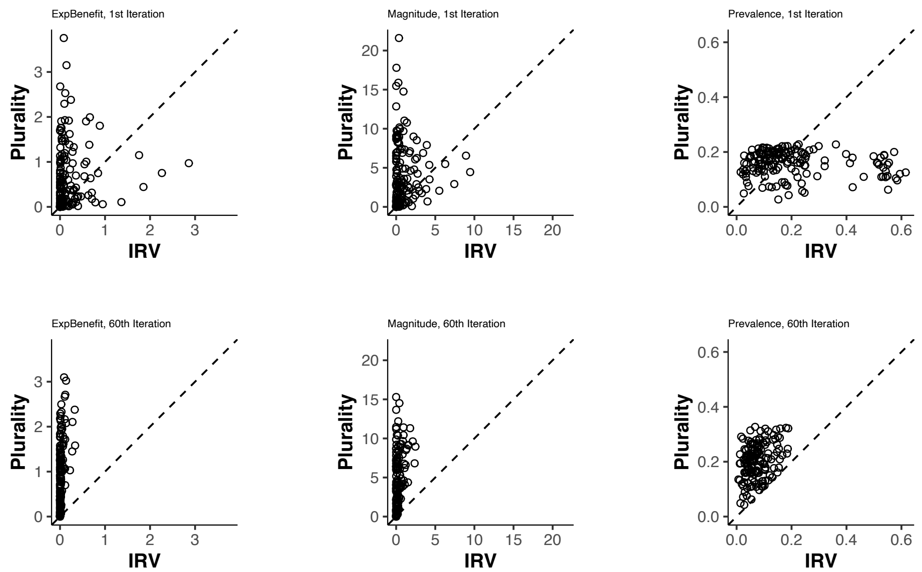


Figure A.24: Expected benefit, magnitude, and prevalence of strategic voting in plurality (vertical axis) vs. IRV (horizontal axis) with medium ($s = 55$) belief precision and medium strategic responsiveness ($\lambda = 0.05$), where each dot is one CSES survey.