Comparing strategic voting incentives in plurality and instant-runoff elections[1]

Andrew C. Eggers[†]

Tobias Nowacki[†]

Abstract

Reformers and researchers often speculate that some voting systems induce less strategic voting than others, but existing research is mostly unhelpful in assessing these conjectures because it is based on unrealistic assumptions about preferences, beliefs, or both. We propose a general approach to assessing strategic voting incentives given realistic preferences and beliefs and we use it to compare strategic voting incentives in three-candidate plurality and instant-runoff (IRV) elections. Drawing on preference data from 160 electoral surveys, we estimate that the average voter's gain in expected utility from being strategic (rather than always voting sincerely) is between 5 and 35 times higher in plurality than in IRV, depending on assumptions about the prevalence of strategic behavior. In IRV, the benefit of voting strategically in response to a poll is smaller on average when one expects other voters to do so, while the reverse is true in plurality. This previously unexplored negative feedback constrains the prevalence and magnitude of strategic voting incentives in IRV.

[1]This version: October 30, 2019. Thanks to Paul Klemperer, David Myatt, Karine van der Straeten, Dimitri Landa, Justin Grimmer, Matias Iaryczower, Gregory Martin, and seminar audiences at Trinity College Dublin, the Toulouse School of Economics, the Democracy and Policy Lab at Stanford University, and the ECPR Standing Group on Strategic Interactions for valuable feedback. Thanks to Mats Ahrenshop for research assistance.

[†]Nuffield College and Department of Politics and International Relations, University of Oxford, UK. email: `andrew.eggers@nuffield.ox.ac.uk`

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

1 Introduction

In recent years, voters in both the United States and the United Kingdom have held referendums on whether to replace the plurality system of election (also known as first-past-the-post) with an instant-runoff voting system (IRV, also known in the US as ranked-choice voting or RCV, in the UK as the alternative vote or AV, in Australia as preferential voting, and in Ireland as STV).[2] In contrast to plurality elections, in which each voter casts a ballot for a single candidate and the winner is the candidate with the most votes, voters in IRV elections rank the candidates and the winner is determined by successively eliminating less-popular candidates.[3] Plurality is the most widely used single-winner election system in the world; IRV variants are far less common but are used to elect legislators at the state and federal level in Australia, presidents of Sri Lanka, India, and Ireland, mayors of Sydney, London, San Francisco and several other American cities, and (as of 2018) members of the U.S. Congress from Maine.

One of the advantages of the IRV system, according to its advocates, is that IRV creates less incentive to vote strategically, i.e. to determine one's optimal vote in light of likely results rather than solely on the basis of one's sincere preferences. According to FairVote, a US advocacy organization, voters in IRV elections "can honestly rank candidates in order of choice without having to worry about how others will vote and who is more or less likely to win."[4] Similarly, UK Deputy Prime Minister Nick Clegg testified in advance of the 2011 referendum that IRV "stops people from voting tactically and second-guessing how everybody else will vote in their area."[5] While quick to point out that strategic voting can be rewarded in IRV, academic researchers have tended to agree with the conjecture that the incentive to vote strategically is lower in IRV

[2] In the 2011 UK referendum, voters rejected the proposal by a 2-1 margin (with over 19 million voters casting a ballot); in a 2016 referendum in the U.S. state of Maine, voters narrowly approved the proposal (with around 750,000 votes cast), leading to IRV being introduced to elect members of the House of Representatives in the 2018 midterm elections.

[3] Most commonly, in each round the candidate receiving the fewest top rankings is eliminated from all ballots until one candidate remains (or, equivalently, until one candidate has a majority of top rankings). In the system used to elect the mayor of London and and the president of Sri Lanka (sometimes called the "contingent vote" or "supplementary vote" system), all candidates but two are eliminated immediately. The other variation within the family of instant-runoff systems is how many candidates voters may rank (up to two in the London mayoral election, up to three in San Francisco, as many as desired in some Australian states, and all in other Australian states). In the three-candidate case (our focus), the elimination procedures are identical and the number of candidates voters rank makes little or no difference.

[4] https://www.fairvote.org/rcv#rcvbenefits, visited 11 June 2019.

[5] "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).

than in plurality. Cox (1997, 95) observes that more information is needed to vote strategically in IRV than in plurality, as does Renwick (2011, pp. 6-7), who concludes that IRV would "reduce but not eliminate incentives for tactical voting" compared to plurality. Dummett (1984, p. 228) sees the difference as less clear-cut, maintaining that "a voter who has understood the workings of the procedure, and who has some information about the probable intentions of the others, will have nearly as much incentive to vote strategically" in IRV as in plurality.

Why should voters or policymakers care about how much a voting system rewards strategic voting? Four reasons stand out. First, to the extent that voters cast ballots that do not reflect their sincere preferences, election results are difficult to interpret (Satterthwaite, 1973). Second, strategic voting requires effort on the part of voters (paying attention to polls, weighing strategic considerations against other values) and/or parties (spreading polling information, convincing voters to act strategically) that might be better devoted to other activities (though see Dowding and Van Hees, 2008). Third, some voters seem to value expressing their preferences honestly (Hamlin and Jennings, 2011; Pons and Tricaud, 2018), and the greater the rewards for strategic voting the more voters may be pushed to forego the expressive value of the vote. Fourth, some types of voters appear to be less able or inclined to vote strategically (Eggers and Vivyan, 2018; Eggers and Rubenson, 2019), and these voters may be more disadvantaged in a voting system that creates greater incentives to vote strategically.[6]

Given these reasons for comparing the strategic incentives created by voting systems (and the claims about these incentives by both activists and academics), it is surprising how little previous research has attempted to measure strategic voting incentives in plurality, IRV, or other voting systems. The Gibbard-Satterthwaite Theorem shows that all reasonable ordinal voting systems could produce a *manipulable* voting result, i.e. a configuration of ballots such that one or more voters would be better off submitting an insincere ballot than a sincere one (Gibbard et al., 1973; Satterthwaite, 1975; Reny, 2001). Following on this, a large literature (mostly in mathematics and computer science) has assessed the likelihood of such a manipulable result for different voting rules given some assumption about the distribution of possible voting outcomes

[6]Although Eggers and Vivyan (2018) and Eggers and Rubenson (2019) provide evidence of discrepancies in strategic voting by e.g. age and income level in plurality elections in the UK and Canada, it remains to be seen whether these discrepancies are smaller in other voting systems.

(e.g. Chamberlin, 1985; Nitzan, 1985; Saari, 1990; Favardin and Lepelley, 2006; Plassmann and Tideman, 2014; Ornstein and Norman, 2014; Miller, 2017). But these studies are of little use for assessing the incentive to vote strategically, even if we accept their assumptions about likely voting outcomes, because they do not take into account uncertainty about others' votes: they essentially ask how likely a voter is to *regret* a sincere vote after the election takes place, not how likely a voter is to *foresee* that an insincere vote would be optimal before the election takes place. (In the next section we discuss why these questions are likely to yield different answers.) Research on the *ex post* manipulability of electoral systems thus has little to say about the *ex ante* incentives voters face.

In this paper, we propose a general framework for assessing strategic voting incentives given realistic preferences and uncertainty, and we use it to compare strategic voting incentives in plurality and IRV elections. Our measure of strategic voting incentives in a given electoral system is the answer to the question, "How much would a voter with typical preferences benefit from voting strategically (i.e. casting the optimal ballot given her preferences and beliefs about likely election outcomes) rather than voting sincerely (i.e. simply casting a ballot that reflects her true preferences)?" The answer to this question depends on what we assume about voters' *preferences* and *beliefs*. For preferences, we use election surveys in the Comparative Study of Electoral Systems (CSES) in which voters are asked to rate each party on a 0-10 scale; this yields preferences for over 220,000 voters in 160 different elections.[7] We then model voters' beliefs about possible election outcomes as a probability distribution satisfying two criteria: first, the precision of the distribution is consistent with the empirical predictability of election outcomes; second, the location of the distribution is consistent with the preference data (i.e. what other voters in the same election want) and a model of how voters might strategically respond to election polls. (The iterative polling algorithm we develop to achieve the second condition is a contribution in itself and may have wide applicability in the study of voting systems.) This approach allows us to measure, for each voter and various assumptions about the prevalence of strategic voting among *other* voters, the expected benefit of strategic voting compared to sincere voting and, ultimately, to compare this benefit across different voting systems.

[7] In each survey we use preferences over the top three parties (in terms of national vote share) only.

We find that, consistent with some of the conjectures noted above, the average incentive to vote strategically is considerably higher in plurality in IRV: depending on how widespread strategic behavior is assumed to be in the electorate, and how precise beliefs are, a voter who switches from being a sincere voter to a strategic voter gains between 5 and 35 times more expected utility in plurality than in IRV. The gap is smaller when beliefs are more precise, suggesting that it is partly explained by the greater difficulty of discerning opportunities for strategic voting in IRV, and larger when other voters are expected to be more strategic (i.e. to respond more to previous polls). The latter point highlights an under-appreciated qualitative difference between strategic voting incentives in plurality and IRV: in an environment where voters respond to public polling information, strategic voting is characterized by positive feedback in plurality (other voters' strategic actions tend to increase my strategic incentives) but negative feedback in IRV (other voters' strategic actions tend to decrease my strategic incentives); this negative feedback implies that insincere voting is unlikely to become widespread in IRV elections, because once some voters engage in it the incentive for others to join in disappears. We also show that IRV is more resistant to strategic voting both because events that reward insincere votes are less likely and because these events tend to be "closer" (i.e. more correlated in probability) to events that punish the same insincere votes.

We emphasize that our focus in this paper is on the *incentive* to vote strategically, not the empirical prevalence or practicability of strategic voting. In order to detect the strategic incentives we measure in this paper, a voter needs basic information about the likely prevalence of each ballot type, a good understanding of the voting system, and the ability to reason strategically. Many voters may lack one or more of these ingredients – especially in IRV, where (as is clear from the exposition below) strategic voting undoubtedly requires more sophisticated reasoning than in plurality. In this sense, our paper sheds light on strategic voting incentives as these incentives might be computed by a voting advice service or perceived by a highly sophisticated party strategist who has well-formed (though imprecise) beliefs about election outcomes and can work through the implications of alternative voting strategies in a given system. Empirical research is necessary to determine whether actual voters can perceive and act on these incentives (e.g. Van der Straeten et al., 2010; Hix, Hortala-Vallve and Riambau-

Armet, 2017), but this paper provides insight on the degree to which these incentives exist.

2 Literature review

As noted in the previous section, there are several reasons why we might prefer voting systems that are less likely to reward voters for misrepresenting their preferences: such systems may produce more intelligible results, require less effort by voters and elites, allow voters to express themselves more fully, and equalize representation between more and less strategically-inclined voters. We know from the Gibbard-Satterthwaite Theorem that all systems (including plurality and IRV) can reward strategic behavior in some circumstances, but perhaps some systems are more resistant than others. In this section we review several attempts to make such comparisons in previous literature.

One way to compare voting systems' susceptibility to strategic voting is to estimate how often a voting system might reward a voter for casting an insincere ballot. Chamberlin (1985), for example, uses Monte Carlo methods to estimate the probability of a manipulable result under four voting systems (including plurality and IRV) assuming four candidates, either 21 or 1,000 voters, and preferences/ballots drawn from either an "impartial culture" (IC)[8] or a four-dimensional spatial model; he then computes the proportion of simulated sincere profiles that constitute coalition-proof Nash equilibria in each system, concluding that IRV is less manipulable than plurality. Favardin and Lepelley (2006) examine a case where multiple groups of voters may manipulate the result and one coalition of manipulators may anticipate another's actions; using preferences drawn from an "impartial and anonymous culture" (IAC),[9] they compute the proportion of sincere profiles that constitute an equilibrium or "quasi-equilibrium" (in which potential manipulators are deterred by counter-threats), with the results suggesting that IRV is somewhat less manipulable than plurality.

Another approach is to assess the *kind* of opportunities for manipulation that voting systems might produce. Dowding and Van Hees (2008) distinguish between sincere manipulation, where

[8] In IC each unique preference order is equally likely to be drawn for each voter, whereas in IAC each unique distribution of voters across preference orderings is equally likely. Thus with 60 voters and six possible preference orderings the distribution $(10, 10, 10, 10, 10, 10)$ is much more likely than $(60, 0, 0, 0, 0, 0)$ under IC but they are equally likely under IAC.

[9] See footnote 8.

a voter strategically moves a candidate to the top of her rank order in order to elect that candidate, and insincere manipulation, where a voter achieves a desired outcome some other way (such as by ranking one candidate higher in order to elect some other candidate). Dowding and Van Hees (2008) prove that all manipulation in plurality is sincere manipulation, but insincere manipulation is possible in IRV (or any runoff system): voters who sincerely prefer the leading candidate and suspect that one opponent is a bigger threat than another can strategically shift their support to the weaker opponent (the "pushover") in order to elected their preferred candidate. (Such manipulation might be considered especially objectionable because it requires more effort from voters, or because it might be beyond many voters' capabilities, or because it makes results particularly unintelligible.) Several recent studies investigate the probability of results that would reward insincere manipulation.[10] Plassmann and Tideman (2014) use an empirically calibrated model of likely preference profiles in three-candidate elections to estimate the probability of monotonicity violations (all of which would reward a form of insincere manipulation) in several voting systems for different numbers of voters; these paradoxes occur in IRV and other systems (but not in plurality), though not surprisingly the frequency of a single voter being able to change the winner through insincere manipulation goes to zero as the postulated electorate grows. Ornstein and Norman (2014) conduct a similar exercise focusing specifically on IRV, warning that the frequency of monotonicity failures should concern IRV advocates.[11]

Even if possibilities for manipulation exist, it might be that manipulation is so complicated in some systems that voters end up voting sincerely instead. A large literature in computer science beginning with Bartholdi, Tovey and Trick (1989) asks whether computing an optimal vote could be so difficult in some systems as to make manipulation impossible. (Faliszewski and Procaccia (2010) provides a useful introduction.) Bartholdi and Orlin (1991) proved that the time necessary to compute the optimal vote in an IRV election increases exponentially in the number of candidates and voters even when one knows all other votes with certainty. Conitzer, Sandholm and Lang (2007) subsequently showed that computing the optimal vote is actually manageable for a fixed number of candidates.

[10]These papers study the probability of monotonicity violations; Dowding and Van Hees (2008) prove that voting systems that are immune to monotonicity violations are also immune to insincere manipulation.

[11]See also Miller (2017).

What is missing from essentially all of the literature mentioned so far[12] is uncertainty: even if there are opportunities for manipulation, and even if voters have the capacity to identify these opportunities from the tabulated results, there may be systems in which voters would rarely if ever have precise enough knowledge of others' votes to make an insincere vote optimal. It is not clear, moreover, whether the incentive for *ex ante* strategic voting is higher when there are more opportunities for *ex post* manipulation. To see this, suppose that a plurality election in a large electorate ends in essentially a three-way tie, with candidate A defeating candidate B by just one vote. Clearly *ex post* manipulation is possible: for example, two supporters of candidate C who strongly prefer B over A could elect B by switching their votes from C to B. But given a poll predicting the same three-way tie, there is little *ex ante* reason for these voters to switch to B: C has just as much of a chance of being tied for first with A as B does, so a switch to B could just as easily tip the result in the wrong direction.

An obvious way to compare strategic voting behavior across systems (given real uncertainty and real preferences) is to look at actual elections.[13] Evidence from plurality elections in the UK suggests that about a third of voters cast an insincere vote when the objective situation seems to call for it (Kiewiet, 2013; Fisher and Myatt, 2017), with almost half doing so when the incentive is strongest (Eggers and Vivyan, 2018). Very little research assesses the prevalence of strategic voting in Australian IRV elections (Cox, 1997, 95), though the widespread assumption appears to be that it is not common.[14] (At any rate, conclusions about IRV based on elections in Australia, which has essentially a two-party system, may not be of much use in predicting its performance elsewhere.) More research assesses strategic voting in French two-round runoff elections, which bear a close resemblance to IRV when there are three candidates;[15] the main

[12] Conitzer, Sandholm and Lang (2007) shows that one could be so uncertain about others' votes that computing the optimal vote would be computationally prohibitive.

[13] One can also turn to the lab (e.g. Blais et al., 2016; Hix, Hortala-Vallve and Riambau-Armet, 2017).

[14] See e.g. Kevin Bonham, "Oh Yes We Do Have Strategic Voting In Australia (Sometimes)", 19 October 2018 http://kevinbonham.blogspot.com/2018/10/oh-yes-we-do-have-strategic-voting-in.html, which comments on the general perception before providing a list of Australian elections (held under IRV) in which manipulation by a small group might have changed the result. Farrell and McAllister (2006) similarly points out a situation where *ex post* manipulation would be possible but does not address the prevalence of strategic voting. Analysis of Google Trends data indicates a strong uptick in attention to strategic voting around elections in the UK, Canada, and the US around elections but not in Australia.

[15] The relevant difference is that in the two-round version there is an incentive to ensure one's candidate gets a majority in the first round (absent from IRV, because a candidate who gets a near-majority of top rankings is essentially certain to win) but less cost to "push-over" voting (because unlike in IRV in the second round one can switch back to one's preferred candidate).

finding is that few voters abandon trailing candidates in the first round and fewer (if any) attempt to elect a preferred candidate by supporting someone else (Blais, 2004; Dolez and Laurent, 2010). In short, empirical studies suggest less strategic voting in IRV elections, but some of the difference may reflect differences in party systems and political culture rather than the voting system itself.

Another strand of research has explored strategic voting as a problem of choice under uncertainty in game theoretic terms (Myerson and Weber, 1993). In most such models, uncertainty arises from voters' turnout decisions (Myerson, 2002) or errors in tabulation (e.g. Laslier, 2009); because these errors are assumed to be independent and the equilibrium analysis focuses on the limit as the size of the electorate increases, in equilibrium the result of the election is essentially known and, perhaps more important for equilibrium analysis, only one type of pivot event is relevant. Another alternative approach, advocated by Myatt (2007) and Fisher and Myatt (2017) and used this in this paper, is to assume that there is aggregate uncertainty about the result (e.g. due to uncertainty about the proportion of types in the electorate) that does not disappear in large electorates. In general, game theoretic approaches emphasize the multiplicity of equilibria, not just in plurality and other scoring rules (Myerson, 2002) but also in runoff elections (Bouton, 2013; Bouton and Gratton, 2015): essentially, when any two candidates are expected to be the only serious contenders, it must be optimal to vote for one of them. This multiplicity of equilibria poses a problem for comparing voting systems, as it is not clear which one(s) to focus on. We are not aware of papers conducting equilibrium analysis of IRV.[16]

3 Our approach

We seek to estimate the prevalence and strength of the incentive to vote strategically in different voting systems given realistic assumptions about voters' preferences and beliefs. How much could voters expect to benefit from being strategic (i.e. consulting their beliefs about likely outcomes and voting accordingly) rather than simply voting according their true preferences in each system? What proportion of voters would optimally submit an insincere vote?

[16]Laurent Bouton and coauthors have a project underway including equilibrium analysis and a voting experiment, but a paper has yet not been released.

Ideally, we would address these questions with a large RCT in which we randomly assign voting systems to a large number of polities, allow time for voters and candidates to adapt to their assigned system, and then compare strategic voting incentives across systems given voters' preferences and reasonable beliefs. Obviously this heroic experiment is not possible. Furthermore, the observational equivalent to this experiment is of limited use because many voting systems of interest are not widely used (including IRV), and countries using different voting systems vary in many other ways that may be difficult to account for.

Instead, we study strategic voting incentives in plurality and IRV by carrying out a thought experiment. We use party ratings from recent election surveys as the basis for preferences. We measure strategic voting incentives for each voter in these surveys given realistic aggregate uncertainty (calibrated using recent election forecasts) and a range of expectations that we infer from a new iterative polling algorithm. (The algorithm traces out a sequence of hypothetical polls; in the first poll voters report sincere votes while in each subsequent poll a growing subset of the voters best-respond to one of the earlier polls, converging toward a strategic voting equilibrium.) We can thus measure strategic voting incentives for realistic preferences given beliefs that are consistent with those preferences and with one of a range of assumptions about how strategically *other* voters might vote.

In using preferences from surveys, our approach offers an alternative to Plassmann and Tideman (2014)'s empirically calibrated spatial models (and Chamberlin (1985)'s much earlier spatial approach) that hews even more closely to the data.[17] Our approach is complementary to complexity research in computational social science: while that strand of research assumes that voters know others' votes and asks whether they could compute the optimal vote, we assume that voters do not know others' votes but can compute the optimal vote given their (uncertain) beliefs. We also offer a new perspective on game theoretic models of elections: while our approach does not produce a general characterization of equilibria, it yields reasonable predictions (including at an equilibrium) from arbitrary preference inputs under realistic aggregate uncertainty, which permits comparison across voting systems even when multiple equilibria (or no equilibria) exist.

[17]Our overall approach is compatible with alternative sources of preferences. Plassmann and Tideman (2014)'s approach has the advantage of being parameterized.

In the remainder of this section we present our approach in more technical terms. The process of deciding on a strategic vote could be summarized in four steps: the paradigmatic strategic voter forms *preferences* about candidates, develops *beliefs* about likely election outcomes, computes *pivot probabilities* (Myerson and Weber, 1993) based on beliefs, and finally computes the *expected utility* of each possible ballot (where the optimal strategic vote is the one that yields the highest expected utility). Our presentation follows these steps in reverse: we show how to compute the expected utility of each ballot in each voting system assuming preferences and pivot probabilities; then we show how to compute pivot probabilities given beliefs about likely election outcomes; next we introduce a new iterative polling algorithm that we use to infer reasonable beliefs for a given voting system from a preference distribution and assumed level of belief precision; and finally we describe the preference distributions and precision parameter that are the empirical inputs of our analysis. Readers who understand the basic approach but are not interested in the technical details may want to skip ahead to Section 4, where we present results.

3.1 Computing the expected utility of each ballot from preferences and pivot probabilities

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 has well-defined common beliefs 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 the voter submits ballot $b \in \mathcal{B}$, 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 for each 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 the sincere ballot, which is a measure of the strategic voting incentive. Studying strategic voting incentives in any voting system given voters' preferences and beliefs is 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.

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) and an election result can 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. We assume throughout that voters consider $\mathbf{v}$ to be a continuous random variable, with beliefs summarized by pdf $f(\mathbf{v})$; this simplifies the analysis by eliminating the possibility of ties.[18] Given a total electorate of size N,[19] the probability that a single ballot

[18]Another common way to deal with ties is to assume they are broken by a pre-determined order.
[19]Thus the n voters may be a sample of the larger electorate.

could change the plurality winner from candidate j to candidate i is then

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

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

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

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

$$p_{j,i} = \pi_j + \pi_{jk} \tag{3}$$

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.[20]

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

[20]Similarly, Myerson and Weber (1993) focuses on the gain in expected utility relative to abstention.

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.[21] 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 round, for each possible way that the first round outcome could determine the winner). Then

[21] 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.

Table 1: Pivot events in IRV

Label	Type	Description	Probability
$ij.2$	Second-round	i and j tie[a] 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: [a]Here and elsewhere 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.

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.

3.2 Computing pivot probabilities from beliefs

The previous section enumerated three pivot probabilities in plurality (π_{ij} for each $i \neq j \in \{A, B, C\}$) and twelve pivot probabilities in IRV ($\pi_{ij,2}, \pi_{ij.ik}, \pi_{ij.kj}, \pi_{ij.ij}$ for each unique pair of distinct candidates i, j from $\{A, B, C\}$). If we can draw samples from the belief distribution $f(\mathbf{v})$, we can use Monte Carlo simulation to estimate each of these: most simply, draw a large number of simulated elections and count the proportion of elections satisfying the associated criteria above (e.g. equation 1 for π_{ij}). This approach has the advantage that it does not rely on a particular parametric form of $f(\mathbf{v})$, but it is computationally intensive to get precise estimates (particularly for very small pivot probabilities).

In what follows we assume Dirichlet beliefs (Fisher and Myatt, 2017), which can be parameterized by an expected result $\bar{\mathbf{v}}$ and a precision parameter s. We will write $f(\mathbf{v}; s\bar{\mathbf{v}})$ to denote the Dirichlet density with expected result $\bar{\mathbf{v}}$ and precision s evaluated at $\mathbf{v}$. (We discuss the choice of s below.) Fisher and Myatt (2017) provide an analytical expression for π_{ij} in three-candidate plurality contests given Dirichlet beliefs.[22] In Appendix Section A we show how to compute each pivot probability in IRV; as validation, we show that Monte Carlo-based estimates converge on our numerical approximation as the number of Monte Carlo simulations increases.

3.3 Inferring reasonable beliefs from preferences and uncertainty

Given a utility matrix $\mathbf{U}$ and the assumption of Dirichlet beliefs with assumed precision level s, we must choose an expected result $\bar{\mathbf{v}}$ at which to center beliefs. What is a reasonable choice? Our point of departure is to assume that $\mathbf{U}$ is representative of the larger electorate (e.g. because it comes from an election survey) and to require that the expected result $\bar{\mathbf{v}}$ be consistent with

[22] Eggers and Vivyan (2018) validate a numerical approximation when there are more than three candidates.

$\mathbf{U}$ and with a coherent account of belief formation.

Consider the simplest coherent account of belief formation: voters will vote sincerely, so that the expected result $\mathbf{v}$ is what would happen if the n voters in $\mathbf{U}$ voted for their favorite candidate (in plurality) or reported their sincere ranking (in IRV). This approach would allow us to assess strategic voting incentives for voters who believe other voters will not act strategically, which is somewhat inconsistent and therefore unsatisfying. At the other extreme we could consider an equilibrium among the voters in $\mathbf{U}$ in which each voter best-responds to the others given uncertainty about the election result in the full electorate. This approach is more consistent, in the sense that it asks about strategic voting incentives assuming others are also strategic, but it too has shortcomings: it is unrealistic to suppose that all voters would vote strategically given the costs; also, there are multiple equilibria in plurality, and equilibria have not yet been characterized in IRV.

To infer reasonable expectations given $\mathbf{U}$ and Dirichlet beliefs (with fixed precision s),[23] we imagine a series of polls in which voters with preferences described by $\mathbf{U}$ (to whom we will refer as *poll respondents*) are repeatedly asked how they plan to vote and (some) poll respondents myopically best-respond to a previous poll. In the first poll the poll respondents report their sincere preference, perhaps because they expect a neutral result; let $\mathbf{v}_1$ denote this result (i.e. the vector of ballot shares given sincere voting). Each poll result thereafter is a mix of the previous poll result and poll respondents' best response to that poll result. That is, letting $\mathbf{v}^*(\mathbf{v}_m, s)$ denote the result when the poll respondents best-respond given beliefs centered at $\mathbf{v}_m$ and precision s, the $m+1$th poll result is

$$\mathbf{v}_{m+1} = \lambda \mathbf{v}^*(\mathbf{v}_m, s) + (1-\lambda)\mathbf{v}_m, \tag{4}$$

where $\lambda \in (0, 1]$ is a mixing parameter indicating the weight on the best response.

Our iterative polling algorithm can be viewed as a method to locate an equilibrium of a voting game among the poll respondents.[24] Specifically, suppose poll respondents believe that

[23]It is not necessary for s to be fixed: one could perform our algorithm assuming that s increases from one poll to the next, for example.

[24]We plan to further theoretically ground the iterative polling algorithm in a companion paper.

the final election result $\mathbf{v}_e$ is drawn from

$$\mathbf{v}_e \sim \text{Dirichlet}(s\mathbf{v}_p), \tag{5}$$

where $\mathbf{v}_p$ is the poll result. That is, these poll respondents believe that the election result is a random variable centered on the poll result (and with randomness due to e.g. unexpected intervening events) and thus that their response in the poll helps determine the election result.[25] If the sequence of poll results converges to a fixed point where $\mathbf{v}^*(\mathbf{v}_M, s) = \mathbf{v}_M$, then $\mathbf{v}_M$ describes a Nash equilibrium of a voting game among the poll respondents: given their beliefs about how poll results produce election results (expression 5), their poll responses are best responses to each other.

The sequence of polling results can also be seen in terms of levels of rationality (Stahl II and Wilson, 1994), where a level-0 agent is non-strategic, a level-1 agent acts strategically but believes others are level-0 agents, a level-2 agent acts strategically and thinks that other agents are either level-1 or level-0 voters, and so on. In this view, the first polling result is the expected election result among level-1 poll respondents who know $\mathbf{U}$ (and believe it to be representative); the second polling result is the expected election result among level-2 poll respondents who know $\mathbf{U}$ and believe that a share λ of voters are level-1; the third polling result is the expected election result among level-3 poll respondents who know $\mathbf{U}$ and believe that a share λ of voters are level-2 and a share $\lambda(1-\lambda)$ are level-1; and so on. Thus polling results further along the sequence assume higher and higher degrees of rationality in the electorate, converging toward a fixed point where all voters are commonly understood to be highly sophisticated; λ determines the assumed mix of types at any particular stage in the sequence.

More simply, the sequence of polling results can be seen as a model of how expectations change over several polls, assuming voters are inattentive and myopic in their response to polls. They are inattentive in the sense that they do not always notice that a poll has taken place: a random fraction λ of voters notice any given poll and update their voting intention, with the

[25] This is admittedly an odd belief: the poll may affect the outcome e.g. by shaping beliefs, but presumably a single poll response does not shift the expected election result in the manner implied by expression 5. One rationale is that the poll respondents *are* the electorate and the official election result is a randomly perturbed version of the poll (e.g. due to imperfect tabulation, à la Laslier (2009). (We thank David Myatt for suggesting this interpretation.)

remaining $1 - \lambda$ not noticing and sticking by their previous voting intention. They are myopic in the sense that, if they do notice poll m, they update their voting intention assuming the expected result is $\mathbf{v}_m$, thus failing to take into account how other voters will respond to polls $m, m+1, \ldots$. In this view, λ is a measure of voters' attentiveness to polls and the length of the sequence is a measure of how many polls have taken place.

Rather than focus on strategic voting incentives at any particular point on this sequence (e.g. the sincere profile or the equilibrium), we will report results for beliefs along the whole sequence. This allows the reader to check the strategic voting incentive for voters who expect no one else to be strategic, voters who expect everyone else to be strategic (and who expect the equilibrium to be determined by something like the iterative process we describe), and voters with beliefs in between those two extremes.

3.4 Inputs: preferences and uncertainty level

The main empirical input to our analysis is a set of 160 national election surveys from 56 unique countries, collected through the Comparative Study of Election Systems (CSES) waves 1-4 (1996-2016).[26] 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. 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 $\mathbf{U}$ matrices, one for each survey. (See Eggers and Vivyan (2018) for a discussion of the suitability of party ratings as utility measures.) The average survey has just under 1,400 respondents who rate all 3 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, thus characterizing incentives for the typical citizen across the countries in the CSES.[27]

[26] 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.

[27] That is, 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

The preference distributions revealed by the CSES party ratings differ widely. Some are strongly single-peaked, with one centrist party receiving the highest or second-highest rating from almost all respondents; in others there is a "divided-majority" arrangement of preferences, with one polarizing party being rated highest or lowest by almost all respondents; in still others the pattern of preferences is more neutral, with e.g. each candidate's supporters being approximately equally likely to give each other candidate their next-highest rating.

The other empirical input is the assumed level of precision of Dirichlet beliefs, s. Rather than assuming one particular value, we conduct all of our analysis for a range of precision parameters informed by recent empirical work on strategic voting in UK parliamentary elections. As a lower limit, we use the level of precision found to be characteristic of voters by Fisher and Myatt (2017) ($s = 10$); as an upper limit we use the level of precision found to be characteristic of forecasters by Eggers and Vivyan (2018) ($s = 85$). As Fisher and Myatt (2017) points out, an observer with an uninformative Dirichlet prior over vote shares who observes an unbiased 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.[28]

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. We focus on the case with precision $s = 85$ (the level of precision associated with UK election forecasters by Eggers and Vivyan (2018)) and mixing parameter $\lambda = .05$ (the mixing parameter in the algorithm). Appendix C contains results for other parameter values; we discuss these robustness checks at many points below.

We begin by illustrating the sequence of hypothetical polling results for every CSES case.

country j in the dataset.

[28]It may therefore seem that $s = 85$ underestimates the precision of forecasters' beliefs, given that typical election surveys have sample sizes of 1000 or more. This underlines the fact that most uncertainty in forecasting is about the representativeness of polls and the influence of events that intervene between polls and the election, not sampling uncertainty.

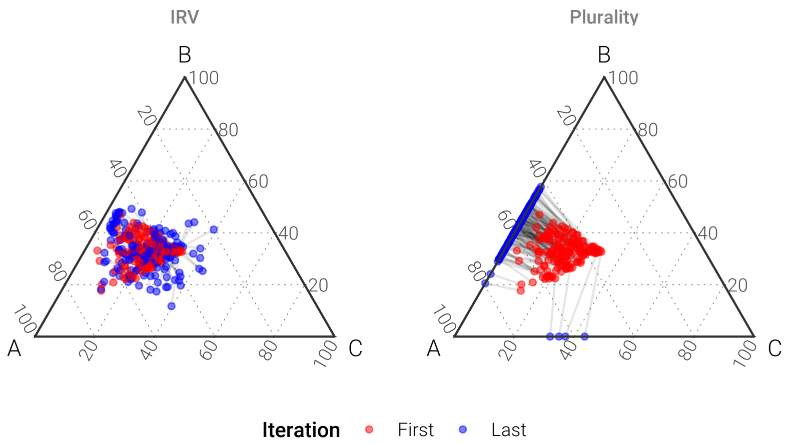


Figure 1: Evolution of ballot share vectors for all 160 CSES election surveys for both IRV (left) and plurality (right), when $s = 85$. Red dots indicate the first hypothetical poll result, blue dots indicate the 250th hypothetical poll result, and gray lines trace the path between the two.

Figure 1 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 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.

Figure 2 uses a different approach to show that convergence takes place in all CSES cases in both systems. Each line shows, for one CSES election survey, the Euclidean distance between the poll result at a given iteration and the survey respondents' best response to that poll result, i.e. the distance between $\mathbf{v}_m$ and $\mathbf{v}^*(\mathbf{v}_m)$. (When this distance is zero the algorithm has fully converged.) In both plurality and IRV the distances go to zero in all cases.

Looking closely at Figure 2 we note small oscillations in IRV. Further investigation shows

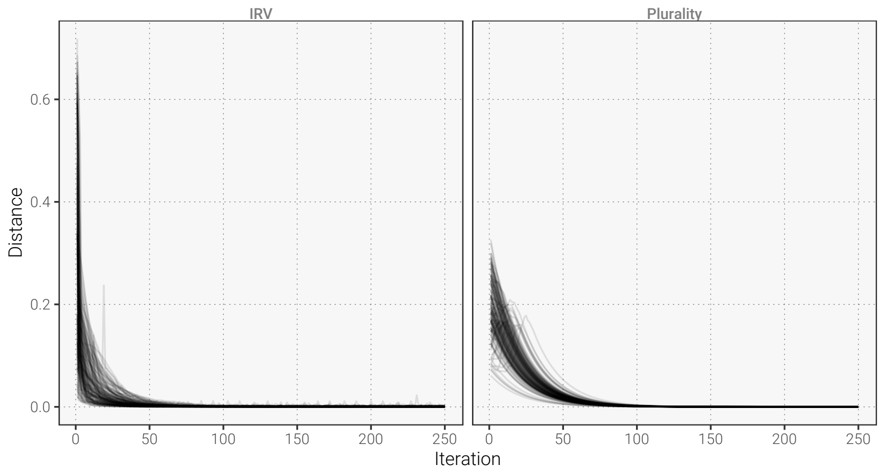


Figure 2: Distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for baseline case with high ($s = 85$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness. Note that in the baseline case, we plot distances on a linear scale to highlight convergence.

that many IRV cases undergo minor oscillations that highlight the negative feedback we will discuss further below; for example, a poll respondent or group of poll respondents choose to desert a leading candidate in poll m, but this desertion decreases the candidate's lead and causes the same respondent(s) to return to the candidate in poll $m+1$.[29] As Figure 18 in the Appendix shows, increasing λ from 0.05 to 0.1 increases the magnitude of these oscillations and (as in plurality) speeds up convergence but does appear to change the destination of the algorithm.

In plurality, the precision parameter s affects how quickly the algorithm converges to a Duvergerian equilibrium: higher precision makes a vote for the trailing candidate more obviously ineffective and thus speeds desertion of this candidate in favor of the leaders. (See Figures 8 and 9 in the Appendix.) In IRV, by contrast, s noticeably affects the location of the equilibrium (as shown by Figure 18). The underlying 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 plurality equilibria a single pivot probability (the probability of a tie between the two frontrunners) dominates all of the others; the choice of s affects how many hundreds or thousands of times larger a tie for first between the frontrunners is than any other tie, but this could only affect the optimal vote for a voter who is nearly indifferent between the frontrunners. Thus a result where two candidates

[29] It may be many more polls before the previous desertion is "forgotten", causing these voters to desert again.

receive all or essentially all of the votes will be a plurality equilibrium at a wide 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 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 single equilibrium. Figure 3 illustrates this for four CSES cases (Australia in 2013, France in 2012, the UK in 2015, and Germany in 2005). 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. (There are discrepancies after 250 iterations, but Figure 20 in the Appendix shows for all CSES cases that all paths from random starting points are still converging toward the endpoint of the baseline algorithm that starts from the sincere profile.) The apparent uniqueness of equilibrium in IRV across 160 cases strongly suggests (though of course cannot prove) the uniqueness of equilibria in IRV and deserves more study.

4.2 Strategic voting incentives

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 utility matrix $\mathbf{U}$ from the CSES. Combining preferences and beliefs as described in Section 3 we can then compute for each CSES case the expected utility matrix $\overline{\mathbf{U}}$ (in its normalized form) for beliefs centered at each expected result in the sequence.[30] We now turn to summarizing what the results say about the incentive to vote strategically in plurality and IRV.

We begin by defining the benefit of strategic voting. Let $\overline{\mathbf{u}}_i$ denote the expected utility

[30] In practice, we compute $\overline{\mathbf{U}}$ at each step of the iterative polling algorithm.

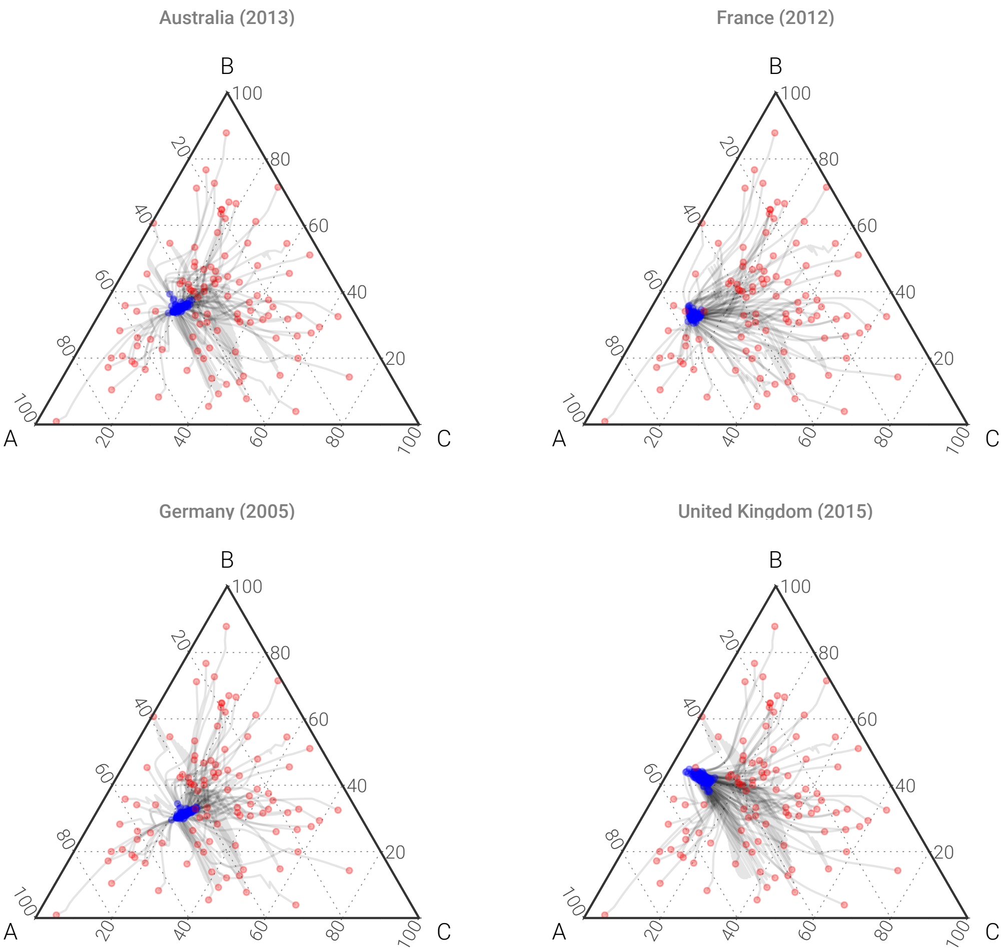


Figure 3: Path of first-preference shares along iterative polling algorithm for select cases in IRV (with baseline parameter values $s = 85$, $\lambda = 0.05$, initialised at random starting points in the vote ternary. Red dots indicate starting point, blue dots indicate first-preference shares after 60 iterations.

vector for voter i (i.e. a row of $\overline{\mathbf{U}}$), with $\overline{u}_i^*$ indicating the maximum of that vector and $\overline{u}_i^s$ indicating the expected utility of the sincere ballot, i.e. the one that is most consistent with the voter's sincere preferences. Then we can define the benefit of strategic voting as

$$\beta_i = \overline{u}_i^* - \overline{u}_i^s.$$

This is the maximum amount of expected utility that the voter can gain by voting strategically (i.e. in light of beliefs about likely outcomes) rather than simply submitting a sincere ballot. Note that β_i is positive if the optimal vote is insincere and zero otherwise.

In Figure 4 we summarize β_i across our 220,000 CSES respondents in plurality and IRV in three different ways. The plots in the left column focus on $E[\beta_i]$ computed within CSES cases (thin lines) and across all CSES respondents (thick lines, weighted as described in footnote 27) separately for plurality (orange) and IRV (blue) at each of the first 60 iterations[31] of the polling algorithm (horizontal axis) and different values of s ($s = 10$ on top, $s = 55$ in middle, $s = 85$ at bottom). Note that β_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 would expect 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 the expected benefit of strategic voting is substantially lower in IRV than in plurality. At $s = 10$ (approximately the level of belief precision Fisher and Myatt (2017) ascribe to UK voters), the benefit 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 $s = 85$ (approximately the level of belief precision Eggers and Vivyan (2018) ascribe to UK election forecasters), 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, near the sincere profile the expected benefit of strategic voting in plurality is around 5 times larger than in IRV (higher at lower s); by the 60th iteration the ratio of expected benefits ranges from around 25 (for $s = 85$) to about 35 (for $s = 10$).

To better understand patterns in the expected benefit of strategic voting, we decompose it into the *magnitude* of the benefit (i.e. how much benefit is there for voters who would benefit from an insincere vote?) and the *prevalence* of the benefit (i.e. what proportion of voters would

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

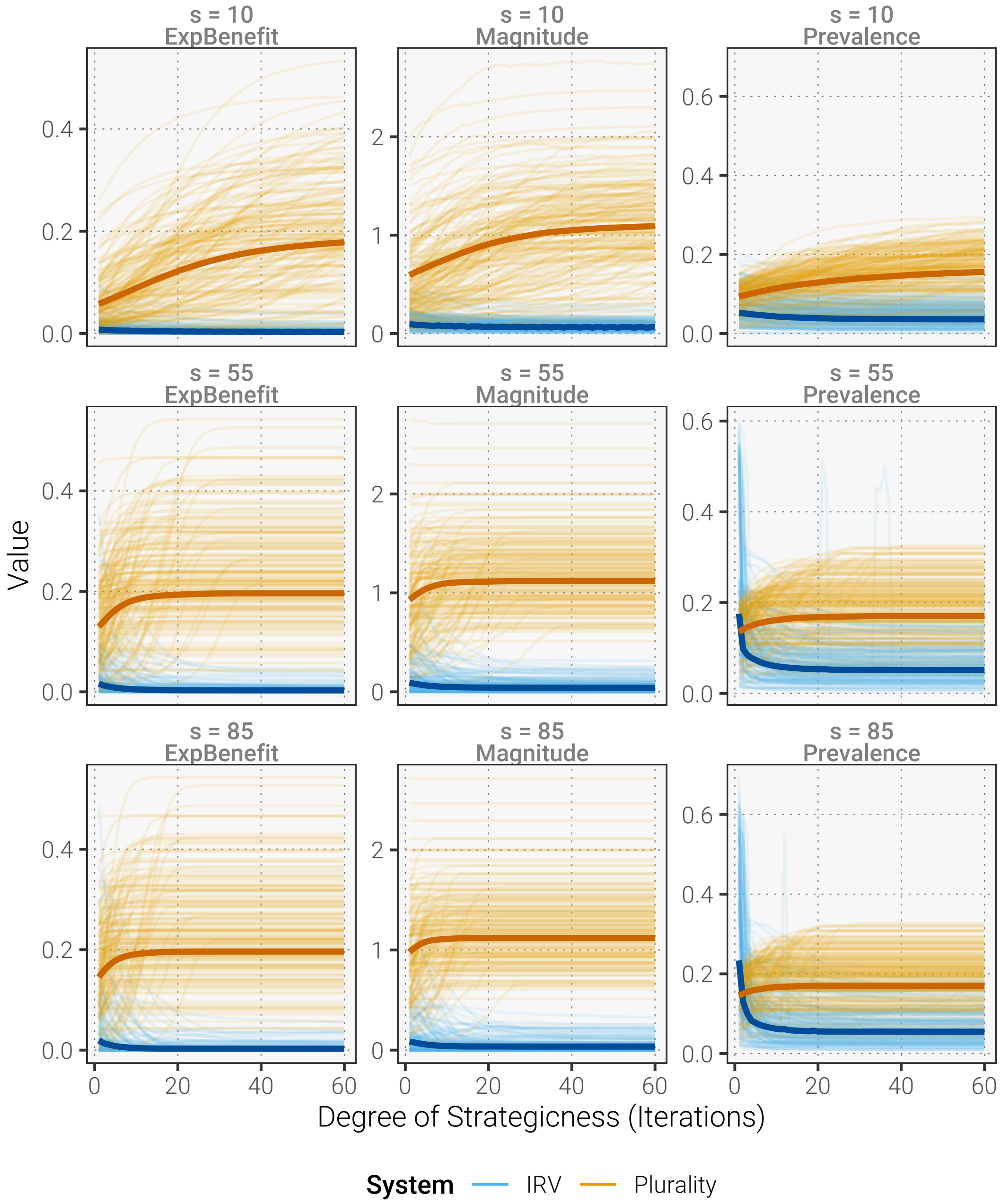


Figure 4: Expected benefit, magnitude, and prevalence of strategic voting

benefit from an insincere vote?). More formally,

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

Thus the magnitude corresponds to the intensive margin of β_i and the prevalence corresponds to the extensive margin of β_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 differences in the expected benefit of strategic voting: voters who expect to benefit from an insincere vote do so by less on average in IRV (magnitude) and there are fewer voters who expect to benefit from an insincere vote in IRV (prevalence). Note, however, that in IRV the prevalence near the sincere profile is fairly high: one-fifth or more of voters optimally submit an insincere vote for $s = 55$ and $s = 85$ when they believe other voters will vote sincerely, which is higher than the equilibrium prevalence in plurality. (The magnitude in both cases is much lower in IRV than in plurality.) This indicates that a strategic voting enthusiast who has fairly precise beliefs and expects other voters in an IRV election to vote sincerely stands a good chance of identifying a (slightly) beneficial insincere vote. But if the same voter expects other voters to do the same, such opportunities vanish, which suggests that strategic voting can never be widespread in IRV. In contrast, the incentive to vote strategically in plurality increases as others respond strategically to polls. We view this difference (the preponderance of negative feedback in IRV vs. positive feedback in plurality) as the fundamental theoretical insight coming from our analysis.

4.3 Why is IRV more resistant to strategic voting than plurality?

We emphasize two factors that help explain why IRV is more resistant to strategic voting than plurality, beginning with role of negative vs. positive feedback just mentioned.

Strategic voting shows negative feedback in IRV and positive feedback in plurality

As noted above, the results in Figure 4 suggest that positive and negative feedback – i.e. dependence of voters' strategic voting incentives on the perceived strategic behavior of other voters – plays an important role in explaining the difference in strategic voting incentives between IRV and plurality. Our iterated polling algorithm can be seen as tracing out a sequence of beliefs about others' strategic-ness: it starts with the belief that others will be purely sincere and converges toward a strategic voting equilibrium. Our results indicate that strategic voting incentives increase in plurality along this sequence and decline in IRV. Why is the case?

The bandwagon logic of strategic voting in plurality is well-understood (e.g. Cox, 1997). If candidate i is trailing in the poll, then candidate i will trail even more when other voters respond to the poll; thus if my best naive response to the poll is to abandon i in favor of my second choice, my incentive to abandon i in favor of my second choice is likely to be even larger when I take into consideration other voters' responses to the poll. Strategic voting in plurality (or more precisely strategic responses to polls in plurality) is thus characterized by positive feedback.[32]

Strategic voting in IRV, by contrast, has a stronger tendency toward negative feedback. We note two components of negative feedback in IRV. The first, and more transparent, is that supporters of the leading candidate may desert that candidate to help the more vulnerable opponent (the "pushover") advance, but as more and more supporters of the leading candidate do this, the leading candidate's lead evaporates (and along with it the incentive to help the pushover advance). The second component is more subtle: efforts by supporters of the leading candidate to help the pushover advance, or by supporters of the pushover to switch to a candidate better able to defeat the leading candidate, tend to make the pushover less of a pushover. Let k be the leading candidate and let i be the pushover, so that $ij.kj$ is the dominant first-round pivot event. What makes i the pushover is that, conditional on a first-round tie for second occurring between i and j, we expect the proportion of i's ballots (i.e. those ranking i first) that rank

[32] Myatt (2007) emphasizes the role of negative feedback in a model where voters facing a strong coordination incentive receive private signals about which candidate is stronger. Feedback in Myatt (2007) refers not to how voters respond to a public poll but how they respond to their private signal about the popularity of candidates: if other voters respond more to their signal (i.e. vote for the candidate they perceive to be more popular) then it is less important than I do.

k second to be smaller than the proportion of j's ballots that rank k second; thus k will get more of a boost in the second round if j is eliminated than if i is eliminated. As kij and kji types tactically rank i first (to help elect k), and as voters with ijk preferences tactically rank j first (to avoid electing k), the expected proportion of i's ballots ranking k second increases and the proportion of j's ballots that rank k second decreases. Thus the strategic responses to a discrepancy in the pattern of lower rankings between i and j tend to eliminate that discrepancy. Together, these two components contribute to the overall pattern of negative feedback evident in Figure 4, in which the incentive to vote strategically in IRV is lower when other voters are more responsive to polls.[33]

This negative feedback in IRV provides a new perspective on Dummett (1984)'s comment that "a voter who has understood the workings of the procedure, and who has some information about the probable intentions of the others, will have nearly as much incentive to vote strategically" in IRV as in plurality. In light of our analysis, it is true that a strategic voting enthusiast might find ample opportunities for strategic voting in IRV: observing a detailed poll on others' vote intentions, a highly informed voter who thinks herself to be the only strategic actor may often recognize chances to do better in expectation by casting an insincere vote. But negative feedback suggests that this strategic voting incentive must remain limited to a small proportion of the electorate, because the opportunities to strategically respond to a poll disappear when one perceives that others will be doing so.

Insincere ballots are less likely rewarded in IRV, and benefits and costs more positively associated

Another way to account for the difference in strategic voting incentives we observe is to examine the probability of events that benefit and punish insincere votes. Our analysis indicates that, on average across voters, events that reward insincere ballots are less likely in IRV than in plurality; also, the benefits and costs of an insincere vote tend to be more positively related in IRV, indicating that the potential benefits of an insincere vote are counteracted more often by

[33] Not all feedback is negative in IRV. For example, when $ij.ij$ is relevant (so leading candidate k expected to lose in the second round), then a top ranking for k is "wasted" and k may lose all support in a dynamic similar to desertion of trailing candidates in plurality.

risks. Both observations contribute to the lower average benefit of strategic voting in IRV.

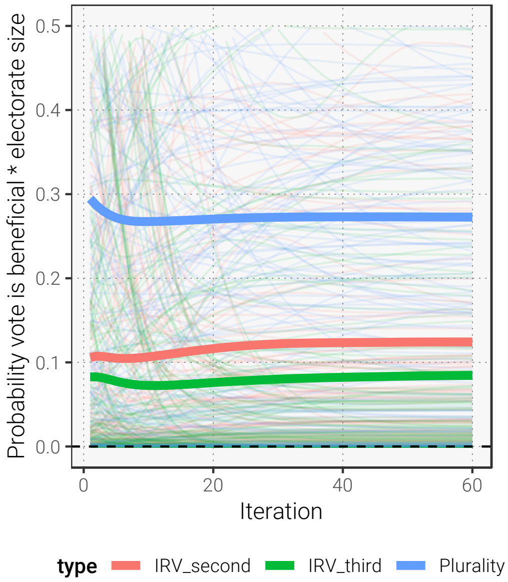
It should not be too surprising that events benefiting insincere votes are less likely in IRV, given previous work on manipulability reviewed in Section 2. The blue line in Figure 5 shows the average probability[34] of ties for first between one's second and third choices in plurality across the first 60 iterations of the algorithm. (For voters whose first choice is A, for example, this is π_{BC}.) The probability dips down over the first several iterations as support for C erodes, making it less likely that A and B supporters would benefit from an insincere vote. The red line in Figure 5 shows the average probability of pivot events that reward a ballot ranking one's second choice first in IRV; the blue line does the same for pivot events that reward a ballot ranking one's third choice first in IRV.[35] The two IRV lines are considerably below the plurality line, indicating that, averaging across voters, the probability of a pivot event rewarding an insincere ballot is higher on average in plurality than in IRV. This exercise is similar in spirit to manipulability results in Chamberlin (1985) and Plassmann and Tideman (2014), except that our assumptions about preferences and beliefs (i.e. the distribution over results) are different.

The right plot in Figure 5 goes further than previous manipulability results by showing that the rewards of insincere voting are also counterbalanced by costs to a greater degree in IRV than in plurality. Having computed the probability of an insincere vote being beneficial for each CSES respondent, we also computed the probability of the same insincere vote being harmful[36] and then computed the correlation in these two probabilities across voters at each iteration. On average, this correlation is negative for plurality (blue line), suggesting that voters who are more likely to see a tie for first between their second and third choice are less likely to see a tie for first involving their first choice. In IRV the correlation is mostly zero for ranking one's second choice first (red line) but positive for ranking one's third choice first (green line); this suggests

[34]For generality, we compute probabilities multiplied by electorate size in each CSES, which allows the reader to compute values for a particular electorate size; alternatively we could use the actual electorate size in each setting.

[35]For a voter with preference order ijk, a ballot ranking j first is rewarded at events $ij.kj$ (where a sincere vote elects k but a ji ballot elects j), $jk.jk$ (where a sincere vote is wasted but a ji ballot elects j), and $jk.ik$ (where a sincere vote is wasted but a ji ballot elects i); a ballot ranking k first is rewarded at events $ij.kj$ (where a sincere vote would elect k but a ki ballot is wasted) and $jk.ji$ (where a sincere vote is wasted but a ki ballot elects i).

[36]For an ijk type in plurality, voting for j is harmful (compared to a sincere vote) at events ij and ik. For an ijk type in IRV, ranking j first is harmful at all first-round pivot events where it is not beneficial plus second-round pivot event $ij.2$; ranking k first is harmful at all first-round pivot events where it is not beneficial plus second-round pivot event $ik.2$.



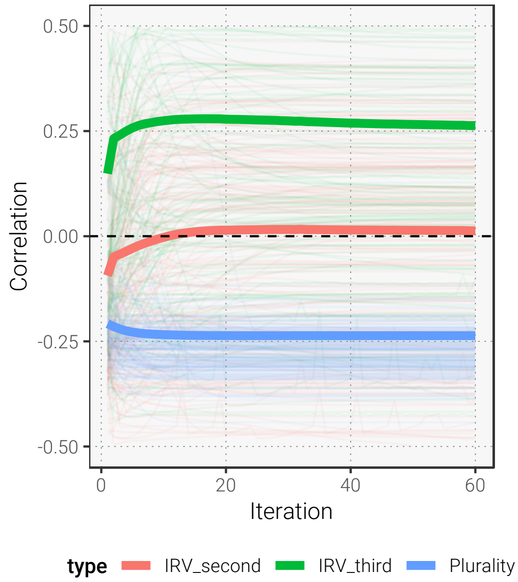


Figure 5: Non-normalized probability that each insincere vote would be rewarded (left) and correlation between probability that each sincere vote would be rewarded and punished (right).

that voters who are more likely to be able to elect their first choice by ranking their third choice first are also more likely to see this tactic backfire when e.g. their first choice and third choice tie in the second round. To make sense of this difference between IRV and plurality, consider that an insincere vote is rewarded only when there is relatively even competition across the three candidates: we require a first-round pivot event to take place, which involves a first-round tie for second such that at least one of the trailing candidates can defeat the leader in the second round. But other pivot events that would punish an insincere vote are also more likely when support is fairly balanced across candidates. Insincere votes can backfire in plurality too, of course, but in plurality insincere votes can be rewarded even when such backfiring is a very remote possibility (as when one candidate is far behind the others). The higher correlation between benefits and costs in IRV (shown by Figure 5) suggests that it is not just the proliferation of pivot events (and the resulting complexity of the decision-making process) but also the *density* of pivot events (and the resulting conflicting incentives) that curb strategic voting in IRV.[37]

[37]This analysis also highlights why measuring the probability of a manipulable result gives an incomplete picture of susceptibility to strategic voting.

5 Conclusion

Our analysis indicates that previous conjectures by Cox (1997) and Renwick (2011) were correct: IRV creates lower incentives for strategic voting than plurality does. We have shown this to be true for typical preferences in recent national elections using a variety of assumptions about the predictability of election results and the prevalence of strategic voting. We suggest that the incentive to vote strategically is lower in IRV partly because strategic voting in that system is subject to negative feedback: typically, the more I expect others to respond strategically (and myopically) to a poll, the less I should do so. This contrasts with the well-known bandwagon effect in plurality elections, where the more other voters desert my preferred candidate the stronger is my incentive to do so. We also find that pivot events that reward insincere ballots are less common in IRV than in plurality and that the costs and benefits of an insincere vote tend to be more positively correlated in IRV.

Our analysis is based on a new framework for measuring and comparing strategic voting incentives across voting systems that can be extended to handle other voting systems, more candidates, different preference data, and different assumptions about how strategic behavior affects beliefs. The framework allows for empirical preference distributions as inputs (e.g. from electoral surveys) and realistic assumptions about aggregate uncertainty. Much work remains, however, to assess the empirical realism of the iterative polling algorithm as a model of elections and compare the equilibria it identifies to analytical results.

References

Bartholdi, John J, Craig A Tovey and Michael A Trick. 1989. "The computational difficulty of manipulating an election." *Social Choice and Welfare* 6(3):227–241.

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.

Chamberlin, John R. 1985. "An investigation into the relative manipulability of four voting systems." *Behavioral Science* 30(4):195–203.

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 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(1):1–15.

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

Eggers, Andrew C and Daniel Rubenson. 2019. "Who votes more strategically in Canada?".

Eggers, Andrew C and Nick Vivyan. 2018. "Who votes more strategically?".

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.

Fisher, Stephen D and D Myatt. 2017. "Strategic voting in plurality rule elections." *Mimeo* .

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 et al. 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.

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.

Kiewiet, D Roderick. 2013. "The ecology of tactical voting in Britain." *Journal of Elections, Public Opinion and Parties* 23(1):86–110.

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.

Miller, Nicholas R. 2017. "Closeness matters: monotonicity failure in IRV elections with three candidates." *Public Choice* 173(1-2):91–108.

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.

Nitzan, Shmuel. 1985. "The vulnerability of point-voting schemes to preference variation and strategic manipulation." *Public choice* 47(2):349–370.

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.

Plassmann, Florenz and T Nicolaus Tideman. 2014. "How frequently do different voting rules encounter voting paradoxes in three-candidate elections?" *Social Choice and Welfare* 42(1):31–75.

Pons, Vincent and Clémence Tricaud. 2018. "Expressive voting and its cost: Evidence from runoffs with two or three candidates." *Econometrica* 86(5):1621–1649.

Renwick, Alan. 2011. The alternative vote: a briefing paper. Technical report.

Reny, Philip J. 2001. "Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach." *Economics Letters* 70(1):99–105.

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.

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 Computing 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: $(v_1, v_2, \dots, v_i + v_j, \dots v_B) \sim \text{Dir}\big(\alpha_1, \alpha_2, \dots, \alpha_i + \alpha_j, \dots \alpha_B\big)$. (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: $v_i \sim \text{Beta}(\alpha_i, \sum_{-i} \alpha)$. (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: $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_B \,|\, v_i) \sim (1-v_i)\text{Dir}(\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_B)$. (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.)

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: If we say that two candidates tie when their vote share differs by less than half a vote,[38] then the probability of second-round pivot event

[38]This is equivalent to saying that we calculate the probability of specific results by rounding continuous vote shares to the closest multiples of $1/N$.

ab can be written

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

This can be factorized as

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

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

$$\int_{s=\frac{-1}{2N}}^{\frac{1}{2N}} \int_0^{\frac{1}{2}} f\left(y - x/2, \frac{1}{2} - y - x/2, \frac{1}{2} + x; s\bar{v}_a, s\bar{v}_{ca}, s(\bar{v}_b + \bar{v}_{cb})\right) dy\, dx$$

which is approximately

$$\frac{1}{N}\int_0^{\frac{1}{2}} f\left(y, \frac{1}{2} - y, \frac{1}{2}; s\bar{v}_a, s\bar{v}_{ca}, s(\bar{v}_b + \bar{v}_{cb})\right) dy.$$

(The approximation is exact if the density is flat in the immediate neighborhood of second-round ties between a and b.) 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 $v_a + v_{ca} = \frac{1}{2}$,

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

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 $s\bar{v}_{cb}$ and $s\bar{v}_b$. 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

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

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

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

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

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

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

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

Putting together the above, we have

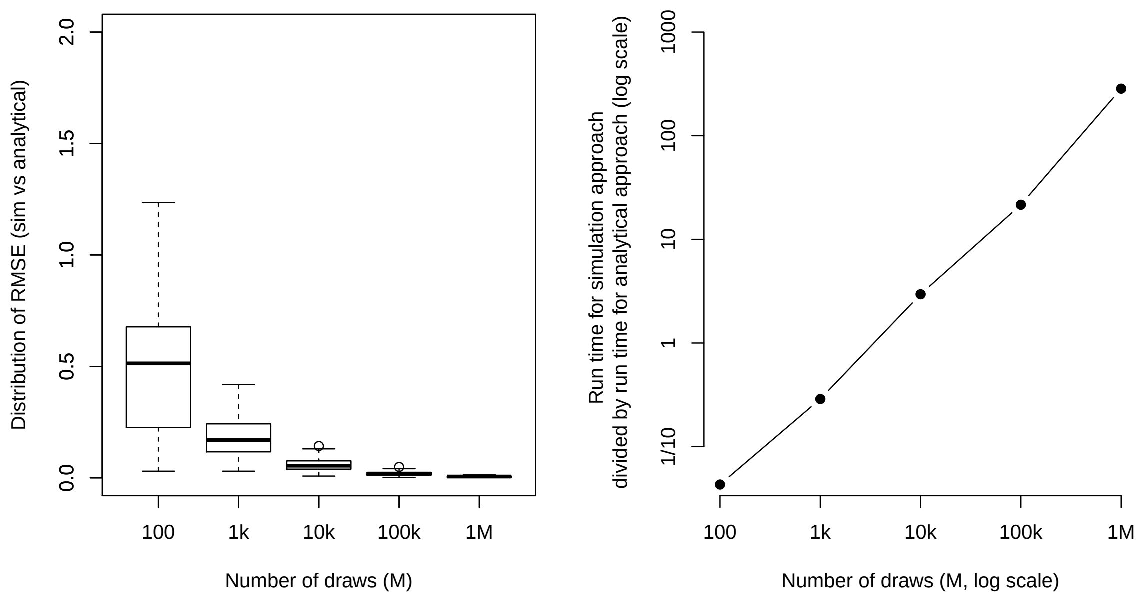
$$\begin{aligned} N\pi_{ab.ab} \approx & \int_{\frac{1}{4}}^{\frac{1}{3}} f\Big(z, z, 1-2z; s\overline{v}_a, s\overline{v}_b, s\overline{v}_c\Big) \times \\ & \int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z}, \frac{z-x}{z}; s\overline{v}_{bc}, s\overline{v}_{ba}\right) dx \times \int_0^{2z-\frac{1}{2}} f\left(\frac{x}{z},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

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

Figure 6: 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

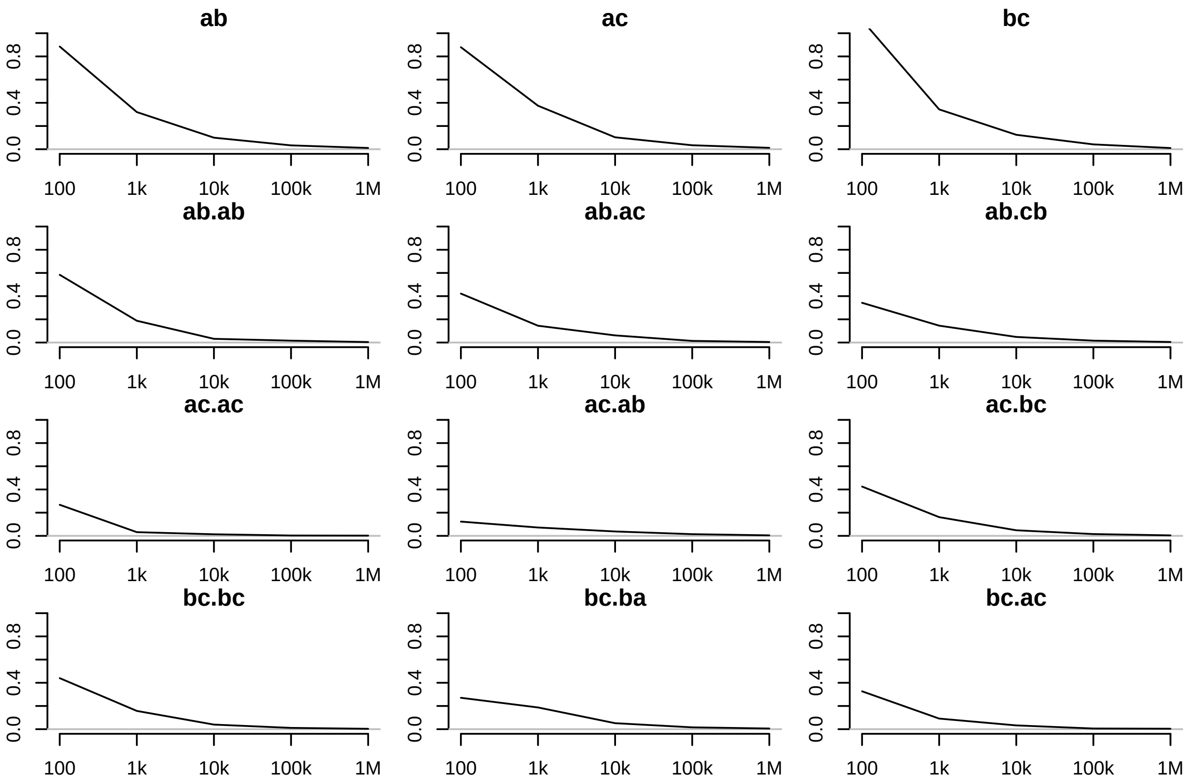
each of these J scenarios there are 12 pivotal probabilities to compute. Let $\mathbf{T}$ denote the $J \times 12$ matrix of pivotal probabilities computed with our numerical approach, and let $\tilde{\mathbf{T}}_M$ denote the $J \times 12$ matrix of pivotal probabilities computed with our simulation 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 7 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.[39]

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 7 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.

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

Figure 7: 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.

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 **??** 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 inbetween 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.[40]

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 add a very small value to the expected utility of voters' sincere votes, such that $\mathbf{EU} = \mathbf{UP}(\bar{\mathbf{v}}, s) + 10^{-10}\mathbf{S}$
2. Any remaining cases where two values in a row are seen as tied for maximum by the computer are resolved in favour of sincerity.

[40] For example, for a voter with sincere preference ABC, voting ABC or ACB amounts to the same 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.

C Convergence of 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 unique equilibrium of ballot shares. In plurality, we can infer the equilibrium behaviour from the vote share paths in Figure 1 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.[41]

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, with the exception of oscillations as mentioned in footnote 20 (for the least converged cases, see Appendix E to see that the patterns of change in best responses behave in regular patterns). 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 IRV ballot share vector (with six items) after the j^{th} iteration and in CSES case k. For example, $\bar{\mathbf{v}}_{0,\text{AUS2013}}$ denotes the ballot shares if every voter in the 2013 Australian sample voted sincerely. Similarly, let $\mathbf{v}_{j,k}^{BR}(s,\lambda)$ denote the ballot shares of everyone's best response in that iteration.

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 Figure 2 and Appendix C.1, as:

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

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.[42]

Convergence onto a fixed point. In Appendix C.1, we report the distance between voters' best response and the result of the algorithm in the previous iteration, and show that the convergence behaviour 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 a poll and a lagged average of the algorithm (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.

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

[42] 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 whether convergence of the polling algorithm also occurs under different parameter values for precision (s) and strategicness (λ). Analogous to Figure 2, we present, for each iteration j and every case k, the distance between the ballot shares of best responses given certain parameter values, to the ballot shares after 250 iterations (using the same parameter values). In contrast to Figure 2 in the main body, however, 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 ($\lambda = 0.05$)

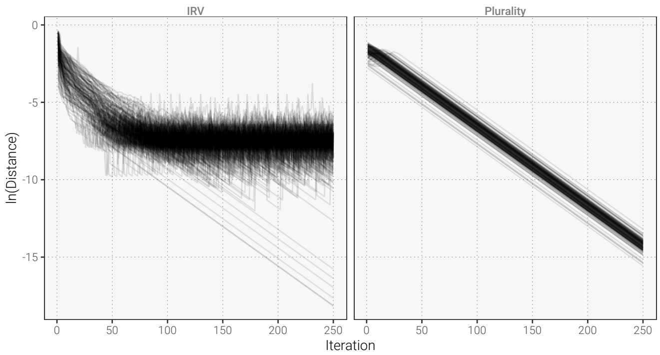


Figure 8: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

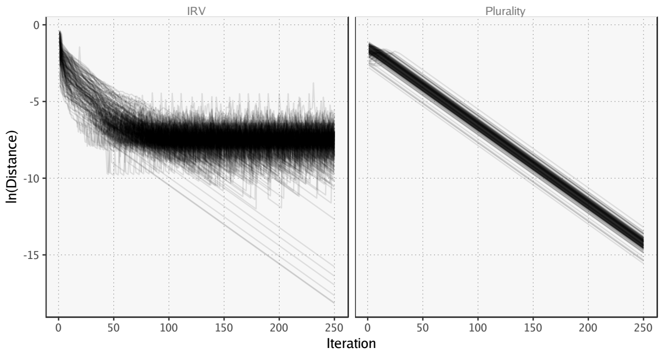


Figure 9: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

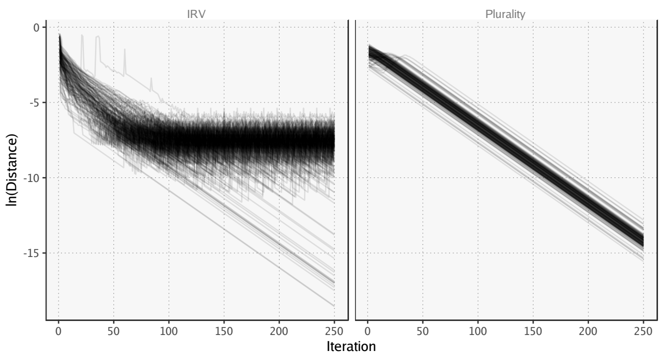


Figure 10: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and medium ($\lambda = 0.05$) strategic responsiveness.

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

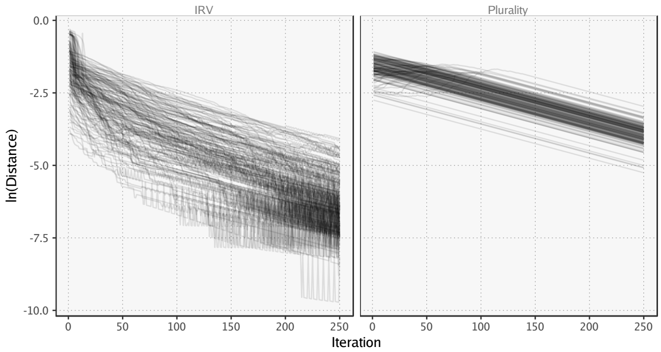


Figure 11: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

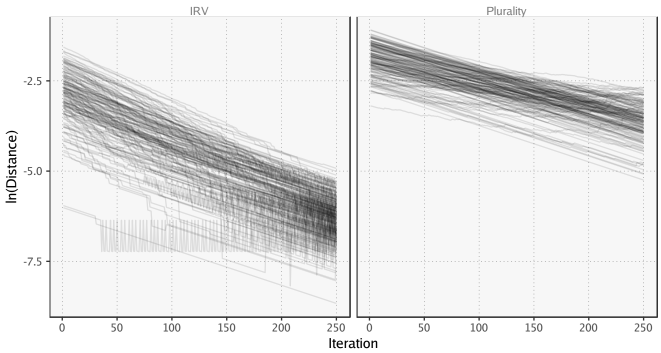


Figure 12: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

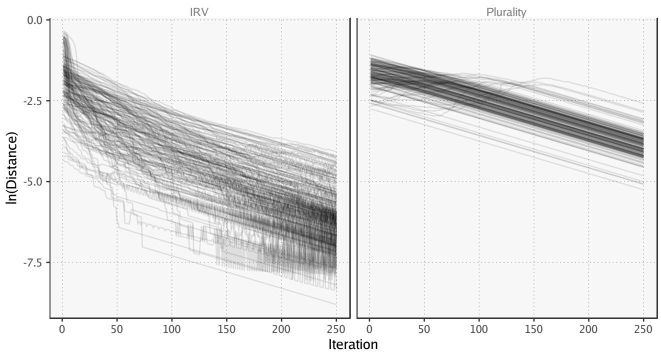


Figure 13: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and low ($\lambda = 0.01$) strategic responsiveness.

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

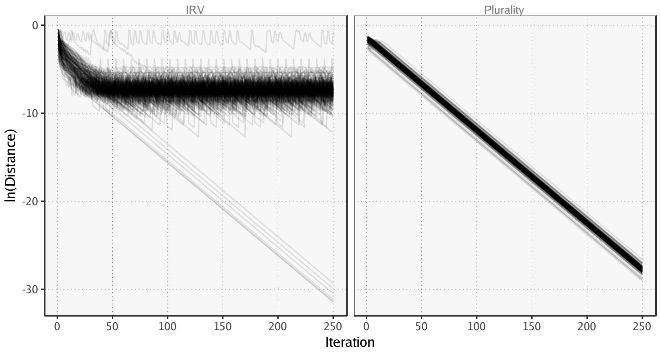


Figure 14: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for high ($s = 85$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

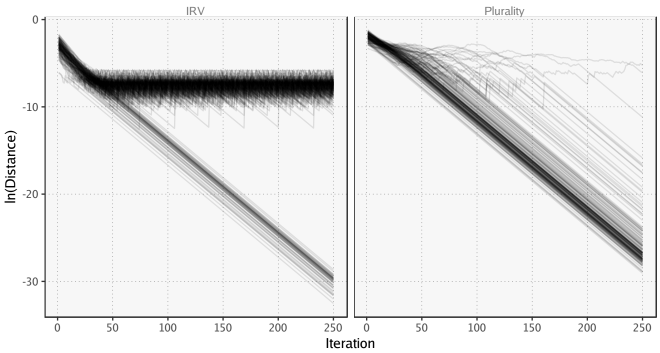


Figure 15: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for low ($s = 10$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

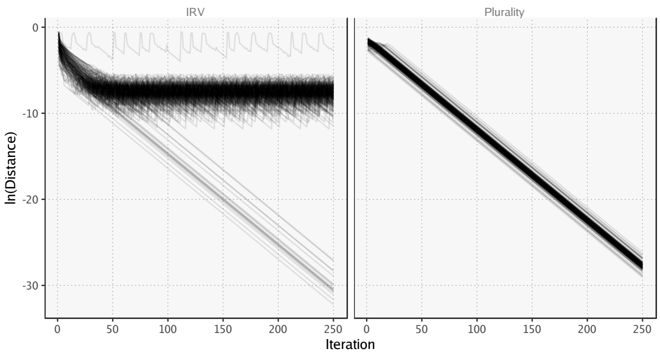


Figure 16: Logged distance between the the shares of voters' best responses after have been given a poll in IRV (left) and Plurality (right), and the vote shares in the previous iteration's poll. Results for medium ($s = 55$) belief precision and high ($\lambda = 0.1$) strategic responsiveness.

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 17) between the resulting best response ballot shares after the poll at time j of the algorithm, and the average of vote shares between the polls at time $j-20$ 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 (in logged values).[43] 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.

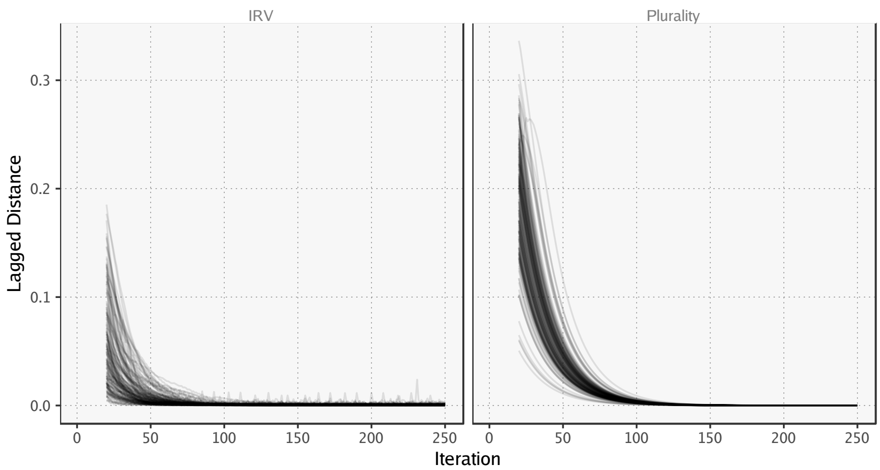


Figure 17: Distance between the shares of voters' best responses after voters have been given a poll in IRV (with $\lambda = .05$ responding strategically), and the average of the respective vote shares 10 to 20 iterations ago.

[43]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 18), as well as after the 250th poll in the case with the same s, but holding $\lambda = 0.05$ (Figure 19).

Formally, the quantities of interest are:

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

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 18), the equilibrium is robust to the choice of λ: when comparing distances across different values of λ, but holding s fixed (Figure 19), we see differences in how quickly the algorithm converges (which is what λ determines by definition), but, ultimately, the distances converge towards zero.

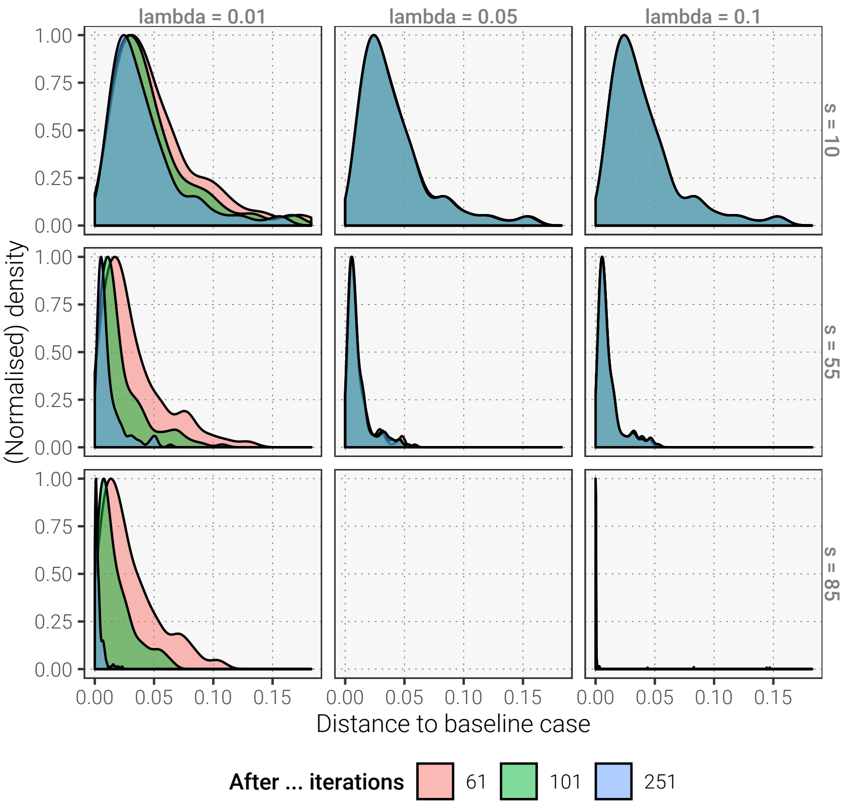


Figure 18: Distribution of Euclidean distances across CSES cases between ballot shares in jth iteration under given parameter combination (information precision, s, and strategic responsiveness, λ) compared to 250th iteration in the baseline case ($s = 85$, $\lambda = 0.05$).

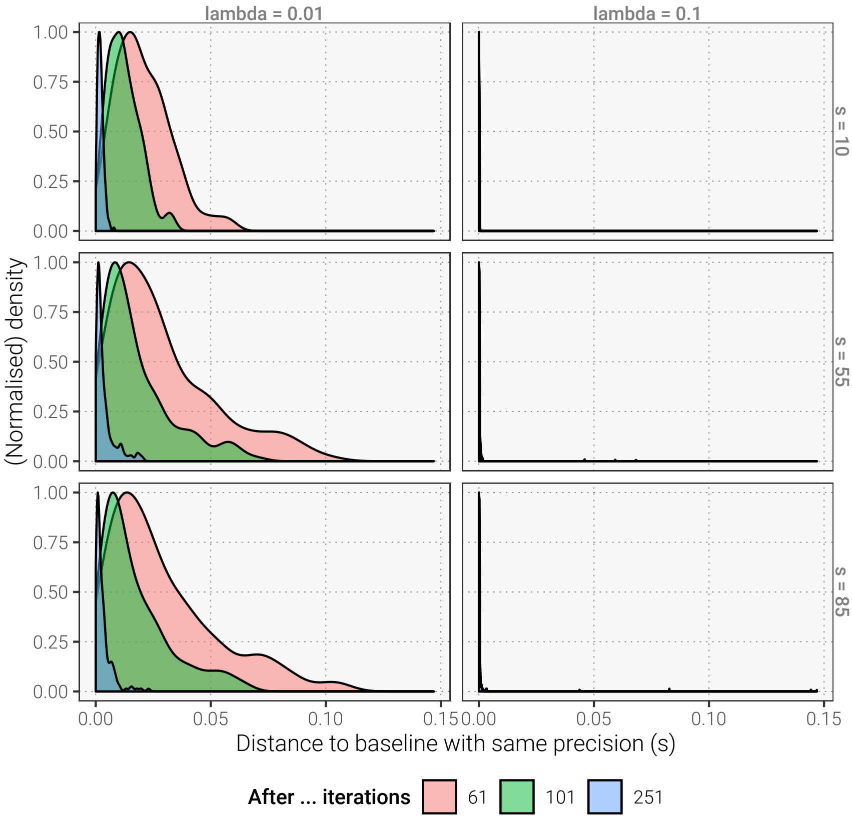


Figure 19: Distribution of Euclidean distances across CSES cases between ballot shares in jth iteration under given parameter combination (information precision, s, and strategic responsiveness, λ) compared to 250th iteration in the case with same s but $\lambda = 0.05$.

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 initialise 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 initialised with the values s, λ, and a starting belief about ballot shares centred 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 centred on $\mathbf{v}_q$, and the ballot share after the 250th iteration where the algorithm was initialised 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 20 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.

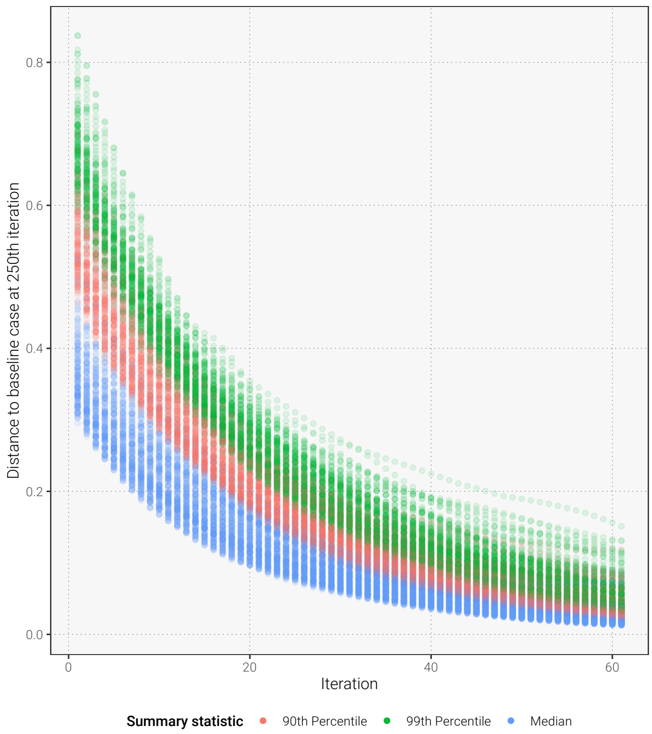


Figure 20: Summary of distances between case-specific distributions of distances between ballot shares after iterations from random starting points, and the converged ballot shares in the baseline case

D Robustness of expected benefit results to parameter values

D.1 Medium strategic responsiveness ($\lambda = 0.05$)

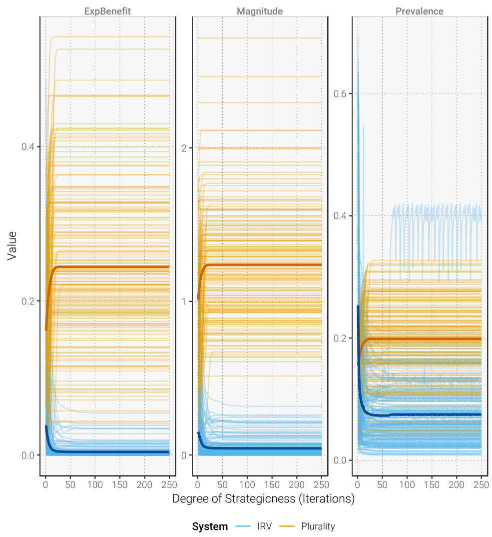
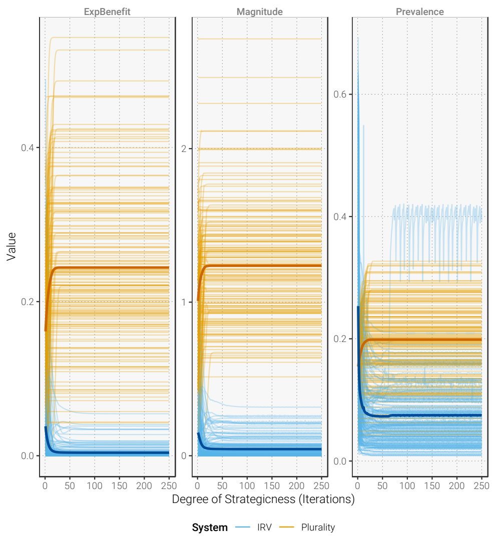


Figure 21: Expected benefit, magnitude, and prevalence of strategic voting with high ($s = 85$) belief precision, and low strategic responsiveness ($\lambda = 0.05$).



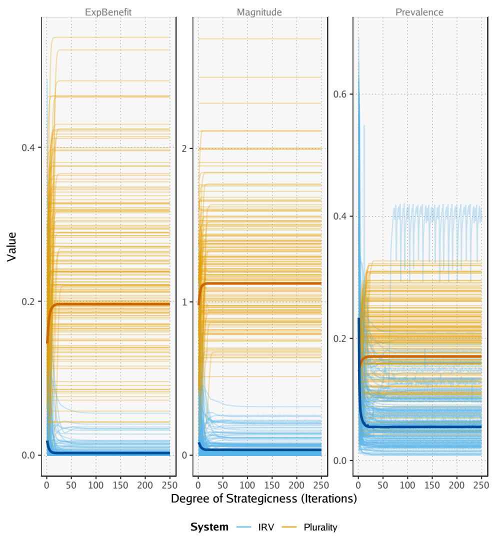


Figure 22: Expected benefit, magnitude, and prevalence of strategic voting with low ($s = 10$, left) and medium ($s = 55$, right) belief precision, and low strategic responsiveness ($\lambda = 0.05$).

D.2 Low strategic responsiveness ($\lambda = 0.01$)

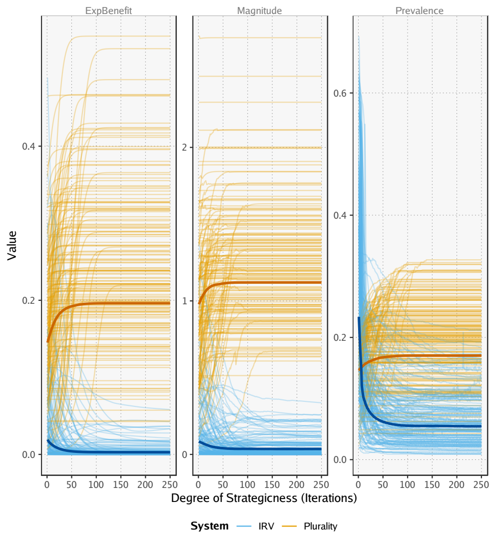


Figure 23: Expected benefit, magnitude, and prevalence of strategic voting with high ($s = 85$) belief precision, and low strategic responsiveness ($\lambda = 0.01$).

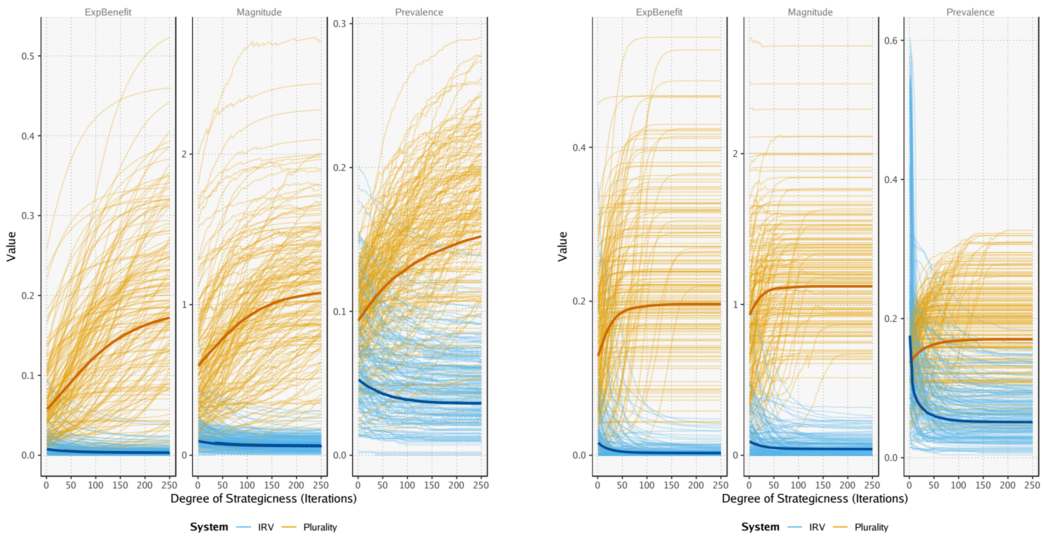


Figure 24: Expected benefit, magnitude, and prevalence of strategic voting with low ($s = 10$, left) and medium ($s = 55$, right) belief precision, and low strategic responsiveness ($\lambda = 0.01$).

D.3 High strategic responsiveness ($\lambda = 0.10$)

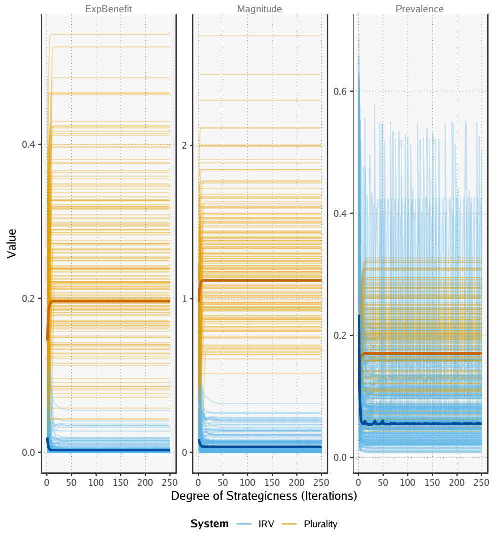


Figure 25: Expected benefit, magnitude, and prevalence of strategic voting with high ($s = 85$) belief precision, and high strategic responsiveness ($\lambda = 0.10$).

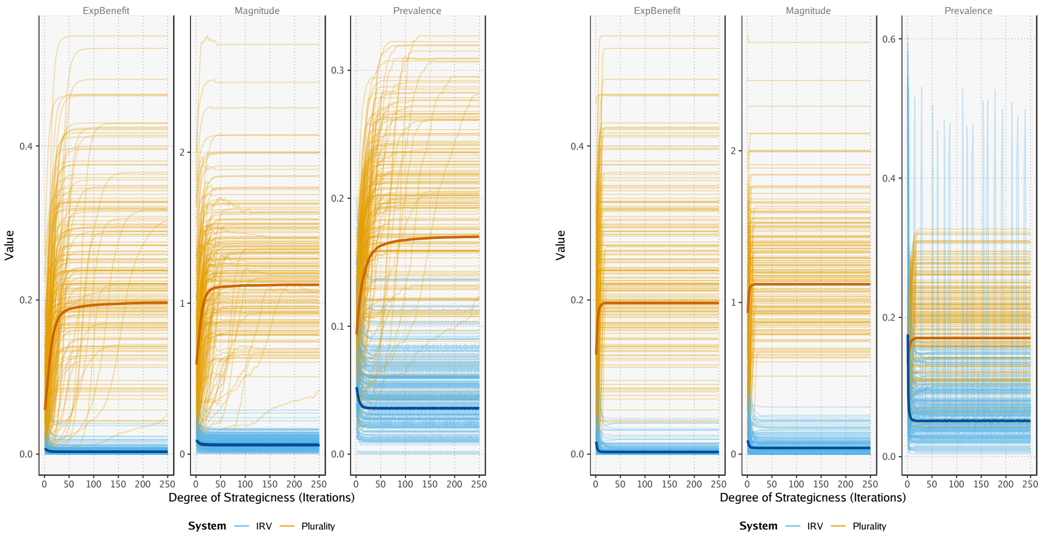


Figure 26: Expected benefit, magnitude, and prevalence of strategic voting with low ($s = 10$, left) and medium ($s = 55$, right) belief precision, and high strategic responsiveness ($\lambda = 0.10$).

E Case-specific strategic voting behavior

Finally, Appendix E offers a case-by-case examination of the prevalence of strategic voting for all cases where the distance between the best response and the outcome of the previous iteration in the baseline case never falls below the value of 0.001.[44]

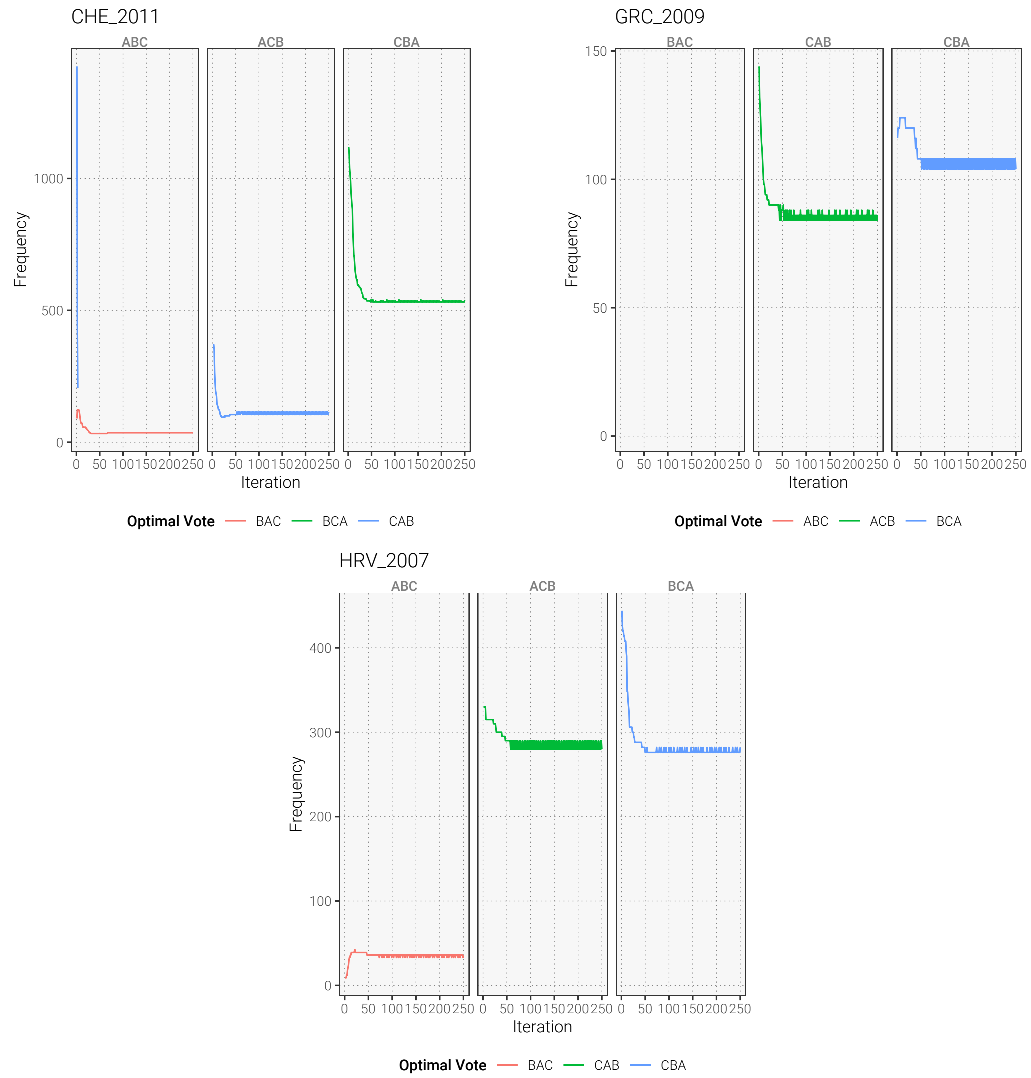


Figure 27: Frequency of optimal strategic votes by case and voter type (sincere preferences) under IRV, as the polling iteration algorithm proceeds. Selected cases shown whose iteration-to-iteration Euclidean distance between vote shares did not converge below a threshold of 0.001 in the baseline case of $s = 85$ and $\lambda = 0.05$. Cases are Czech Republic 2011 (top left), Greece 2009 (top right) and Croatia 2007 (bottom).

[44] Appendix E only reports the three qualifying cases obtained when $s = 85$ and $\lambda = 0.05$.