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. We use this approach to compare strategic voting incentives in three-candidate plurality (FPTP) and instant-runoff (IRV) elections, drawing on preference data from 160 electoral surveys. We show that the incentive to vote strategically is lower in IRV than FPTP overall and on both the intensive and extensive margins. We also show that the difference is bigger when the electorate is assumed to be more strategic: the more other voters respond strategically to polls, the greater the rewards of voting strategically in FPTP and the smaller the rewards of doing so in IRV. This previously unexplored negative feedback constrains the prevalence and magnitude of strategic voting incentives in IRV.

[1]This version: June 18, 2019. Thanks to Paul Klemperer, David Myatt, Karine van der Straeten, Dimitri Landa, and seminar audiences at Trinity College Dublin, the Toulouse School of Economics, and the ECPR Standing Group on Strategic Interactions for valuable feedback.

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

†Stanford University. email: tobias.nowacki@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 first-past-the-post system of election (FPTP, also known as plurality) 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 FPTP 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] FPTP 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, and mayors of Sydney, London, San Francisco and several other American cities.

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, Deputy Prime Minister Nick Clegg testified in advance of the 2011 UK 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

[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", HC 358-i, 22 October 2010 (link). Testimony given before the Political and Constitutional Reform Committee on 15 July 2010.

lower in IRV than in FPTP. Cox (1997, 95) observes that more information is needed to vote strategically in IRV than in FPTP, as does Renwick (2011, pp. 6-7), who concludes that IRV would "reduce but not eliminate incentives for tactical voting" compared to FPTP. 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 FPTP.

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 FPTP, 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 FPTP 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).[7] 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. Research on the *ex post* manipulability of electoral systems thus has little to say about the *ex ante* incentives voters face, and thus does not allow us to compare incentives for strategic voting across voting systems.

In this paper, we propose a general framework for assessing strategic voting incentives in the presence of uncertainty and use it to compare strategic voting incentives in FPTP 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.[8] Next, we 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 one of a range of assumptions about other voters' strategic inclination. This approach allows us to measure, for each voter and a given assumption about how strategic *other* voters are expected to be, the expected benefit of strategic voting compared to sincere voting and, ultimately, to compare this benefit across different voting systems.

We find that, consistent with some of the conjectures noted above, the incentive to vote strategically is considerably lower in IRV than in FPTP. This is true for any level of uncertainty

[7] A closely related set of papers assesses the probability of results that violate monotonicity and other desirable properties of choice rules (Plassmann and Tideman, 2014; Ornstein and Norman, 2014; Miller, 2017).

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

(within the range of empirically plausible values that we consider) and no matter how widespread strategic behavior by other voters is (within the model of strategic responses to polls that we consider). Decomposing the expected benefit of voting strategically, we observe that IRV is more resistant to strategic voting both because the probability of benefiting from a strategic vote is lower and because the magnitude of that benefit (when it exists) is smaller. We also find that the probability of an insincere vote being beneficial is lower on average in IRV than in FPTP and that costs and benefits are more positively correlated across voters in IRV, suggesting that incentives are more conflicting in IRV. We find that the expected benefit from strategic voting is lower in IRV regardless of how strategic other voters are expected to be, but the gap between FPTP and IRV is larger the more strategic other voters are expected to be. This highlights an under-appreciated qualitative difference between strategic voting incentives in FPTP and IRV: strategic voting is subject to positive feedback in FPTP (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 emphasize that our focus in this paper is on the *incentive* to vote strategically given reasonable uncertainty about others' votes, 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, which previous researchers have noted is more complicated to manipulate than FPTP and other voting systems.[9] 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

[9] In a framework where voters are certain about each others' votes, Bartholdi, Tovey and Trick (1989) use computational complexity theory to assess the difficulty of manipulating an outcome in various voting systems; Bartholdi and Orlin (1991) argues that manipulation in IRV is uniquely complex. Cox (1997) and Renwick (2011) also note that strategies in IRV are more logically complicated than strategies in plurality, which suggests that some voters may find it too difficult to even attempt.

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 A framework for measuring strategic voting incentives

What does it mean to say that one voting system creates greater strategic voting incentives than another? Such conjectures are commonly made but rarely made precise. In this section we offer a definition of strategic voting incentives and describe in informal terms a tractable framework for measuring and comparing these incentives across voting systems. The next section formalizes the framework, and in subsequent sections we apply it to compare strategic voting incentives in FPTP and IRV elections.

We consider an election in which many voters will collectively choose a single winner from a set of several candidates or alternatives; this could be an election for the leader of a country or organization, an election for a representative to a legislative body, or the choice of a policy alternative from several possibilities. We assume that voters have preferences over which candidate wins and that these preferences can be formulated as von Neuman-Morgenstern *utilities*, i.e. scores that indicate not just preferences over candidates but also, when combined with probabilities and converted to expected utilities, track preferences in the presence of uncertainty.[10] We also assume that voters have beliefs about how the election may turn out, which we characterize as a distribution over possible election results.

We highlight two distinct approaches to voting. *Sincere voting* means casting the ballot that is most consistent with one's preferences regardless of the circumstances. (For example, sincere voting in FPTP means always voting for one's favorite candidate.) *Strategic voting* means casting the ballot that yields the maximum expected utility, given one's beliefs about possible election results. The expected utility from a particular ballot ultimately depends on the probability of the winner depending on a single ballot in various ways, given one's beliefs.

[10]More formally, let $\mathbf{u}$ be a vector of utilities over alternatives, and let $\mathbf{p}_1$ and $\mathbf{p}_2$ be any pair of lotteries over these alternatives. $\mathbf{u}$ is a von Neumann-Morgenstern utility function for an agent if and only if $\mathbf{u} \cdot \mathbf{p}_1 \geq \mathbf{u} \cdot \mathbf{p}_2$ whenever the agent weakly prefers the first lottery ($\mathbf{p}_1$) over the second ($\mathbf{p}_2$).

In some circumstances the expected-utility maximizing ballot is the one that is most consistent with one's preferences, but not always (such as when one is better off voting for one's second choice in a FPTP contest because one's first choice is very unlikely to be in the running for first place).

Given a particular voter with preferences over candidates and beliefs about possible election outcomes, we can define the **benefit of strategic voting** as

$$\begin{array}{c}\text{Benefit of strategic}\\ \text{voting}\end{array} = \begin{array}{c}\text{Expected utility from}\\ \textbf{strategic}\text{ voting}\end{array} - \begin{array}{c}\text{Expected utility from}\\ \textbf{sincere}\text{ voting.}\end{array}$$

Because strategic voting by definition maximizes expected utility, the benefit of strategic voting cannot be negative; it is positive when strategic voting and sincere voting imply different votes and zero otherwise. The benefit of strategic voting measures the extent to which a voter would be better off undertaking strategic voting (i.e. casting the expected-utility-maximizing vote) as opposed to simply sincere voting (i.e. reporting one's sincere preference).

Given preferences and beliefs for a collection of voters and a means of computing the expected utility of each possible ballot, it is straightforward to compute each voter's benefit of strategic voting (as defined above) and characterize this by e.g. taking the average. The key question is what to assume about the inputs, i.e. the joint distribution of preferences and beliefs: what combination of preferences and beliefs are characteristic of a given voting system?

To measure the preferences and beliefs that typical of a given voting system, we would ideally force a large random subset of polities to adopt that voting system, allow time for voters and candidates to adapt to the system, and measure the joint distribution of preferences and beliefs for all voters across polities forced to use the system. By forcing different random subsets to adopt different voting systems (i.e. running an RCT), we could straightforwardly compare the average benefit of strategic voting across systems. Obviously this heroic experiment is not possible. Furthermore, the observational equivalent to this experiment is of limited use: no single-winner voting system other than perhaps FPTP is used widely enough that we can credibly infer a joint distribution of preferences and beliefs typical of that system. Given the impracticability of both the heroic experiment and its observational equivalent, how can we compare strategic voting benefits across voting systems?

Our approach is to use empirical preference data from a collection of national electoral surveys[11] and infer reasonable beliefs that are consistent with the preference data and empirical evidence about the predictability of election results. In each survey, the voters are asked to rate the main parties competing in the election on a 0-10 scale; we use these ratings as proxies for utilities. We treat this data as a representative sample of preference distributions for which strategic voting incentives should be measured. Of course these preferences are collected in conjunction with elections held under different voting systems, and party entry and position-taking (and therefore voter preferences) are undoubtedly a function of the voting system. We use the same distribution of preferences (or, more precisely, the same distribution of distributions of preferences) for each system, and thus readers should interpret our comparison of strategic voting incentives across systems as capturing "first-order" differences, before parties respond to the incentives created by the different systems.[12] Put more simply, our analysis ultimately asks: what are typical strategic voting incentives given preference distributions like the ones observed in recent national elections?[13]

We next need to specify beliefs about likely election outcomes. Needless to say, we do not have a direct measure of what each survey respondent believes would happen if an election were held in their country under each possible voting system. Instead we attempt to infer beliefs for each survey (not each voter) that are reasonable given the preferences in the survey and empirical evidence about how well voters can predict election results.

As the basis for our model of beliefs, we follow Fisher and Myatt (2017) in using a Dirichlet distribution, which can be parameterized simply as an expected result (i.e. an expected proportion of voters casting each distinct ballot) and a precision parameter. The Dirichlet distribution has additional properties that make it easy to work with; these properties are discussed in the Appendix where we show how pivot probabilities are estimated for IRV elections.

Next, we need to choose a precision parameter to capture how predictable election results are. Rather than assuming one particular value, we conduct all our analysis for a range of precision

[11]We use the Comparative Study of Electoral Systems (CSES) http://www.cses.org, described further below.

[12]Notably, strategic voting incentives in plurality elections could cause minor parties to exit political competition, which drives strategic voting incentives to zero.

[13]In our main results we weight a voter in country j by the population of country j divided by the number of surveys we have from country j; we thus try to make our results representative of the typical voter in recent elections, rather than the typical recent election.

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); as an upper limit we use the level of precision found to be characteristic of forecasters by Eggers and Vivyan (2018).[14]

Finally, we must specify the expected election result for each election survey in each voting system. Here we encounter a thorny problem. As a basic principle, we want to specify an expected result that makes sense given the distribution of preferences in the survey. To do so, however, we need to make assumptions about how voters vote, given their preferences. The simplest possibility is to assume that voters vote sincerely, in which case the expected result can be read directly from the election survey. (For example, in FPTP the expected vote share for party a would simply be the proportion of survey respondents who most prefer a.) Once we allow strategic behavior, however, we need to make more assumptions in order to infer an expected result from a distribution of preferences.

Our approach is to imagine a sequence of vote-intention polls among the voters in each election survey.[15] In the first poll the voters report their sincere preference. In the second poll a random fraction of voters report their optimal vote given beliefs centered at the first poll (which reported sincere preferences), while the rest continue to report a sincere vote; in the third poll, a new random fraction of voters report their optimal vote given beliefs centered at the second poll (which was a combination of sincere preferences and strategic responses to the first poll), while the rest report the vote they reported in the second poll; and so on. This iterative process may eventually converge on a strategic voting equilibrium. Rather than evaluating strategic voting incentives at this equilibrium or another particular point in this sequence, we compute strategic voting incentives for beliefs centered at each of the hypothetical polls in the sequence. This may allow us to satisfy readers with different views on the likely extent of strategic behavior, but it also allows us to show how strategic voting incentives depend on the strategic orientation of other voters which, as we show below, is an important and under-appreciated difference between FPTP and IRV elections.[16]

[14]Writing the Dirichlet parameters as $s\mathbf{v}$, where s is precision and $\mathbf{v}$ is location, the low value is $s = 10$ and the high value is $s = 85$.

[15]The algorithm is described more technically in the next section.

[16]One could alter this process by making voters more sophisticated in the way that they respond to polls. For

Section summary

We define the benefit of strategic voting for a voter as the difference in the voter's expected utility from strategic voting and sincere voting, given the voter's preferences and beliefs. Because the benefit of strategic voting depends on preferences and beliefs, any measure of the benefit of strategic voting in a given voting system (or comparison across systems) depends on what is assumed about the distribution of voter preferences and how voters form beliefs about likely election outcomes given those preferences. Our approach uses party ratings from recent national election surveys to represent typical distributions of preferences. For each survey and for an assumed level of belief precision, we trace out a sequence of possible expected results that might emerge as voters respond strategically to a series of polls. Different readers may be interested in the benefit of strategic voting at different points along this sequence; comparing benefits at two points along the sequence allows us to characterize the dependence of the benefit of strategic voting on the assumed prevalence of strategic behavior.

3 The framework: technical version

We consider an election with K candidates and B admissible ballots, where B depends on the voting system in operation. (In FPTP, $B = K$; in IRV or another ordinal system, $B = K!$.) Let $\mathbf{v} = \{v_1, v_2, \ldots, v_B\}$ denote an election result, with element v_b indicating ballot b as a share of all ballots. Let $f(\mathbf{v})$ denote beliefs about possible election results. We assume beliefs can be specified in terms of an expected result $\overline{\mathbf{v}}$ and a precision parameter s, as in the case of the Dirichlet distribution (Fisher and Myatt, 2017).

To compute the optimal strategic vote given preferences and beliefs, it is useful to convert beliefs into a matrix containing the probability of each candidate being elected as a function of the marginal ballot. In particular, let $\mathbf{P}$ denote a $K \times B$ matrix

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,B} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,B} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K,1} & \cdots & \cdots & p_{K,B} \end{bmatrix}$$

example, we might allow that some voters anticipate that other voters will respond strategically to the poll, or we might allow voters to coordinate in limited ways with other voters.

where element $p_{k,b}$ indicates the probability of candidate k being elected when the marginal ballot is b. Abusing standard matrix notation slightly, let $\mathbf{P}(\overline{\mathbf{v}}, s)$ denote the $\mathbf{P}$ matrix for a particular voting system given beliefs centered at $\overline{\mathbf{v}}$ with precision parameter s.

Let $\mathbf{U}$ denote an $N \times K$ matrix of von Neumann-Morgenstern utilities for N voters (indexed by i). (In our empirical analysis, $\mathbf{U}$ is the matrix of party ratings from a single voter survey.) Note that

$$\overline{\mathbf{U}} \equiv \mathbf{UP}(\overline{\mathbf{v}}, s)$$

is an $N \times B$ matrix of expected utilities, with element $\overline{u}_{i,b}$ denoting the expected utility for voter i from casting ballot b (given beliefs centered at $\overline{\mathbf{v}}$ with precision s).

Let $\mathbf{S}$ denote the sincere vote matrix, which is an $N \times B$ matrix with element $s_{i,b}$ being 1 if ballot b most closely reflects voter i's sincere preference and 0 otherwise. Let the function $g(\cdot)$ convert an arbitrary matrix into a vector containing the "row maximums" of the matrix; that is, if $\mathbf{X}$ is an $I \times J$ matrix with representative element $x_{i,j}$, then

$$g(\mathbf{X}) = \{\max_j x_{1,j}, \max_j x_{2,j}, \ldots, \max_j x_{I,j}\}.$$

Then the benefit of strategic voting for each of the N voters, given beliefs centered at $\overline{\mathbf{v}}$ and precision parameter s, can be written as the N-length vector

$$\mathbf{b}\left(\overline{\mathbf{v}}, s\right) = g\left(\mathbf{UP}(\overline{\mathbf{v}}, s)\right) - g\left(\mathbf{S} \circ \mathbf{UP}(\overline{\mathbf{v}}, s)\right),$$

where $\circ$ indicates the Hadamard (or entrywise) product. In words, $\mathbf{b}\left(\overline{\mathbf{v}}, s\right)$ is the difference in expected utility for each voter between a strategic vote and sincere vote, given beliefs centered at $\overline{\mathbf{v}}$ and precision parameter s. For voters with preferences represented by $\mathbf{U}$ and beliefs represented by $\overline{\mathbf{v}}$ and s, the mean of $\mathbf{b}$ is a measure of the expected benefit of strategic voting.

To generate beliefs that are reasonable given $\mathbf{U}$ and s, we introduce an iterative polling algorithm that yields a sequence of expected results $\{\overline{\mathbf{v}}_1, \overline{\mathbf{v}}_2, \ldots\}$ that might be observed as voters respond strategically to a series of vote-intention polls. Let the function $h(\cdot)$ convert an arbitrary matrix into a vector, the first element of which indicates the proportion of rows in which the first column contains the row maximum, the second element of which indicates the proportion of rows in which the second column contains the row maximum, etc; that is, if $\mathbf{X}$ is an $I \times J$ matrix with representative element $x_{i,j}$ and $\mathbb{1}\{\cdot\}$ is the indicator function, then

$$h(\mathbf{X}) = \frac{1}{N}\left\{\sum_i \mathbb{1}\{x_{i,1} = g(\mathbf{X})_i\}, \sum_i \mathbb{1}\{x_{i,2} = g(\mathbf{X})_i\}, \ldots, \sum_i \mathbb{1}\{x_{i,J} = g(\mathbf{X})_i\}\right\}.$$

Then our iterative polling algorithm starts with

$$\overline{\mathbf{v}}_1 = h(\mathbf{S}),$$

which is the proportion of voters for whom each ballot is the sincere vote, and at each subsequent step j we have

$$\overline{\mathbf{v}}_j = \lambda h\left(\mathbf{UP}(\overline{\mathbf{v}}_{j-1}, s)\right) + (1-\lambda)\overline{\mathbf{v}}_{j-1}$$

where λ determines the extent to which voters in one poll respond to the previous poll. The algorithm can be said to converge when $|\overline{\mathbf{v}}_j - \overline{\mathbf{v}}_{j-1}| < \varepsilon$.

Given preferences $\mathbf{U}$ and an assumed precision level s, all that is necessary to compute strategic voting incentives (across the sequence of hypothetical polls) for a given voting system is a way to compute $\mathbf{P}(\overline{\mathbf{v}}, s)$. This can be done with analytical, numerical, or simulation methods; the Appendix provides details for the case of FPTP and IRV.

4 Strategic voting incentives in FPTP and IRV elections: qualitative characterization

We now turn from the general framework to the specific voting systems examined in this paper. Our objective is to provide intuition and concepts that further motivate the comparison and help in interpreting the results on FPTP and IRV that we present below.

As is well known (e.g. Cox, 1997; Fey, 1997), strategic voting in FPTP elections involves abandoning trailing candidates in order to avoid wasting one's vote. If candidates A, B, and C compete in a FPTP election, and C is expected to trail far behind the other two, then (for reasonable specifications of beliefs) it is far more likely that A and B will tie for first than that C

will tie with either A or B for first. This implies that a vote for C is "wasted", in the sense that it is much less likely to determine the election outcome than a vote for A or B, and a strategic voter maximizes expected utility by voting for either A or B. For voters in this situation whose first preference is C, sincere voting implies voting C while strategic voting implies voting for their second choice (A or B).

In IRV (as its advocates like to point out), there is no need to abandon a hopeless candidate. Recall that voters rank the candidates in IRV elections; in the case of just three candidates, the candidate with the fewest top rankings is eliminated, after which the winner is the remaining candidate ranked higher on the majority of all ballots. If candidates A, B, and C compete in an IRV election, and it is known that C will finish last in top rankings, then a ballot that ranks C first is not "wasted": after C is eliminated, a ballot that ranks the candidates CBA is a vote for B just as much as any other ballot that ranks B above A.

IRV thus eliminates the need to abandon a trailing candidate to avoid wasting one's vote, but it creates other strategic voting incentives that do not exist in FPTP. Suppose, for example, that C is expected to finish last in top rankings but there is some chance that B and C will tie for second, such that either candidate could be eliminated. Suppose further that voter preferences are expected to be fairly single-peaked, with B a centrist candidate, meaning that most ballots that rank A or C first are expected to rank B second, while ballots that rank B first are expected to rank A or C second in equal proportion. This arrangement of preferences implies that if B and C tie for second, B is much more likely to defeat A than C is. This situation creates incentives for strategic voting. A voter with sincere preference order CBA may realize that, if B and C tie for second, the likely effect of a sincere CBA vote is to eliminate B and thereby cause the election of A; this suggests that a vote of BCA would be better. A voter with sincere preference ABC or ACB, on the other hand, may realize that, if B and C tie for second, a sincere vote is wasted (i.e. it does not determine whether B or C is eliminated), and a vote of CAB would be better: by helping to eliminate B, such a vote makes it more likely that A wins.

In short, while the logic of strategic voting in FPTP ("abandon a trailing candidate to avoid wasting one's vote") does not apply in IRV, two other distinct logics of strategic voting do apply. The first one, illustrated by the CBA type in the scenario in the previous paragraph, is "abandon a trailing candidate to avoid helping one's least-favored candidate".[17] The second one, illustrated by the ABC and ACB types in that scenario, is "abandon a leading candidate to avoid wasting one's vote".

This discussion of strategic voting incentives in FPTP and IRV illiustrates that strategic voting in IRV is more complex and requires more information than strategic voting in FPTP, as observed by Cox (1997) and Renwick (2011). There are more ways to vote in an IRV election than in an FPTP election with the same number of candidates (e.g. with three candidates, six vs. three). Working out how a single ballot could affect the outcome from a tally of ballots submitted requires one step in FPTP (check whether any pair of candidates is tied for first) but at least three steps in IRV (convert the tally into a tally of first-preference votes, see whether any pair of candidates is tied for second in first preferences, and reconvert to determine who would win depending on which candidate is eliminated). This complexity has been shown formally in the computational social choice literature (Bartholdi and Orlin, 1991) and is in part the basis for Cox (1997)'s and Renwick (2011)'s conjectures that there would be less strategic voting in IRV than in FPTP.

Supposing this complexity were no obstacle (e.g. because strategic voting advice services were available), what can we say about the extent to which voters might expect to benefit from strategic voting in IRV compared to FPTP? Ultimately the answer depends on assumptions about voters' preferences and beliefs; the framework introduced above and applied below offers a principled basis for such assumptions. With that caveat, we offer a few observations that guide our empirical analysis below.

First, it seems reasonable to expect that the probability of events that reward insincere votes is lower in IRV than in FPTP. As illustrated above, an insincere vote is rewarded in IRV when two candidates tie for second in top rankings and the election winner depends on which one advances; by contrast, an insincere vote is rewarded in FPTP when two candidates tie for first. It seems reasonable to assume that ties for second in IRV are roughly as likely as ties for first in FPTP. But a tie for second rewards an insincere vote in IRV only when the election winner

[17] In both FPTP and IRV, "trailing" means not expected to finish first (in first preferences), not "expected to finish last".

depends on which candidate advances, which implies that events that reward insincere votes in IRV will be less likely than events that reward insincere votes in FPTP.

Second, we suspect that strategic voting incentives are more conflicted in IRV than in FPTP, meaning that in IRV the benefits of an insincere vote are more likely to be accompanied by countervailing costs. To see why, recall that an insincere vote is rewarded in IRV when (i) a tie for second in top rankings is likely and (ii) one or both of the trailing candidates could defeat the leader; for both conditions to hold, expected first-preference support for the candidates must be fairly balanced. (The least-supported candidate cannot be too far behind, or else (i) would not hold, and the most-supported cannot be too far ahead, or else (ii) would not hold.) This requirement that expected support be fairly balanced across candidates in IRV suggests that, when an event that would reward an insincere vote seems likely, events that would punish the same insincere vote will also seem likely. Consider the example above in which a tie for second between B and C is likely, creating the opportunity for an ABC type to help A win by voting CAB. If C is stronger than expected and B weaker than expected, then the election might ultimately be a tie between A and C, in which case a vote of CAB would backfire by electing C instead of A. The key point is that the CAB vote is more likely to help when A is not far ahead of B and C (because this is when B could beat A but C could not), but this is also when it is more likely that A and C are more likely to end up tied, such that the CAB vote backfires. Insincere votes can backfire in FPTP too, of course (e.g. a vote of B by a CBA type backfires if the election ends up being a tie between C and B or between C and A), but in FTPT an insincere vote can be rewarded even when such backfiring is a very remote possibility (such as when the trailing candidate is far behind). This suggests that, on average, the benefits of an insincere vote are less positively related to the risks in FPTP than in IRV.

Third, voters' strategic incentives depend on other voters' strategic behavior in a different way in the two systems. Suppose a poll takes place reporting voters' voting intentions, and some voters respond strategically assuming the result will resemble the poll. In FPTP, this strategic response of other voters tends to amplify the strategic incentives implied by the poll: if candidate C is trailing in the poll, then candidate C will trail even more when other voters respond to the poll; thus if my best naive response to the poll is to abandon C in favor of my second choice, my incentive to abandon C in favor of my second choice is likely to be even larger when I take into consideration other voters' responses to the poll. Thus in FPTP, strategic responses to polls are characterized by *positive feedback* (or strategic complements). Strategic voting in IRV, by contrast, has a stronger tendency toward *negative feedback* (or strategic substitutes). Consider again the example in which a tie for second between B and C is likely, creating the opportunity for an ABC type to help A win by voting CAB. If a poll suggests that such a vote would be beneficial, and some ABC voters follow suit, then another ABC voter's incentive to do so is diminished, for two reasons. First, as voters abandon A, it becomes more likely that A would finish tied for second, which makes voting CAB less attractive. Second (and more subtly), an ABC type's reason for voting CAB is that C is a weaker opponent for A than B (because the proportion of B's ballots ranking A second is larger than the proportion of C's ballots ranking A second), but C becomes stronger as an opponent for A as more ABC types vote CAB (because an increasing proportion of C's ballots rank A second), which means that the original motivation for the insincere vote tends to evaporate. In fact, the general pattern in IRV is that, when voters strategically respond to a discrepancy between the second preferences of two candidates, this discrepancy tends to disappear. Not all feedback is positive in IRV,[18] but the importance of negative feedback for strategic desertion of the leading candidate in IRV (compared to the predominance of positive feedback in FPTP) suggests that, as expectations adjust over several polls, the incentive to vote strategically may decline in IRV and increase in FPTP.

Section summary

FPTP and IRV create different types of incentives to cast an insincere vote. The logic of strategic voting in FPTP is to "abandon a trailing candidate to avoid wasting one's vote", while in IRV it is "abandon a trailing candidate to avoid helping one's least-favored candidate" or "abandon

[18]Suppose A, the candidate expected to receive the most top rankings, would probably lose to either B or C (because each voter's supporters tend to rank the other second), so that ABC types strategically vote BAC and ACB types strategically vote CAB; these defections from A make it more likely that A loses to both B and C, which could lead to A's support declining further. Suppose B, the candidate expected to finish second in top rankings, is less likely to defeat A, the expected leader, than C is (because B's supporters tend to list C second while C's tend to be split between A and B), which leads BCA types to strategically vote CBA; these desertions make a tie for second between B and C more likely, which could lead to B's support declining further (although the desertions also tend to narrow the discrepancy in second preferences that motivated them).

a leading candidate to avoid wasting one's vote". While any statement about strategic voting incentives depends on preferences and beliefs, we conjecture that (1) the probability of an insincere vote being beneficial is lower in IRV than in FPTP, (2) the probability of an insincere vote being beneficial is more positively related to the probability of it being costly in IRV than in FPTP, and (3) strategic voting incentives depend positively on the extent to which other voters vote strategically in FPTP and negatively in IRV.

5 Data

The main empirical input to our analysis is a set of 160 national election surveys from the Comparative Study of Election Systems (CSES) waves 1-4 (1996-2016).[19] 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 the numerical ratings in each survey for the three largest parties based on national vote share. 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. The surveys cover 56 unique countries. Because these countries differ widely in population, and some countries have more surveys in the dataset than others, we weight voter i in country j by $\frac{w_i N_j}{n_j}$, where w_i is the survey weight assigned to respondent i, N_j is country j's population, and n_j is the number of surveys from country j in the dataset; thus the results characterize the typical citizen across the countries in the CSES.

We add a small amount of random noise to the party ratings so that there is a unique sincere vote for every voter. We also label the parties in a consistent way across cases, with A indicating the party most preferred by the largest proportion of respondents and C indicating the party most preferred by the smallest proportion of respondents.

The ratings in the surveys indicate a wide variety of preference distributions. 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,

[19]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 lack sufficient preference data.

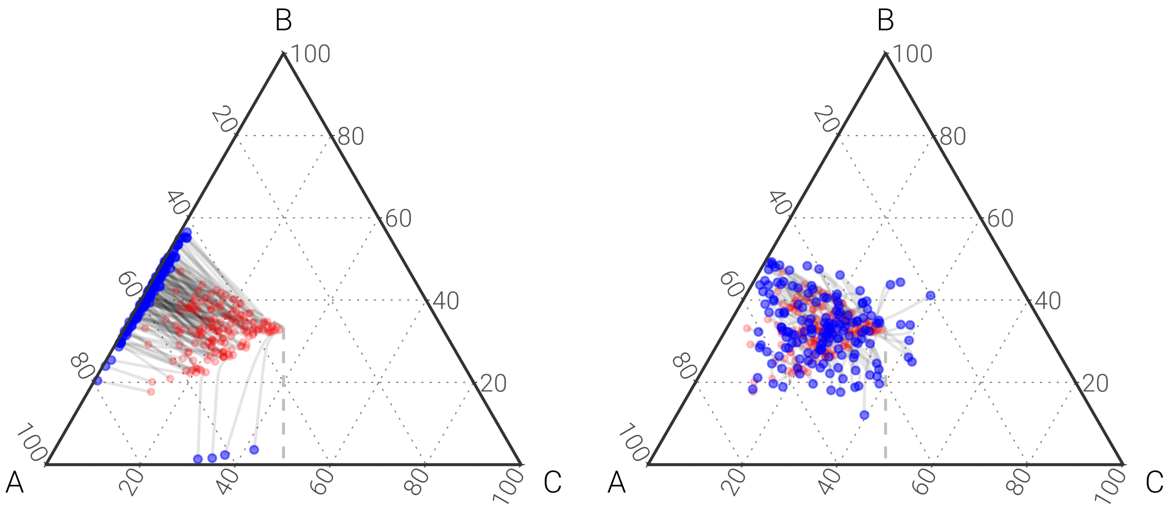

Figure 1: Evolution of ballot share vectors for all 158 CSES election surveys for both FPTP (left) and IRV (right), when $s = 85$. Red dots indicate the first hypothetical poll result, blue dots indicate the 60th hypothetical poll result, and gray lines trace the path between the two.

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.

6 Results

6.1 Iterative polling algorithm

We begin by describing the outcome of the iterative polling algorithm, which provides the sequence of expected results that forms the basis of beliefs in our main analysis. Figure 1 uses a ternary diagram to report the share of votes in FPTP (left) and first-preference votes in IRV (right) in the first hypothetical poll (red dots), i.e. the sincere profile, and the 60th hypothetical poll (blue dots); a gray line shows the path between them, i.e. the results of intervening polls. In the case shown we assume a precision level of $s = 85$ (the level associated with UK election forecasters by Eggers and Vivyan (2018)) and a proportion of voters strategically responding to the previous poll of $\lambda = .05$.

In FPTP, the iterative polling algorithm traces a path directly from the sincere profile to a (nearly-)Duvergerian equilibrium in which two parties receive all the votes. In almost all cases, the algorithm ends at a poll in which the two parties with the most sincere preferences (A and B) receive essentially all of the votes. (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.) The endpoint of the algorithm in some cases is precisely at the boundary of the ternary diagram, with the trailing party receiving zero support; in other cases, a handful of voters continue to support the trailing party, either because the algorithm had not quite converged after 60 iterations or because these voters are essentially indifferent between the leading parties. (When assuming less precise beliefs, i.e. lower values of s, the proportion supporting the trailing party after 60 iterations is larger because the trailing party is perceived to have a slightly larger chance of being in the running, which slows convergence and increases the number of holdouts.) The choice of λ affects how many iterations it takes to get to the edge of the ternary diagram, but does not otherwise affect the results. In short, the sequence of hypothetical poll results in FPTP can essentially be considered the path from the sincere profile to the Duvergerian equilibrium that is closest to the sincere profile, i.e. the one in which the two candidates who lead in sincere preferences receive all votes.

In IRV, the iterative polling algorithm tends to stop at an "interior" point, i.e. one where all candidates receive some first-preference support. This is consistent with the negative feedback characteristic of IRV, pointed out in the previous section: when voters respond to an impetus to cast an insincere ballot, that impetus tends to diminish. As a result of this negative feedback, the choice of λ matters more in IRV than in FPTP. Larger λ (especially in conjunction with higher s) can create wild swings as e.g. a large proportion of the voters who prefer the leading candidate, A, abandon that candidate, such that A no longer leads in the polls and the strategic voting incentives change entirely. At $\lambda = .05$ such swings are more restrained and by the 60th hypothetical poll, almost all cases have essentially converged.[20] It appears that in IRV, as in

[20] In many cases there remains a minor oscillation as e.g. one or two voters optimally vote insincerely in response to one poll, sincerely in response to the next, and so on indefinitely; their own vote choice moves the poll result sufficiently to change their vote choice in their next poll. In other cases more significant shifts takes place (visible in the right column of Figure 2) when a small shift in the composition of expected support for a trailing party (e.g. from 4 CAB votes and 2 CBA votes to 4 CAB votes and 1 CBA vote) that triggers an exodus from a leading party to the trailing party. These jumps could be eliminated by incorporate a prior belief so that small

FPTP, the sequence of hypothetical poll results can be seen as the path from the sincere profile to the closest equilibrium, although further analysis is necessary given the absence of a general characterization of strategic voting equilibrium in IRV elections.

6.2 Expected benefit, magnitude, and prevalence

The left column of Figure 2 reports the expected benefit of strategic voting in each CSES election survey (thin lines) as well as on average across CSES survey respondents (thick lines) in FPTP (orange) and IRV (blue) at different levels of belief precision ($s = 10$, top; $s = 55$, middle; $s = 85$, bottom). The horizontal axis indicates the step of the iterative polling algorithm that forms the basis for assumed beliefs (i.e. the "Iteration"): at the furthest left point (iteration 1), voters expect other voters to vote sincerely; by the furthest right point (iteration 60), most cases are in a strategic voting equilibrium, as described above. Expected benefit is expressed here 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 FPTP. 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 FPTP 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. The overall difference in level is consistent with our conjectures that events benefiting an insincere vote are on average less likely in IRV and that voters face more countervailing incentives (both of which we test directly below); the fact that this gap increases with

movements in the poll results would not change beliefs so dramatically.

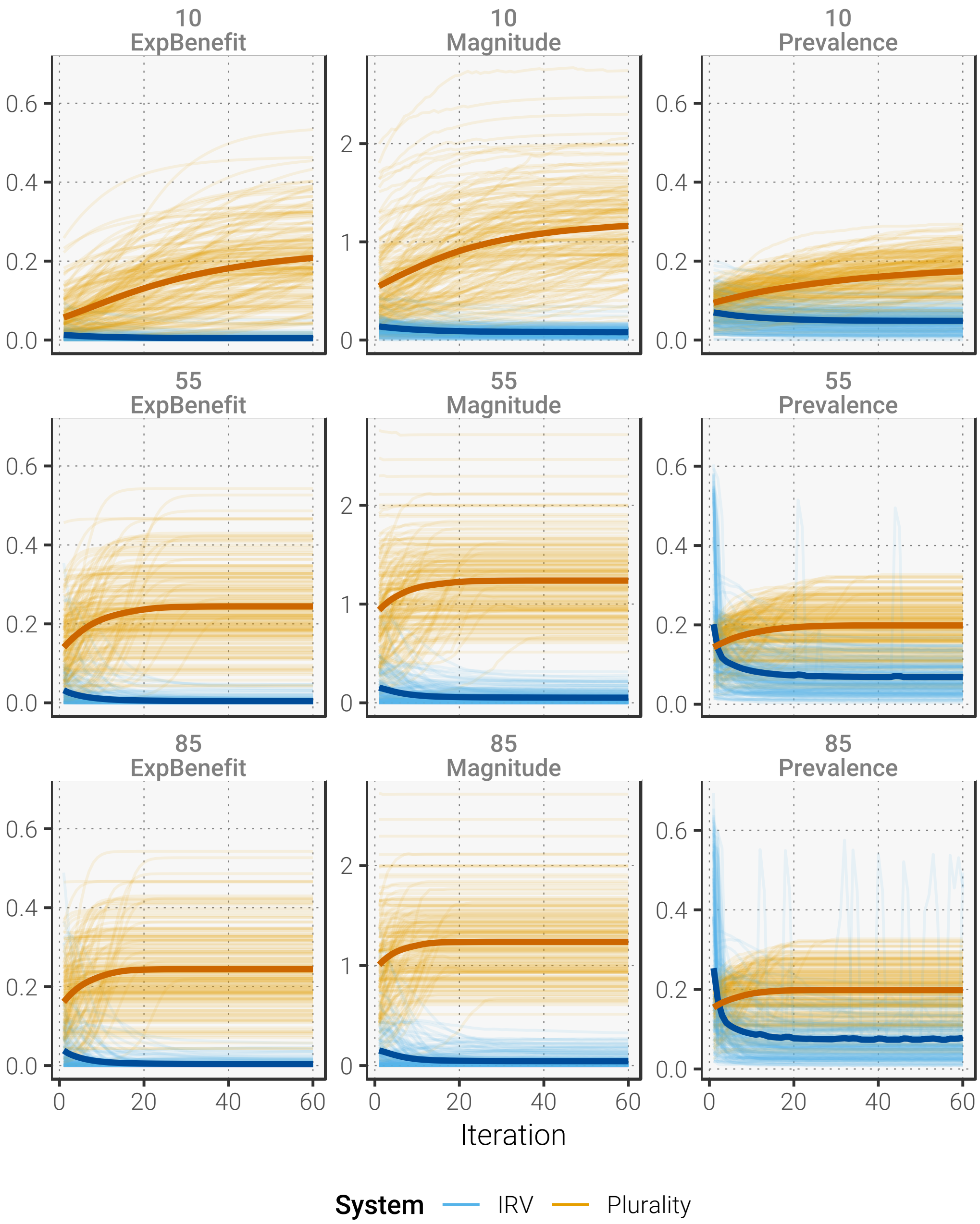

Figure 2: Expected benefit, magnitude, and prevalence of strategic voting

more iterations of the polling algorithm is consistent with our observation that strategic voting in FPTP is characterized by positive feedback while strategic voting in IRV is characterized by negative feedback.

The second and third columns of Figure 2 help us understand the difference in expected benefit by decomposing the expected benefit into two parts. Let b_i denote the benefit of strategic voting for the ith voter. Then the expected benefit can be decomposed as

$$E[b_i] = E[b_i \mid b_i > 0] \times E[\mathbb{1}\{b_i > 0\}],$$

where the first component can be referred to as the *magnitude* of strategic voting benefits and the second component can be referred to as the *prevalence* of strategic voting benefits. (The magnitude captures the intensive margin of strategic voting benefits while the prevalence captures the extensive margin.) The second and third columns of Figure 2 show how the magnitude and prevalence of strategic voting benefits differs between FPTP and IRV across levels of belief precision and across hypothetical polls. They indicate that magnitude and prevalence both play a role: the expected benefit of strategic voting is lower in IRV both because voters who expect to benefit from an insincere vote do so by less on average (magnitude) and because fewer voters expect to benefit from an insincere vote (prevalence). Note, however, that at $s = 55$ and $s = 85$ the prevalence of strategic voting benefits is *higher* on average in IRV than in FPTP for beliefs close to the sincere profile; the prevalence in IRV subsequently plummets as voters respond strategically to previous polls. This illustrates the role of negative feedback in strategic voting in IRV. At high belief precision, supporters of the leading candidate (who are, by definition, the most numerous) perceive an opportunity to avoid wasting their vote by ranking another candidate first; the incentive to do so evaporates after a few iterations when the leading candidate loses some support and the discrepancies between other candidates' second preferences narrow.

6.3 Testing additional conjectures

We hypothesized above that the probability of an insincere vote being rewarded would be lower in IRV than FPTP, basically because an insincere vote is rewarded in IRV only when two events coincide (a tie for second in first preferences and additional conditions) but only one event is necessary in FPTP (a tie for first). The diagram at left in Figure 3 suggests that this is the case. For each voter in each election survey and each iteration of the polling algorithm, we compute the probability of events that reward an insincere vote, and we average these over voters in the survey (thin lines) and over all voters (thick lines). (These probabilities are not normalized for electorate size; to do so, divide the value by the postulated electorate size, so that e.g. a value of .1 in an electorate of 1 million indicates a probability of 1 in 10 million.) The diagram indicates that the probability of one's second and third choices tying for first in FPTP (blue line) is higher than the probability of any event that rewards ranking one's second choice first in IRV (red line) or ranking one's third choice third in IRV (green line).

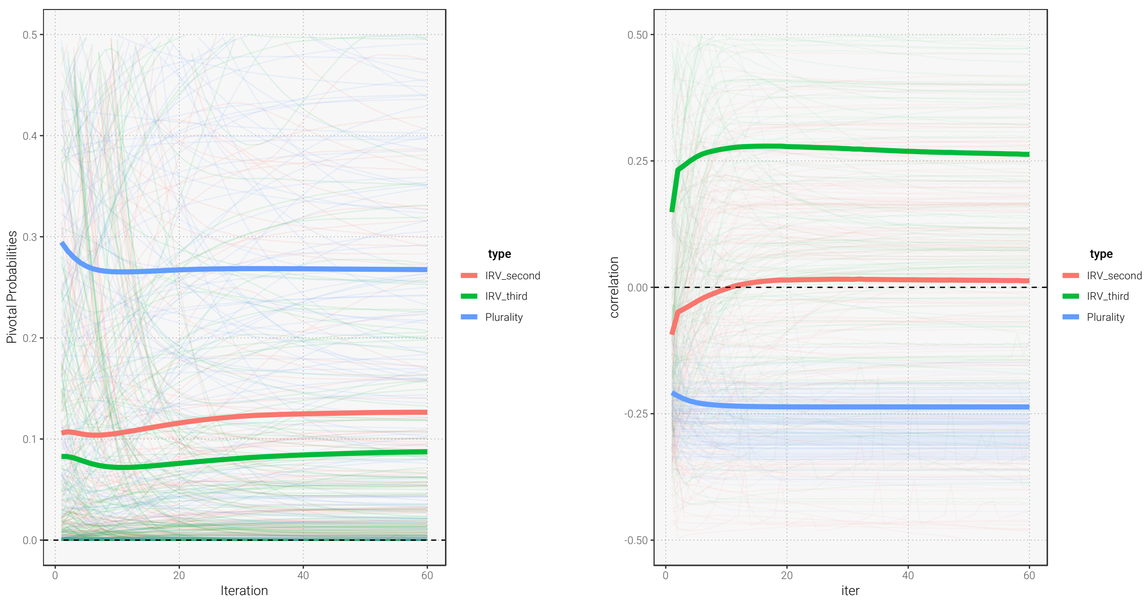

Figure 3: Non-normalized probability that each insincere vote would be rewarded (left) and correlation between probability that each sincere vote would be rewarded and punished (right).

We also hypothesized that the probability of an insincere vote being beneficial would be more positively correlated with the probability of it being harmful in IRV than in FPTP, basically because an insincere vote could be rewarded in IRV only when the candidates have fairly even support, which is also when it could be harmful. The diagram at right in Figure 3 suggests this is indeed the case: on average, the correlation is negative for FPTP (blue line), close to zero for ranking one's second choice first in IRV (red line), and positive for ranking one's third choice first in IRV (green line).

These tests suggest that at least part of the reason for the lower expected benefit of strategic voting in IRV than in FPTP, and for the lower magnitude and prevalence of strategic voting benefits, is that an insincere vote is less likely to benefit the voter and, when it is likely to benefit the voter, this benefit is more likely to be counteracted by risks.

7 Conclusion

Our analysis indicates that previous conjectures by Cox (1997) and Renwick (2011) were correct: IRV creates lower incentives for strategic voting than FPTP 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. 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.

To motivate and explain our findings for FPTP and IRV, we described and analyzed the logic of strategic voting in FPTP and IRV. We argued that events rewarding insincere votes are both less likely in IRV and "closer" to events punishing those same votes in IRV than in FPTP, and we find this to be the case. We also highlighted the important role of positive and negative feedback in strategic voting: if a poll suggests an opportunity to benefit from an insincere vote in FPTP, this opportunity becomes more attractive when other voters respond to it in FPTP but less attractive in IRV.

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 FPTP. 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.

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.

Bouton, Laurent. 2013. "A theory of strategic voting in runoff elections." *American Economic Review* 103(4):1248–88.

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

Cox, Gary W. 1997. *Making votes count: strategic coordination in the world's electoral systems.* Cambridge University Press.

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?".

Favardin, Pierre and Dominique Lepelley. 2006. "Some further results on the manipulability of social choice rules." *Social Choice and Welfare* 26(3):485–509.

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

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

Gibbard, Allan et al. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41(4):587–601.

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.

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

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.

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.

A Computing the P matrix in FPTP and IRV

A.1 FPTP

Consider the case of a FPTP election involving candidates $\{a, b, c\}$. Let π_a denote the probability that a wins by more than one vote (so that the marginal ballot is irrelevant, with π_b and π_c defined correspondingly) and let π_{ab} denote the probability that a and b finish tied for first (with π_{ac} and π_{bc} defined correspondingly). Assume that ties are decided by alphabetic precedence (so e.g. a beats b if they tie for first) and that any candidate is just as likely to win by one vote over another candidate as to tie for first with that candidate.[21] Then we can write

$$\mathbf{P} \;=\; \begin{bmatrix} \pi_a + 2(\pi_{ab} + \pi_{ac}) & \pi_a + \pi_{ac} & \pi_a + \pi_{ab} \\ \pi_b + \pi_{bc} & \pi_b + 2(\pi_{ab} + \pi_{bc}) & \pi_b + \pi_{ab} \\ \pi_c + \pi_{bc} & \pi_c + \pi_{ac} & \pi_c + 2(\pi_{ac} + \pi_{bc}) \end{bmatrix}$$

For example, $p_{a,a}$ (the top-left item) indicates that a wins given a marginal ballot of a if (i) a would win given any marginal ballot (π_a), (ii) a is tied with either b or c for first ($\pi_{ab} + \pi_{ac}$), or (iii) a trails b or c by one vote (again, $\pi_{ab} + \pi_{ac}$); similarly, $p_{a,b}$ (the top-center item) indicates that a wins given a marginal ballot of b if (i) a would win given any marginal ballot (π_a) or (ii) a is tied with c for first (π_{ac}). Next we can separate $\mathbf{P}$ into two parts

$$\mathbf{P} = \begin{bmatrix} 2(\pi_{ab} + \pi_{ac}) & \pi_{ac} & \pi_{ab} \\ \pi_{bc} & 2(\pi_{ab} + \pi_{bc}) & \pi_{ab} \\ \pi_{bc} & \pi_{ac} & 2(\pi_{ac} + \pi_{bc}) \end{bmatrix} + \begin{bmatrix} \pi_a & \pi_a & \pi_a \\ \pi_b & \pi_b & \pi_b \\ \pi_c & \pi_c & \pi_c \end{bmatrix} = \tilde{\mathbf{P}} + \mathbf{Q},$$

where each element of $\tilde{\mathbf{P}}$ can be interpreted as the probability that the outcome depends on a single ballot *and* a given candidate is elected given a particular marginal ballot. Then the matrix of expected utilities from each ballot can be expressed

$$\overline{\mathbf{U}} \equiv \mathbf{UP} = \mathbf{U}\tilde{\mathbf{P}} + \mathbf{UQ} = \mathbf{U}\tilde{\mathbf{P}} + \text{constant},$$

where the last equality obtains because the columns of $\mathbf{Q}$ are identical by definition. Because the strategic voting incentive is defined as the difference between the expected utility of the optimal vote and the expected utility of the sincere vote, the value of the constant does not matter.

One can use either simulations or analytical approaches to estimate $\tilde{\mathbf{P}}$. Following Fisher and Myatt (2017), we compute the elements of $\tilde{\mathbf{P}}$ by integrating the Dirichlet distribution along the relevant vote share loci. For example, in the plurality case we compute π_{ab} by integrating along the locus where $v_a = v_b > v_c$ and multiplying by $\frac{1}{N}$, where N is the size of the electorate (because a single ballot for a could push a ahead of b when $v_b - v_a \in \left[0, \frac{1}{N}\right]$). That is,

$$\pi_{ab} = \frac{1}{N} \int_{x=\frac{1}{3}}^{\frac{1}{2}} \mathrm{Dir}\big(x, x, 1 - 2x; s\overline{\mathbf{v}}\big)\, dx,$$

where $\mathrm{Dir}\big(x, x, 1 - 2x; s\overline{\mathbf{v}}\big)$ is the Dirichlet density with parameter vector $s\overline{\mathbf{v}}$ evaluated where $v_a = x$, $v_b = x$, and $v_c = 1 - 2x$. The procedure is analogous for π_{ac} and π_{bc}.

A.2 IRV

We use a similar procedure for computing $\tilde{\mathbf{P}}$ in IRV elections, although it is somewhat more difficult because (as discussed above) there are more ballot types and more ways in which the

[21] Both assumptions are common (e.g. Bouton, 2013; Fisher and Myatt, 2017) and simplify the exposition.

marginal ballot can affect the outcome. *To be continuted.*