Ranked Choice Voting and Political Polarization

Nathan Atkinson[a] **and Scott C. Ganz**[b]

This manuscript was compiled on September 5, 2023

American states and municipalities are increasingly adopting ranked choice voting. In particular, advocates favor a particular form of Ranked Choice Voting: Instant Runoff Voting (IRV), arguing that it can ameliorate extremism and lead to less polarized outcomes. We study the relationship between IRV and political polarization. We show that because polarization affects both voters and candidates, there is a non-monotonic relationship between polarization and the representativeness of outcomes. We show how increases in polarization are likely to reduce the representativeness of outcomes under IRV. Strategic positioning by moderates can ameliorate—but will not eliminate—extremist outcomes. Moreover, we show that if one political party grows more extreme than the other, outcomes under IRV will consistently favor the less extreme party.

Ranked Choice Voting | Elections | Political Economy | ...

Polarization on both sides of the political spectrum is exacerbating divisions in American society, potentially jeopardizing the stability of the democratic framework (1–3). In this political climate, the search for electoral innovations to counterbalance this polarization has become a priority. Voting systems, as the underlying framework of democratic governance, can either promote representative or extremist outcomes.

Amidst this growing polarization, a growing contingent of policymakers and voters alike are advocating for the implementation of Instant Runoff Voting (IRV), an alternative electoral system, as a panacea for extremism. Under IRV, voters rank-order their candidates, and the process simulates a sequence of runoff elections. In any given round, the candidate with the fewest votes is eliminated from the election, and any voters who ranked that candidate first have their votes transferred to their next most preferred candidate. The process continues until a candidate captures a majority of the votes. The system, already adopted in Maine and Alaska, and on the ballot in Oregon (2024) and Nevada (2026), is often hailed as a countermeasure against extremist electoral outcomes.

The international stage offers few definitive answers, with Australia's usage of IRV revealing little evidence that it results in electoral moderation relative to other democracies (4). Furthermore, the formal social choice literature, perhaps surprisingly, offers little axiomatic grounding for IRV (5). That is, compared to other voting mechanisms, relatively little is known about how IRV behaves. Indeed, many experts dismiss it entirely because of its well-documented failure of the monotonicity criterion (5, 6). Consequently, the theoretical landscape remains murky on IRV's ability to produce representative electoral outcomes, particularly in polarized contexts (7, 8).

Against this backdrop, we develop a simple spatial model to better understand the interaction between IRV and ideological polarization. We consider a balanced electorate, with two partisan candidates—one drawn from each of the left and right—and a third candidate is positioned at the location of the median voter. By fixing the location of the median voter, we can judge electoral representativeness based on deviations from the median. We find that electoral representativeness depends critically on the distribution of the electorate.

We identify a non-monotonic relationship between extremism and electoral outcomes. More specifically we show that extremism has effects both on who voters vote for and on which candidates run for office, each of which work in opposite directions. First, increased extremism leads to voters preferring relatively more extreme candidates. This means that for any given partisan candidate, that candidate will receive more support from a relatively extreme distribution of voters than under a less extreme distribution of voters. This increases the probability that a partisan candidate wins the election, leading to more extreme outcomes. On the other hand, increased extremism also leads to more extreme candidates running for office. For a given pool of voters, these candidates get less support than would

Author affiliations: [a]University of Wisconsin; natkinson@wisc.edu; [b]Georgetown McDonough School of Business, American Enterprise Institute; scott.ganz@georgetown.edu

more moderate candidates. This has the effect of decreasing the probability that a partisan candidate wins the election, leading to more moderate outcomes. Which of these effects dominates depends critically on how the electorate polarizes.

After deriving necessary conditions for when outcomes will grow more or less representative as voters grow more extreme, we then explore sufficient conditions on voter distributions in order to better understand how real-world changes in polarization are likely to affect the representativeness of IRV. We find that under fairly general conditions that increasing polarization diminishes the probability of a moderate candidate's victory under IRV while increasing the expected ideological gap between the winning candidate and the median voter. That is, when electorates become polarized, a candidate located at the median will often fail to capture enough first-choice votes to progress to the next round of the election.

Given that a representative candidate who is preferred to every other candidate in pairwise comparison can fail to win the election because she fails to garner enough first choice votes, we consider a strategic moderate candidate who chooses any position between the two partisan candidates. We find that in order to win the election, the moderate will need to adopt increasingly extreme positions as polarization increases.

Finally, we consider the impact of asymmetric polarization. Our findings indicate that in such scenarios, IRV tends to favor the less extreme party. However, asymmetric polarization is no antidote to extremism, and we show that as the asymmetry grows, the average winner becomes more extreme in the opposite direction.

Out results coupled with empirical research on polarization suggest that IRV may not be an effective method to elect representative candidates in polarized electorates.

Definitions and Preliminaries

We develop a simple spatial model where the electorate is comprised of left-leaning voters and right-leaning voters. Each voter is described by an ideal point in $\mathbb{R}$. We constrain the mass of left-leaning voters to be equal to the mass of right-leaning voters, so that the median is always located at 0. The goal of the model is to provide a simple framework for understanding how changes in polarization might affect outcomes under IRV.

Consider an absolutely continuous probability density function $f(x)$ defined over $\mathbb{R}$, with associated cumulative distribution function $F(x)$. Let $F(0) = \frac{1}{2}$, ensuring that thereby ensuring a median voter at the origin. For ease of discussion, we define $f_L(x) = f(x \mid x < 0)$ and $f_R(x) = f(x \mid x > 0)$ to indicate the distributions of left-leaning and right-leaning voters, respectively.

We define extremism according to the usual stochastic order (also known as first order stochastic dominance) (9): $g_R(x) \neq f_R(x)$ is more extreme than $f_R(x)$ if $F_R(X) \geq G_R(x)$ for all x. That is, voters are more likely to take on larger—more extreme—values under g than under f. We will denote this $g_R(x) >_{st} f_R(x)$. We define this analogously for $f_L(x)$.

In this analysis, we employ a citizen-candidate model (10), with a candidate set $C = \{L, M, R\}$. Here, L represents a left-leaning candidate sampled from $f_L(x)$ and R a right-leaning candidate sampled from $f_R(x)$. We hold a moderate candidate, $M = 0$, fixed, corresponding to the median of the voter distribution. This construction offers a valuable analytical benchmark, given that M would defeat all other candidates in any pairwise contest, thereby serving as the most democratically representative choice (11).

Voters are assumed to rank candidates sincerely, based on ideological proximity. The voting proceeds under the rules of Instant-Runoff Voting (IRV). We break ties in favor of M then L then R (this has no effect on the results but simplifies the analysis). Let V^c symbolize the proportion of voters ranking candidate c as their top choice, and P_c represent the probability of candidate c winning the election.

Voter Extremism

We begin our discussion by acknowledging two central points in line with Black's median voter theorem (12). First, should candidate M advance to the second round of voting, her victory is guaranteed due to the pivotal role of the median voter. Second, if M fails to advance, the remaining candidate—either L or R—closest to the origin will win, by virtue of capturing the median voter's support.

Given these points, our analytical focus narrows to the first-round vote shares, as they are determinative of M's advancement to the second round. Consider a right-leaning voter with an ideal point $v > 0$. Such a voter will always rank either R or M first, as M is always closer than L: $|v - M| < |v - L|$.

To understand the choice between M and R, consider a hypothetical voter located at $\frac{1}{2}R$. This voter would be precisely indifferent between R and M. Voters with ideal points less than $\frac{1}{2}R$ will prefer M over R given distribution $f(x)$, while those with ideal points greater than $\frac{1}{2}R$ will do the opposite. We define $V^M_{f_R}$ as the proportion of right-leaning voters who prefer M over R, thereby yielding $V^M_{f_R} = F_R(\frac{R}{2})$. The fraction of right-leaning voters captured R is given by $1 - F_R(\frac{R}{2})$.

The expectation of the proportion of right-leaning voters favoring M over R, given a voter distribution $f_R(x)$, is thus given by:

$$E[V^M_{f_R}] = \int_0^\infty f(x)F\left(\frac{x}{2}\right)dx. \quad [1]$$

Next, we turn our attention to how changes to the voter distribution impact this expectation. Let f_R and g_R represent two different voter distributions. Then $E[V^M_{f_R}] \geq E[V^M_{g_R}]$ if and only if:

$$\int_0^\infty f(x)F\left(\frac{x}{2}\right)dx \geq \int_0^\infty g(x)G\left(\frac{x}{2}\right)dx. \quad [2]$$

We will write this condition as $f(x) \prec g(x)$. There is no tight relationship between this necessary condition and extremism. Various dynamics come into play when considering only right-leaning voters. A shift in voter distribution toward the extreme right can have both beneficial and detrimental effects on M's expected vote share. Specifically, a rightward shift can lead to more extreme candidates being drawn, thereby distancing the average candidate from the median voter and favoring M. Simultaneously, such a shift could move the pivotal voter to the right, undermining M's chances. These effects balance when 2 holds with equality.

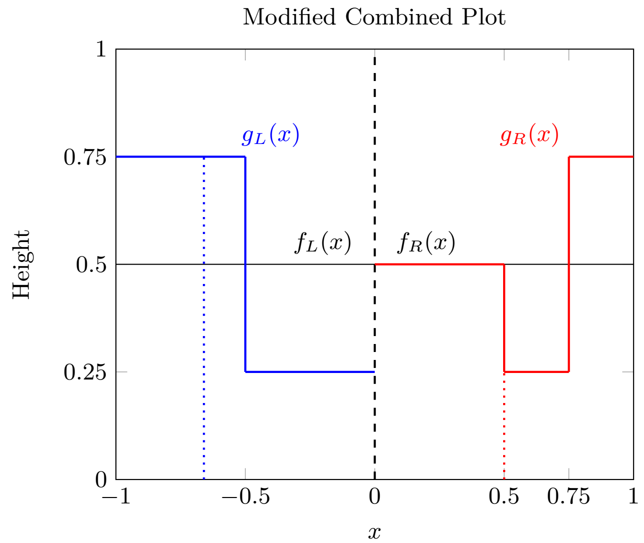

Fig. 1. Examples of how increases in extremism can increase or decrease the probability that the moderate wins the election. $f_L(x) \prec g_L(x)$. $f_R(x) \not\prec g_R(x)$.

Consider for example Figure 1. Let $f(x)$ be a uniform distribution on $[-1, 1]$, represented by the black horizontal line at 0.5. Given this voter distribution, the indifferent voter on the left will be at $\frac{L}{2}$ and the indifferent voter on the right will be at $\frac{R}{2}$. Therefore any voter with ideal point $|v| > \frac{1}{2}$ will always vote for a partisan candidate, because $|v - C| < |v - M|$ for $C \in \{L, R\}$, and any $|v| > \frac{1}{2}$. Therefore only voters with ideal points $|v| \le \frac{1}{2}$ will make their decision based on the realization of the partisan candidates.

Now consider the changing the distribution of voters. First consider a change to the left side of the distribution, represented by the blue curve, $g_L(x)$ which is more extreme than $f_L(x)$, generated by moving half of the voters on $[-\frac{1}{2}, 0]$ to $[-1, -\frac{1}{2}]$. This moves the median left-leaning voter from $-\frac{1}{2}$ to $-\frac{2}{3}$. Moreover, a full three quarters of left-leaning voters now have an ideal point of $v < \frac{1}{2}$ and will always vote for the partisan candidate over the moderate candidate, lowering the vote share of M. Next, consider the red distribution, $g_R(x)$, which is more extreme on the right, generated by moving half of the voters on $[\frac{1}{2}, \frac{3}{4}]$ to $[\frac{3}{4}, 1]$. In this case, relatively extreme voters grew more extreme, but relatively moderate voters remained unchanged. This leaves the median of right-leaning voters at $\frac{1}{2}$, meaning that the set of swing voters on the right remains the same, but the candidate pool has grown more extreme. This will increase the moderate's vote share because the moderate will capture a higher vote share when $R \in [\frac{3}{4}]$ than when $R \in [\frac{1}{2}, \frac{3}{4}]$.

This illustrates the complex dynamics of how IRV changes with extremism. Shifting relatively moderate voters to be more extreme can be expected to lead to more extreme outcomes, whereas shifting relatively extreme voters to be more extreme can be expected to lead to more moderate outcomes. Taking this in reverse leads to a paradoxical result similar to IRV's failure of the monotonicity criterion (6): the moderate candidate can perform worse when the most extreme voters moderate their views.

Specific Distributional Assumptions. The effectiveness of IRV in generating representative outcomes hinges on certain characteristics of the voter distribution, encapsulated by equation Eq. (2). We now look at the consequences of two unique distributional transformations that affect extremism: one that involves the rising extremism among already-extreme partisans and another that accounts for a general spread of extremism among all voters, and show how these lead to differing outcomes under IRV.

We begin by examining the assumption that rising extremism is primarily the result of relatively extreme partisans becoming increasingly extreme. We model this by considering $f_R(x)$ bounded on $[0, 1]$. Define If $G_R(x) = F_R(x)$ for $x \le \frac{1}{2}$, and $G_R(x) \le F_R(x)$ for all $x > \frac{1}{2}$, which satisfies $g(x) >_{st} f(x)$. This means that $F(\frac{x}{2}) = G(\frac{x}{2})$ on $x \in [0, 1]$. Therefore, following equation 2, the moderate's vote share increases if and only if $\int_{\frac{1}{2}}^{1} f(x)F\left(\frac{x}{2}\right) dx \le \int_{\frac{1}{2}}^{1} g(x)F\left(\frac{x}{2}\right) dx$. Which holds by first order stochastic dominance because $F(x)$ is weakly increasing, and $g_R(x) >_{st} f_R(x)$ on $[\frac{1}{2}, 1]$ (9). Therefore if extreme voters grow more extreme, IRV grows more representative (this is the case of $g_R(x)$ in figure 1). This also leads to the result that if extreme voters grow less extreme, then IRV becomes less representative.

Alternatively, we consider the case where all voters, not just the extreme partisans, grow more extreme. We model this as a linear transformation, $g_R(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right)$ for some $a \ge 1$ and $b \ge 0$, which satisfies $g(x) <_{st} f(x)$. Rewriting the right hand side of (2) gives $\int_0^\infty g(x)G\left(\frac{x}{2}\right) dx = \int_0^b g(x)G(\frac{x}{2})dx + \int_b^\infty g(x)G(\frac{x}{2})dx$. Note that for $b > 0$, the first integral is equal to 0. Focusing on the remaining integral and substituting $\int_b^\infty \frac{1}{a} f\left(\frac{x-b}{a}\right) F\left(\frac{x-b}{2a}\right) dx = \int_0^\infty f(u)F\left(\frac{u}{2}\right) du$, yields

$$\int_0^\infty f(x)F\left(\frac{x}{2}\right) dx \ge \int_0^\infty f(u)F\left(\frac{u}{2}\right) du. \qquad [3]$$

The inequality holds weakly for $a \ge 1$ and strictly for $b > 0$ since the right-hand side excludes the region from 0 to b. That is, if more moderate voters are also growing more extreme, then IRV can be expected to grow less representative.

Beyond these transformations, it is also possible to look at specific parameterized distributions. For example, if the distribution $f_R(x)$ is log-normal(μ, σ^2), then the distance from the median to the expected outcome will be decreasing in μ.

Which world are we in?. Understanding the efficacy of IRV in electing representative outcomes therefore depends on understanding political polarization and how it is developing. At the congressional level, polarization has been increasing exponentially for decades with no sign of abating (13). At the same time, the ideal point estimates of House candidates who won their primaries have grown increasingly polarized (14).

At the voter level, the data also support increasing polarization. Data from Pew suggests that it is not just the extremes moving apart. In 1994 Pew found that 64% of Republicans were more conservative than the median democrat and 70% of Democrats were more liberal than the median Republican. By 2017 the numbers was 95% and 97% (15). The gap in political values across partisans has grown substantially. Figure 2 shows how the distributions of the parties have grown apart over decades, with a substantial divergence between the medians of the two parties. This hollowing out of the center has been observed in many areas.

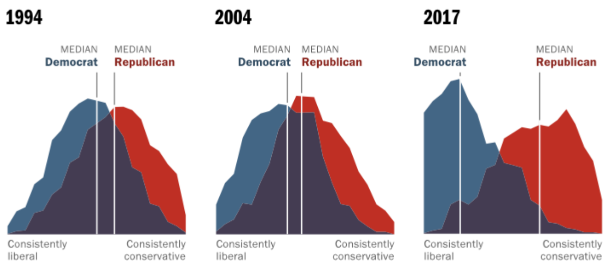

Fig. 2. Ideological consistency based on a scale of 10 political values questions. Source: Pew Research (15).

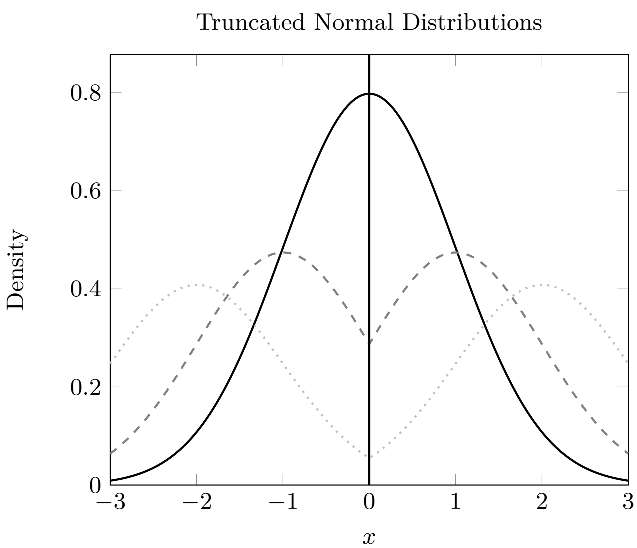

Fig. 3. Truncated Normal Distributions.

Political activism and voting behavior is U-shaped, with those holding the most extreme views the most likely to contribute to political candidates and vote (16). The partisan divisions between voters have never been larger (15). Ideal point estimation of individual voters shows a strongly bimodal distribution of ideal points (17). Taken together, the data are consistent with a bimodal distribution of voters that is growing more polarized. We now turn our focus towards how IRV responds to symmetric polarization, holding the assumption that extremism is of the form $f_(x) \prec g(x)$.

Symmetric Polarization

We have thus far explored the effect of extremism. we now turn our attention to polarization. In a comparative sense, we say that a distribution $g(x)$ is *more polarized* than $f(x)$ if it satisfies two conditions:

1. Either $g_L(x) >_{st} f_L(x)$ or $g_R(x) >_{st}$
2. Neither $f_L(x) >_{st} g_L(x)$ nor $f_R(x) >_{st} g_R(x)$

That is, $g(x)$ is more polarized than $f(x)$ if at least one of the distribution is more extreme under $g(x)$ than under $f(x)$. In this section we consider symmetric polarization, where $f_L(x) = f_R(-x)$, and we later explore asymmetric polarization, where $f_L(x) \neq f_R(-x)$.

As a portion of the electorate grows more extreme, $g_r >_{st} f_r$, voters are more likely to rank a given partisan candidate above the moderate candidate (Eq. (2)). Because the same is true for both sides of the voter distribution, we can see that increases in symmetric polarization will lower the moderate candidate's expected first round vote share, and will therefore decrease the probability that the moderate candidate progresses to the second round of the election. Therefore, as symmetric polarization increases, the probability that the moderate candidate wins is weakly decreasing. For sufficiently extreme polarization, the probability that the moderate candidate wins can go to 0.

In addition to lowering the probability that the moderate wins the election, increasing polarization also leads to more extreme outcomes in those cases where the moderate loses the election. That is, letting W be the location of the winning candidate, and let $\delta = |W|$ be the deviation between the median voter and the winning candidate, then $E[\delta(f(x))|W \neq M] < E[\delta(g(x))|W \neq M]$ for $g \succ f$. Taken together, the lower probability of the moderate winning and the more extreme locations of partisan candidates means that outcomes under IRV are stictly increasing in the level of polarization, $E[\delta(f(x))] < E[\delta(g(x))]$ for any $g(x) \succ f(x)$.

The Effect of a Moderate Candidate. The question facing votes is whether or not IRV is an improvement on the current plurality voting system. That is, will IRV tend to elect candidates that are more representative relative to plurality?

We can begin by noting that in any given three candidate election, IRV will perform weakly better than plurality rule (8). If M captures a plurality in the first round, $V^M > V^R$ and $V^M > V^L$, then IRV and plurality both elect M. Suppose instead, without loss of generality, that L captures the most votes in the first round. Under plurality, L wins, but under IRV either M or R is eliminated. If R is eliminated, then M wins under IRV, improving upon the plurality result. If M is eliminated, then whichever of L or R is closer to 0 will win. If this is L, then IRV and plurality perform identically. If this is R, then IRV improves upon plurality.

However, following Durverger's Law, a plurality election will tend towards two-candidate elections (18). That is, in two-party systems, voters are unlikely to vote for a moderate candidate because they don't want to waste their votes. Much of the interest in IRV, therefore, is based on the claim that it will change the candidate pool. That is, instead of a two-candidate plurality election, voters face a three-or-more candidate election under IRV.

To assess this claim, we compare an election with just candidates L and R to an IRV election with L, R, and M (note that with two candidaes, IRV and pluarlity rule are equivalent). Consider the election without candidate M. In this case, the winning candidate will either always be R or L. Letting W be the location of the winning candidate, $|W|$ is the deviation between the median voter and the winning candidate.

Let $f_R(x)$ be a normal distribution with mean μ and standard deviation $\sigma^2 = 1$, truncated below at 0. Let $f_L(x)$ be defined similarly with mean $-\mu$, but truncated above at 0. Figure 3 plots this for $\mu \in \{0, 1, 2\}$. We now consider the effect of increasing μ on election outcomes. Figure 4 plots the outcomes of these elections as a function of μ. The black line is the probability that the moderate candidate wins the election. The blue line is the average distance from the median voter to the more moderate of L and R. That is,

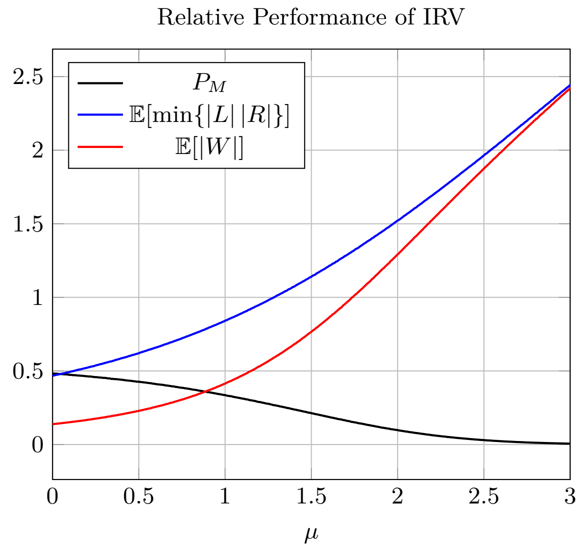

Fig. 4. Electoral outcomes as function of symmetric polarization.

assuming a two candidate election between L and R, the blue line is the average distance from the median to the location of the winning candidate. The red line is the average distance from the median voter to the winning candidate under IRV.

First note the inverse relationship between the probability that the moderate wins the election and the average distance to the more moderate partisan candidate. That is, IRV frequently chooses the moderate candidate when there is relatively little loss from choosing a partisan candidate. However, as the loss from choosing a partisan candidate increases, the probability of choosing one of those partisan candidates is increasing.

Next note the gap between the red and blue curves. This is the relative gain of adding a moderate candidate to an election between L and R. When polarization is low, $\mu = 0$, the distribution of voters is a standard normal curve, and IRV performs better than plurality. For low levels of polarization, IRV continues to perform significantly better than plurality. However, as polarization increases past a certain level, the relative gain of IRV over plurality diminishes. This is because there are fewer and fewer relatively moderate voters, so that the moderate candidate grows increasingly unlikely to progress past the first round of the election. Once polarization is sufficiently extreme, the results under IRV and plurality are identical, and there is no gain from adding a moderate candidate.

Strategic Entry. The previous section shows that as an electorate grows more polarized, candidates located at the median are less likely to be elected under IRV, because they simply are not the first choice of enough voters. Can the moderate do better by choosing a non-median position? That is, instead of $m = 0$, let $m \in [L, R]$. We change the model so that a moderate candidate can choose a location m after observing the realizations of L and R. We also assume that if the moderate colocates at the same location as another candidate, M captures all of the voters towards the center, and the other candidate captures all of the voters towards the extreme. The moderate's objective is to choose as moderate of a policy as possible conditional on winning the election. Without loss of generality, assume that $|L| < |R|$. The moderate's objective is therefore

$$\min_{m \in [L,R]} |m| \tag{4}$$

$$s.t.\;\; F\left(\frac{m+R}{2}\right) - F\left(\frac{m+L}{2}\right) \geq 1 - F\left(\frac{m+R}{2}\right) \tag{5}$$

where the left side of the constraint is the vote share of M, and the right side is the vote share of R. That is, M needs to get at least as many votes as R in order to progress to the second round of the election.

Observe that there is no guarantee that there exists an m that optimizes this program. For example, suppose that $f(x)$ is uniformly distributed on $[-1, 1]$, and $L = -0.4$ and $R = 0.5$ The moderate candidate's vote share is therefore $V^M(m) = \frac{1}{4}\big(R - L\big) = 0.225$ for any $m \in [L, R]$.* But V^L is bounded below by 0.3 (if $m = L$) and V^R is bounded below by 0.25 (if $m = R$). Therefore there does not exist any $m \in [L, R]$ such that the moderate progresses past the first round of the election.

For other values of L and R it may be possible that the moderate candidate can win the election. Nonetheless, we will now show that increasing polarization will lead to the moderate candidate choosing more extreme positions in order to win the election. For analytic tractability, we focus here on the case of a linear transformation of the voter distribution: $g_R(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right)$ for some $a \geq 1$ and $b \geq 0$.

Suppose that there exists a value m_f^* that optimizes the above problem for distribution f. Now consider $g \succ f$. For each pair of candidates drawn from $f(x)$, (L_f, R_f), there is a corresponding pair drawn from g. That is, $L_g = aL_f - b$ and $R_g = aL_R + b$.

Now consider without loss of generality an optimal $m_f^* > 0$. Optimally requires that $V^m = V^L > V^R$ or $V^m = V^R > V^L$. That is, the moderate must be tied with one of the other candidates, otherwise she could choose some $m_f' < m_f^*$. Following the transformation to g, consider $m_g = am_f^* + b$. The location of the two pivotal voters has moved from $\frac{1}{2}(C_f + M_f)$ to $\frac{1}{2}a(C_f + M_f)$ for $C \in \{L, R\}$. Observe that the moderate's vote share relative to R is unchanged because after the transformations, $G(\frac{R_g + m_g}{2}) = F(\frac{R_f + m_f}{2})$. The moderate's vote share relative to L however is given by:

$$G_R\left(\frac{1}{2}a(|L_f + M_f|)\right) = F_R\left(\frac{1}{2}(|L_f + M_f|) - \frac{b}{a}\right) \tag{6}$$

$$\leq F_R\left(\frac{1}{2}(|L_f + M_f|)\right) \tag{7}$$

$$\text{for } C \in \{L, R\}.$$

Where we can focus on just the right side of the distribution because of symmetry. We see therefore that the moderate's vote share is weakly decreasing even after choosing a more extreme position.

Observe that the moderate could never choose a more moderate position. That is, if the moderate could win with a position $m_g < m_f^* + b$, then $m_f' = \frac{m_g - b}{a} < m_f^*$ would also win, contradicting the optimality of m_f^*.

Moreover we can see that the vote share from choosing $m = 0$ is weakly decreasing in symmetrical polarization: $G_R(\frac{R_g}{2}) = F_R(\frac{1}{2}(R_f - \frac{b}{a})) \leq F_R(\frac{1}{2}R_f)$.

* $F(x) = \frac{x+1}{2}$.

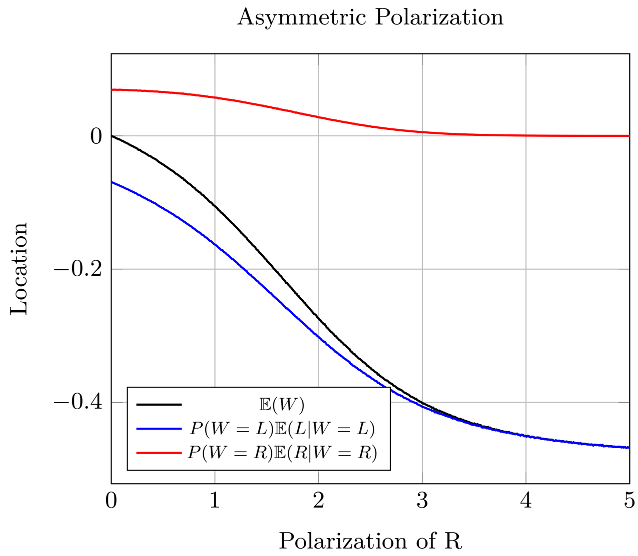

Fig. 5. Asymmetric Polarization

We can therefore see that as the population becomes more extreme, the expected policy position taken by a strategic moderate will tend to grow more extreme under IRV voting. That is, if m_f^* optimizes (4), then $E[m_g^*] > E[m_f^*]$.

Asymmetric Polarization

Under symmetric polarization, the probability that the moderate wins is weakly decreasing in the level of polarization. Moreover, the distance from the median voter to the winning candidate is increasing in the degree of polarization. We now turn our attention to the question of asymmetric polarization. That is, fixing one half of the distribution, what is the effect of increasing polarization on election outcomes?

One of the key features of elite polarization is its apparent asymmetry—while both parties have polarized, Republicans appear to have polarized more than Democrats (19). At the voter level, ideal point estimates of extreme conservatives are farther from the median voter than extreme liberals (17). In this section we fix the $f_L(x)$ and consider just changes that make $f_R(x)$ more extreme.

Under symmetric polarization, outcomes under IRV are symmetrical. That is, even though the expected outcome grows more extreme in the level of polarization, the outcomes grow equally extreme on both the left and the right. That is, in expectaiton candidates from the left and right are equally likely to win and equally extreme as one another. However, for asymmetric polarization, IRV will be tend to favor of the less extreme party.

Recall that when the right distribution grows more extreme, the right candidate captures a weakly higher share of first round votes. That is, $g_R \succ f_R$, implies that $E[V^M_{g_R}] \leq E[V^M_{f_R}]$. By capturing more first round votes, R is therefore more likely to beat M in the first round. This has no direct effect on the contest between L and M. That is, the indifferent left-leaning voter will be located at $\frac{1}{2}L$, regardless of the the distribution of right-leaning voters. However, because the moderate candidate can capture votes from both the left and right, the decrease in the expected share of right-leaning voters decreases the moderate's overall first round vote share. That is, holding f_L fixed, then a change from f_R to $g_R \succ f_R$ leads to the following changes to the three candidates' first-round vote shares:

$$E[V^M|f_L, g_R] - E[V^M|f_L, f_R] \tag{8}$$
$$= \frac{1}{2}\left(E[V^M_{f_L}] - E[V^M_{f_L}] + E[V^M_{g_R}] - E[V^M_{f_R}]\right) \leq 0 \tag{9}$$
$$E[V^L|f_L, g_R] - E[V^L|f_L, f_R] \tag{10}$$
$$= 1 - \frac{1}{2}\left(E[V^M_{f_L}] - E[V^M_{f_L}]\right) = 0 \tag{11}$$
$$E[V^R|f_L, g_R] - E[V^R|f_L, f_R] \tag{12}$$
$$= 1 - \frac{1}{2}\left(E[V^M_{g_R}] - E[V^M_{f_R}]\right) \geq 0. \tag{13}$$

Changing the right-leaning voters from f_R to g_R decreases the moderate's expected first round vote share, while leaving the expected vote share for L unchanged. This increases the probability that M has the fewest first round votes and does not progress.

By lowering the moderate's expected vote share, asymmetric polarization lowers the probability that the moderate wins, and leads to more elections where the final round consists of L and R. Recall that in an election without M, then candidate L wins if and only if $|L| \leq |R|$. And because $g_R \succ f_R$, the expected right candidate under g will be more extreme than that under f. This means that in expectation, L will be closer to the median voter than R, and therefore left-leaning candidates will win more elections than right-leaning candidates when the distribution of right leaning voters grows asymmetrically more extreme.

Not only does asymmetric polarization lead to more victories for left-leaning candidates, but the average winning candidate on the left will be more extreme. We can see that the shift from f_R to g_R lowers the moderate's expected vote share, but the left-leaning candidate's vote share remains unchanged. This in turn means that the left-leaning candidate needs fewer votes under g_R than under f_R to beat M and progress to the second round of the election. Therefore a more extreme left-leaning candidate is more likely to progress under g_R than under f_R.Taken together, as asymmetric polarization increases, the expected election outcome grows more extreme in the opposite direction.

We now look at an example using the same functional form as in the example for symmetric polarization. Let $f_R(x)$ be a normal distribution with parameters μ and $\sigma^2 = 1$, truncated below at 0. Let $f_L(x)$ be defined similarly, truncated above at 0, and with a fixed mean of 0. That is, we consider changes to just the distribution of right-leaning voters while leaving the distribution of left-leaning voters unchanged. Figure 5 plots the expected outcomes under IRV. The black line is the expected location of the winning candidate, that is $E[W] = P(W=L)\mathbb{E}(L|W=L) + P(W=R)\mathbb{E}(R|W=R)$. The blue and red lines decompose this into the weighted conditional expectations of the left candidate and the right candidate.

When the electorate is symmetrically polarized at $\mu = 0$, the expected winner is located at 0. That is, while IRV leads to non-median outcomes in expectation, the distribution of winners is symmetric. As right-leaning voters grow more extreme, the probability that R wins is decreasing. If right voters grow extreme enough, the probability of R winning goes to 0. This increases the probability that L wins, and at the same time, the average winner on the left grows more extreme. Therefore the weighted conditional expectation of the left candidate is increasing in absolute value. Finally, because the probability of the moderate winning is decreasing as well, outcomes as a whole move farther to the left.

Concluding Comments

IRV is quickly growing in popularity, and millions of voters will soon be faced with the choice of whether or not to adopt IRV. Advocates claim that it will lead to moderation, yet offer little empirical or theoretical evidence for that claim.

Our primary finding is that IRV is not a meaningful buffer against extremism. Specifically, our model indicates that as polarization intensifies through shifting or stretching the distribution of voters, the odds of a moderate candidate winning decline. That is, a candidate located at the median voter's ideal point often struggles to gather enough first-choice votes to even survive the initial round. Strategic positioning by the moderate can increase the odds of winning, but at the cost of the moderate adopting increasingly extreme positions to secure victory.

In the context of asymmetric polarization, we find that IRV favors the less extreme party. However, this in turn means that it can be expected to produce winners who are increasingly extreme in the opposite direction, thereby failing to represent the median voter. If it is true that the Republican electorate has become more polarized than the Democratic electorate, then our findings support the claim that IRV may favor Democrats over Republicans.

Given these findings, we argue that IRV is not the electoral panacea it is often proclaimed to be. Alternative methods exist that guarantee the election of moderate candidates in polarized environments (7).

1. SA Levin, HV Milner, C Perrings, The dynamics of political polarization (2021).
2. S Iyengar, Y Lelkes, M Levendusky, N Malhotra, SJ Westwood, The origins and consequences of affective polarization in the united states. *Annu. review political science* **22**, 129–146 (2019).
3. N McCarty, *Polarization: What everyone needs to know®.* (Oxford University Press), (2019).
4. JD Mussel, H Schlechta, Australia: No party convergence where we would most expect it. *Party Polit.* p. 13540688231189363 (2023).
5. PK Pattanaik, Positional rules of collective decision-making. *Handb. social choice welfare* **1**, 361–394 (2002).
6. JH Smith, Aggregation of preferences with variable electorate. *Econom. J. Econom. Soc.* pp. 1027–1041 (1973).
7. N Atkinson, SC Ganz, EB Foley, Beyond the spoiler effect: Can ranked choice voting solve the problem of political polarization? *Univeristy Ill. Law Rev.* (Forthcoming).
8. R Holden, J Quiggin, On the optimality of ranked-choice voting. *Available at SSRN* (2022).
9. M Shaked, JG Shanthikumar, *Stochastic orders.* (Springer), (2007).
10. MJ Osborne, A Slivinski, A model of political competition with citizen-candidates. *The Q. J. Econ.* **111**, 65–96 (1996).
11. RA Dahl, *On democracy.* (Yale university press), (2020).
12. D Black, On the rationale of group decision-making. *J. political economy* **56**, 23–34 (1948).
13. C Andris, et al., The rise of partisanship and super-cooperators in the us house of representatives. *PloS one* **10**, e0123507 (2015).
14. T Economist, Far-left candidates did poorly in the democratic primaries. *The Econ.* (year?).
15. PR Center, The partisan divide on political values grows even wider. (2017).
16. PR Center, Political polarization in the american public. (2014).
17. SB Warner, Toward an ideological common space: Extending bonica's cfscores to the citizen level. *Polit. Behav.* pp. 1–22 (2023).
18. M Duverger, *Political parties: Their organization and activity in the modern state.* (Metheun & Co. Ltd.), (1959).
19. NE Leonard, K Lipsitz, A Bizyaeva, A Franci, Y Lelkes, The nonlinear feedback dynamics of asymmetric political polarization. *Proc. Natl. Acad. Sci.* **118**, e2102149118 (2021).