A Group-Based Theory of Electoral Turnout

Dan Bernhardt* Laurent Bouton† Stefan Krasa‡ Francesco Squintani§

March 20, 2025

Abstract

We propose a model of costly voting in a canonical spatial setting, where citizens have both expressive and instrumental voting motives. We build on the ethical voter model with the main novelty that voting groups/coalitions form endogenously. Our individual-based notion of group formation requires that citizens do not free ride on each others, but also cannot be forced to join a coalition they do not benefit from. We first prove the existence of coalitional voting equilibria in general settings. We then derive how the underlying primitives—e.g., voting costs, elite polarization, and the distributions of voter bliss points—interact with the extent of expressive vs. instrumental voting motives to affect equilibrium coalition formation, voter turnout and turnout rates. Finally, we explore the consequences of these primitives for candidate platform choices. Many of our comparative static predictions find support in the data, and we provide testable insights into how voter motives influence voting coalition formation and candidate platform choices.

*Department of Economics, University of Illinois and Department of Economics, University of Warwick, danber@illinois.edu
†Department of Economics, Georgetown University, and NBER. lb910@georgetown.edu
‡Department of Economics, University of Illinois, skrasa@illinois.edu
§Department of Economics, University of Warwick, f.squintani@warwick.ac.uk

1 Introduction

Voters are at the heart of the democratic decision-making process. They influence policy directly through referenda and ballot initiatives, and indirectly by electing representatives and holding them accountable. In 2024 alone, approximately 70 nations—representing about half of the world's population—held elections to choose their national leaders, including presidents and national assembly members. Beyond these national-level elections, citizens often participate in local elections to select state legislators, mayors, school board members, etc. Despite the ubiquity and importance of voters around the world, scholars remain divided on how to formally model voter behavior, and existing models deliver predictions at odds with empirical evidence. A consequence is that addressing questions such as the optimal design of voting laws is impaired, as the answers rely on the appropriateness of the underlying model. It also undermines the validity of other political economy models: even when these models are not explicitly focused on voters, accurately accounting for voter behavior is essential to capture voters' influence on other actors and dimensions of the democratic decision-making process.

The lack of consensus does not imply the absence of voting models. On the contrary, many model variants exist, including the so-called expressive voting, pivotal voting, and group-based voting models. However, each of these models has serious weaknesses. As we elaborate in the literature review, the expressive voting and pivotal voting models both fail to align with empirical evidence: expressive voters are unrealistically insensitive to their electoral environment, while pivotal voters have essentially no reason to vote, i.e., they suffer from the paradox of voting. Group-based models avoid these pitfalls, as they can predict high voter turnout that is responsive to the characteristics of the electoral environment. However, these models simply assume that the electorate is exogenously divided into large groups or coalitions. The lack of criteria to determine why groups form makes it impossible to determine in which situations, if any, the existence of such large groups is reasonable.

This paper develops a tractable model of voting within the canonical spatial model pioneered by Downs (1957). The model incorporates voters with both expressive and ethical motives.[1] The novelty of our approach is that voting groups/coalitions form endogenously: we do not exogenously impose the group to which a citizen belongs, and hence which fellow citizens affect their decision of whether to vote. We characterize the equilibria of our endogenous-group model of voting, and derive comparative statics results which strongly align with voting patterns identified by empirical studies. We also show how the tractability of our model allows it to be embedded in other political economy models. In particular, we embed it

[1]Research has indeed shown that voters (i) derive a direct "expressive" benefit from voting for a candidate whose platform closely aligns with their preferences (see, e.g., Spenkuch 2018, Pons and Tricaud 2018, Ujhelyi et al. 2021), and (ii) may choose to bear the cost of voting out of loyalty to like-minded individuals, a sense of civic duty toward fellow citizens, or social pressure (see, e.g., Gerber et al. 2008, Feddersen, Gailmard, Sandroni 2009, Funk 2010, Della Vigna et al. 2017).

in a traditional model of electoral competition between two policy-motivated candidates and analyze how strategic voting shapes candidates' political positioning and election outcomes.

Our model assumes the presence of (left and right) partisan and non-partisan citizens. Partisan citizens are non-strategic and show up in random numbers on election day to vote for their preferred candidate. Non-partisan citizens with horizontally-differentiated policy preferences rationally decide whether to vote or abstain, and voting is costly. A non-partisan citizen's payoff depends on both her voting choice and the election outcome–receiving a direct "expressive" payoff from voting for a candidate and an "instrumental" payoff derived from the winning candidate's electoral platform. The "instrumental" payoff decreases with the distance between the non-partisan's preferred policy and the winning candidate's platform; the "expressive" payoff decreases in the distance between her bliss-point and the platform of the candidate she votes for. Importantly, non-partisan citizens do not act independently of each other, but rather internalize the welfare of like-minded voters by possibly coordinating into groups or coalitions that collectively vote for the same candidate. Such "ethical" considerations may lead to coalitions that include some citizens whose expressive benefits from voting would not, on their own, outweigh the cost of voting.

We develop a new individual-based notion of group coordination. We adopt the insight from the ethical voting literature (e.g., Feddersen and Sandroni (2006); Coate and Conlin (2004)) that voters may incur large psychic costs from free-riding on the costly voting efforts of others by abstaining. However, distinct from the classical ethical voting formulations where coordination is facilitated by a social planner who seeks to maximize group welfare, our theory of voting coalition formation is posed at the individual level. We ask: when would an individual necessarily not feel a personal moral obligation to join a coalition? and when would an individual necessarily feel a personal moral obligation to join a coalition?

We first argue that an individual *should not feel* a personal moral obligation to join a voting coalition if she would be better off abstaining even if doing so meant that the coalition did not form (in which case only voters whose expressive benefit exceeds the voting cost would support the candidate). Imposing this "no coercion" condition ensures that any voter who joins a candidate's coalition benefits from its formation. We then argue that an individual *should feel* a personal moral obligation to join a voting coalition if she is one of a set of voters who, by joining, would all be made strictly better off vis-à-vis abstaining and have no coalition, save for the minimal expressive one, form. If such voters were not to join, then they would be free-riding on the original coalition by not participating, and hence incur a psychic ethical cost. Imposing this "no free riding" condition means that enlarging a coalition cannot make all members of the larger coalition strictly better off.

Our coalitional equilibrium concept only imposes these two basic conditions. Because it is not a refinement of Nash equilibrium, standard existence results cannot be directly applied. We start by establishing the existence of a coalitional equilibrium in several settings: (i) when the two candidate platforms and the

distribution of voter bliss points are symmetric around the median voter's bliss point; (ii) when the random partisan vote advantage is uniformly distributed with sufficiently large support, and (iii) (trivially) when expressive motives are sufficiently strong.

We then characterize the unique coalitional equilibrium for two opposite situations: when citizens have purely instrumental motives, and when they have purely expressive motives. Beyond simplifying the analysis, this allows us to highlight the qualitative consequences of the nature of voter motives. It is also useful to understand the mixed motive case, which proves to be a weighted combination of these two opposite cases. We also work under the assumption that the distribution of the random net partisan vote is uniform. This greatly simplifies the analysis because, in that case, the behavior of instrumentally-motivated citizens on the left is unaffected by the exact nature of strategic competition on the right and vice versa.

We find that for expressively-motivated citizens, incentives to abstain are the same for moderates and non-moderates. Expressive citizens on both sides of a candidate's platform can be "alienated." For instrumentally-motivated citizens, the situation is different: while moderates may abstain due to a lack of differentiation between the two candidates, the non-moderates are never alienated because (i) they are more reactive to differences in platforms that are further away from their bliss-point, and (ii) they do not bear an expressive cost of voting for a candidates whose policy is far from their bliss-point. The result that non-moderates are never alienated is the extreme version of a feature of strategic coalition that holds quite generally: with instrumental motives, moderates are less willing to participate than non-moderates.

We then derive how varying the primitives of the electoral environment impacts non-partisan turnout for the two opposite cases of purely instrumental motives and purely expressive motives. We focus on the comparative statics of turnout that speak to empirical studies identifying a clear causal impact of specific features of the electoral environment.

First, we show that, consistent with empirical findings and regardless of the nature of citizen preferences, higher voting costs reduce non-partisan turnout. In contrast, various other primitives have sharply different effects on coalition formation, according to whether instrumental or expressive preferences dominate. For example, increasing elite polarization, i.e., a symmetric spreading of candidate platforms, increases turnout if citizens have purely instrumental preferences, but does the opposite if citizens have purely expressive preferences. With instrumental preferences, the intuition for the increased turnout reflects that shifting platforms further apart raises the stakes of winning the election. With expressive preferences, in contrast, the result reflects that the set of citizens close enough to the candidate's platform for whom the expressive benefits of voting outweigh the cost is shifted toward the tail where there are fewer citizens.

Similarly, the impact of shifting the distribution of non-partisan citizens toward the candidate on the right hinges on the nature of voter preferences. With purely instrumental preferences, it increases voter turnout on the right and decreases it on the left. This reflects that, ceteris paribus, the larger is the coalition

that a citizen might join, the greater is the benefit from being a member. Thus, the coalition supporting R must expand, while the opposite must hold for candidate L. With purely expressive preferences, a shift in the support leads to a rise in turnout for one candidate and a fall for the other, but which candidate gains support and which one loses depends on the dispersion in the distribution of voter ideal points.

We also explore the impact of differences in support for the two candidates in the electorate on the turnout rates. Previous theoretical research has established the existence of an underdog effect, i.e., supporters of the runner-up turn out at a higher rate than supporters of the front-runner. This prediction has been tested in the laboratory and most studies (Duffy and Tavits 2008, Grober and Schram 2010, Kartal 2015, and Agranov et al. (2018)) find the opposite result: supporters of the front-runner turn out at a *higher rate* than supporters of the runner-up–the so-called *bandwagon effect*. Our model predicts such a bandwagon effect if the instrumental motives are sufficiently strong. Indeed, instrumental voters are subject to a discouragement effect when they anticipate that they have fewer co-supporters in the electorate, as the benefit of forming a voting coalition then decreases, and to an encouragement effect when they anticipate that they have more co-supporters in the electorate.

We conclude analysis by endogenizing candidate platform choices along the lines of Wittman (1983) and Calvert (1985) within our coalitional voting framework. We first show that as long as voting costs are not too high, elite polarization is greater when citizens have purely expressive preferences than when they have purely instrumental preferences. To understand the intuition, remember that, with expressive preferences, only citizens sufficiently close to a candidate's platform participate. Hence, the cost to a candidate of adopting a more extreme platform, i.e., of moving closer to his ideal policy, is limited. While such a move to the extreme alienates some moderates, it also convinces some non-moderates to vote. By contrast, with instrumental preferences, only moderates abstain. Hence, a move to the extreme may push more moderates to abstain, but without the benefit of convincing more non-moderates to vote. The impact on the probability of winning is thus larger, which limits the incentives of candidates to move outward.

Then, we explore the impact of a change in voting costs on polarization. With expressive preferences, increases in voting costs lead to an increase in polarization. This is because, following an increase in voting costs, the marginal moderate and non-moderate voters get closer to the candidate's platform. With a single-peaked density over voter ideal points, this reduces the net number of voters lost by choosing a more extreme platform, reducing further the cost of locating more extremely. In contrast, when voters' instrumental preferences are sufficiently strong, elite polarization *decreases* in voting costs. This occurs because raising voting costs decreases voting of moderates, making the cutoff voters on each side more extreme. As voters are risk averse, this, in turn, implies that the candidate L's marginal strategic coalition member is more sensitive to changes in candidate R's platform than R's marginal coalition member. In particular, if R locates more moderately, he always gains more votes on the left than he may lose on the right

(and, whether he gains or loses votes on the right depends on other primitives). Hence, if candidate R moves marginally to the center he gains more votes after the cost increase, incentivizing moderation.

1.1 Literature

Our paper is closely related to the literature on costly voting. As mentioned previously, various models have been proposed in the literature. The expressive voting model assumes that voters enjoy the act of voting for an alternative they like, and that they compare this benefit to their cost of voting to decide whether to vote. While research has shown that voters derive a direct "expressive" benefit from voting for a candidate whose platform closely aligns with their preferences,[2] this model's predictions are contradicted by the facts.[3] Nevertheless, this approach to voters' behavior is not clearly dominated by the workhorse model of strategic voting, i.e. the pivotal voter model. That model considers fully rational and instrumental agents, who turn out to vote only if the probability that their ballot is pivotal in changing the identity of the winner of the election is sufficiently large in comparison to the cost of voting (Riker and Ordeshook (1968), Palfrey and Rosenthal (1983), Palfrey and Rosenthal (1985), Myerson (1998), Castanheira (2003), Myatt (2012)). The pivotal voter model is afflicted by a serious weakness: the vanishingly low turnout levels it predicts– highlighting the *paradox of voting*.[4,5]

In reaction to the weaknesses of those widely used models, the so-called group-based voting models have been proposed. There are three main variants of those models: (i) ethical voter models assume that individuals are ethical, or rule-utilitarian, in the sense that they decide whether to vote depending on what is optimal for individuals in their group or for the electorate as a whole (Harsanyi (1977), Feddersen and

[2]There is clear evidence that voters receive a direct payoff from supporting a particular candidate while also caring about the election outcome. Empirical studies based on election data (e.g., Fujiwara (2011); Pons and Tricaud (2018); Spenkuch (2018)) and laboratory experiments (e.g., Forsythe et al. (1996); Duffy and Tavits (2008); Esponda and Vespa (2014); Bouton et al. (2015)) support this dual motivation. The original concept of expressive voting was introduced by Brennan and Lomasky (1993).

[3]There are various pieces of evidence that do not square well with the predictions of the expressive voting model. First, the literature comparing each voter's actual vote to her preferences finds a fraction of voters casting misaligned votes (i.e. voting for another candidate than their most-preferred one) ranging from 3% to 17% (see, e.g., Alvarez and Nagler 2000, Blais, Nadeau, Gidengil, and Nevitte 2001 and papers cited therein). Second, some studies (Kawai and Watanabe 2013, Spenkuch 2016) find that a larger fraction of voters are strategic (estimates range from 40% to 85%), but they do not necessarily cast a misaligned vote. Finally, the experimental literature finds evidence of strategic voting both in two-candidate elections (for the costly voting side, see e.g., Palfrey and Levine 2007; for the Condorcet jury theorem side, see e.g., Guarnaschelli et al. 2000, Battaglini et al. 2008, 2010, Goeree and Yariv 2011, Bhattachary et al. 2014, Bouton et al. 2017a) and in multicandidate elections (see e.g. Forsythe et al. 1993 and 1996, Fisher and Myatt 2001, Morton and Rietz 2008, Bouton et al. 2016, 2017b).

[4]Models with aggregate uncertainty about the distribution of preferences predict a turnout rate that converges to zero at a much slower rate that models without aggregate uncertainty (Good and Mayer (1975), Castanheira (2003), Myatt 2012). Yet, for large electorates, these models still predict counterfactually low turnout rate.

[5]Another weakness of the pivotal voter model is its predictions of the behavior of voters when there are more than two candidates in the race. See the discussions in, e.g., Fey (1997), Bouton et al. (2015), and Bouton and Ogden (2021).

Sandroni (2006), and Coate and Conlin (2004)); (ii) mobilization voting models assume that individuals' turnout decisions depend on how much mobilization effort leaders exert (Uhlaner (1989), Morton (1991), Shachar and Nalebuff (1999)); and (iii) social pressure models assume that individuals' decisions to vote are influenced by the pressure of their peers, e.g., they want to appear prosocial to others or avoid a costly punishment from their peers (Ali and Lin (2013), Levine and Mattozzi (2020)). All these variants predict high turnout levels (and hence provide an explanation for the paradox of voting), and produce various comparative statics that are empirically sound. While direct tests of those models on observational data are encouraging in comparison to both the purely expressive voting model and the pivotal voter model (Coate and Conlin (2004), Coate et al. (2008)), their absolute performance, even in the small elections being analyzed, are not fully satisfying.[6]

Another issue with the group-based voting models is the assumption that the group to which a voter belong is determined exogenously.[7] For instance, in the ethical voting model of Feddersen and Sandroni (2006), voters consider the welfare of the whole electorate, whereas they only consider the welfare of voters with similar preferences in the model of Coate and Conlin (2004). Similarly, in Shachar and Nalebuff (1999), leaders can target mobilization effort only to their supporters. Beyond the fact that it is unsatisfactory to have a theory of voting that relies on voters belonging to groups but that does not explain how those groups are formed, a perhaps bigger issue is that those models' predictions are sensitive to the specific assumptions made about which individuals belong to which groups. To see this, just notice that the ethical voter model is equivalent to the pivotal voter model when each individual forms her own group.[8] The ability of group-based voting model to escape the paradox of voting thus requires that the electorate is partitioned in sufficiently large groups of voters. But we do not know in which situations, if any, this is a reasonable assumption.

Our voting model embraces the group-based approach without exogenously determining the group to which someone belongs. Instead, voting groups and coalitions form endogenously following two simple rules: no free-riding and no coercion. Based on these rules, our model predicts which voters will form groups, and under which conditions those groups are sufficiently large to guarantee a high turnout rate.[9]

Our model delivers various comparative statics that qualitatively align with systematic patterns in voting behavior identified in the empirical literature. The ones we deem most important, because of the prominent

[6]Using experimental data, Merlo and Palfrey (2018) also show sub-par performances of the ethical voter models in terms of concealed parameter recovery (a kind of external validation analysis).

[7]One exception is Herrera and Martinelli (2006). They consider a model in which group leaders emerge endogenously out of the population. These leaders then compete for followers. Note however that there is still an exogenous dimension to the groups in this model since leaders are exogenously and randomly assigned a group of voters that they can influence. Their effort then determine the share of individuals in their group that turn out to vote.

[8]See also Section 5.2 in Bouton and Ogden (2021) for relevant results about multicandidate elections.

[9]Bernhardt et al. (2024) uses a closely related group-based approach to analyze strategic voting in a multi-candidate setting where voting is costless. Their model also produces empirically sound predictions.

position they occupy in the voting literature, concern the impact on turnout of a change in the cost of voting, in elite polarization or the stake of the election, in the platform of one of the two candidates, and in the closeness of the election (measured in terms of the support for the two candidates in the electorate). We discuss this further below, when we present the relevant results. We also explore the impact of features of the electoral environment on the turnout rates of supporters of the two candidates. Our finding that a bandwagon effect exists when voters a instrumentally motivated, which in stark contrast with the typical prediction of an underdog effect by extant models in the literature, helps explain otherwise puzzling findings of the empirical literature.

Finally, our analysis is also related to the literature on the strategic choice of platforms by candidates (Downs 1957, Wittman (1983), Calvert (1985), Lindbeck and Weibull 1987, Dixit and Londregan 1996, Persson and Tabellini 1999, 2000, Grossman and Helpman 2005, and Stromberg 2008, Bierbrauer et al.2022). Given that we assume a continuum of citizens with spatially differentiated bliss points, our model is more closely related to the "spatial politics" strand of that literature (Downs, 1957; Calvert, 1985; Wittman, 1983; Ledyard, 1984; Adams and Merrill, 2003). Most of these studies assume that voters are expressive (see, e.g., Calvert (1985) and Wittman (1983), for a model with costless voting, and Adams and Merrill (2003), for a model with costly voting).[10] These models predict full convergence to the bliss point of the (expected) median when candidates are purely office-motivated, and partial convergence (i.e., between the bliss point of the median voter and the candidate's) when candidates are policy-motivated. Our model delivers similar predictions in terms of convergence of platforms. Yet, it produces substantially different comparative statics. For instance, our model predicts that elite polarization (i.e., the distance between the platforms of the two candidates in equilibrium) may actually *decrease* with the cost of voting (if the voters' instrumental motives are sufficiently strong). This has important implications for the evaluation of policies influencing the cost of voting. For instance, Oprea et al. (2024) argue, based on traditional models of spatial politics with expressive voters, that an important argument in favor of compulsory voting (a policy which decreases the cost of voting) is that it should decrease elite polarization. We show that, when taking the strategic behavior of voters into account, this conclusion can be completely reversed.

2 Model

There are two candidates, $i = L, R$, with policy positions x_i, and a continuum of citizens. There are two types of citizens—partisans and non-partisans. A partisan for candidate i always votes for i regardless of the policy positions taken. Thus, the partisan support for a candidate i is summarized by the number $\rho_i \geq 0$ of i's partisan supporters. Without loss of generality, we restrict attention to the distribution, G, over the net

[10] One exception is Ledyard (1984), which embeds the pivotal voter model in a model of spatial politics.

partisan turnout for candidate R, $\rho = \rho_R - \rho_L$.

Non-partisans are distinguished by their ideal policy bliss points, y, which we refer to as the non-partisan's type. Let Φ be the cdf over voter types y in $\mathbb{R}$. With slight abuse of notation we also use Φ to denote the probability measure. For some results, we need Φ to satisfy:

Assumption 1 *The distribution of citizen ideologies, Φ, is single peaked, symmetric around zero, and log concave.*

Non-partisan have both expressive and instrumental voting considerations. Each non-partisan y chooses whether or not to vote, $v \in \{0, 1\}$. If non-partisan y votes for a candidate with policy x then she receives expressive utility from voting for that candidate. She also receives an instrumental payoff derived from the policy x_w of the winning candidate. Let $\beta \in [0, 1]$ be the weight on a sophisticated voter's expressive preferences. Both the expressive utility of voting and the instrumenal payoff from the winning policy have a quadratic form.[11] Non-partisan y's utility from voting decision v is[12]

$$u_y(v, x, x_w) = -v\beta(x - y)^2 - (1 - v)\beta A - (1 - \beta)(x_w - y)^2 - vc,$$

where $c \geq 0$ is the cost of voting, and $-A < 0$ is an expressive payoff from abstention.[13]

Given that there is a continuum of citizens, an individual citizen cannot influence the electoral outcome on their own. However, non-partisans may come together to form a voting coalition with some other members, instead of abstaining, in order to increase the probability that their preferred candidate wins. The timing of the game is as follows. First, non-partisans choose whether to abstain or to be part of a voting coalition S_i for a candidate $i = L, R$. After coalitions are formed, voting occurs and the uncertainty over the net extent of partisan turnout, ρ, is realized. Candidate L wins if and only if $\Phi(S_L) > \Phi(S_R) + \rho$. The probability that candidate L wins is thus $G(\Phi(S_L) - \Phi(S_R))$.

A voter coalition $S_i \subset \mathbb{R}$ for candidate $i = L, R$ is given by the set of all non-partisan types who vote for candidate i. Non-partisans with ideologies $y \notin S_L \cup S_R$ abstain. We can decompose a voter coalition into two groups: citizens whose expressive payoff from voting exceeds their voting costs, and citizens who must be motivated by strategic considerations to vote.

Denote the set of citizens whose expressive payoff from voting for candidate i exceeds the payoff from abstention and the payoff from voting for the other candidate by E_i:

$$E_i = \{y : -\beta(x_i - y)^2 - c \geq \max\{-\beta A, -\beta(x_{-i} - y)^2 - c\}\}. \quad (1)$$

[11] The quadratic form facilitates explicit solution. Qualitatively, our analysis only hinges on non-partisans being globally strictly risk averse.

[12] As discussed in Appendix 5.2, we can also integrate candidate valence into our model.

[13] Note that voters' choices would be unchanged if we were to add a positive constant to both A and the payoffs from voting outcomes as only the net payoffs determine decision making; in effect we normalize this constant to zero.

Citizens in E_i will always vote for candidate i, as it maximizes their payoffs. In principle, E_i could include a citizen at position $y = (x_L + x_R)/2$ who is indifferent between the two candidates. This occurs when x_L and x_R are sufficiently close. It eases presentation and analysis to rule out this case in order to avoid having to augment proofs to accommodate this special case. Thus, we assume that there is sufficient platform differentiation between candidates.

Assumption 2 *Voting costs satisfy $c \leq A$ and $x_R - x_L > 2\sqrt{A - c}$.*

We say that coalition S_i is strategic if $S_i \neq E_i$. Citizens $y \in S_i \setminus E_i$ join the coalition for strategic (and ethical) reasons. Our coalitional equilibrium concept is formulated at the individual voter level, determining whether an individual joins a coalition or abstains. Our coalitional concept identifies when an individual should or should not feel an ethical obligation to join a voting coalition. Specifically, an individual *should not feel* a personal moral obligation to join a voting coalition if he would be better off abstaining even if doing so meant that the strategic coalition did not form (in which case only voters whose expressive benefit exceeds the voting cost would support the candidate). Conversely, an individual *should feel* a personal moral obligation to join a voting coalition if he is one of a set of voters who, by joining, would all be made strictly better off vis-à-vis abstaining and have no coalition form. Imposing these "no coercion" and "no free riding" conditions ensures that (i) any voter who joins a candidate's coalition benefits from its formation, while (ii) enlarging a coalition cannot make all members of the larger coalition strictly better off, else by abstaining they would be free-riding on the voting efforts of others and hence incur a large psychic cost.

To define those conditions formally, we first need to compute citizen y's payoff when joining a coalition and when the coalition does not form. Citizen y's payoff from voting for candidate L is

$$-\beta(x_L - y)^2 - (1-\beta)\Big(G(\Phi(S_L) - \Phi(S_R))(x_L - y)^2 + (1 - G(\Phi(S_L) - \Phi(S_R)))(x_R - y)^2\Big) - c. \tag{2}$$

The first term is citizen y's expressive payoff from voting for candidate L, the second terms capture citizen y's (win-probability weighted) instrumental payoffs when candidates L and R win, and the final term is the voting cost. The payoff to citizen y from abstaining (when only expressive voters join the coalition in support of candidate L) is

$$-\beta A - (1-\beta)\Big(G(\Phi(E_L) - \Phi(S_R))(x_L - y)^2 + (1 - G(\Phi(E_L) - \Phi(S_R)))(x_R - y)^2\Big) \tag{3}$$

The net benefit to citizen y from having coalition S_L form is given by the difference (2) – (3):

$$\begin{aligned} \gamma_L(y, S_L, S_R) \;=\; & \beta(A - (x_L - y)^2) - c \\ & + (1-\beta)\left(G(\Phi(E_L) - \Phi(S_R)) - G(\Phi(S_L) - \Phi(S_R))\right)\Big((x_L - y)^2 - (x_R - y)^2\Big) \end{aligned} \tag{4}$$

Similarly, y's net benefit from having coalition S_R form is

$$\begin{aligned}\gamma_R(y, S_L, S_R) &= \beta(A - (x_R - y)^2) - c \\ &\quad +(1-\beta)\left(G(\Phi(S_L) - \Phi(E_R)) - G(\Phi(S_L) - \Phi(S_R))\right)\left((x_L - y)^2 - (x_R - y)^2\right).\end{aligned} \tag{5}$$

The no free-riding condition formalizes what it means for someone to behave "ethically." An individual y who obtains a net benefit from having coalition S_i form, instead of having only expressive voters in E_i vote for candidate i, feels an ethical obligation to support the other coalition members by voting for i. Specifically, someone who is not in a coalition must prefer not having the coalition form to having it expand to include her:

Definition 1 (No Free Riding) *A coalitional voting equilibrium (S_L, S_R) satisfies no free riding if and only if for $i = L, R$ there exists $\varepsilon > 0$ such that for all sets B with $B \cap (S_L \cup S_R) = \emptyset$ and $\Phi(B) < \varepsilon$ all members of B are strictly worse off joining coalition S_i, i.e., $\gamma_i(y, S_i \cup B, S_{-i}) < 0$ for all $y \in B$.*

The no-coercion condition ensures that no voter feels an ethical obligation to join a coalition for a candidate i if doing so makes that voter worse off in expectation than if she abstained even when abstaining means that candidate i only receives votes from voters for whom the expressive net benefit of voting for candidate i exceeds the expressive payoff $-A$ from abstention.

Definition 2 (No Coercion) *A voter coalition S_i satisfies no coercion if all of its members are at least as well off in expectation if the coalition forms than if it does not. That is, $\gamma_i(y, S_L, S_R) \geq 0$ for all $y \in S_i$.*

Finally, we define a coalitional voting equilibrium:

Definition 3 (Coalitional Voting Equilibrium) *Coalitions S_i, $i = L, R$ comprise a coalitional voting equilibrium if and only if the coalitions satisfy no coercion and no free riding.*

3 Equilibrium Analysis

We organize the equilibrium analysis in five parts. First, we prove preliminary results about some necessary features of equilibrium voting coalitions. These preliminary results are key to prove and understand our main results. Second, we identify different sets of (fairly general) conditions that guarantee the existence of a coalitional voting equilibrium. Third, we characterize the unique coalitional voting equilibrium for two extreme, but highly relevant cases: purely expressive preferences ($\beta = 1$) and purely instrumental preferences ($\beta = 0$). The realistic setting with mixed motives, i.e., $\beta \in (0, 1)$, largely inherits the properties

from these two extreme scenarios. Fourth, we prove some comparative static results about the impact of features of the electoral environment on turnout. Finally, we explore the impact of these features on turnout rates, with a special focus on so-called bandwagon and underdog effects.

3.1 Preliminary Results

First, note that, for fixed S_L and S_R, the net benefit $\gamma_i(y, S_L, S_R)$ is continuous and quadratic in a citizen's ideal policy y. On the one hand, the expressive payoff from voting is quadratic in y. On the other hand, the instrumental payoff from voting is increasing in $|y|$ as more extreme voters benefit more from the increase in the probability that their preferred candidate wins if the coalition forms. For $|y|$ sufficiently large, the quadratic impact on expressive payoffs dominates. Second, note that, for a given citizen y, expanding the coalition to which she belongs increases her net benefit $\gamma_i(y, S_L, S_R)$. This is because joining a larger coalition is more valuable to her as the impact of the coalition on the electoral outcome is larger.

Lemma 1 *Let (S_L, S_R) be a coalitional voting equilibrium, with $S_i = [\underline{y}_i, \bar{y}_i]$. Then:*

1. *For $\beta > 0$, coalition S_i, $i = L, R$, is a closed interval with $S_i \supset E_i$.*
2. $\bar{y}_L < (x_L + x_R)/2 < \underline{y}_R$.
3. *If y is on the boundary of S_i then $\gamma_i(y, S_L, S_R) = 0$.*

The results in Lemma 1 stem from the fact that $\gamma_i(y, S_L, S_R)$ is continuous and quadratic in y, combined with the no free-riding and no coercion conditions. The third result, i.e., that the marginal coalition member is indifferent between voting for her preferred candidate or abstaining and having no coalition form, is particularly important to understand the next steps of the analysis.

We next consider the impact of expanding the coalition on the net benefit of its marginal member.

Lemma 2 *Let (S_L, S_R) be a coalitional voting equilibrium, and let $S_i = [\underline{y}_i, \bar{y}_i]$. Then*

$$\frac{\partial}{\partial y}\gamma_i(y, [y, \bar{y}_i], S_{-i})|_{y=\underline{y}_i} \geq 0 \geq \frac{\partial}{\partial y}\gamma_i(y, [\underline{y}_i, y], S_{-i})|_{y=\bar{y}_i}.$$

Lemma 2 says that starting from an equilibrium, expanding the coalition (either to the left or to the right) must have a negative impact on the net benefit of the marginal member. This requires that, at the equilibrium, the direct negative effect of moving y on the net benefit $\gamma_i(y, S_L, S_R)$ must exceed the indirect positive effect on $\gamma_i(y, S_L, S_R)$ from expanding the coalition. Otherwise, the no free-riding condition would require more voters to join the coalition.

3.2 Equilibrium Existence

From Lemma 1, it follows that establishing the existence of a coalitional voting equilibria requires proving that there exists a pair (S_L, S_R), with $S_i = [\underline{y}_i, \bar{y}_i]$, such that

$$\gamma_L\left(\underline{y}_L, S_L, S_R\right) = 0 = \gamma_L\left(\bar{y}_L, S_L, S_R\right), \tag{6}$$

$$\gamma_R\left(\underline{y}_R, S_L, S_R\right) = 0 = \gamma_R\left(\bar{y}_R, S_L, S_R\right), \tag{7}$$

that also satisfy no coercion and no free riding.

We first show that coalitional voting equilibria always exist in symmetric settings.

Proposition 1 *Suppose that Assumption 1 holds. Let $x_L = -x_R$ and the distribution of net partisans, G, be symmetric around zero. Then a symmetric coalitional voting equilibrium exists.*

The key to the argument is that increases in S_i serve to raise the value for citizens of having a strategic coalition form for candidate i, thereby increasing the bite of the no free riding condition. Iterating on the monotone mapping describing the marginal (indifferent) coalitional member, and applying Tarski's fixed point theorem then delivers existence.

Alternatively, we can relax the symmetry requirement on the distribution of citizens' ideologies, Φ, and on candidate locations, and instead establish existence of a coalitional equilibrium when the distribution over net partisan turnout G is uniform.

Assumption 3 *The distribution G over net partisan turnout is uniform on $\left[-\frac{1}{2\alpha}, \frac{1}{2\alpha}\right]$, where $\alpha \leq 0.5$.*

Proposition 2 *Suppose that Assumption 3 holds. Then a coalitional voting equilibrium exists.*

As with Proposition 1, the proof of the existence result uses Tarski's fixed point theorem.

In most of our remaining analysis, we will specialize to a uniform distribution G over net partisan turnout. Doing so greatly simplifies the analysis and improves expositional clarity. The analysis is simpler because, when G is uniform, the conditions describing S_L and S_R are separable and do not depend on each other. More precisely, in an equilibrium, S_L satisfies no free riding and no coercion for any arbitrary (not necessarily equilibrium) $S'_R \subset [(x_L + x_R)/2, \infty)$. Similarly, S_R satisfies no free riding and no coercion for any arbitrary $S'_L \subset (-\infty, (x_L + x_R)/2]$. To see this, consider the equation (6) that describes the 'responses' $\underline{y}_L$ and $\bar{y}_L$ by coalition supporters of L to S_R (that is, it determines the strategic coalition S_L). When G is a uniform distribution with sufficient support, $G(\Phi(E_L) - \Phi(S_R)) - G(\Phi(S_L) - \Phi(S_R))$ reduces to $\alpha(\Phi(E_L) - \Phi(S_L))$.

That is, S_R drops out of the conditions pinning down $\underline{y}_L$ and $\bar{y}_L$. Substituting the uniform distribution on $\left[-\frac{1}{2\alpha}, \frac{1}{2\alpha}\right]$ for $\alpha \leq 0.5$ into (6) yields that $\underline{y}_L$ and $\bar{y}_L$ satisfy:

$$\beta(A - (x_L - y)^2) - c - (1-\beta)\alpha\left(\Phi(S_L) - \Phi(E_L)\right)\left((x_L - y)^2 - (x_R - y)^2\right) = 0, \tag{8}$$

which does not depend on S_R. Analogously, if $S_R = [\underline{y}_R, \bar{y}_R] \neq \emptyset$ then $\bar{y}_R$ and $\underline{y}_R$ solve

$$\beta(A - (x_R - y)^2) - c + (1-\beta)\alpha\left(\Phi(S_R) - \Phi(E_R)\right)\left((x_L - y)^2 - (x_R - y)^2\right) = 0, \tag{9}$$

which does not depend on S_L. This reflects the insight that with a uniform distribution, anything that shifts, say, the support of S_R on the interior does not affect the size of S_R, and hence does not affect the winning probability, which is what enters the strategic considerations of citizens on the left. This, in turn, implies that the size of S_L does not vary with such shifts (and vice versa).[14]

3.3 Equilibrium Characterization

In this section, we characterize the unique equilibrium for two different cases: purely instrumental motives ($\beta = 0$) and purely expressive motives ($\beta = 1$). Focusing on these extreme cases helps with understanding the specific impact of the different motives on voters' behavior. Moreover, analyzing these cases helps shed intuition on outcomes for the intermediate case with mixed motives ($\beta \in (0, 1)$). That case indeed proves to be mostly a weighted combination of the forces at work when $\beta = 0$ and $\beta = 1$.

Purely Expressive Preferences

When voters have purely expressive preferences ($\beta = 1$), we obtain a closed-form solution for the equilibrium voting coalitions:[15]

Proposition 3 *Suppose $\beta = 1$. Then, there exists a unique coalitional equilibrium (S_L, S_R) defined by*

$$S_i = E_i = \left[x_i - \sqrt{A - c}, x_i + \sqrt{A - c}\right]. \tag{10}$$

This implies that, when non-empty, the expressive voter sets E_L and E_R are symmetric around x_L and x_R. This is because incentives to abstain are the same for moderates and non-moderate expressive voters. Thus, expressive voters on both sides of a candidate's platform can be "alienated."

[14] Note that this insight does not rely on the symmetry of G.

[15] As we show in the proof, this result does not require specific assumptions about the forms of the distributions G and Φ.

Purely Instrumental Preferences

Now consider purely instrumental voter preferences ($\beta = 0$). As we have seen above, the instrumental payoff from voting is increasing in the degree of extremism $|y|$ of a citizen's ideology. It follows that the voting coalitions are unbounded in the extreme (see the proof of Propositoin 4 below): if S_L is non-empty then it takes the form $S_L = (-\infty, \bar{y}_L]$, and if S_R is non-empty then it takes the form $S_R = [\underline{y}_R, \infty)$. That is, with purely instrumental voter preferences, non-moderate voters are never "alienated."

The expressions (8) and (9) that pin down $\bar{y}_L$ and $\underline{y}_R$ simplify to:

$$\xi_L(\bar{y}_L) = c/\alpha, \tag{11}$$

$$\xi_R(\underline{y}_R) = c/\alpha, \tag{12}$$

where

$$\xi_L(\bar{y}_L) = \Phi(\bar{y}_L)(x_R - x_L)(x_L + x_R - 2\bar{y}_L); \tag{13}$$

$$\xi_R(\underline{y}_R) = (1 - \Phi(\underline{y}_R))\,(x_R - x_L)\,(2\underline{y}_R - x_L - x_R). \tag{14}$$

To state our characterization result, it is useful to introduce a few pieces of notation. Let $\bar{y}_L^*$ be the unique solution to $\xi_L(\bar{y}_L^*) = c/\alpha$ that satisfies $\xi_L'(\bar{y}_L^*) \leq 0$, and let $\underline{y}_R^*$ be the unique solution to $\xi_R(\underline{y}_R^*) = c/\alpha$ that satisfies $\xi_R'(\underline{y}_R^*) \geq 0$. Define $\bar{c}_L = \max_{y<0.5(x_L+x_R)} \alpha\xi_L(y)$ and $\bar{c}_R = \max_{y>0.5(x_L+x_R)} \alpha\xi_R(y)$.

Proposition 4 *Suppose that Assumption 3 holds and that $\beta = 0$. Then, there exists a unique coalitional equilibrium (S_L, S_R). S_i is non-empty if and only if $c \leq \bar{c}_i$.*

1. *If $c \leq \bar{c}_L$ then $S_L = (-\infty, \bar{y}_L^*]$, where $\bar{y}_L^* < (x_L + x_R)/2$.*
2. *If $c \leq \bar{c}_R$ then $S_R = [\underline{y}_R^*, \infty)$, where $\underline{y}_R^* > (x_L + x_R)/2$.*

Log-concavity of the distribution Φ implies that $\xi(\bar{y}_L)$ is single peaked. Thus, the equation $\xi_i(y) = c/\alpha$ has exactly two solutions $y_{1,i} < y_{2,i}$ if and only if c is not too large. Suppose that ξ is maximized at $\bar{y}_L$. Then a solution to equation (11) exists if and only if $\xi(\bar{y}_L) \geq c/\alpha$, or equivalently if $c \leq \bar{c}_L$, where $\bar{c}_L = \alpha\xi(\bar{y}_L)$. If $c < \bar{c}_L$ then equation (11) has two solutions, $\bar{y}_{L1} < \bar{y}_{L2}$. In a coalitional equilibrium, by Lemma 2, it must be that $\bar{y}_L^* = \bar{y}_{L2}$—else marginal increases of the coalition would make all members, including the new ones, strictly better off, violating no free-riding. Single-peakedness of ξ implies that $\xi'(\bar{y}_L^*) < 0$ when $c < \bar{c}_L$.

An interesting feature of the equilibrium when $\beta = 0$ and G is distributed uniform is that non-moderate voters are never "alienated." This result is the extreme version of a general feature of strategic coalition: moderates are less willing to participate than non-moderates. In particular, the marginal non-moderate strategic coalition member is located further from the candidate's platform than the marginal moderate strategic coalition member (i.e., $x_L - \underline{y}_L > \bar{y}_L - x_L$ and $x_R - \underline{y}_R < \bar{y}_R - x_R$):

Proposition 5 *If the strategic coalitions S_L and S_R are non-empty, then S_L and S_R are asymmetric around x_L and x_R, with $x_L - \underline{y}_L > \bar{y}_L - x_L$ and $\bar{y}_R - x_R > x_R - \underline{y}_R$.*

This result simply reflects that a non-moderate coalition member is hurt by more than a moderate when her less-preferred candidate wins, as he is further away. This contrasts sharply with outcomes when citizens have purely expressive preferences.[16]

3.4 Comparative Statics of Turnout

We now vary the primitives of the electoral environment and contrast the consequences for equilibrium outcomes for the two opposite cases of purely instrumental motives ($\beta = 0$), where strategic considerations dominate, and of purely expressive motives ($\beta = 1$), where those considerations are completely absent. We investigate the impacts on non-partisan turnout which, for candidate i, is given by the number $\Phi(S_i)$ of non-partisans in the coalition S_i. The results follow directly from Propositions 3 and 4.

While our model produces many comparative static results, for the sake of conciseness, we focus on four of them here: the impact of (i) changes in voting costs, c, (ii) changes in elite polarization, $x_R - x_L$, (iii) unilateral shifts in one candidate's platform i.e., either x_L or x_R, and (iv) shifts in the support for candidate R, captured by a shift in the distribution of non-partisan ideologies to the right by μ (i.e., $\Phi_\mu(x) = \Phi(x - \mu)$). This focus is driven by our desire to connect our model with empirical studies identifying a clear causal impact of specific features of the electoral environment on voting outcomes. The other results are relegated to Appendix 5.

We start with the impact of the cost of voting:

Result 1 *Increasing the cost of voting, c, decreases turnout for both $\beta = 0$ and $\beta = 1$.*

With purely expressive preferences ($\beta = 1$), equation (10) immediately implies that E_L and E_R decrease when voting costs increase. We get an analogous result for purely instrumental preferences, $\beta = 0$. To see this, recall that the boundaries of S_L and S_R are given by the solutions to the equations $\xi_L(\bar{y}_L) = c/\alpha$ and $\xi_R(\underline{y}_R) = c/\alpha$ that satisfy $\xi'_L(\bar{y}_L) < 0$ and $\xi'_R(\underline{y}_R) > 0$. Increasing c to c' increases the right-hand side of those equations without affecting the left-hand side. This, in turn, means that $\bar{y}_L$ and $\underline{y}_R$ must both move further away from zero, i.e., turnout falls.

[16] In Adams and Merill (2003) non-participation of moderates is generated by their indifference assumption, i.e., if the difference in utility between candidate L and R is less than some ϵ, then a voter abstains. In our model, both expressive and instrumental motives may drive abstention of moderates. Consider a citizen close to $(x_L + x_R)/2$. From an instrumental motive standpoint, she cares little about who wins. From an expressive motive standpoint, she may be too far from the platform of her preferred candidate to be willing to bear the cost of voting. Concerning the abstention of non-moderates, their notion of alienation reduces to our notion if citizens only have expressive preferences, i.e., if $\beta = 1$.

The left panel of Figure 1 graphically illustrates the simple intuition (which does not hinge on the value of β): if the cost of voting increases, then the net benefit of the marginal coalition voter must increase to retrieve indifference of the marginal member. As starting from equilibrium, expanding the size of the coalition reduces the net benefit of the marginal coalition voter, this means that the coalition must shrink.

This result is in line with empirical evidence about the positive, causal impact on turnout of features of the electoral environment that reduce voting costs. These include voting laws such as early voting laws and all-mail voting laws (see, e.g., Kaplan and Yuan (2020), Thompson et al. (2020), and Yoder et al. 2021), inclement weather on election day (see, e.g., Gomez et al. 2007), proximity to the polling station (see, e.g., Cantoni 2019, Tomkins et al. 2023, and Bagwe et al. 2024), and lax registration requirements (see, e.g., Braconnier et al. 2017).

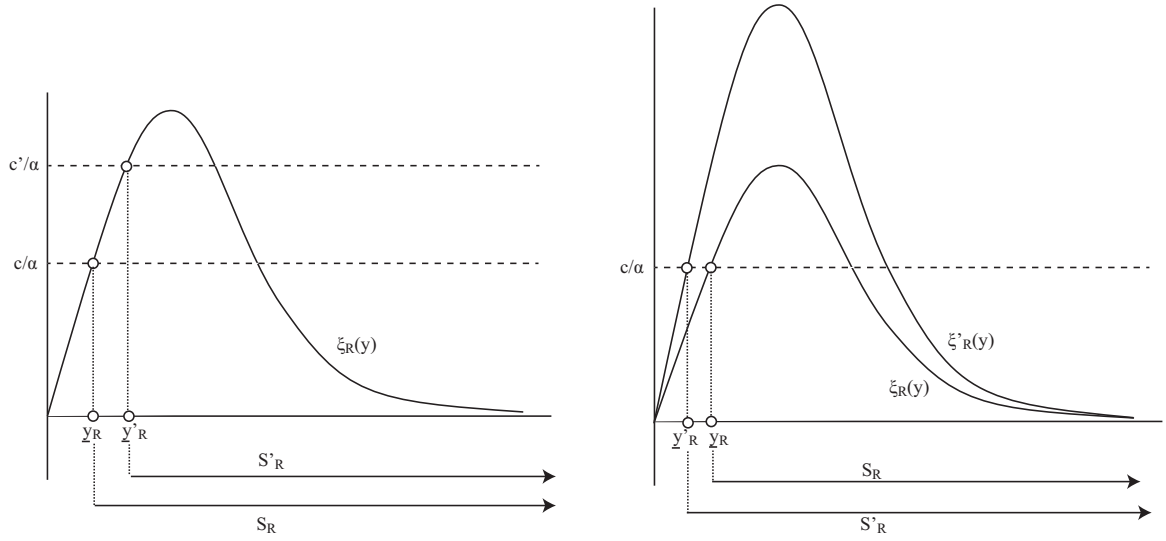


Figure 1: *Left Panel:* Increasing voting costs from c to c' reduces S_R to S'_R for $\beta = 0$. *Right Panel:* Increasing elite polarization, $x_R - x_L$ increases S_R to S'_R for $\beta = 0$.

Our second result focuses on the impact of increased elite polarization, i.e., increased separation between the candidates' platforms x_R and x_L.

Result 2 *Suppose that $x_L < 0 < x_R$. Increasing elite polarization $x_L - x_R$ while keeping $x_L + x_R$ fixed increases turnout when $\beta = 0$, but decreases turnout when $\beta = 1$.*

With purely instrumental preferences, shifting x_R to the right and x_L to the left by the same amount increases $x_R - x_L$, increasing, e.g., $\xi_R(y)$. As a consequence, $\underline{y}_R$ must decrease (since $\xi'_R(\underline{y}_R) > 0$ to get back to equilibrium). The right panel of Figure 1 illustrates this decrease to $\underline{y}'_R$. Intuitively, the marginal strategic voter has more at stake when the candidates locate further apart. In contrast, with purely expressive

preferences, moving x_L to the left and x_R to the right just shifts E_L and E_R to the left and right, respectively. The single-peakedness of Φ then implies that $\Phi(E_L)$ and $\Phi(E_R)$ decrease.

The empirical literature provides (suggestive) evidence of a positive effect of elite polarization on turnout (see, e.g., Dassonneville and Cakir 2021 and references therein). This in line with the prediction of our model as long as instrumental incentives are sufficiently strong. Interpreting elite polarization as the stake of the election for any given voter, this result also aligns with the general pattern that turnout increases with the importance of the office for which candidates are running. For instance, in the US, turnout is systematically higher in Presidential than in Senatorial or House elections (this is true both comparing across races for any given on-cycle election due to voter roll-off, and comparing across races in different electoral cycles, with turnout dropping substantially in mid-term elections).

Our third result concerns the impact of one candidate becoming more extreme:

Result 3 *Suppose that $x_L < 0 < x_R$ and that x_R becomes more extreme while x_L does not change. For $\beta = 1$, turnout for candidate L is unchanged, while turnout for candidate R decreases. For $\beta = 0$, turnout for candidate L rises, while turnout for candidate R can either increase or decrease depending on parameters.*

The results for purely expressive preferences ($\beta = 1$) follow immediately from the characterization of E_L and E_R (the former is independent of x_L and the latter is centered, and hence increasing, in x_R) as well as the fact that the distribution Φ is single peaked.

For purely instrumental preferences ($\beta = 0$), we need to understand the impact of the increase of x_R on $\bar{y}_L$ and $\underline{y}_R$. Note that $\xi_L(y)$ is increasing in x_R. Hence, following the same argument as for the previous result, the relevant solution of the equation $\xi_L(\bar{y}_L) = c/\alpha$ must increase. Intuitively, candidate R's victory is a worse outcome for leftist voters when x_R is more extreme, thus boosting strategic turnout. In contrast, the impact on $\underline{y}_R$ is ambiguous. Recall that $\xi_R(y) \equiv (1 - \Phi(y))(x_R - x_L)(2y - x_L - x_R)$. It is immediate that $\partial\xi_R/\partial x_R$ is positive if $y > x_R$ and negative otherwise. Intuitively, candidate R's victory is a worse outcome for the marginal coalition voter $\underline{y}_R$ if a shift of x_R to the right means moving away from $\underline{y}_R$, but it is a better outcome for that voter if it means moving closer to $\underline{y}_R$.[17]

Under the assumption that, on average, candidates are more extreme than the most moderate citizen voting for them, these results align with the findings of Hall and Thompson (2018) about the causal impact of a more ideologically extreme candidate on turnout. Using a regression discontinuity design on the result

[17] A naïve intuition might suggest that x_R is always larger than $\underline{y}_R$ because a voter whose ideal point is given by the policy x_R should want to be part of a coalition supporting R. However, when candidate policies are not very differentiated, i.e., x_L and x_R are quite close, citizen $y = x_R$ may not care enough about who wins to feel an ethical obligation to vote. In contrast, given the quadratic loss assumption, citizens who are sufficiently extreme care more about who wins. If there are sufficiently many such extreme voters, i.e., if the standard deviation of Φ is sufficiently large, then these extreme voters are incentivized to form a coalition for R.

of the primaries in US House races, they find that the quasi-random nomination of a more extreme candidate (instead of a more moderate one) decreases the turnout of supporters and increase the turnout of opponents in the general elections.

Now consider the effect of moving the distribution of non-partisans by μ to the right to $\Phi_\mu(x) = \Phi(x-\mu)$.

Result 4 *Suppose the distribution of non-partisan ideologies is shifted to the right. For $\beta = 1$, turnout for L decreases and turnout for R increases. For $\beta = 0$, there is no effect on E_L and E_R, and hence the impact on turnout for L and R depends on the distribution of voter ideal points, Φ, and the positions of the candidates, x_L and x_R.*

For purely instrumental preferences ($\beta = 0$), inspection reveals that increasing μ is equivalent to reducing $x_L + x_R$, leaving $x_R - x_L$ unaffected, thereby decreasing ξ_L and increasing ξ_R. Thus, both $\bar{y}_L$ and $\underline{y}_R$ must decrease. For purely expressive preferences ($\beta = 1$), it is clear from the inspection of equation (10) that the shifts does not affect E_L and E_R. This is because expressive citizens do not weigh win probabilities in their decision-making. In terms of turnout, the shift (generically) leads to a rise in the turnout for one candidate and a fall for the other, but which candidate gains support and which one loses depends on the distribution of voter ideal points and the platforms of the candidates.

To see the intuition of the result for instrumental preferences, consider the turnout for candidate L. The move of x_R to the left means that the marginal coalition voter for L (before the move) finds a victory of candidate R a less bad outcome. This reduces her incentives to join the coalition. The same-sized move of x_L to the left may mean that this marginal coalition voter finds a victory of candidate L a better outcome (if $x_L > \bar{y}_L$), which would increase her incentives to join the coalition. However, even when the two effects push in opposite direction, the impact of the change in x_R dominates due to the quadratic loss assumption (as R is further away from the marginal member than L), so the overall effect on the incentives to join the coalition is negative. The same logic yields that turnout for R falls.

These results relate to the literature on the effect of a change in the closeness of the electoral support for the two candidates on turnout. Using different empirical strategies, this literature finds effects of closeness ranging from non-significant to positive but small (Moskowitz and Schneer 2019 and Ainsworth et al. 2024 on US data, and Bursztyn et al. 2024 on Swiss data). Our model proposes an explanation for these findings: the effect of increasing closeness by shifting one candidate closer to the other on overall turnout is small simply because it affects turnout for L and R in opposition directions. This stand in stark contrast with predictions of other voting models, which typically predict a positive impact of closeness on the turnout of supporters of both candidates, and hence on overall turnout.

3.5 Comparative Statics of Turnout Rate: Bandwagon and Underdog Effects

Previous theoretical research on costly voting have established the existence of an underdog effect, i.e., supporters of the runner-up turn out at a higher rate than supporters of the front-runner. For instance, using the pivotal voter model, Palfrey and Rosenthal (1985) assume that a particular citizens is a candidate L supporters with probability p and a R supporter with probability $1-p$. Voters have a cost drawn from some distribution F, and the Nash equilibrium is determined by cost cutoffs c_A and c_B for A and B supporters, respectively. If $p < 0.5$, then $c_A > c_B$, i.e., participation rates are higher among supporters of the runner-up.[18] For group-based voting models, Feddersen and Sandroni (2006) also predict an underdog effect.

This prediction has been tested in the laboratory and most studies (Duffy and Tavits 2008, Grober and Schram 2010, Kartal 2015, and Agranov et al. (2018)) find the opposite result: supporters of the front-runner turn out at a *higher rate* than supporters of the runner-up—the so-called *bandwagon effect*.[19] Agranov et al. (2018) argue that a modification of the pivotal voter model in which voters like to vote for the winner, as in Callander (2008), organizes their data well, but that neither other variants of the pivotal voter model (e.g., Grillo (2017)) nor the ethical voter model (Feddersen and Sandroni 2006) do. As we show below, our model proposes an alternative explanation for these findings in support of the bandwagon effect: instrumental voters are subject to a discouragement effect when they anticipate that they have fewer co-supporters in the electorate as the benefit of forming a voting coalition then decreases.[20]

We define turnout rates r_i for candidates $i = L, R$ as follows:

$$r_L = \frac{\Phi(S_L)}{\Phi((x_L + x_R)/2)} \text{ and } r_R = \frac{\Phi(S_R)}{1 - \Phi((x_L + x_R)/2)}. \tag{15}$$

The denominators are the number of citizens whose bliss point is closer to a particular candidate, i.e., the number of citizens who expressively prefer that candidate. This is the closest analogue in our model to the exogenous probability p that a particular voter is an L supporter.

To model changes in the number of potential L and R supporters we shift the distribution Φ of voters as in Result 4. From this result, we know that, for the case of purely instrumental preferences ($\beta = 0$), a shift of

[18] Different variants of the pivotal voter model producing bandwagon have been proposed. For instance, Callander (2007)'s variant produces a bandwagon effect because voters like to vote for the winner. Grillo (2017) obtains bandwagon effects by assuming that voters are risk averse, and Leontiou et al. (2023) by assuming loss aversion.

[19] The only exception we are aware of is Levine and Palfrey (2005), which find evidence of an underdog effect in a laboratory experiment. Note however that, as discussed in Agranov et al. (2018), the effect they observe is quite small.

[20] Testing the underdog effect using observational data is particularly challenging as it requires to measure the turnout *rate* of supporters of the front-runner and the runner-up. The only relevant analysis using observational data that we are aware of is in the Appendix of Bouton et al. (2024). Using a differences-in-differences approach relying on changes in competitiveness at the precinct level following the 2020 redistricting cycle, they find fairly precise null effects of partisan alignment (i.e., underdog or bandwagon effect) on turnout. Our model provides an explanation for these findings: the bandwagon effect stemming from instrumental incentives is compensated by the underdog effect that may stem from expressive incentives.

the voter distribution that is favorable to candidate R reduces the turnout for L, $\Phi(S_L)$, which has a negative effect on the turnout rate. However, the shift also reduces the number of potential L supporters, $\Phi((x_L + x_R)/2)$, which ceteris paribus would increase the turnout rate of L supporters. Proposition 6 below shows that, starting from a symmetric situation, the first effect always dominates: turnout for the disadvantaged candidate falls, while it rises for the advantaged candidate. The proposition also contrasts this outcome with the case of purely expressive voter preferences. From Result 4, we know that following a shift in the distribution Φ in favor of candidate R, the turnout for L, $\Phi(E_L)$, goes down. Since the shift also reduces the number of potential supporters for candidate L, the overall effect on the turnout rate r_L depends on the extent of mass polarization, as measured by the standard deviation σ of Φ. The same is true for the effect on the turnout rate r_R.

Proposition 6 *Suppose that G and Φ are symmetric around 0 and $x_R = -x_L$. Let $\Phi_\mu(x) = \Phi(x - \mu)$, and increase μ marginally thereby providing an advantage to candidate R. Then:*

1. *For $\beta = 0$, under Assumption 3 and $c < \bar{c}$, the turnout rate r_L falls and the turnout rate r_R rises.*
2. *For $\beta = 1$, if the standard deviation of Φ is sufficiently large, then the turnout rate r_L rises and the turnout rate r_R falls. The opposite holds if the standard deviation of Φ is sufficiently small, Φ has infinite support, and $\lim_{x\to-\infty} \phi(x)/\Phi(x) > 2\phi(0)$ holds. All normal distributions satisfy this condition.*

Proposition 6 implies that purely instrumental voters are subject to a discouragement effect in the sense that if the voter distribution shifts away from them, then their participation rate drops. This is because, when the voter distribution shifts away from candidate L, making L the (expeceted) runner up, then strategic voting for L becomes more difficult to support. The benefit of strategic voting is measured by $\Phi(S_L)$ and its impact on L's winning probability. If there are fewer voters in S_L because the distribution shifts away, then the benefit of strategic voting decreases. In contrast, the cost to an individual citizen of voting remains unchanged. Hence, the turnout for L must fall.

4 Endogenous Platforms

In this section, we show how the nature of strategic voters affects candidates' choices of platforms. We endogenize the platforms of the candidates in a classical Calvert-Wittman style model of electoral competition between two policy-motivated candidates, contrasting outcomes when voters hold purely instrumental ($\beta = 0$) and purely expressive ($\beta = 1$) preferences. We maintain Assumption 3 on the distribution of partisan voters, and focus on a symmetric setting.

Candidates have quadratic preferences, with the ideal points denoted by θ_i, $i = L, R$, where $\theta_L = -\theta_R$. Let $P(x_L, x_R)$ denote the probability that candidate L wins given platform choices x_L and x_R by candidates L and R, respectively. Then candidate R solves

$$\max_{x_R} -P(x_L, x_R)(x_L - \theta_R)^2 - (1 - P(x_L, x_R))(x_R - \theta_R)^2.$$

For purely instrumental preferences, we have shown that

$$P(x_L, x_R) = G(\Phi(\bar{y}_L) - (1 - \Phi(\underline{y}_R))),$$

where $\bar{y}_L$ and $\underline{y}_R$ solve $\xi_L(\bar{y}_L) = c/\alpha$ and $\xi_R(\underline{y}_R) = c/\alpha$, and ξ_L and ξ_R are given by (13) and (14). For purely expressive preferences, we have shown that

$$P(x_L, x_R) = G(\Phi(E_L) - \Phi(E_R)),$$

where E_i is given by (10).

We first characterize how the nature of voter preferences shapes the choice of platforms by the candidates.

Proposition 7 *There exists $\bar{c} > 0$ such that if voting costs $c < \bar{c}$ then candidates choose more extreme platforms when $\beta = 1$ than when $\beta = 0$.*

To understand the intuition for why candidates adopt more extreme platforms when voters have purely expressive preferences, note that the effect of a move of a candidate's platform closer to his bliss point can be decomposed into two parts. First, there is a (negative) impact on the probability of winning the election. Second, there is a (positive) impact on the candidate's utility of winning with a platform closer to his bliss point. The second effect is not affected by the behavior of voters, as it only depends on the utility of the candidates. The comparison of the first effect for the cases with purely instrumental and purely expressive preferences thus drives the result. The key is that, for the purely expressive case, there are both moderates and non-moderate citizens who are alienated. A move of the candidate's platform closer to her bliss point then alienates additional moderate citizens but also convince some non-moderate citizens to vote instead of abstain. These opposite effects limit the impact of the change in platform on the probability that the candidate wins, which frees the candidates to move closer to their bliss points.[21] By contrast, for the purely instrumental case, only moderate citizens are alienated, while all sufficiently non-moderate ones join the voting coalition. A move of the candidate's platform closer to his bliss point then alienates additional moderate citizens, but does not help convince any non-moderate ones to join the coalition. The impact on

[21] For the case of uniform Φ, the two effects exactly compensate and there is thus no impact of the plaftorm change on the probability of winning. Candidates then always adopt their ideal points as policies.

the probability of winning is thus larger, which limits the incentives of the candidates to move closer to their bliss point, reducing polarization.

One direct implication of Proposition 7 is that, if citizens' behavior is (at least partly) driven by instrumental incentives, then there will be partial but not full convergence of candidates' platforms. This is in line with empirical evidence, which shows that (i) candidates' platforms rarely fully converge (i.e., there is polarization) and that (ii) candidates are reactive to voters' preferences in the sense that their platforms move with the position of the expected median (Ansolabehere, Snyder, and Stewart 2001, Bonica 2014, Di Tella et al. 2023).

4.1 Comparative Statics of Polarization

We are now in position to explore how features of the electoral environment affect political polarization. In this section, we only present one of our comparative static results–about the impact of voting costs. This result is in stark contrast with the predictions of existing model and thus illustrates well how including strategic voters in the analysis of candidates' behavior can dramatically affect its conclusions. Other comparative static results are relegated to Appendix 5.

The following Proposition shows that a change in voting costs affects polarization differently whether expressive or instrumental incentives drive the behavior of voters.

Proposition 8 *There exists $\bar{c} > 0$ such that if voting costs $c < \bar{c}$ then an increase in voting costs decreases elite polarization ($x_R - x_L$) if $\beta = 0$ while the reverse is true if $\beta = 1$.*

To see the impacts of increasing voting costs, first note that with purely expressive voter preferences, increasing c reduces $z \equiv \sqrt{A - c}$, and hence the size of the expressive voting coalitions E_L and E_R. This, in turn, reduces $\frac{\partial P(x_L, x_R)}{\partial x_R}$, i.e., choosing more extreme platforms reduces the winning probability by less than before the increase in c. The reason is that, since there are fewer non-partisan voters, partisan voters matter more for the outcome of the election, and platform choice does not affect their participation. Indeed, when c is sufficiently large, only partisans vote and candidates are then unrestricted in their platform choice.

Now consider increasing voting costs in the purely instrumental case. To simplify the explanation, let us consider a case for which, before the change in c, $x_L \in S_L$ and $x_R \in S_R$. As we have seen above, increasing c reduces both S_L and S_R. Everything else equal, this implies that the marginal opponent is further away from, say, x_R, and the marginal supporter closer to x_R. Due to the quadratic loss assumption, this implies that $\bar{y}_L$ becomes more sensitive to a shift in x_R (i.e., $\frac{\partial \bar{y}_L}{\partial x_R}$ increases) and that $\underline{y}_R$ becomes less sensitive (i.e., $\frac{\partial \underline{y}_R}{\partial x_R}$ decreases). Importantly, the change is larger for the former than the latter. Hence, overall, we have that the change in the probability that R wins following an increase in x_R increases (i.e., $\frac{\partial \bar{y}_L}{\partial x_R} + \frac{\partial \underline{y}_R}{\partial x_R}$ increases). This

means that for a candidate, moving her platform outward, closer to her bliss point, becomes more costly in terms of probability of victory. This gives candidates incentives to adopt more moderate platforms, i.e., to polarize less.

Proposition 8 implies that, if voters are instrumentally motivated, the strategic reaction of candidates amplifies the negative effect of a change in c on turnout and turnout rates. Indeed, Result 2 (in Appendix 5) shows that both turnout and turnout rates are increasing in elite polarization, which the proposition shows is decreasing in c.

Proposition 8 also has interesting policy implications which contrast with common beliefs in the literature. In particular, Oprea et al. (2024) argue, using models of electoral competition without strategic voters, that a strong argument in favor of compulsory voting (a policy which is equivalent to a stark reduction in voting costs) is that it should help decrease elite polarization. By contrast, our model predicts that compulsory voting could actually *increase* polarization if voters have instrumental motives. We are not aware of any existing empirical evidence that could directly inform this theoretical debate. Such work could either explore the impact of the introduction of compulsory voting on elite polarization, or other changes in the electoral environment which influence the voting costs of (most) voters.

5 Conclusion

This paper addresses the longstanding challenge of modeling voter behavior by proposing a novel endogenous-group model that integrates expressive and instrumental motives within the canonical spatial model. By endogenizing the formation of voting coalitions, we offer a more realistic and empirically consistent framework for understanding voter turnout and decision-making. Our model not only overcomes the limitations of existing models but also provides a clear mechanism for how citizens coordinate their voting behavior in response to the characteristics of the electoral environment.

References

Adams, J. and S. Merrill (2003). Voter turnout and candidate strategies in American elections. *Journal of Politics 65*, 161–189.

Agranov, M., J. K. Goeree, J. Romero, and L. Yariv (2018). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association 16*(3), 825–856.

Ali, S. N. and C. Lin (2013, May). Why people vote: Ethical motives and social incentives. *American Economic Journal: Microeconomics 5*(2), 73–98.

Bernhardt, D., S. Krasa, and F. Squintani (2024). Political competition and strategic voting in multi-candidate elections. Working Paper, University of Warwick.

Bouton, L., M. Castanheira, and L.-S. Aniol (2015). Multicandidate elections: Aggregate uncertainty in the laboratory. *Games and Economic Behavior 101*, 132–150.

Bouton, L. and B. G. Ogden (2021). Group-based voting in multicandidate elections. *The Journal of Politics 83*(2), 468–482.

Brennan, G. and L. Lomasky (1993). *Democracy and Decision*. Cambridge University Press.

Callander, S. (2008). Majority rule when voters like to win. *Games and Economic Behavior 64*(2), 393–420. Special Issue in Honor of Michael B. Maschler.

Calvert, R. L. (1985). Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence. *American Journal of Political Science 29*, 69–95.

Castanheira, M. (2003). Victory margins and the paradox of voting. *European Journal of Political Economy 19*(4), 817–841.

Coate, S. and M. Conlin (2004). A group rule-utilitarian approach to voter turnout: Theory and evidence. *American Economic Review 94*(5), 1476–1504.

Coate, S., M. Conlin, and A. Moro (2008). The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda. *Journal of Public Economics 92*(3-4), 582–596.

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

Duffy, J. and M. Tavits (2008). Beliefs and voting decisions: a test of the pivotal voter model. *American Journal of Political Science 52*(3), 603–618.

Esponda, I. and E. Vespa (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Review 6*(4), 180–202.

Feddersen, T. and A. Sandroni (2006). A theory of participation in elections. *American Economic Review 96*(4), 1271–1282.

Forsythe, R., T. Rietz, R. Myerson, and R. Weber (1996). An experimental study of voting rules and polls in three-candidate elections. *International Journal of Game Theory 25*, 355–383.

Fujiwara, T. (2011). A regression discontinuity test of strategic voting and duverger's law. *Quarterly Journal of Political Science 6*, 197–233.

Good, I. J. and L. S. Mayer (1975). Estimating the efficacy of a vote. *Behavioral Science 20*(1), 25–33.

Grillo, A. (2017). Risk aversion and bandwagon effect in the pivotal voter model. *Public Choice 172*(3), 465–482.

Harsanyi, J. (1977). Morality and the theory of rational behavior. *Social Research 44*(4), 623–655.

Herrera, H. and C. Martinelli (2006). Group formation and voter participation. *Theoretical Economics 1*, 461–481.

Kaplan, E. and H. Yuan (2020, January). Early voting laws, voter turnout, and partisan vote composition: Evidence from ohio. *American Economic Journal: Applied Economics 12*(1), 32–60.

Ledyard, J. (1984). Pure theory of two-candidate competition. *Public Choice 44*, 7–41.

Levine, D. and A. Mattozzi (2020). Voter turnout with peer punishment. *American Economic Review 110*(10), 3298–3314.

Levine, D. and T. Palfrey (2005). The paradox of voter participation? A laboratory study. Princeton.

Merlo, A. and T. Palfrey (2018). External validation of voter turnout models by concealed parameter recovery. *Public Choice 176*, 297–314.

Morton, R. (1991). Groups in rational turnout models. *American Journal of Political Science 35*(3), 758–776.

Myerson, R. B. (1998). Population uncertainty and poisson games. *International Journal of Game Theory 27*, 375–392.

Oprea, A., L. Martin, and G. H. Brennan (2024). Moving toward the median: Compulsory voting and political polarization. *American Political Science Review*, 1–15.

Palfrey, T. R. and H. Rosenthal (1983). A strategic calculus of voting. *Public Choice 41*, 7–53.

Palfrey, T. R. and H. Rosenthal (1985). Voter participation and strategic uncertainty. *American Political Science Review 79*, 62–78.

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

Riker, W. H. and P. C. Ordeshook (1968). A theory of the calculus of voting. *The American political science review 62*(1), 25–42.

Shachar, R. and B. Nalebuff (1999). Follow the leader: Theory and evidence on political participation. *American Economic Review 89*(3), 525–547.

Spenkuch, J. L. (2018). Expressive vs. strategic voters: An empirical assessment. *Journal of Public Economics 165*, 73–81.

Thompson, D. M., J. A. Wu, J. Yoder, and A. B. Hall (2020). Universal vote-by-mail has no impact on partisan turnout or vote share. *Proceedings of the National Academy of Sciences 117*(25), 14052–14056.

Uhlaner, C. J. (1989). Rational turnout: The neglected role of groups. *American Journal of Political Science 33*(2), 390–422.

Wittman, D. (1983). Candidate motivation: A synthesis of alternative theories. *American Political Science Review 77*, 142–157.

Appendix A: Proofs

Proof of Lemma 1. Let $S_i \neq \emptyset$, and consider $\underline{y}, \bar{y} \in S_i$ with $\underline{y} < \bar{y}$. Consider any y between $\underline{y}$ and $\bar{y}$. Without loss of generality suppose that $|\bar{y} - x_i| \leq |\underline{y} - x_i|$. This implies that $|y - x_i| \leq |\underline{y} - x_i|$. No coercion implies that $\gamma_i(\underline{y}, S_i, S_{-i}) \geq 0$, and hence $\gamma_i(y, S_i, S_{-i}) \geq 0$. But then no free riding implies that $y \in S_i$. Thus, S_i is an interval.

We next show that S_i does not contain points that are arbitrarily close to $\bar{x} = (x_L + x_R)/2$. Suppose by way of contradiction that there exists a sequence y_n, $n \in \mathbb{N}$ with $\lim_{n\to\infty} y_n = \bar{x}$ and $y_n \in S_L$ for all n. No coercion implies $\gamma_i(y_n, S_L, S_R) \geq 0$ for all n. Thus, continuity of γ_i implies that $0 \leq \lim_{n\to\infty} \gamma_i(y_n, S_L, S_R) = \gamma_i(\bar{x}, S_L, S_R) = \beta(A - (x_L - \bar{x})^2) - c$. Thus, $A - (x_L - \bar{x})^2 \geq c$, which implies $\bar{x} - x_L \geq \sqrt{A - c}$. Substituting $\bar{x} = (x_L + X_R)/2$ and simplifying yields $2\sqrt{A - c} \geq x_R - x_L$ which contradicts Assumption 2. The argument for S_R is analogous.

To show closedness, let y_n be a sequence in S_L with $\lim_{n\to\infty} y_n = y$. No coercion implies $\gamma_L(y_n, S_L, S_R) \geq 0$. Hence, by continuity, $\gamma_L(y, S_L, S_R) \geq 0$. Further, no free riding implies that $y \in S_L$, as we have shown that $y < \bar{x}$ and hence $y \notin S_R$, which would be the only impediment for y to join coalition S_L. Closedness of S_R is established identically.

Next, let y be a boundary point of S_i. Then no coercion implies that $\gamma_i(y, S_L, S_R) \geq 0$. Suppose by way of contradiction that $\gamma_i(y, S_L, S_R) > 0$. Then continuity of γ_i in y implies that there exists a $y' < \bar{x}$, $y' \notin S_i$, with $\gamma_i(y', S_L, S_R) > 0$. This violates no free riding. ■

Proof of Lemma 2. $\gamma_i(y, [y, \bar{y}_i], S_{-i})$ is continuously differentiable in y. Suppose by way of contradiction that $\frac{\partial}{\partial y}\gamma_i(y, [y, \bar{y}_i], S_{-i}) < 0$. For example, consider coalition S_L. Then there exists a $\delta > 0$ such that extending S_L to $S_L \cup [\underline{y}_L - \delta, \underline{y}_L)$ makes all new members strictly better off, contradicting no free riding. The arguments for expanding S_R to the left and S_L and S_R to the right are analogous. ■

Proof of Proposition 1. Substituting $S_L = -S_R$ into (4) and using symmetry yields

$$\gamma_L(y, S_L) = \beta(A - (x_L - y)^2) - c + 4(1 - \beta)\left(G(\Phi(S_L \setminus E_L)) - \frac{1}{2}\right) x_L y. \tag{16}$$

Note that the last summand is non-negative, and strictly positive if $\Phi(S_L) > \Phi(E_L)$. Thus, increasing S_L increases $\gamma_L(y, S_L)$.

Consider the set of all intervals $\mathcal{I}$ in $\mathbb{R}_-$ that contain E_L. Endow $\mathcal{I}$ with the natural order. Next, define a mapping $T : \mathcal{I} \mapsto \mathcal{I}$ as follows. Let $[\underline{y}, \bar{y}] \in \mathcal{I}$ ($\underline{y}$ can be $-\infty$), and let $\underline{y}' < \underline{y}$ be a root of $\gamma_L(\underline{y}', [\underline{y}', \bar{y}]) = 0$

with $\gamma_L(y,[y,\bar{y}]) < 0$ for all $y \in (\underline{y}' - \delta, \underline{y}')$ for some $\delta > 0$. If $\gamma_L(\underline{y}',[\underline{y}',\bar{y}]) \geq 0$ for all $\underline{y}'$ then we define $\underline{y}' = -\infty$. Similarly, let $\bar{y}' > \bar{y}$ be a root of $\gamma_L(\bar{y}',[\underline{y},\bar{y}']) = 0$ in $\mathbb{R}_-$ with $\gamma_L(y,[\underline{y},y]) < 0$ for all $y \in (\bar{y}',0)$. Then define $T([\underline{y},\bar{y}]) = [\underline{y}',\bar{y}']$.

Next, note that T is monotone, i.e., if $S \subset S'$ then $T(S) \subset T(S')$. This follows because γ_L increases as S is increased. Because $\mathcal{I}$ is a complete lattice, Tarski's fixed point theorem implies that there exists a maximal fixed point S^*. Because $T(S^*) = S^*$, no coercion is satisfied. Because S^* is maximal, the equilibrium satisfies no free riding. ■

Proof of Proposition 2. Consider (8). It is immediate that this equation is independent of S_R. It remains to prove that a solution S_L exists or that $S_L = \emptyset$, and that S_L satisfies no free riding. This argument, follows as in the proof of Proposition 1 by applying Tarski's fixed point theorem. ■

Proof of Proposition 3. If voters have purely expressive preferences, $\beta = 1$, then, regardless of the form of G and Φ, the net benefit of a citizen $y < (x_L + x_R)/2$ from having coalition S_L form is:

$$\gamma_L(y, S_L, S_R) = A - (x_L - y)^2 - c,$$

and the net benefit of a citizen $y > (x_L + x_R)/2$ from having coalition S_R form is:

$$\gamma_R(y, S_L, S_R) = A - (x_R - y)^2 - c.$$

Solving $A - (x_L - y)^2 - c = 0$, yields $\underline{y}_i = x_i - \sqrt{A - c}$ and $\bar{y}_i = x_i + \sqrt{A - c}$ for $i = L, R$. Assumption 2 ensures that $\bar{y}_L < (x_L + x_R)/2 < \underline{y}_R$. It is immediate that the strategic coalitions S_L and S_R coincide with the expressive coalitions. Thus, $S_i = E_i = \left[x_i - \sqrt{A - c}, x_i + \sqrt{A - c}\right]$. ■

Proof of Proposition 4. Suppose that $S_L \neq \emptyset$. Then no coercion implies

$$\alpha \Phi(S_L)(x_R - x_L)(x_L + x_R - 2y) \geq c, \tag{17}$$

for all $y \in S_L$. The left-hand side of (17) is decreasing in y. Thus, equality can only hold at a single y, and the inequality is strict for all $y' < y$. Hence, S_L takes the form $(-\infty, \bar{y}_L^*]$. Further, no free riding implies that (17) must hold with equality at $y = \bar{y}_L^*$.

Consider ξ_L in (13). By construction (11) holds if and only if $\xi_L(\bar{y}_L) = c/\alpha$. Note that $\xi_L(\bar{y}_L) = 0$ for $\bar{y}_L = \frac{x_L + x_R}{2}$. Further, $\xi_L'(\bar{y}_L) = \phi(\bar{y}_L)(x_L + x_R - 2\bar{y}_L) - 2\Phi(\bar{y}_L)$. Thus, $\xi_L'(\bar{y}_L) < 0$ if and only if

$$\frac{x_L + x_R}{2} < \frac{1}{\phi(\bar{y}_L)/\Phi(\bar{y}_L)} + \bar{y}_L. \tag{18}$$

(18) is clearly satisfied for $\bar{y}_L = \frac{x_L+x_R}{2}$. The right-hand side of (18) is increasing in $\bar{y}_L$ since $\phi(\bar{y}_L)/\Phi(\bar{y}_L)$ is decreasing by the log-concavity of Φ. Thus, (18) is violated if $\bar{y}_L$ becomes sufficiently negative. Further, log concavity implies that ξ_L' changes its sign once from positive to negative as $\bar{y}_L$ is increased towards $\frac{x_L+x_R}{2}$.

Let $\bar{c}_L = \max_{\bar{y}_L \in (-\infty, \frac{x_L+x_R}{2}]} \alpha \xi_L(\bar{y}_L)$. For $c < \bar{c}_L$, there are two solutions, $\bar{y}_{1L} < \bar{y}_{2L}$, to the condition $\xi_L(\bar{y}_L^*) = c/\alpha$ on $\bar{y}_L \in (-\infty, \frac{x_L+x_R}{2})$. Solution $\bar{y}_{1L}$ is not admissible, because $\xi_L'(\bar{y}_{1L}) > 0$ and hence $\frac{\partial}{\partial y}\gamma_L(y, [-\infty, y], S_R)|_{y=\bar{y}_{1L}} > 0$, violating Lemma 2. The unique equilibrium coalition is $S_L = (-\infty, \bar{y}_{2L}]$, because $\xi_L'(\bar{y}_{2L}) < 0$ and hence $\frac{\partial}{\partial y}\gamma_L(y, [-\infty, y], S_R)|_{y=\bar{y}_{2L}} < 0$. When $c = \bar{c}_L$, there is a unique $\bar{y}_L^* \in (-\infty, \frac{x_L+x_R}{2})$ such that $\xi_L(\bar{y}_L^*) = c/\alpha$, and $\frac{\partial}{\partial y}\gamma_L(y, [-\infty, y], S_R)|_{y=\bar{y}_L^*} = \xi_L'(\bar{y}_L^*) = 0$. The unique equilibrium coalition S_L is $(-\infty, \bar{y}_L^*]$. For $c > \bar{c}_L$, there does not exist any $\bar{y}_L^* \in (-\infty, \frac{x_L+x_R}{2})$ such that $\xi_L(\bar{y}_L^*) = c/\alpha$, and hence no equilibrium strategic coalition: $S_L = \emptyset$.

The argument for S_R is analogous. ■

Proof of Proposition 5. Consider S_L. Rewrite (8) as:

$$\beta\left(A - (x_L - y)^2\right) - c + (1-\beta)\alpha\Phi(S_L \setminus E_L) 2(x_R - x_L)\left(\frac{x_L + x_R}{2} - y\right) = 0.$$

Note that $\underline{y}_L \leq x_L - \tau < x_L + \tau \leq \bar{y}_L$. Hence, $\frac{x_L+x_R}{2} - \underline{y}_L > \frac{x_L+x_R}{2} - \bar{y}_L$, so the final term must be higher for $y = \underline{y}_L$ than for $y = \bar{y}_L$. Hence, the first term must be lower for $y = \underline{y}_L$ than for $y = \bar{y}_L$, that is to say $\left(x_L - \underline{y}_L\right)^2 > (x_L - \bar{y}_L)^2$, i.e., $|\underline{y}_L - x_L| > |\bar{y}_L - x_L|$. Analogous arguments for S_R yield that $\underline{y}_R \leq x_R - \tau < x_R + \tau \leq \bar{y}_R$, together with (9), imply that $|x_R - \bar{y}_R| > |x_R - \underline{y}_R|$. ■

Lemma 3 *Suppose that $x\phi'(x)/\phi(x)$ is decreasing in x, which holds for normal distributions. Let $\Phi_s(x) = \Phi(x/s)$. Then there exists $\bar{s} \geq 0$ such that turnout decreases for $s > \bar{s}$, and increases for $s < \bar{s}$ as mass polarization, s, is increased when $\beta = 1$. Further, $\bar{s} > 0$ if Φ is a normal distribution.*

Proof of Lemma 3. Let $S_R = [\underline{y}, \bar{y}]$. The measure of voters for candidate R is $\Phi(\bar{y}/s) - \Phi(\underline{y}/s)$. The derivative with respect to s is

$$\frac{\underline{y}}{s^2}\phi\left(\frac{\underline{y}}{s}\right) - \frac{\bar{y}}{s^2}\phi\left(\frac{\bar{y}}{s}\right) = \frac{\phi\left(\frac{\underline{y}}{s}\right)}{s^2}\left(\underline{y} - \bar{y}\frac{\phi\left(\frac{\bar{y}}{s}\right)}{\phi\left(\frac{\underline{y}}{s}\right)}\right). \tag{19}$$

Note that

$$\lim_{s\to\infty} \underline{y} - \bar{y}\frac{\phi\left(\frac{\bar{y}}{s}\right)}{\phi\left(\frac{\underline{y}}{s}\right)} = \underline{y} - \bar{y} < 0. \tag{20}$$

Further

$$\frac{\partial}{\partial s}\left(\underline{y}-\bar{y}\frac{\phi\left(\frac{\bar{y}}{s}\right)}{\phi\left(\frac{\underline{y}}{s}\right)}\right)=\frac{\bar{y}}{s^2}\frac{\bar{y}\phi'\left(\frac{\bar{y}}{s}\right)\phi\left(\frac{\underline{y}}{s}\right)-\underline{y}\phi'\left(\frac{\underline{y}}{s}\right)\phi\left(\frac{\bar{y}}{s}\right)}{\phi^2\left(\frac{\underline{y}}{s}\right)}=\frac{\bar{y}\phi\left(\frac{\bar{y}}{s}\right)\phi\left(\frac{\underline{y}}{s}\right)}{s\phi^2\left(\frac{\underline{y}}{s}\right)}\left(\frac{\bar{y}}{s}\frac{\phi'\left(\frac{\bar{y}}{s}\right)}{\phi\left(\frac{\underline{y}}{s}\right)}-\frac{\underline{y}}{s}\frac{\phi'\left(\frac{\underline{y}}{s}\right)}{\phi\left(\frac{\bar{y}}{s}\right)}\right). \tag{21}$$

Because $x\phi'(x)/\phi(x)$ is decreasing in x and $\bar{y}>\underline{y}$, it follows that (21) is negative. This, in turn, implies that the sign of (19) can change at most once, i.e., if the sign of (19) is negative for small s then it is always negative. Otherwise, it becomes zero for some $\bar{s}$ and negative for $s>\bar{s}$.

Finally, note that for normal distributions

$$\lim_{s\to 0}\underline{y}-\bar{y}\frac{\phi\left(\frac{\bar{y}}{s}\right)}{\phi\left(\frac{\underline{y}}{s}\right)}=\underline{y}>0. \tag{22}$$

Thus, there exists a unique value $\bar{s}>0$ for which (19) becomes zero. ■

The following lemma is used for the proof of Proposition 6.

Lemma 4 *Let Assumption 3 hold with $\beta=0$ and $c<\bar{c}_i$. Suppose that Φ is symmetric with mean 0 and standard deviation 1. Let $x_L=-x_R$. For $\sigma>0$, define $\Phi_\sigma(x)=\Phi(x/\sigma)$, so that Φ_σ has standard deviation σ. Then S_L and S_R are non-empty if and only if $c\leq 4x_R\bar{m}\sigma\alpha$.*

Proof of Lemma 4. By symmetry it is sufficient to show the result for $S_L=(-\infty,\bar{y}_L]$. For $S_L\neq\emptyset$, we must have $\xi_L(\bar{y}_L)=c/\alpha$. Note that $\xi_L(\bar{y}_L)=-4x_R\bar{y}_L\Phi_\sigma(\bar{y}_L)$. Substituting $z=\bar{y}_L/\sigma$ yields $-4x_R\sigma z\Phi(z)=c/\alpha$. Note that $\lim_{z\to-\infty}z\Phi(z)=0$ because Φ has a bounded first moment. Thus, if z is sufficiently negative then $\xi_L(z)=-4x_R\sigma z\Phi(z)<c/\alpha$. Next note that symmetry implies that there exists a $\tilde{z}<0$ such that $-\tilde{z}\Phi(\tilde{z})=\bar{m}$. Thus, $\xi_L(\tilde{z})=4x_R\sigma\bar{m}\geq c/\alpha$ by assumption. The intermediate value theorem yields the result. ■

Proof of Proposition 6. Consider first the case where $\beta=0$. Differentiating the equilibrium condition $\xi_L(\bar{y}_L(\mu))=c/\alpha$ with respect to μ at $\mu=0$ with symmetrically-situated candidates, $x_R=-x_L$, yields

$$\Phi\left(\bar{y}_L(\mu)\right)\frac{\partial\bar{y}_L(\mu)}{\partial\mu}\Big|_{\mu=0}+\phi\left(\bar{y}_L(\mu)\right)\left(\frac{\partial\bar{y}_L(\mu)}{\partial\mu}\Big|_{\mu=0}-1\right)\bar{y}_L(\mu)=0, \tag{23}$$

which we solve for

$$\frac{\partial\bar{y}_L(\mu)}{\partial\mu}\Big|_{\mu=0}=1-\frac{\Phi\left(\bar{y}_L\right)}{\Phi\left(\bar{y}_L\right)+\bar{y}_L\phi\left(\bar{y}_L\right)}. \tag{24}$$

Corollary 4 and Lemma 4 imply that y_L is strictly decreasing in μ if $c<rx_R\bar{m}\sigma\alpha$. Thus,

$$\Phi\left(\bar{y}_L\right)+y_L\phi\left(\bar{y}_L\right)>0. \tag{25}$$

Differentiating $r_L = \Phi\left(\frac{\bar{y}_L - \mu}{\sigma}\right)/\Phi(-\mu)$ with respect to μ at $\mu = 0$ and denoting $\partial \bar{y}_L(\mu)/\partial \mu_{|_{\mu=0}}$ by $\bar{y}'_L$ yields

$$\frac{\partial r_L}{\partial \mu}\Big|_{\mu=0} = \frac{\phi(\bar{y}_L)\left(y'_L - 1\right)\Phi(0) + \phi(0)\Phi(\bar{y}_L)}{\Phi(0)^2}.$$

We also know that

$$\frac{\partial y_L}{\partial \mu}\Big|_{\mu=0} = 1 - \frac{\Phi(\bar{y}_L)}{\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L)}$$

Plugging that into $\frac{\partial r_L}{\partial \mu}|_{\mu=0}$, we obtain

$$\frac{\partial r_L}{\partial \mu}|_{\mu=0} = \frac{\phi(\bar{y}_L)\left(-\frac{\Phi(\bar{y}_L)}{\Phi(\bar{y}_L)+y_L\phi(\bar{y}_L)}\right)\Phi(0) + \phi(0)\Phi(\bar{y}_L)}{\Phi(0)^2}$$

which has the same sign as

$$-\phi(\bar{y}_L)\left(\frac{\Phi(y_L)}{\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L)}\right)\Phi(0) + \phi(0)\Phi(\bar{y}_L) = \Phi(\bar{y}_L)\left(\phi(0) - \frac{1}{2}\left(\frac{\phi(\bar{y}_L)}{\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L)}\right)\right). \tag{26}$$

Note that $\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L) > 0$, else $\frac{\partial y_L}{\partial \mu}|_{\mu=0} > 0$, which cannot hold in equilibrium. Therefore, the sign of (26) does not change when multiplied by $\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L)$, which yields

$$\Phi(\bar{y}_L)\left(\phi(0)\left(\Phi(\bar{y}_L) + y_L\phi(\bar{y}_L)\right) - \frac{1}{2}\phi(\bar{y}_L)\right). \tag{27}$$

The sign of (27) is the same as

$$\phi(0)\Phi(\bar{y}_L) + \phi(\bar{y}_L)\left(y_L\phi(0) - \frac{1}{2}\right) = \Phi(\bar{y}_L)\left(\phi(0) + \frac{\phi(\bar{y}_L)}{\Phi(\bar{y}_L)}\left(y_L\phi(0) - \frac{1}{2}\right)\right).$$

Because $y_L < 0$, the log concavity of Φ yields $\phi(\bar{y}_L)/\Phi(\bar{y}_L) \geq \phi(0)/\Phi(0)$. Thus, given $y_L\phi(0) - \frac{1}{2} < 0$ (as $y_L < 0$), we have

$$\begin{aligned}
\Phi(\bar{y}_L)\left(\phi(0) + \frac{\phi(\bar{y}_L)}{\Phi(\bar{y}_L)}\left(y_L\phi(0) - \frac{1}{2}\right)\right) &\leq \Phi(\bar{y}_L)\left(\phi(0) + \frac{\phi(0)}{\Phi(0)}\left(y_L\phi(0) - \frac{1}{2}\right)\right) \\
&= \Phi(\bar{y}_L)\left(\phi(0) + 2\phi(0)\left(y_L\phi(0) - \frac{1}{2}\right)\right) \\
&= \Phi(\bar{y}_L)\, 2y_L\phi(0)^2 < 0.
\end{aligned}$$

Thus, $\frac{\partial r_L}{\partial \mu}|_{\mu=0} < 0$, i.e., the participation rate, r_L, of L supporters decreases as μ is increased. Conversely, decreasing μ increases the turnout rate for candidate L. The result for candidate R follows by symmetry.

Finally, suppose that $\beta = 1$. It is immediate that $\Phi_{\mu,\sigma}(E_L)$ decreases as μ is increased. Let $E_L = [\underline{y}, \bar{y}]$. The turnout rate is

$$r_L = \frac{\Phi\left(\frac{\bar{y}-\mu}{\sigma}\right) - \Phi\left(\frac{\underline{y}-\mu}{\sigma}\right)}{\Phi\left(\frac{-\mu}{\sigma}\right)}. \tag{28}$$

Then

$$\frac{\partial r_L}{\partial \mu}\Big|_{\mu=0} = \frac{4\phi(0)\left(\Phi\left(\frac{\bar{y}}{\sigma}\right) - \Phi\left(\frac{\underline{y}}{\sigma}\right)\right) - 2\left(\phi\left(\frac{\bar{y}}{\sigma}\right) - \phi\left(\frac{\underline{y}}{\sigma}\right)\right)}{\sigma} = \frac{4\phi(0) - 2\frac{\phi\left(\frac{\bar{y}}{\sigma}\right) - \phi\left(\frac{\underline{y}}{\sigma}\right)}{\Phi\left(\frac{\bar{y}}{\sigma}\right) - \Phi\left(\frac{\underline{y}}{\sigma}\right)}}{\sigma\left(\Phi\left(\frac{\bar{y}}{\sigma}\right) - \Phi\left(\frac{\underline{y}}{\sigma}\right)\right)}. \tag{29}$$

L'Hôpital's rule implies that

$$\lim_{\sigma\to\infty} \frac{\phi\left(\frac{\bar{y}}{\sigma}\right) - \phi\left(\frac{\underline{y}}{\sigma}\right)}{\Phi\left(\frac{\bar{y}}{\sigma}\right) - \Phi\left(\frac{\underline{y}}{\sigma}\right)} = \lim_{\sigma\to\infty} \frac{-\frac{\bar{y}}{\sigma^2}\phi'\left(\frac{\bar{y}}{\sigma}\right) + \frac{\underline{y}}{\sigma^2}\phi'\left(\frac{\underline{y}}{\sigma}\right)}{-\frac{\bar{y}}{\sigma^2}\phi\left(\frac{\bar{y}}{\sigma}\right) + \frac{\underline{y}}{\sigma^2}\phi\left(\frac{\underline{y}}{\sigma}\right)} = \lim_{\sigma\to\infty} \frac{-\bar{y}\phi'\left(\frac{\bar{y}}{\sigma}\right) + \underline{y}\phi'\left(\frac{\underline{y}}{\sigma}\right)}{-\bar{y}\phi\left(\frac{\bar{y}}{\sigma}\right) + \underline{y}\phi\left(\frac{\underline{y}}{\sigma}\right)} = 0, \tag{30}$$

because Φ is single peaked and symmetric so $\phi'(0) = 0$, and the denominator converges to $\phi(0)(\underline{y} - \bar{y}) \neq 0$. Thus, the numerator of (29) becomes strictly positive.

Now consider the numerator of (29) as $\sigma \to 0$. Because $\underline{y} < \bar{y}$ it follows that

$$\lim_{\sigma\to 0} \frac{\phi\left(\frac{\bar{y}}{\sigma}\right) - \phi\left(\frac{\underline{y}}{\sigma}\right)}{\Phi\left(\frac{\bar{y}}{\sigma}\right) - \Phi\left(\frac{\underline{y}}{\sigma}\right)} = \lim_{\sigma\to\infty} \frac{\phi\left(\frac{\bar{y}}{\sigma}\right)}{\Phi\left(\frac{\bar{y}}{\sigma}\right)} \tag{31}$$

as $\phi(\underline{y}/\sigma)$ and $\Phi(\underline{y}/\sigma)$ are of smaller order than $\phi(\bar{y}/\sigma)$ and $\Phi(\bar{y}/\sigma)$, respectively. The assumption in the statement therefore implies that (29) is strictly negative. It is easy to verify that this condition holds for a normal distribution, as $\lim_{x\to-\infty} \phi(x)/\Phi(x) = \infty$. ■

Proof of Proposition 7. For $c = 0$ and $\beta = 1$ Proposition 3 implies that $E_i = \left[x_i - \sqrt{A}, x_i + \sqrt{A}\right]$ if $x_R \geq \sqrt{A}$. Candidate L's winning probability is therefore $G(\Phi(E_L) - \Phi(E_R))$ The first-order condition for $\beta = 1$ and $c = 0$ for candidate R and evaluating at $x_L = -x_R$ yields

$$-4g(0)\left(\phi\left(x_R - \sqrt{A}\right) - \phi\left(x_R + \sqrt{A}\right)\right)x_R\theta_R + \theta_R - x_R = 0,$$

which implies

$$x_R = \frac{\theta_R}{1 + 4g(0)\left(\phi\left(x_R - \sqrt{A}\right) - \phi\left(x_R + \sqrt{A}\right)\right)\theta_R}. \tag{32}$$

If $x_R < \sqrt{A}$ then $E_R = [0, x_R + \sqrt{A}]$ and the first-order condition becomes

$$4g(0)\phi\left(x_R + \sqrt{A}\right)x_R\theta_R + \theta_R - x_R > 0,$$

which means that it is optimal for candidate R to move past $\sqrt{A}$, which implies that the equilibrium x_R is given by (32)

In contrast, if $\beta = 0$ then (13) and (14) yield $S_L = (-\infty, \bar{x}]$ and $S_R = [\bar{x}, \infty)$, where $\bar{x} = (x_L + x_R)/2$. Candidate L's winning probabilities are therefore $G(\Phi(E_L) - \Phi(E_R))$ and $G(2\Phi(\bar{x}) - 1)$, respectively. The first-order condition for candidate R taken at $x_R = -x_L$ is therefore

$$-4g(0)\phi(0)x_R\theta_R + \theta_R - x_R = 0,$$

which implies

$$x_R = \frac{\theta_R}{1 + 4g(0)\phi(0)\theta_R}. \tag{33}$$

Because Φ is single peaked, $\phi(0) > \phi\left(x_R - \sqrt{A}\right) - \phi\left(x_R + \sqrt{A}\right)$ and hence (33) is strictly smaller than (32). Given that $x_L = -x_R$ in both cases, this implies that elite polarization is lower with purely instrumental preferences when $c = 0$. The monotonicity result in Proposition 8 concludes the proof that elite polarization is strictly smaller for instrumental preferences when $c < \bar{c}$. ■

Proof of Proposition 8. *Purely instrumental case: $\beta = 0$.* The coalitions are given by $S_L = (-\infty, \bar{y}_L]$ and $S_R = [\underline{y}_R, +\infty)$, where $\bar{y}_L$ and $\underline{y}_R$ satisfiy (11) and (12). Hence, $\underline{y}_R$ solves:

$$\alpha(1 - \Phi(\underline{y}_R))\,(x_R - x_L)\left(2\underline{y}_R - x_L - x_R\right) - c = 0,$$

or, using $x_L + x_R = 0$,

$$4\alpha x_R \underline{y}_R (1 - \Phi(\underline{y}_R)) - c = 0. \tag{34}$$

Here,

$$\left.\frac{\partial \bar{y}_L}{\partial x_R}\right|_{-x_L = x_R} = -\left.\frac{\partial \gamma_L(\bar{y}_L, S_L, S_R)/\partial x_R}{\partial \gamma_L(\bar{y}_L, S_L, S_R)/\partial \bar{y}_L}\right|_{-x_L = x_R} = \frac{1}{2x_R}\frac{(1 - \Phi(\underline{y}_R))(x_R + \underline{y}_R)}{1 - \Phi(\underline{y}_R) - \underline{y}_R\phi\left(\underline{y}_R\right)}$$

$$\left.\frac{\partial \underline{y}_R}{\partial x_R}\right|_{-x_L = x_R} = -\left.\frac{\partial \gamma_R(\underline{y}_R, S_L, S_R)/\partial x_R}{\partial \gamma_R(\underline{y}_R, S_L, S_R)/\partial \underline{y}_R}\right|_{-x_L = x_R} == \frac{1}{2x_R}\frac{(1 - \Phi(\underline{y}_R))(x_R - \underline{y}_R)}{1 - \Phi(\underline{y}_R) - \underline{y}_R\phi\left(\underline{y}_R\right)}$$

By Lemma 2, the admissible solutions $\bar{y}_L$ and $\underline{y}_R$ must be such that $\frac{\partial}{\partial y}\gamma_L(y, (-\infty, y], S_R)|_{y=\bar{y}_L} \leq 0$ and $\frac{\partial}{\partial y}\gamma_R(y, [y, +\infty), S_L)|_{y=\underline{y}_R} \geq 0$. Hence, we have $x_R\left[\underline{y}_R\phi\left(\underline{y}_R\right) - (1 - \Phi(\underline{y}_R))\right] \leq 0$, i.e., $\underline{y}_R\phi\left(\underline{y}_R\right) \leq (1 - \Phi(\underline{y}_R))$. Further,

$$\left.\frac{\partial \bar{y}_L}{\partial x_R}\right|_{-x_L = x_R} \geq \left.\frac{\partial \underline{y}_R}{\partial x_R}\right|_{-x_L = x_R} \geq 0,$$

and

$$\left.\frac{\partial \bar{y}_L}{\partial x_R}\right|_{-x_L = x_R} + \left.\frac{\partial \underline{y}_R}{\partial x_R}\right|_{-x_L = x_R} = \frac{1 - \Phi(\underline{y}_R)}{1 - \Phi(\underline{y}_R) - \underline{y}_R\phi\left(\underline{y}_R\right)}$$

Candidate R solves

$$\max_{x_R}\left\{EU_R(x_R, x_L) = -\left[\frac{1}{2} + \alpha\left[\Phi(\bar{y}_L) - (1 - \Phi(\underline{y}_R))\right]\right](x_L - \theta_R)^2 - \left[\frac{1}{2} - \alpha\left[\Phi(\bar{y}_L) - (1 - \Phi(\underline{y}_R))\right]\right](x_R - \theta_R)^2\right\}$$

$$\begin{aligned}\frac{\partial}{\partial x_R}EU_R(x_R, x_L) \;=\; & -(x_R - \theta_R) + 2\alpha\left[\Phi(\bar{y}_L) - (1 - \Phi(\underline{y}_R))\right](x_R - \theta_R) \\ & -\alpha\left[\phi(\bar{y}_L)\frac{\partial \bar{y}_L}{\partial x_R} + \phi(\underline{y}_R)\frac{\partial \underline{y}_R}{\partial x_R}\right]\left[(x_L - \theta_R)^2 - (x_R - \theta_R)^2\right]\end{aligned}$$

The FOC is:

$$\begin{aligned}\theta_R - x_R &= 4x_R\theta_R\alpha\phi(\underline{y}_R)\left(\frac{\partial\bar{y}_L}{\partial x_R} + \frac{\partial\underline{y}_R}{\partial x_R}\right) \\ &= 4\alpha x_R\theta_R\phi(\underline{y}_R)\frac{1-\Phi(\underline{y}_R)}{1-\Phi(\underline{y}_R) - \underline{y}_R\phi\left(\underline{y}_R\right)}\end{aligned} \tag{35}$$

or

$$(\theta_R - x_R)\left(1 - \underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}\right) = 4\alpha x_R\theta_R\phi(\underline{y}_R) \tag{36}$$

which implies that $1 - \underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)} > 0$.

Again, the left-hand side is the marginal benefit for diverging from zero close to θ_R, whereas the right-hand side is the marginal cost, in terms of probability of losing the election $\alpha\phi(\underline{y}_R)\left(\frac{\partial\bar{y}_L}{\partial x_R} + \frac{\partial\underline{y}_R}{\partial x_R}\right)$ times the cost if losing, $(x_L - \theta_R)^2 - (x_R - \theta_R)^2 = 4x_R\theta_R$.

We want the implicit function theorem to imply that the symmetric equilibrium platform x_R increases in c and θ_R for all c sufficiently small. Solve for x_R in equation (34),

$$x_R = \frac{c}{4\alpha\underline{y}_R(1-\Phi(\underline{y}_R))},$$

and substitute into equation (35) to obtain:

$$\gamma(\underline{y}_R, c) = \left(\theta_R - \frac{c}{4\alpha\underline{y}_R(1-\Phi(\underline{y}_R))}\right)\left(1 - \underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}\right) - c\frac{\theta_R\phi(\underline{y}_R)}{\underline{y}_R(1-\Phi(\underline{y}_R))} = 0. \tag{37}$$

For small c,

$$\frac{d\underline{y}_R}{dc} = -\frac{\partial\gamma(\underline{y}_R, c)/\partial c}{\partial\gamma(\underline{y}_R, c)/\partial\underline{y}_R} = -\frac{-\frac{1-\underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}}{4\alpha\underline{y}_R(1-\Phi(\underline{y}_R))} - \frac{\theta_R\phi(\underline{y}_R)}{\underline{y}_R(1-\Phi(\underline{y}_R))}}{\theta_R\frac{d}{d\underline{y}_R}\left(1 - \underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}\right)} = -\frac{-\frac{1-\underline{y}_R\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}}{4\alpha\underline{y}_R(1-\Phi(\underline{y}_R))} - \frac{\theta_R\phi(\underline{y}_R)}{\underline{y}_R(1-\Phi(\underline{y}_R))}}{-\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)} - \underline{y}_R\frac{d}{d\underline{y}_R}\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)}} < 0,$$

from our increasing hazard rate condition, $\frac{d}{d\underline{y}_R}\frac{\phi\left(\underline{y}_R\right)}{1-\Phi(\underline{y}_R)} > 0$.

Totally differentiating (34) yields

$$4\alpha\underline{y}_R(1-\Phi(\underline{y}_R))\frac{dx_R}{dc} + 4\alpha x_R\left(1-\Phi(\underline{y}_R)\right)\left(1 - y_R\frac{\phi\left(y_R\right)}{1-\Phi(\underline{y}_R)}\right)\frac{dy_R}{dc} = 1,$$

and because $4\alpha\underline{y}_R(1-\Phi(\underline{y}_R)) > 0$, $4\alpha x_R\left(1-\Phi(\underline{y}_R)\right)\left(1 - y_R\frac{\phi(y_R)}{1-\Phi(\underline{y}_R)}\right) > 0$ and $\frac{dy_R}{dc} < 0$, it follows that $\frac{dx_R}{dc} > 0$. Thus, the platform x_R increases in c.

Purely expressive case: $\beta = 1$. With $x_L = -x_R$, $\frac{x_L+x_R}{2} = 0$. For $\beta = 1$, the sets of voters who participate takes

the form

$$E_L = [x_L - z, \min\{x_L + z, 0\}], \text{ and } E_R = [\max\{x_R - z, 0\}, x_R + z], \tag{38}$$

where $z \equiv \sqrt{A - c}$. Suppose first that $z \geq x_R$. Then candidate L wins if $\Phi(0) - \Phi(x_L - z) \geq \Phi(x_R + z) - \Phi(0) + \rho$, where ρ is the realized net partisan advantage for candidate R. Thus, candidate L's winning probability is

$$P(x_L, x_R) = G(2\Phi(0) - \Phi(x_L - z) - \Phi(x_R + z)). \tag{39}$$

Candidate R solves

$$\max_{x_R} -P(x_L, x_R)(x_L - \theta_R)^2 - (1 - P(x_L, x_R))(x_R - \theta_R)^2 \tag{40}$$

Let $EU_R(x_R, x_L)$ be the value of the objective.

Taking the first order condition at $x_L = -x_R$ yields

$$\begin{aligned}
\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L = x_R} &= -\frac{\partial}{\partial x_R} P(x_L, x_R)[(x_L - \theta_R)^2 - (x_R - \theta_R)^2] - 2(1 - P(x_L, x_R))(x_R - \theta_R)\Big|_{-x_L = x_R} \\
&= -g(2\Phi(0) - \Phi(x_L - z) - \Phi(x_R + z))|_{-x_L = x_R} \left[2\phi(0) - \phi(x_R + z)\right] 4\theta_R x_R - (x_R - \theta_R) \\
&= -4g(0)\left[2\phi(0) - \phi(x_R + z)\right]\theta_R x_R - (x_R - \theta_R) = 0,
\end{aligned} \tag{41}$$

using $\Phi(0) - \Phi(x_L - z) = \Phi(x_R + z) - \Phi(0)$ implied by the symmetry of Φ at 0.

In words, the marginal benefit $\theta_R - x_R$ for raising x_R closer to θ_R must be met by the marginal cost, which is given by the marginal increase $g(0)\left[\phi(0) - \phi(x_R + z)\right]$ in the risk of losing the election times the consequent utility loss $(x_L - \theta_R)^2 - (x_R - \theta_R)^2 = 4\theta_R x_R$.

Differentiating the left-hand side of (41) with respect to x_R yields:

$$\frac{\partial}{\partial x_R}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L = x_R}\right) = -4g(0)\left[2\phi(0) - \phi(x_R + z)\right] x_R - 4g(0)\left[-\phi'(x_R + z)\right]\theta_R x_R - 1 < 0,$$

where the inequality follows from the single-peakedness of Φ, which implies that $\phi'(x_R + z) < 0$.

Differentiation of equation (41) with respect to θ_R and z and the implicit function theorem imply that x_R decreases in z, whereas x_R increases in θ_R if $g(0)$ and $\phi(0)$ are sufficiently small, because

$$\frac{\partial}{\partial z}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L = x_R}\right) = -4g(0)\left[-\phi'(x_R + z)\right]\theta_R x_R < 0,$$

$$\frac{\partial}{\partial \theta_R}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L = x_R}\right) = 1 - 4g(0)\left[2\phi(0) - \phi(x_R + z)\right] x_R.$$

Now consider the case where $x_L + z < 0 [= 0] < x_R - c$. Candidate L wins if $\Phi(x_L + z) - \Phi(x_L - z) \geq \Phi(x_R + z) - \Phi(x_R - z) + \rho$, where ρ is the realized net partisan advantage for candidate R. Thus, candidate L's winning probability is

$$P(x_L, x_R) = G(\Phi(x_L + z) - \Phi(x_L - z) + \Phi(x_R - z) - \Phi(x_R + z)). \tag{42}$$

Accounting for this difference, candidate R's optimization problem is the same as above. The first-order condition at $x_L = -x_R$ is now:

$$\begin{aligned}
&\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L=x_R} = -\frac{\partial}{\partial x_R} P(x_L, x_R)[(x_L-\theta_R)^2 - (x_R-\theta_R)^2] - 2(1-P(x_L,x_R))(x_R-\theta_R)\Big|_{-x_L=x_R} \\
= \;& -g(\Phi(x_L+z) - \Phi(x_L-z) + \Phi(x_R-z) - \Phi(x_R+z))|_{-x_L=x_R} \left[\phi(x_R-z) - \phi(x_R+z)\right] 4\theta_R x_R - (x_R-\theta_R) \quad (43) \\
= \;& -4g(0)\left[\phi(x_R-z) - \phi(x_R+z)\right]\theta_R x_R - (x_R-\theta_R) = 0, \quad (44)
\end{aligned}$$

because $-x_L = x_R$ and symmetry of Φ at zero imply $\Phi(x_R+z) = \Phi(-x_L+z) = 1-\Phi(x_L-z)$ and $\Phi(x_R-z) = \Phi(-x_L-z) = 1-\Phi(x_L+z)$.

Differentiating the left-hand side of (44) with respect to x_R now gives:

$$\frac{\partial}{\partial x_R}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L=x_R}\right) = -4g(0)\theta_R[\left[\phi'(x_R-z) - \phi'(x_R+z)\right] x_R + \phi(x_R-z) - \phi(x_R+z)] - 1 < 0,$$

as long as $\left[\phi'(x_R-z) - \phi'(x_R+z)\right] x_R + \phi(x_R-z) - \phi(x_R+z) > 0$. Noting that $\phi(x_R-z) > \phi(x_R+z)$ by single-peakedness, a sufficient condition is that $\phi'(x_R-z) > \phi'(x_R+z)$.

Differentiating (44) with respect to θ_R and z and the implicit function theorem imply again that x_R decreases in z, whereas x_R increases in θ_R if $g(0)$ and $\phi(0)$ are sufficiently small:

$$\frac{\partial}{\partial \theta_R}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L=x_R}\right) = -4g(0)\left[\phi(x_R-z) - \phi(x_R+z)\right] x_R + 1$$

$$\frac{\partial}{\partial z}\left(\frac{\partial}{\partial x_R} EU_R(x_R, x_L)\Big|_{-x_L=x_R}\right) = 4g(0)\left[\phi'(x_R-z) + \phi'(x_R+z)\right]\theta_R x_R < 0.$$

To verify that the second-order condition holds, note that the second derivative of candidate R's optimization problem (40) is

$$-\alpha\frac{\partial^2 P}{\partial x_R^2}\left((x_L-\theta_R)^2 - (x_R-\theta_R)^2\right) - 4\alpha\frac{\partial P}{\partial x_R}(\theta_R - x_R) - 1. \quad (45)$$

Note that

$$\frac{\partial P}{\partial x_R} = \alpha\left(\phi(x_R-z) - \phi(x_R+z)\right) > 0 \;\text{ and }\; \frac{\partial^2 P}{\partial x_R^2} = \alpha\left(\phi'(x_R-z) - \phi'(x_R+z)\right). \quad (46)$$

Hence the second term in (45) is strictly negative. Further, the first term is negative if the density is concave between θ_L and θ_R. Alternatively, the first term becomes irrelevant if the density is sufficiently flat. ■

Proof of Proposition 9. For $\beta \to 0$, it must be that $\underline{y}_L \to -\infty$ because $\underline{y}_L = -\infty$ for $\beta = 0$. As there is vanishing mass in the tail of Φ, we can approximate the comparative statics by assuming that $S_L = (-\infty, \bar{y}_L)$.

Consider the indifference condition (8) for the marginal coalition member, and define:

$$\hat{\xi}_L(\bar{y}_L) = \beta(A - (x_L - \bar{y}_L)^2) - c + (1-\beta)\alpha\Phi((-\infty, \underline{e}_L] \cup [\bar{e}_L, \bar{y}_L])\,(x_R - x_L)\,(x_L + x_R - 2\bar{y}_L) = 0.$$

The implicit function theorem yields

$$\begin{aligned}\frac{\partial}{\partial\beta}\bar{y}_L &= -\frac{\partial\xi_L(\bar{y}_L)/\partial\bar{y}_L}{\partial\xi_L(\bar{y}_L)/\partial\beta} \\ &= -\frac{-2\beta\,(\bar{y}_L - x_L) + (1-\beta)\alpha\phi(\bar{y}_L)\,(x_R - x_L)\,(x_L + x_R - 2\bar{y}_L) - 2(1-\beta)\alpha\Phi(S_L \setminus E_L)\,(x_R - x_L)}{A - (x_L - \bar{y}_L)^2 - \alpha\Phi(S_L \setminus E_L)\,(x_L + x_R - 2\bar{y}_L)}.\end{aligned}$$

For $\beta \to 0$,

$$\frac{\partial}{\partial\beta}\bar{y}_L \to -2\alpha\,(x_R - x_L)\,\frac{\phi(\bar{y}_L)\left(\frac{x_L + x_R}{2} - \bar{y}_L\right) - \Phi(S_L \setminus E_L)}{A - (x_L - \bar{y}_L)^2 - \alpha\Phi(S_L \setminus E_L)\,(x_L + x_R - 2\bar{y}_L)}$$

The denominator is negative, because $A < (x_L - \bar{y}_L)^2$, as $\bar{y}_L > \bar{e}_L$, and the second term is negative.

Using $x_R = -x_L$, we obtain

$$\lim_{\beta\to 0}\frac{\partial}{\partial\beta}\bar{y}_L \propto -\bar{y}_L\phi(\bar{y}_L) - \Phi(S_L \setminus E_L) < 0.$$

Hence, the participation rate $r_L = \frac{\Phi(\bar{y}_L)}{\Phi(0)}$ also decreases in β. ■

Proof of Proposition 10. We prove the first statement. The second one follows analogously. Note that increasing μ decreases $\Phi(S \setminus E_L)$ for any interval $S = [\underline{s}, \bar{s}] \supset E_L$ with $S \leq (x_L + x_R)/2$. In particular

$$\frac{\partial\Phi(S \setminus E_L)}{\partial\mu} = -(\phi(\bar{s} - \mu) - \phi(\bar{e}_L - \mu)) - (\phi(\underline{e}_L - \mu) - \phi(\underline{s} - \mu)) < 0, \tag{47}$$

where the inequality follows because $\phi'(x) > 0$ for $x < (x_L + x_R)/2$.

Now consider $S_L(\mu)$ and let $\mu' > \mu$. Given distribution $\Phi(x-\mu')$ rather than $\Phi(x-\mu)$ reduces the value of strategic voting for coalition $S_L(\mu)$. Let $h_{\mu'}$ be defined as in the proof of Proposition 2 when the distribution of types is $\Phi(x-\mu')$. Then $h_{\mu'}(S_L(\mu)) \subset S_L(\mu)$. Tarski's fixed point theorem implies that there exists a fixed point $\underline{S}_{\mu'}$ of $h_{\mu'}$ with $\underline{S}_{\mu'} \subset S_L(\mu)$.

Suppose by way of contradiction that the maximal fixed point $\bar{S}_{\mu'}$ of $h_{\mu'}$ is not a subset of $S_L(\mu)$. If we decrease μ' back to μ an analogous argument as in (47) implies that the measure of $\bar{S}_{\mu'} \setminus E_L$ increases. Thus, $h_\mu(\bar{S}_{\mu'}) \supset \bar{S}_{\mu'}$. The Tarski fixed point theorem implies that there exists a maximal fixed point $\bar{S}_\mu$ for h_μ where $\bar{S}_\mu$ strictly contains $\bar{S}_{\mu'}$. However, because the largest fixed point of h_μ is $S_L(\mu)$, this is a contradiction. ■

Appendix B: Additional Comparative Statics

5.1 Exogenous Platforms

Using Propositions 3 and 4, we can investigate the impact of mass polarization. Define $\Phi_s = \Phi(x/s)$. If Φ has standard deviation σ then Φ_s has standard deviation $s\sigma$. Increasing s corresponds to increasing the dispersion of citizen ideologies and hence increases mass polarization.

Result 5 *Suppose $x_L = -x_R$ and Φ is symmetric and mass polarization is increased by replacing $\Phi(x)$ with $\Phi(x/s)$ where $s > 1$. For $\beta = 0$ this causes S_i to increase, as well as both $\Phi(S_i)$ and r_i to increase (for $c_i < \bar{c}_i$, where $\bar{c}_i$ is given in Proposition 4). For $\beta = 1$, this does not affect E_i, but the effect on $\Phi(E_i)$ and r_i depends on parameters.*

With symmetry $\xi_{R,s}(y) = 4x_R y(1 - \Phi(y/s))$. It is immediate that $\xi_{R,s}$ is increasing in s for $y > 0$, so we can again conclude that $\bar{y}_R$ decreases and hence the interval of types S_R increases. Further, the number of centrist citizens decreases, i.e., $1 - \Phi_s(y)$ increases as s is increased, and hence $1 - \Phi(S_R)$ increases. Because the number of R supporters does not change, the turnout rate rises. For $\beta = 1$ it is immediate that the sets E_L and E_R do not change. We show in Lemma 3 in the Appendix that turnout and turnout rates increase are single-peaked functions of mass polarization s, increasing when s is small and decreasing when s is large.

5.2 Endogenous Platforms

TO BE ADDED.

Appendix C: Candidate Valence

Letting ω_i be the valence of candidate i, we have that the utility of voter y utility from voting decision v is

$$u_y(v, x, x_w, \omega, \omega_w) = -v\beta\Big((x - y)^2 - \omega\Big) - (1 - v)\beta A - (1 - \beta)\Big((x_w - y)^2 - \omega_w\Big) - vc.$$

ADD PARAGRAPH ABOUT EXISTENCE AND CHARACTERIZATION RESULTS ROBUST TO INTRODUCING VALENCE.

We can also explore the effect of changes in candidate valence on turnout and turnout rates. Suppose the candidates have non-trivial valences, ω_L and ω_R. Then ξ_R becomes

$$\tilde{x}_R(y) = (1 - \Phi(y))\Big((x_R - x_L)(2y - x_L - x_R) + \omega_R - \omega_L\Big).$$

Thus, $\tilde{\xi}_R(y) > \xi_R(y)$ if candidate R's net valence advantage $\omega_R - \omega_L > 0$. In this case the insight from the right panel of Figure 1 applies, i.e., S_R increases in size. The opposite holds for S_L. We have

Result 6 *For $\beta = 0$ only net valence matters: if the net valence $\omega_R - \omega_L$ of candidate R rises, then S_R and turnout rate r_R increase while S_L and r_L decrease. In contrast, for $\beta = 1$, S_i and r_i increase if and only if ω_i increases.*

Appendix D: Changes in the Weight on Expressive Preferences, β

In the core of the paper, we have analyzed the benchmark case of purely expressive ($\beta = 1$) and purely instrumental ($\beta = 0$) preferences. In this Appendix, we investigate how varying the preference parameter β affects coalition formation. Our first result shows that the size of the equilibrium coalition is not monotone in β.

Proposition 9 *Let Assumption 3 hold and suppose $c < \bar{c}_i$. For β close to zero, voter participation rates decrease in β. In contrast, for β close to one, voter participation rates increase in β.*

Proposition 6 analyzed changes in turnout when the distribution of voter types Φ shifts in the favor of candidate R when$\beta = 0$ and when $\beta = 1$. For arbitrary β, we now provide conditions under which the voting coalition S_L shrinks due to the shift in Φ if there is strategic voting, i.e., $S_L \setminus E_L \neq \emptyset$.

Proposition 10 *Suppose the distribution of types is given by $\Phi_\mu(x) = \Phi(x - \mu)$, where $\phi'(x) > 0$ for $x < 0$ and $\phi'(x) < 0$ for $x > 0$.*

Suppose that $\mu \geq (x_L + x_R)/2$. Let $S_L(\mu)$ be the set of strategic voters when the peak of the distribution is μ. Then $S_L(\mu')$ is a strict subset of $S_L(\mu)$ for $\mu' > \mu$.

To understand the intuition, consider a coalition S_L. Because G is uniform, only the measure of voters in $S_L \setminus E_L$ matters for strategic coordination. The marginal voters are exactly indifferent between having coalition S_L, and increasing candidate L's vote share by $\Phi_\mu(S_L \setminus E_L)$. If the distribution Φ_μ is shifted to the right by increasing μ to μ', single peakedness implies that the number of voters in $S_L \setminus E_L$ decreases. This, in turn, decreases the benefit of strategic voting, and hence no coercion is violated for the marginal members of S_L. The new marginal members will be strictly interior of the old coalition. The Tarski fixed point theorem in fact implies that the new strategic coalition is strictly contained in the original S_L.

Finally, consider arbitrary $\beta \in (0,1)$ and the distribution of voter types is given by a family of normal distributions $\Phi_{\mu,\sigma}$. Consider the impact of varying the extent of mass polarization, σ. If σ is very small or very large, then there is no benefit to strategic voting. In particular, if σ is very small, then most probability mass is concentrated around zero, where voters are close to indifferent when policy positions are symmetric around zero. Similarly, S_L collapses to E_L for all sufficiently large σ. This is because for a given $\beta \in (0,1)$, sufficiently extreme voters are better off abstaining. For sufficiently large σ further increases in σ mean that the same set S_L would contain fewer voters, implying a reduced net benefit of strategic voting.