Social Choice and Welfare (2020) 55:369–398
https://doi.org/10.1007/s00355-020-01250-6

ORIGINAL PAPER

A theory of strategic voting with non-instrumental motives

Christopher Li[1] · **Ricardo Pique**[2]

Received: 5 August 2018 / Accepted: 6 March 2020 / Published online: 17 March 2020

Abstract
Empirical studies have documented non-instrumental motives for voting. However, the theoretical literature on strategic voting has largely ignored these motives. In this paper, we examine voter behavior in multi-candidate elections in the presence of ethical, expressive, and instrumental concerns. Voters in our model derive utility from both the election outcome and the action of voting. A fraction of voters are ethical, who follow a group-welfare maximizing voting rule. The rule may require them to misalign their votes, that is, to vote for a candidate who is not their most preferred. We characterize the optimal rule for the ethical voters and provide comparative statics with respect to various electoral parameters. In particular, we find that the degree of misaligned voting is increasing in the importance of the election but is non-monotonic in the popularity of the Condorcet loser.

1 Introduction

In multi-candidate elections, voters are sometimes better off voting for a candidate who is not their most preferred, since doing so may prevent a less preferred candidate from winning. In these situations, strategic voters, those who maximize their expected payoff from voting, will have an incentive to misalign their votes.

Strategic voting has received significant attention from both the theoretical and empirical literature. Existing theories by and large build on the rational voting

A previous version of the paper was circulated under the title "Strategic Voting with Ethical Voters". We thank the editor and two reviewers for suggestions that improved the paper. We also thank Sandeep Baliga, Daniel Diermeier, Georgy Egorov, Timothy Feddersen, Samir Mamadehussene, Kym Pram, Alvaro Sandroni, Michael Ting, and participants at the 2017 SPSA meeting for helpful comments.

✉ Christopher Li
cli5@fsu.edu

Ricardo Pique
rpique@ryerson.ca

[1] Florida State University, Tallahassee, FL 32306, USA

[2] Ryerson University, Toronto, ON M5B 2K3, Canada

paradigm.[1] In these models, voter utility depends only on the election outcome, and strategic behavior is dictated by the likelihood that a single vote alters the outcome. Little consideration is given to how other motives for voting, for example ethical and expressive ones, may impact voting behavior,[2] even though empirical evidence suggests such concerns play an important role (Blais et al. 2000; Coate and Conlin 2004; Funk 2010; DellaVigna et al. 2016; Pons and Tricaud 2018; Spenkuch 2018). Some scholars have studied turnout decisions in the presence of ethical motives (Feddersen and Sandroni 2006a, b) and their findings are more consistent with voting data than previous theories (Coate and Conlin 2004). Thus, going beyond instrumental motives seems fruitful in the study of strategic voting in multi-candidate elections.

In this paper, we explore strategic voting in a model where voters have ethical, expressive, and instrumental concerns. In particular, we consider an election with three candidates and a large electorate. There are two "centrist" candidates and an "extremist" candidate. The electorate is comprised of a minority of extremist supporters and a majority of centrist supporters, who prefer either of the centrist candidates to the extremist. In sum, the extremist is the Condorcet loser, though he has sufficient support so that misaligned voting among centrist voters is necessary for his defeat.

Centrist supporters care both about the election outcome and who they actually vote for. These are the instrumental and expressive concerns, respectively. Centrist voters derive heterogeneous expressive benefits from voting for the two centrist candidates. Moreover, an unknown fraction of centrist supporters are ethical or "rule-utilitarian" in the spirit of Harsanyi (1977) and Feddersen and Sandroni (2006a, b). These voters follow a voting rule that maximizes the aggregate welfare of centrist voters.[3] Such a rule may prescribe some ethical voters to misalign their votes. In determining the degree of vote misalignment, ethical centrists trade off lower expressive benefits for a higher probability of centrist victory.

Our model is tractable and delivers comparative statics of misaligned voting with respect to electoral parameters such as the relative importance of instrumental and expressive concerns, and the popularity of the extremist candidate. We show that, when the popularity of the extremist is low, ethical voters are willing to sacrifice their expressive benefits and coalesce behind one of the two centrist candidates. On the other hand, if the extremist's popularity is sufficiently high, ethical voters refrain from misaligned voting. In this case, the extremist's victory is certain.

The main insight of our model is that the degree of misaligned voting is non-monotonic in the popularity of the extremist. Misaligned voting is increasing for moderate popularity levels. However, ethical centrists have no incentive to engage

[1] See, for example, Bouton (2013), Cox (1997), Feddersen (1992), Fey (1997), Myatt (2007), Myerson and Weber (1993), Myerson (2002), Palfrey (1989), Piketty (2000).

[2] Hamlin and Jennings (2011, 2019) provide an excellent discussion of the foundation of expressive political behavior as well as a survey of the literature.

[3] In other words, ethical centrists identify their fellow centrists as a "group" and vote to maximize the group's welfare.

in misaligned voting when the extremist's popularity is too high (but still remains a minority). Intuitively, when the support for the extremist increases, the probability of a centrist victory decreases. Ethical voters partially offset this by increasing misaligned voting insofar the loss in expressive utility is lower than the gain in instrumental utility. This is the case when the extremist's popularity is below some threshold. When the extremist's popularity is sufficiently high, misaligned voting does not improve the probability of victory sufficiently to compensate for the loss in expressive benefits. Therefore, ethical voters revert to sincere voting.

Empirical evidence so far has failed to yield a consistent verdict on the relationship between misaligned voting and the popularity of the Condorcet loser.[4] This may be due to the difficulty in identifying misaligned voting,[5] the lack of exogenous variation in electoral parameters, and differences in electoral contexts. While there is a lack of causal evidence on how misaligned voting responds to the popularity of the Condorcet loser, the sometime conflicting correlational evidence found across contexts could be the result of a non-monotonic relationship between vote misalignment and extremist support.

In addition to the non-monotonicity result, we show that misaligned voting is increasing in the importance of the election for centrists, as measured by the weight on the instrumental component of their utility. This result follows intuitively from the trade-off between instrumental and expressive motives faced by ethical voters. This conclusion fits one of the findings in Spenkuch (2018), who shows that more voters abandoned non-contender candidates in the critical 2005 German federal election, which followed a non-confidence vote in the Parliament, compared to the less important 2009 election.

This paper complements previous theoretical studies of strategic voting. In these models, voter behavior is predominantly driven by pivotal vote considerations. Seminal works by Palfrey (1989), Cox (1994, 1997) and Myerson and Weber (1993) identify conditions for misaligned voting. In their studies, voters either abandon all but two candidates (i.e., Duvergerian equilibrium) or split their votes in equilibrium.[6] More recent works by Piketty (2000) and Myatt (2007) provide a theoretical foundation for partial misaligned voting by taking into account incomplete information. The latter argues that misaligned voting is increasing in the popularity of the

[4] For example, Fisher (2000) uses survey data from English constituencies to provide evidence on the intuition of Cain (1978) that "tactical" voting is greater in marginal constituencies, where extremist support should not be very high. However, he finds that tactical voting is increasing in the margin of victory, which depends positively on extremist support. On the other hand, using survey data from the 1988 Canadian election, Blais and Nadeau (1996) demonstrate a positive correlation between misaligned voting and the closeness of the race between the second and third choices of the strategic voting (i.e., when the margin of victory is lower). Similarly, Fujiwara (2011) provides causal evidence on the effect of a single-ballot versus a dual-ballot plurality system using Brazilian data and finds that the decrease in votes for the top two candidates due to a change to a dual-ballot system is stronger in close elections. That is, a higher degree of misaligned voting occurs in marginal elections.

[5] Degan and Merlo (2009) provides conditions under which the researcher can disentangle sincere from strategic voting.

[6] Fey (1997) provides a rationale based on dynamic stability for selecting the Duvergerian Equilibrium.

Condorcet loser.[7] Our paper contributes to this literature by presenting an alternative approach that incorporates non-instrumental motives for voting.

Our model also contributes to an emerging theoretical literature on how ethical, or "rule utilitarian", behavior affects political outcomes. In two seminal papers, Harsanyi (1977, 1992) developed the concept of "rule utilitarian" and applied it to study voter turnout. Feddersen and Sandroni (2006a, b) build on Harsanyi's idea and present a seminal theory of turnout in large elections where voters are ethical and have heterogeneous preferences. Using data from Texas liquor referenda, Coate and Conlin (2004) find support for the predictions of Feddersen and Sandroni (2006a, b). More recently, studies have used ethical motivations to account for other behaviors. For example, Piolatto and Schuett (2015) use ethical voters to explain the demand for political news. Our study contributes to this literature by extending the ethical voter framework to the analysis of strategic voting.

In a contemporaneous paper, Bouton and Ogden (2018) explore a model of multicandidate elections with ethical voters. Our approaches and results differ in key aspects. In our model, ethical voters face expressive considerations, which generates the key trade-off in our framework. Also, Bouton and Ogden (2018) predict that misaligned voting is increasing in the extremist's popularity while our novel observation is that this relationship is non-monotonic.

The rest of this paper is organized as follows. Section 2 describes the model. Section 3 presents our main results. Section 4 discusses several extensions to the model. Section 5 concludes.

2 The model

We consider a model of large elections with three candidates. To approximate a large election, we assume a unit continuum of voters. Voting is costless,[8] and the winner of the election is decided by plurality. For convenience, we refer to one of the candidates as the "extremist" and to the other two as the "center-left" and the "center-right."

There is a known measure k_e of voters who will vote for the extremist.[9] This simplifying assumption allows us to focus on the strategic incentives of centrist voters and it is found in other models of strategic voting (see for example Myerson and Weber 1993; Myatt 2007). For the problem to be non-trivial, we impose that $\frac{1}{3} < k_e < \frac{1}{2}$. The upper bound ensures that the extremist supporters are a minority (i.e., the extremist candidate is the Condorcet loser). The lower bound implies that some misaligned voting by the centrist supporters is necessary to defeat the

[7] On the other hand, Cain (1978) provides an informal argument on how misaligned voting disappears when the electoral support for the Condorcet loser is high.

[8] This assumption is relaxed in Sect. 4.2.

[9] The main result does not change qualitatively if we allow for uncertainty in the turnout of the extremist voters (see Sect. 4.3).

extremist. That is, the extremist will win with certainty if centrist supporters evenly split their votes among the two centrist candidates.

The centrist supporters, of measure $1 - k_e$, prefer either centrist candidate to the extremist. Specifically, centrist supporters obtain utility $w > 0$ if either centrist candidate wins, and receive zero utility if the extremist wins. Thus, w is the instrumental value of a centrist victory. In future discussions, we also refer to w as the importance of the election. The assumption of an homogeneous w is made to simplify exposition. In Sect. 4.1, we show that our insights hold when centrist voters have different values of w for the two centrist candidates.

In addition to instrumental utility, centrist supporters derive expressive utility from voting for the two centrist candidates. Formallly, centrist voters have "bliss points" that are uniformly distributed on the interval [0, 1], and the center-left and center-right candidates are located at 0 and 1, respectively. A voter's expressive utility for voting for a candidate is determined by the distance between the voter's bliss point and the location of the candidate. A voter with bliss point $x \in [0, 1]$ (voter x for short) receives an expressive benefit of $\theta(1 - x)$ for voting for the center-left and θx for voting for the center-right.[10] The scalar θ measures the intensity of expressive motive for voting.

In sum, voter x's expected (personal) utility of voting for the center-left and the center-right are, respectively, $wp + \theta(1 - x)$ and $wp + \theta x$, where p is the probability that a centrist candidate wins the election. Note that if voter x acts based on her personal utility, she votes for the center-left if $x < \frac{1}{2}$ and for the center-right otherwise.[11] We refer to this as sincere voting.

So far, we have incorporated both instrumental and expressive motives for voting. We also assume that each centrist voter is ethical or "rule-utilitarian" with probability q_c. The precise value of q_c is unknown, but it is common knowledge that q_c is drawn from the uniform distribution on [0, 1]. As in Harsanyi (1977) and Feddersen and Sandroni (2006a, b), ethical voters follow a voting rule that maximizes the expected aggregate welfare of centrist supporters.[12] In Sect. 4.5, we consider the alternative setting in which centrists are divided into two ethical voter groups, the center-right and center-left, based on voters' bliss points, and each group follows their own ethical voting rule.[13]

[10] The centrist voters derive zero expressive benefit from voting for the extremist.

[11] Without loss of generality, the voter with bliss point $\frac{1}{2}$ votes for the center-right.

[12] One can formalize the notion of ethical voters by assuming that they obtain a sufficiently large utility for following the ethical voting rule. See Feddersen and Sandroni (2006a) for a discussion.

[13] In this case, one would need to impose a consistency condition similar to the notion of Nash Equilibrium (see Feddersen and Sandroni 2006a, b). We show that when the fraction of ethical voters is the same for both groups, the predictions of the baseline framework will hold. The analysis is more complicated if we allow the fractions of ethical voters to be independent draws. Nonetheless, we can show that for certain ranges of parameters, the pattern of misasligned voting predicted in our baseline framework (i.e., non-monotonicity with respect to extremist support) can be supported in equilibrium.

Without loss of generality, we focus on voting rules that are defined by a threshold on the interval [0, 1].[14]

Definition 1 A voting rule is a cut-off $\sigma_c \in [0, 1]$ that instructs ethical voters with bliss points $x \geq \sigma_c$ to vote for the center-right and those with bliss points $x < \sigma_c$ to vote for center-left.

The *ethical voting rule*, σ_c^*, is the voting rule that maximizes the expected aggregate welfare of centrist voters. That is,

$$\sigma_c^* \in \mathrm{argmax}_{\sigma_c} wP(\sigma_c) + \int_0^1 q_c \cdot B_E(\sigma_c) + (1 - q_c) \cdot B_S dq_c \tag{1}$$

where

- $P(\sigma_c) \equiv \Pr\left[q_c \max\left\{\sigma_c, 1 - \sigma_c\right\} + \frac{1}{2}(1 - q_c) \geq \frac{k_e}{1-k_e}\right]$ is the probability of a centrist victory.[15]
- $B_E(\sigma_c) \equiv \int_0^{\sigma_c} \theta(1 - x)dx + \int_{\sigma_c}^1 \theta x dx$ is the aggregate expressive benefit of the ethical centrists.
- $B_S \equiv \int_0^1 \max\{\theta x, \theta(1 - x)\}dx = \frac{3}{4}\theta$ is the aggregate expressive benefit of those centrists who vote sincerely.

The ethical voting rule σ_c^* exists since the objective function is continuous on a compact domain. Note that the objective function is symmetric around $\frac{1}{2}$: the expected aggregate welfare under σ_c is the same as that under $1 - \sigma_c$. Thus, ethical voting rules always come in pairs. For simplicity, we restrict attention to the ethical voting rule σ_c^* in the interval $[0, \frac{1}{2}]$. In a more general setting where centrist voters have heterogeneous instrumental utilities, the ethical voting rule is generically unique (see Sect. 4.1). Alternatively, one can introduce uncertainty about which centrist candidate ethical voters will rally for. For example, suppose that an opinion poll takes places before the election in which interviewees express their sincere opinion on who their most preferred candidate is. If the poll sample is unbiased, both centrist candidates will be the most popular centrist with probability equal to one half. It can then be assumed that, if ethical voters misalign their vote, they will do so in favor of the most popular centrist candidate according to the poll.[16]

There is *misaligned voting* when $\sigma_c < \frac{1}{2}$. Here, ethical voters with bliss point $x \in [\sigma_c^*, \frac{1}{2}]$ vote for the center-right instead of their most preferred candidate, the center-left. We refer to $m \equiv \frac{1}{2} - \sigma_c^*$ as the degree of misaligned voting.

[14] For any voting rule that is not of the threshold type, there is a threshold type rule that gives the same probability of a centrist victory and a higher aggregate expressive benefit.

[15] The probability is implicitly defined by the distribution of q_c.

[16] In this case, all three candidates would have a positive probability of winning before the poll takes place.

3 Results

To simplify exposition, we define $\tilde{k}_e \equiv \frac{k_e}{1-k_e}$. Since $\tilde{k}_e$ is a monotonic transformation of k_e, it retains the interpretation of the popularity of the extremist. Note that the restriction on k_e translates to $\frac{1}{2} < \tilde{k}_e < 1$.

Proposition 1 below states that ethical voters misalign their votes when the extremist has moderate levels of popularity but they vote sincerely if the extremist is sufficiently popular (but still a Condorcet loser).

Proposition 1 *There exists a threshold on $\tilde{k}_e$, denoted $\bar{k}_e$, that is less than* 1, *such that*

- *if $\tilde{k}_e < \bar{k}_e$, ethical voters misalign their vote* (i.e. $\sigma_c^* < \frac{1}{2}$). *In particular, $\sigma_c^* < 1 - \tilde{k}_e$.*
- *if $\tilde{k}_e > \bar{k}_e$, ethical voters vote sincerely* (i.e. $\sigma_c^* = \frac{1}{2}$),
- *if $\tilde{k}_e = \bar{k}_e$, then ethical voters are indifferent between voting sincerely and misaligning their votes.*

In deciding whether to vote against their preferences, the ethical voters balance the instrumental value of a centrist victory and the expressive benefit of voting. Misaligned voting improves the probability of a centrist victory but there is an opportunity cost in terms of foregone expressive benefits for ethical voters with bliss points $x \in [\sigma_c^*, \frac{1}{2}]$. These voters derive greater expressive utility from voting for the center-left but are instructed by the rule to vote for the center-right. Generally, a minimum level of misaligned voting is needed for it to change the election outcome. For example, when $\tilde{k}_e = \frac{3}{4}$, at least a quarter of the ethical centrists need to misalign their votes in order for centrists to have a positive probability of victory. Thus, there is a fixed cost associated with (effective) misaligned voting. When the extremist's popularity is not too high (i.e., $\tilde{k}_e < \bar{k}_e$), ethical voters are willing to bear this fixed cost and misalign their votes to increase the chance of a centrist victory. When the extremist's popularity is very high (i.e., $\tilde{k}_e > \bar{k}_e$), the fixed cost for misaligned voting becomes prohibitively high.[17] Ethical voters are unwilling to bear this cost and they will vote sincerely.

The following corollary describes how the threshold $\bar{k}_e$ depends on w (i.e., the importance of the election) and θ (i.e., the intensity of expressive benefits).

Corollary 1 *The threshold $\bar{k}_e$ is increasing in w and decreasing in θ. Specifically, it is decreasing in the ratio $\frac{\theta}{w}$.*

Thus, the threshold at which ethical voters switch from misaligned voting to sincere voting is increasing in the instrumental value of voting and decreasing in the expressive value of voting. This is intuitive given our previous discussion on the

[17] Obtaining a positive probability victory requires almost half of ethical voters to misalign their votes.

trade-off faced by ethical voters. When ethical voters care more about defeating the extremist (i.e., higher values of w), they are more willing to bear the cost of misaligned voting. On the other hand, when the intensity of expressive motive increases (i.e., higher values of θ), ethical voters face a greater cost of voting against their true preferences. Hence, they are less willing to engage in misaligned voting.

Two empirical implications emerge from the previous results. First, misaligned voting is more likely to occur in elections in which the outcome has a large impact on public welfare. Second, misaligned voting is more likely to arise when centrist voters are relatively homogeneous in ideology or when there is little differentiation between centrist candidates in platforms or valence. In these cases, the foregone expressive benefit due to voting against preference should be low.

The next set of results provides a more detailed characterization of misaligned voting.

Proposition 2 *Full misaligned voting (i.e. $\sigma_c^* = 0$) occurs if and only if $\frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e \leq 1 - \frac{\theta}{16w}$.*

The proposition states that, conditional on misaligned voting being optimal (i.e., $\bar{k}_e = 1 - \frac{\theta}{16w} < \tilde{k}_e$),[18] ethical voters fully misalign their votes when the extremist is sufficiently popular (i.e., $\frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e$). Intuitively, if the extremist's popularity is low, the marginal benefit of misaligned voting is low since centrists are already in a strong position. Therefore, ethical voters would only engage in partial misaligned voting (i.e., $\sigma_c^* > 0$).[19] Note that the condition $\frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e \leq 1 - \frac{\theta}{16w}$ holds only if $\frac{\theta}{w} \leq \frac{8}{3}$. In other words, a necessary condition for full misaligned voting is that the importance of the election is sufficiently high or the intensity of expressive benefits is sufficiently low.

Figure 1 below maps out regions in the space of electoral parameters (i.e., $\tilde{k}_e$ and $\frac{\theta}{w}$) where misaligned or sincere voting occurs.

Proposition 3 shows that the degree of misaligned voting is increasing in the importance of the election and the extremist's popularity, and is decreasing in the intensity of expressive benefits.

Proposition 3 *When misaligned voting is optimal (i.e. $\tilde{k}_e < \bar{k}_e$), the degree of misaligned voting, $m(w, \theta, \tilde{k}_e) \equiv \frac{1}{2} - \sigma_c^*$, is increasing in w and $\tilde{k}_e$, and is decreasing in θ.*

Proposition 3 together with Proposition 1 implies a non-monotonic relationship between misaligned voting and the popularity of the extremist. The degree of misaligned voting is increasing in the extremist's popularity, $\tilde{k}_e$, conditional on $\tilde{k}_e \leq \bar{k}_e$.

[18] Whenever the inequality $\frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e \leq 1 - \frac{\theta}{16w}$ holds, it is the case that $\bar{k}_e = 1 - \frac{\theta}{16w}$. For details, see the proof for Proposition 2.

[19] Recall that there is misaligned voting for $\tilde{k}_e$ below $\bar{k}_e$. It follows that ethical voters would engage in misaligned voting when $\tilde{k}_e < \frac{1}{2} + \frac{\theta}{8w}$, but only partially.

However, once the extremist's support exceeds $\bar{k}_e$, ethical voters revert to sincere voting.

The comparative statics of misaligned voting with respect to w and θ are intuitive and in line with Corollary 1. When w increases, centrist supporters are more concerned with defeating the extremist. Therefore, ethical voters have a stronger incentive to misalign their votes. The intensity of the expressive benefit, θ, determines the opportunity cost of misaligned voting. When θ increases, so does the cost. This leads to less misaligned voting.

Proposition 4 below presents comparative statics of the (ex-ante) probability of a centrist victory with respect to the electoral parameters.[20] We show that the probability of a centrist victory is increasing in the importance of the election, and decreasing in both the intensity of the expressive benefit and the extremist's popularity.

Proposition 4 *The probability of a centrist victory given the ethical voting rule, $P(\sigma_c^*)$, is increasing in w and decreasing in θ and $\tilde{k}_e$.*

The intuition behind the comparative statics with respect to θ and w is straightforward. We have shown that the degree of misaligned voting is increasing in w and decreasing in θ. Since the probability of a centrist victory is increasing in misaligned voting, it must be increasing in w and decreasing in θ. On the other hand, the effect of an increase in the extremist's popularity is less obvious. It decreases directly the likelihood of a centrist victory, but the ethical voters may have a greater incentive to misalign their votes (see Proposition 3). It happens that the ethical voters' response is of a second order effect. Hence, the probability of a centrist victory is decreasing in the extremist's popularity.

3.1 Empirical evidence

In the real world, voter decisions are driven by multiple considerations (e.g., expressive, ethical, and instrumental), which can be difficult to disentangle empirically. Survey data provide some insights into voters' motives, but, due to its self-reported nature, it is not necessarily conclusive. As voter ideological positions are not directly observable and must be inferred, distinguishing between different motivations is not trivial. For example, Degan and Merlo (2009) show that the hypothesis that voters vote sincerely is not falsifiable when using single election data. Moreover, when using data for the same electorate across elections, the number of elections needs to be greater than the number of dimensions of the policy space for the hypothesis to be testable.

In spite of these obstacles, there is evidence that ethical and expressive concerns drive voter behavior in large elections. Coate and Conlin (2004) show that an

[20] For a given realization of q_c, the centrists win or lose with certainty. The ex-ante uncertainty about a centrist victory is driven by the uncertainty about q_c.

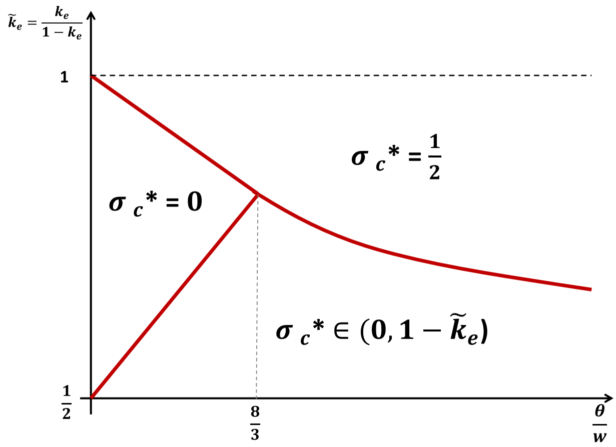

Fig. 1 Ethical voting rule

structurally estimated group rule-utilitarian model provides a good fit of Texas liquor referenda data. Using the same data, Coate et al. (2008) find that a simple model based on expressive motives provides a better fit than a pivotal-voter model. Moreover, Pons and Tricaud (2018) find support for expressive motives by showing how the presence of a third candidate decreases the vote share of the two leading candidates in French parliamentary elections. DellaVigna et al. (2016) uses a field experiment to infer that, in the 2010 Congressional election, voters in the Chicago suburbs had an expressive value of voting (voting "to tell others") of close to US $15.

While the expressive and ethical motives appear to be important, instrumental concerns remain a factor. Voting models that incorporate multiple considerations may provide a better fit of electoral data. Spenkuch (2018) takes advantage of Germany's electoral system to estimate that close to two thirds of voters do not behave according to the pivotal voter model while close to a third do not behave according to expressive considerations.[21]

While existing evidence suggests that various motives play a role in voter behavior, it does not offer a direct validation of our model. An ideal experiment to falsify our main result would be one in which the support for the extremist (or Condorcet loser), $\frac{1}{3} < k_e < \frac{1}{2}$, varies exogenously across multiple electoral constituencies in the same election.[22] Our model predicts a non-monotone relationship between k_e and the degree of misaligned voting. Ceteris paribus, this should translate into a non-monotone relationship between k_e and the ratio of votes of the most-voted centrist party to the votes of the least-voted centrist party.

[21] The author concludes that "voters cannot be neatly classified into strategic and sincere "types"."(Spenkuch 2018,74)

[22] We require an election in which three parties run candidates in these multiple constituencies, where the winner is decided under a first-past-the-post system. Supporters of two of these parties should have an incentive to misalign their vote, while the third party, the extremist must not. That is, ideally, we require a setting similar to Myatt (2007)'s "beat the conservative" game.

Unfortunately, finding a plausibly exogenous variation in k_e is difficult. Extremist support depends on unobservable socio-demographic characteristics, which may be correlated with other key parameters such as the fraction of ethical voters and the intensity of the expressive motive.[23]

Existing studies do not offer a conclusive answer as to whether our model fits electoral behavior due to the lack of causal evidence on how misaligned voting responds to the popularity of the Condorcet loser. Most of the empirical evidence on the matter is suggestive at best. For example, using survey data in English constituencies, Fisher (2000) finds a positive correlation between misaligned or "tactical" voting and the margin of victory. Since the margin of victory should be increasing in extremist support, one can interpret this finding as a positive correlation between misaligned voting and the popularity of the extremist. On the contrary, Blais and Nadeau (1996) use Canadian data to find that close races (i.e., races with small margins of victory) are positively correlated with more misaligned voting. Fujiwara (2011) also finds support for a negative relationship between misaligned voting and extremist support. Using data from Brazilian mayoral elections, he provides causal evidence on the effect of a single-ballot versus a dual-ballot plurality system and finds that the top two candidates experience a bigger drop in their vote shares due to a change to a dual-ballot system in areas with close elections. This means that more misaligned voting occurs in marginal elections. Using both voting and survey data from multiple UK elections, Kiewiet (2013) finds that, in some elections, strategic voting increases with the closeness of the election while in others it decreases. The lack of a consistent message may reflect the difficulty in casually identifying misaligned voting, the lack of exogenous variation in electoral parameters, and differences in electoral contexts. But it may also suggest that there is a non-monotonic relationship between vote misalignment and extremist support.

Finally, regarding the effect of the relative importance of instrumental motives, it is worth noting that the findings in Spenkuch (2018) support our result in Proposition 3. Using German data, the author finds that voters were more likely to abandon non-contender candidates in the critical 2005 federal election than in the less important 2009 election. This supports our conclusion that misaligned voting is increasing in the importance of the election.

4 Extensions

4.1 Heterogeneity in instrumental utility

In the benchmark model, centrist supporters receive the same instrumental utility, w, regardless of which centrist candidate wins. In this section, we explore an extension

[23] Measuring of ethical and expressive concerns independently, say using laboratory experiments, may help resolve this issue, but such data can be problematic. For example, it could be difficult to generate a sense of duty and group affiliation that voters receive when voting in an election.

of our model where centrists derive heterogeneous instrumental benefits from victory, and the benchmark results are shown to hold qualitatively.

Keeping other elements of the benchmark model the same, let the centrists' instrumental utilities for the centrist candidates be described by two measurable functions $w_l : [0,1] \to \mathbb{R}_{++}$ and $w_r : [0,1] \to \mathbb{R}_{++}$. Specifically, $w_l(x)$ and $w_r(x)$ are the instrumental utilities that voter x receives when the center-left and the center-right candidate wins, respectively. Note that this formulation allows for very general correlation between the instrumental and expressive motives, e.g., voters with strong expressive motive also have strong instrumental motive. The ethical voting rule takes into account the distribution of instrumental utilities among the centrists. The aggregate welfare for centrists is now:

$$\int_0^1 q_c \cdot B_E(\sigma_c) + (1-q_c) \cdot B_S dq_c + P(\sigma_c) \cdot \begin{cases} \int_0^1 w_l(x)dx & \text{if } \sigma_c > \frac{1}{2} \\ \int_0^1 w_r(x)dx & \text{if } \sigma_c \leq \frac{1}{2} \end{cases} \tag{2}$$

where $B_E(\sigma_c)$ and B_S are defined in formula (1). Unlike in the benchmark setting, centrists' aggregate welfare depends on whether the center-left or the center-right wins. Specifically, if $\sigma_c > \frac{1}{2}$, then the center-left always obtains more votes than the center-right. In this case, the aggregate instrumental utility is given by $\int_0^1 w_l(x)dx$. Similarly, if $\sigma_c < \frac{1}{2}$, the center-right always obtains more votes than the center-left, and the aggregate instrumental utility is $\int_0^1 w_r(x)dx$.

In the benchmark model, voting rule σ_c induces the same aggregate welfare as $1-\sigma_c$. That is, ethical voters are indifferent between which centrist candidates to misalign their votes for. In the current setting, ethical voters prefer to misalign their vote for the candidate whose victory gives the highest average instrumental utility. Formally,

Proposition 5 *The ethical voting rule σ_c^* is in the interval $[0, \frac{1}{2}]$ if $\int_0^1 w_r(x)dx > \int_0^1 w_l(x)dx$ and is in the interval $[\frac{1}{2}, 1]$ if $\int_0^1 w_r(x)dx < \int_0^1 w_l(x)dx$.*

This result is intuitive. As in the benchmark case, the expressive component of the welfare function is symmetric in σ_c, and the direction of misaligned voting under the ethical voting rule can only depend on the distribution of instrumental utilities. Clearly, it is optimal for ethical voters to misalign their votes for the candidate whose victory induces the highest aggregate instrumental utility.

Except for breaking the indifference between voting rules, ethical voters face the same trade-off as in the benchmark setting. Indeed, treating $\max\{E(w_l), E(w_r)\}$ as w, one may proceed with the analysis as before. The characterization of the ethical voting rule and the comparative statics with regard to θ and $\tilde{k}_e$ would be the same as in the benchmark case. The comparative statics with respect to the importance of the election also go through given that one uses $\max\{\mathbb{E}(w_r), \mathbb{E}(w_l)\}$ in place of w.

4.2 Turnout decision

In this section, we consider an extension of the model where centrist voters face stochastic turnout costs.[24] Let centrist voters be indexed by a tuple (x, t) (the voter's type), where x is the voter's bliss point, and $\theta t > 0$ is the voter's turnout cost. We let the intensity of expressive motives θ be a scaling factor only to simplify notation; it is not material for the results. Because voting is now costly, centrist voters may choose to abstain. Formally, the (personal) utility of a type (x, t) voter is now given by:

$$\begin{array}{cl} p \cdot w & \text{if abstain} \\ p \cdot w + \theta(1 - x - t) & \text{if vote for center-left} \\ p \cdot w + \theta(x - t) & \text{if vote for center-right.} \end{array}$$

We assume that voter types are drawn uniformly from $[0, 1] \times [0, 1]$. In the benchmark setting, we imposed that $\frac{1}{2} \leq \tilde{k}_e \leq 1$ so the problem is non-trivial. Here, abstentions due to turnout costs means that the extremist can win even if his vote share is less than $\frac{1}{3}$. Thus, we need to revise the restriction on $\tilde{k}_e$ to $\frac{3}{8} < \tilde{k}_e < 1$. The new lower bound for $\tilde{k}_e$ reflects the fact that a quarter of non-ethical centrists will abstain.[25]

The aggregate welfare of the centrists depends on the expressive benefits and turnout costs. A priori, it is not clear what the form of voting rules is e.g., whether they are thresholds. Thus, we broaden the definition of a voting rule, which is now a pair of sets $L, R \subset [0, 1] \times [0, 1]$, where voters of type $(x, t) \in L$ vote for the center-left, voters of type $(x, t) \in R$ vote for the center-right, and voters of type $(x, t) \in (L \cup R)^c$ abstain. Given this, the aggregate welfare of centrists can be written as:

$$wP(L, R) + \int_0^1 q_c \cdot C_E(L, R) + (1 - q_c) \cdot C_S dq_c$$

where

- $P(L, R)$ is the probability that a centrist candidate wins the election given the voting rule.
- $C_E(L, R) = \int_{(x,t) \in L} \theta(1 - x - t) dxdt + \int_{(x,t) \in R} \theta(x - t) dxdt$ is the per-capita expressive utility net of turnout cost faced by ethical voters under the voting rule.
- $C_S = \int_{\max\{1-x,x\}>t} \theta(\max\{1 - x, x\} - t) dxdt$ is the per-capita expressive utility net of turnout cost faced by non-ethical (i.e., sincere) voters.[26]

[24] We maintain the assumption of full turnout by the extremist supporters.

[25] The lower bound on $\tilde{k}_e$ is the minimum support the extremist needs in order to win when centrists vote sincerely.

[26] Note that ideological voters turnout if and only if $\max\{1 - x, x\} > t$ and vote for the candidate that provides the greatest expressive utility.

Interestingly, even though voter type is two-dimensional, the ethical voting rule can be characterized by a one-dimensional threshold. As in the benchmark model, we assume, without loss of generality, that ethical voters misalign their votes for the center-right.[27]

Proposition 6 *The ethical voting rule is characterized by a cut-off* $0 \leq \sigma_c^* \leq \frac{1}{2}$ *such that voter* (x, t)

- *votes for the center-right if* $\sigma_c^* \leq x$ *and* $t \leq 1 - 2\sigma_c^* + x$.
- *votes for the center-left if* $x < \sigma_c^*$ *and* $t < 1 - x$.
- *abstains otherwise.*

Figure 2 below provides an illustration of the ethical voting rule. In general, voters who misalign their votes under the ethical voting rule fall into one of two categories: (1) voters who would turnout and vote for the center-left based on their personal utility, and (2) voters who would abstain based on their personal utility. The opportunity cost of misaligned voting for voters in the first category is the difference between the expressive benefit of voting for the center-left and that of voting for the center-right. This depends only on the voter's bliss point and is equal to $1 - 2x$. On the other hand, the opportunity cost of misaligned voting for voters in the second category is $t - x$, which is the voter's turnout cost minus the expressive utility from voting for the center-right. To maximize aggregate welfare, the ethical voting rule equates both opportunity costs. This leads to the set of restrictions in Proposition 6.

We will not provide a detailed characterization of the ethical voting rule with turnout costs. The structure of the problem (i.e., the objective function for the ethical voters) takes the same form as before and therefore our insights would continue to hold. In particular, the objective function is of the same form as (1) with $P(\sigma_c)$, $C_E(\sigma_c)$ and C_S now defined as follows:[28]

$$P(\sigma_c) = \max\left\{0, 1 - \frac{\tilde{k}_e - \frac{3}{8}}{\frac{5}{8} - \sigma_c - \sigma_c^2}\right\}$$

$$C_E(\sigma_c) = \frac{\theta}{2}\left(\sigma_c - \sigma_c^2 + \frac{\sigma_c^3}{3}\right) + \frac{\theta}{2}\left(\frac{1}{3} - (1 - 2\sigma_c)^2 - \frac{\sigma_c^3}{3} - (1 - 2\sigma_c)^2\sigma_c\right)$$

$$C_S = \frac{7\theta}{24}.$$

Given the objective function, one can verify that our main results (i.e., Propositions 1 and 3) hold. In particular, it is still the case that misaligned voting is optimal for moderate values of $\tilde{k}_e$, and sincere voting is optimal when $\tilde{k}_e$ is sufficiently high.

[27] The symmetry of the objective function implies that ethical voters are indifferent between misaligning their votes for the center-left or the center-right.

[28] The derivation of these expressions is given in the Appendix.

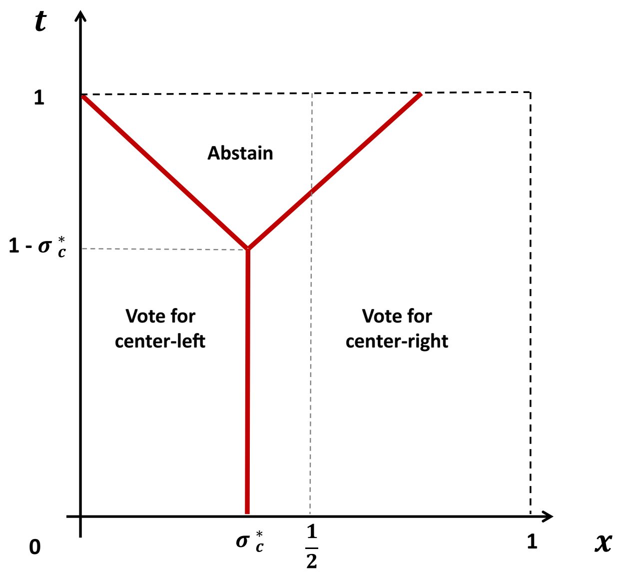

Fig. 2 Ethical Voting Rule

Intuitively, the trade-off faced by ethical voters is the same as before. A positive level of misaligned voting is required to obtain a centrist victory and this level is increasing in the extremist's popularity. When the extremist's support is sufficiently high, the instrumental utility under misaligned voting does not compensate for the loss in expressive benefits. Hence, ethical voters revert to sincere voting.

4.3 Random extremist turnout

In the baseline model, we assumed full turnout by the extremist voters. We relax this assumption in this section. For tractability, we assume that the turnout of extremist voters is a uniform random variable τ with support $[0, k_e]$. Note that the center-right now wins whenever

$$q_c \geq \frac{\frac{\tau}{1-k_e} - \frac{1}{2}}{\frac{1}{2} - \sigma_c}.$$

We show below that a main observation of the benchmark model—the non-monotonicity of misaligned voting with respect to extremist popularity—continues to hold qualitatively. Specifically, the degree of misaligned voting is non-monotonic in the upper-bound of the extremist turnout if the intensity of expressive motives is sufficiently high relative to the instrumental motive.[29] If the intensity of expressive motives is low, then the degree of misaligned voting is always increasing in the extremist's popularity.

Proposition 7 *If $\frac{\theta}{w} \leq \frac{1}{2}$, then the degree of misaligned voting m is increasing in the extremist's popularity (i.e., $\frac{1}{2} - \sigma_c^*$ is increasing in $\tilde{k}_e$). If $\frac{\theta}{w} > \frac{1}{2}$, then there exists a threshold $\bar{k}$ such that the degree of misaligned voting is continuously increasing in $\tilde{k}_e$ if $\tilde{k}_e \leq \bar{k}$, and is continuously decreasing in $\tilde{k}_e$ if $\tilde{k}_e > \bar{k}$.*

4.4 General bliss point distribution

In this section, we consider more general distributions of voter ideological positions. Suppose the bliss points of the centrists are distributed according to CDF G over the interval [0, 1]. The two key components of the ethical voter's objective function, probability of a centrist of victory, $P(\sigma_c)$, and the aggregate expressive benefit of the ethical centrist, $B_E(\sigma_c)$, are now:

- $P(\sigma_c) = \Pr\left[\max\left\{q_c G(\sigma_c) + (1-q_c)G(\frac{1}{2}),\ q_c\big(1-G(\sigma_c)\big) + (1-q_c)\Big(1 - G(\frac{1}{2})\Big)\right\} > \tilde{k}_e\right]$
- $B_E(\sigma_c) = \int_0^{\sigma_c} \theta(1-x)dG(x) + \int_{\sigma_c}^1 \theta x dG(x)$

Unlike in the benchmark setting, there is no ambiguity about the direction of misaligned voting so long as G is not symmetric around $\frac{1}{2}$. However, contrary to what may seem obvious, the direction of misaligned voting is not simply pinned down by whether the median centrist voter is left or right leaning (i.e., whether $G(\frac{1}{2})$ is greater or less than $\frac{1}{2}$).[30] Suppose $G(\frac{1}{2})$ is less than but close to $\frac{1}{2}$ (i.e., the center-right has slightly more support than the center-left), but a substantial number of center-left voters have extreme ideologies (i.e., with bliss points close to 0) while a substantial number of center-right voters are moderate (i.e., with bliss points close to $\frac{1}{2}$). It could be optimal for the ethical voters to align for the center-left candidate since it is less costly for the center-right voters to misalign their votes because of their moderate stance. However, we can derive a sufficient condition on the "skewness" of the distribution that ensures the ethical voters strictly prefer to misalign their votes for the candidates with more support. We state the result for the case where the distribution of voters is skewed to the right.

Proposition 8 *If $G(x) \geq 1 - G(1-x)\ \forall x \geq \frac{1}{2}$ and the inequality is strict for $x = \frac{1}{2}$. Then, the ethical voters have a strict incentive to misalign their votes for the*

[29] Unlike in the benchmark setting, however, this relationship between the degree of misaligned voting and extremist turnout is continuous.

[30] Specifically, for a fixed cutoff σ_c, the fraction of centrists who vote for the center-left depends not only on $G(\frac{1}{2})$ but also on $G(\sigma_c)$. Also, the aggregate expressive cost, $B_E(\sigma_c)$, is not pinned down by $G(\frac{1}{2})$ alone.

center-right candidate (i.e., the ethical voting rule σ_c^* is strictly less than $\frac{1}{2}$)conditional on misaligned voting being optimal.

Unfortunately, the comparative statics with respect to various parameters are no longer straightforward given a general distribution of bliss points. Indeed, even the concavity of the ethical voter's objective function is no longer guaranteed.

One way to allow for more generality while preserving tractability is to allow a general distribution of bliss points for the ideological voters but maintaining that the ethical voters' bliss points are uniformly distributed.[31] In this case, the direction of misaligned voting would be pinned down by the location of the median ideological voter. Specifically, letting G_{id} be the CDF of the bliss points of ideological voters, then ethical voters prefer to misalign their votes for the center-right (left) candidates iff $G_{id}(\frac{1}{2}) < (>)\frac{1}{2}$. Intuitively, ethical voters prefer to misalign their votes for the candidates with the higher support from the ideological voters since there is no cost differential for them in terms of misaligning votes in favor of the center-left vs. the center-right. In addition to the uniqueness of ethical voting rule, the comparative statics from the benchmark setting go through. Since the ideological voters' behavior is fixed, the marginal benefit and marginal cost of misaligned voting by the ethical voters would be independent of the distribution of ideological voters.[32] Consequently, the ethical voters' incentives at the margin are unaffected by the distribution of ideological voters, and the comparative statics of misaligned voting with respect to the various electoral parameters are the same as in the benchmark setting.

4.5 Two ethical centrist groups

In this subsection, we depart from the assumption that the ethical voters treat centrist voters as one "group" when optimizing the voting rule. Specifically, we suppose now that there are two ethical voter groups: the center-left, made up of voters with bliss points $x \in [0, \frac{1}{2}]$, and the center-right, made up of voters with bliss points $x \in [\frac{1}{2}, 1]$.[33] Each center-left and center-right voter is ethical, or "rule-utilitarian", with probability q_l and q_r, respectively. As before, the centrist voters, of measure $1 - k_e$, prefer either centrist candidate to the extremist. To simplify notation, we renormalize the value of a centrist victory for the two groups to be 2w.

Let $\sigma_l \in [0, \frac{1}{2}]$ and $\sigma_r \in [\frac{1}{2}, 1]$ be cut-off type voting rules followed by center-left and center-right ethical voters, respectively.[34] Then, the expected aggregate welfare of center-left and center-right voters are respectively:

[31] In the benchmark model, the two types of voters share the same distribution of bliss points.

[32] This can be seen from the objective function (1). The vote-share threshold for centrist victory and the aggregate expressive benefit for ideological voters would depend on the distribution of ideological voters but not on the voting rule.

[33] We maintain the assumption that the centrist voters' expressive "bliss points" are uniformly distributed on the interval [0, 1].

[34] Specifically, center-left ethical voters vote for the center-left candidate if and only if $x \le \sigma_l$. Center-right ethical voters vote for the center-right candidate if and only if $x \ge \sigma_r$.

$$G_l(\sigma_l,\sigma_r) = wP(\sigma_l,\sigma_r) + \int_0^1 q_l\left(\int_0^{\sigma_l}\theta(1-x)dx + \int_{\sigma_l}^{\frac{1}{2}}\theta x dx\right) + (1-q_l)\cdot B_S dq_l \tag{3}$$

$$G_r(\sigma_l,\sigma_r) = wP(\sigma_l,\sigma_r) + \int_0^1 q_r\left(\int_{\frac{1}{2}}^{\sigma_r}\theta(1-x)dx + \int_{\sigma_r}^{1}\theta x dx\right) + (1-q_r)\cdot B_S dq_r, \tag{4}$$

where

- $P(\sigma_l,\sigma_r) \equiv \Pr\left[\max\left\{q_l\sigma_l + q_r(\sigma_r - \frac{1}{2}) + \frac{1}{2}(1-q_l), q_l(\frac{1}{2}-\sigma_l) + q_r(1-\sigma_r) + \frac{1}{2}(1-q_r)\right\} \geq \tilde{k}_e\right]$ is the probability of a centrist victory.
- $B_S \equiv \int_0^{\frac{1}{2}} \max\{\theta x, \theta(1-x)\}dx = \frac{3}{8}\theta$ is the aggregate expressive benefit for each centrist group.

Ethical voters are assumed to follow a voting rule which maximize the expected welfare of their group, subject to the behavior of other voters. Hence, as in (Feddersen and Sandroni 2006b), the analysis will focus on ethical rules which satisfy the following consistency requirement:

Definition 2 A pair (σ_l^*, σ_r^*) is a consistent rule profile if $G_l(\sigma_l^*, \sigma_r^*) \geq G_l(\sigma_l, \sigma_r^*)$ for all $\sigma_l \in [0, \frac{1}{2}]$, and $G_r(\sigma_l^*, \sigma_r^*) \geq G_r(\sigma_l^*, \sigma_r)$ for all $\sigma_r \in [\frac{1}{2}, 1]$.

Next, we consider the case that the share of ethical voters is the same in the two groups, i.e., $q_l = q_r = q_c$, where q_c is drawn from the uniform distribution on [0, 1]. We then consider the case where q_l and q_r are independent draws from the uniform distribution on [0, 1].

Common Draw of q_l and q_r Suppose that $q_l = q_r = q_c$, with q_c drawn from the uniform distribution, we argue the pair $(\sigma_l^* = \sigma_c^*, \sigma_r^* = \frac{1}{2})$, which is equivalent to the optimal voting rule in our baseline framework, is a consistent profile. This would imply that our comparative statics would continue to hold.

Proposition 9 *$\sigma_l^* = \sigma_c^*$ and $\sigma_r^* = \frac{1}{2}$ is a consistent rule profile.*

Intuitively, when the fractions of ethical voters are the same for both groups, it is not optimal for both groups to misalign their votes at the same time i.e., whenever one group misaligns their votes, it is optimal for the other group to vote sincerely. It follows that the group that is misaligning their votes faces essentially the same trade-off as the ethical voters in the original framework. Therefore, the patterns of misaligned voting remain the same.

Independent Draws of q_l and q_r The case where the fractions of ethical voters in the two groups are uncorrelated is more complex. Specifically, it is no longer straightforward to argue in general that the two groups would not find it optimal to misalign their votes simultaneously. However, we can demonstrate that for a range of parameters, the pair ($\sigma_l^* = \sigma_c^*, \sigma_r^* = \frac{1}{2}$) is a consistent profile. Moreover, misaligned voting exhibits non-monotonicity.

Proposition 10 *For parameters where $\sigma_c^* = \frac{1}{2}$ and $\sigma_c^* < \epsilon$ for some $\epsilon > 0$ small, $\sigma_l^* = \sigma_c^*$ and $\sigma_r^* = \frac{1}{2}$ are a consistent rule profile.*

Recall that misaligned voting is increasing in $\tilde{k}_e$ (i.e.,σ_c^* is decreasing) when σ_c^* is close to 0. At the same time, for sufficiently large $\tilde{k}_e$, there is no incentive to misalign votes (i.e., $\sigma_c^* = \frac{1}{2}$). Thus, Proposition 10 implies that the non-monotonicity of misaligned voting with respect to extremist support is still present in the current context even though we cannot characterize the consistent rule profiles for all parameters.

5 Conclusion

In this paper, we examine strategic voting in multi-candidate elections when voters have ethical and expressive concerns in addition to instrumental ones. The model is parsimonious and provides a clear mechanism of when and to what degree misaligned voting occurs. We also show how misaligned voting can vary significantly with changes in the importance of the election, the intensity of the expressive motive, and the popularity of the Condorcet loser. The novel insight is that misaligned voting is non-monotonic in the popularity of the Condorcet loser.

Our paper illustrates how ethical and expressive motives can help provide additional insights regarding voting patterns. The model can be the basis for studying other important issues in electoral politics. For example, by adapting the model to a dynamic environment, one can examine the incentives of centrist parties to form coalitions. In general, ethical agent models can prove useful in explaining how individuals act when faced with collective action problems such as political protests. This study also highlights the need for more empirical evidence on misaligned voting. While our results can reconcile some of the existing evidence, more research is needed to test the different implications of our model.

Mathematical appendix

First, we include some preliminary calculations used in the proofs. Take the expected aggregate welfare of centrist voters (i.e., the objective function in Eq. (1)) is denoted by

$$F(\sigma_c) \equiv wP(\sigma_c) + \int_0^1 q_c \cdot B_E(\sigma_c) + (1 - q_c) \cdot B_S dq_c.$$

where $B_E(\sigma_c)$ is the per capita expressive benefit derived by ethical voters and B_S is the per capita expressive benefit derived by sincere voters. First, it can be shown that:

- $B_E(\sigma_c) = \theta - \theta(\sigma_c^2 - \sigma_c + \frac{1}{2}) = \theta(\frac{1}{2} - \sigma_c^2 + \sigma_c)$, and
- $B_S = \frac{3}{4}\theta$.

Therefore, it follows that:

$$\int_0^1 q_c \cdot B_E(\sigma_c) + (1 - q_c) \cdot B_S dq_c = \frac{\theta}{2}\left(\frac{1}{2} - \sigma_c^2 + \sigma_c\right) + \frac{3}{8}\theta$$

Now, lets compute $P(\sigma_c)$, the probability of a centrist victory induced by voting rule σ_c. Since we restrict our attention on $\sigma_c \in [0, \frac{1}{2}]$, $P(\sigma_c)$ is equivalent to the probability that the center-right receives more votes than the extremist. That is:

$$\begin{aligned} P(\sigma_c) &= \Pr\left((1 - k_e)\left[(1 - q_c)\frac{1}{2} + q_c(1 - \sigma_c)\right] \geq k_e\right) \\ &= \Pr\left(q_c \geq \frac{\frac{k_e}{1-k_e} - \frac{1}{2}}{\frac{1}{2} - \sigma_c}\right) \end{aligned}$$

Since q_c is distributed uniformly on [0, 1], it follows that:

$$P(\sigma_c) = \max\left\{0, 1 - \frac{\frac{k_e}{1-k_e} - \frac{1}{2}}{\frac{1}{2} - \sigma_c}\right\}$$

Therefore, the objective function $F(\sigma_c)$ can be written as:

$$F(\sigma_c) = w \cdot \max\left\{0, 1 - \frac{\frac{k_e}{1-k_e} - \frac{1}{2}}{\frac{1}{2} - \sigma_c}\right\} + \frac{\theta}{2}\left(\frac{1}{2} - \sigma_c^2 + \sigma_c\right) + \frac{3}{8}\theta. \tag{5}$$

We will use Eq. 5 in the following proofs.

Proposition 1

Proof Since

$$1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - \sigma_c} \leq 0 \ \forall \sigma_c \in \left[1 - \tilde{k}_e, \frac{1}{2}\right],$$

the objective function becomes $F(\sigma_c) = \frac{\theta}{2}(\frac{1}{2} - \sigma_c^2 + \sigma_c) + \frac{3}{8}\theta$ within the interval $[1 - \tilde{k}_e, \frac{1}{2}]$, and it achieves its maximum at $\frac{1}{2}$. Therefore, the choice domain of the maximization problem can be reduced to $[0, 1 - \tilde{k}_e) \cup \{\frac{1}{2}\}$.

The solution to the maximization problem can be found by first finding the local optimal $\tilde{\sigma}_c$ on the interval $[0, 1 - \tilde{k}_e]$. If $\tilde{\sigma}_c = 1 - \tilde{k}_e$, then the global optimum is $\sigma_c^* = \frac{1}{2}$ since $F(1 - \tilde{k}_e) < \frac{3}{4}\theta = F(\frac{1}{2})$. Otherwise, a comparison between $F(\tilde{\sigma}_c)$ and $F(\frac{1}{2})$ determines the global optimum.

It is easy to verify that the second order derivative of F within the interval $[0, 1 - \tilde{k}_e]$ is negative because $\tilde{k}_e > \frac{1}{2}$ and $\sigma_c < \frac{1}{2}$. Consequently, the first order condition is sufficient and necessary for $\tilde{\sigma}_c$.

Let $F^*(w, \theta, \tilde{k}_e) = \max_{\sigma_c \in [0,1-\tilde{k}_e]} F$. By Berge's theorem, F^* is continuous in its arguments, in particular it is continuous in $\tilde{k}_e$. We will show below the existence of a $\tilde{k}_e$ for which F^* exceed $\frac{3}{4}\theta$ and another $\tilde{k}_e$ for which F^* is less than $\frac{3}{4}\theta$. This will allow us to apply the intermediate value theorem and conclude that there exists $\bar{k}_e$ such that $F^*(w, \theta, \bar{k}_e) = \frac{3}{4}\theta = F(\sigma_c = \frac{1}{2})$. This $\bar{k}_e$ is the threshold we desire because F^* is monotonic in $\tilde{k}_e$.

It is straightforward to see that for $\tilde{k}_e$ sufficiently low, $F^*(w, \theta, \tilde{k}_e) > \frac{3}{4}\theta = F(\sigma_c = \frac{1}{2})$. When $\tilde{k}_e$ sufficiently low, a positive probability of a centrist victory can be obtained with an infinitesimal loss of expressive benefit. Next, we would like to show that the reverse inequality holds for sufficiently high $\tilde{k}_e$. Note that the inequality $F^*(w, \theta, \tilde{k}_e) < \frac{3}{4}\theta$ is necessary and a sufficient condition for $\frac{1}{2}$ being optimum. Hence, a sufficient condition for the optimality of $\frac{1}{2}$ is:

$$\frac{3}{4}\theta \geq \max_{\sigma_c \in [0,1-\tilde{k}_e]} \left\{wP(\sigma_c)\right\} + \max_{\sigma_c \in [0,1-\tilde{k}_e]} \left\{\frac{\theta}{2}\left(\frac{1}{2} - \sigma_c^2 + \sigma_c\right)\right\} + \frac{3}{8}\theta$$

Because $\frac{1}{2} = \operatorname{argmax}_{\sigma_c} \frac{\theta}{2}\left(\frac{1}{2} - \sigma_c^2 + \sigma_c\right)$, $\max_{\sigma_c \in [0,1-\tilde{k}_e]} \left\{\frac{\theta}{2}\left(\sigma_c^2 - \sigma_c + \frac{1}{2}\right)\right\} = \frac{3}{8}\theta - \epsilon$ for some positive ϵ. Additionally, $\max_{\sigma_c \in [0,1-\tilde{k}_e]} \left\{wp(\sigma_c)\right\} = w(2 - 2\tilde{k}_e)$. Rewriting the sufficient condition, we have:

$$\frac{3}{4}\theta \geq w(2 - 2\tilde{k}_2) + \frac{3}{4}\theta - \epsilon \iff \tilde{k}_e \geq 1 - \frac{\epsilon}{2w}$$

Thus, for $\tilde{k}_e$ sufficiently high, $\frac{1}{2}$ is the optimum and equivalently, $F^*(w, \theta, \tilde{k}_e) < \frac{3}{4}\theta$. □

Corollary 1

Proof Recall from the proof of Proposition 1 that the threshold $\bar{k}_e$ is the solution to the equation

$$F^*(w, \theta, \bar{k}_e) = \frac{3}{4}\theta = F\left(\sigma_c = \frac{1}{2}\right). \tag{6}$$

where $F^*(w, \theta, \tilde{k}_e) = \max_{\sigma_c \in [0,1-\tilde{k}_e]} F$ is the maximum aggregate social welfare conditioned on positive levels of misaligned voting. Moreover, by the Berge's maximum theorem, $F^*(\cdot,\cdot,\cdot)$ is continuous in its arguments. Moreover, if we rewrite equality 6 as $F^*(w,\theta,\bar{k}_e) - \frac{3}{4}\theta = 0$, then by the implicit function theorem, we have that $\bar{k}_e$ is a continuous function of w and θ. Observe that by the envelop theorem, $F^*(w,\theta,\bar{k}_e) - \frac{3}{4}\theta$ is increasing in w , decreasing in θ and decreasing in $\tilde{k}_e$. it follows that $\bar{k}_e$ must be increasing in w and decrease in θ. Finally, note that the expression $\left(F^*(w,\theta,\bar{k}_e) - \frac{3}{4}\theta\right)\frac{1}{w}$, seen as a function of $\frac{\theta}{w}$, is decreasing in $\frac{\theta}{w}$. And therefore it follows that $\bar{k}_e$ must be decreasing in $\frac{\theta}{w}$. □

Proposition 2

Proof Note first that, $F(0) \geq F(\frac{1}{2}) \iff w\left(1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2}}\right) + \frac{\theta}{4} \geq \frac{3}{8}\theta \iff 1 - \frac{\theta}{16w} \geq \tilde{k}_e$. This implies that $1 - \frac{\theta}{16w} \geq \tilde{k}_e$ is sufficient for misaligned voting. Furthermore, $\frac{\partial F(0)}{\partial \sigma_c} = -\frac{w(\tilde{k}_e - \frac{1}{2})}{\frac{1}{4}} + \frac{\theta}{2} \leq 0 \iff \frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e$. This means that for $\frac{1}{2} + \frac{\theta}{8w} \leq \tilde{k}_e \leq \frac{\theta}{16w} \geq \tilde{k}_e, \sigma^* = 0$. □

Proposition 3

Proof We shall use the monotone comparative statics results from Milgrom and Shannon (1994): If the cross derivative of F with respect to the choice variable (i.e. σ_c) and the parameter of interest, and if the cross partial is positive (negative), then the solution is increasing (decreasing) in that parameter.

Observe that for voting rules $\sigma_c \in [0, 1 - \tilde{k}_e)$, the probability of centrist victory is $P(\sigma_c) = 1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - \sigma_c}$. Therefore the aggregate welfare (see expression (5)) is

$$F(\sigma_c) = w \cdot \left(1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - \sigma_c}\right) + \frac{\theta}{2}\left(\frac{1}{2} - \sigma_c^2 + \sigma_c\right) + \frac{3}{8}\theta.$$

From this, one can easily verify that $\frac{\partial^2 F}{\partial w \partial \sigma_c} < 0, \frac{\partial^2 F}{\partial \theta \partial \sigma_c} > 0$ and $\frac{\partial^2 F}{\partial \tilde{k}_e \partial \sigma_c} < 0$. □

Proposition 4

Proof The result with respect to w and θ follows immediate from Proposition 1 and 3 . We will focus instead on the comparative statics with respect to $\tilde{k}_e$. When $0 < \sigma_c^* < \frac{1}{2}$, the first order conditions characterizes the optimum and the following expression is obtained: $\sigma_c^* = \frac{1}{2} - \sqrt[3]{\frac{w}{\theta}\left(\tilde{k}_e - \frac{1}{2}\right)}$. Thus, we see that $\frac{\partial \sigma_c^*}{\partial \tilde{k}_e} = -\frac{w}{3\theta}\left(\frac{w}{\theta}(\tilde{k}_e - \frac{1}{2})\right)^{-\frac{2}{3}}$. From the expression for the probability of the center winning the election, we obtain:

$$\frac{\partial P(\sigma_c^*)}{\partial \tilde{k}_e} = -\frac{\frac{1}{2} - \sigma_c^* + (\tilde{k}_e - \frac{1}{2})\frac{\partial \sigma_c^*}{\partial \tilde{k}_e}}{(\frac{1}{2} - \sigma_c^*)^2}$$

Replacing our previous values for σ_c^* and $\frac{\partial \sigma_c^*}{\partial \tilde{k}_e}$, we obtain that the numerator is equal to $\frac{2}{3}\sqrt[3]{\frac{w}{\theta}\left(\tilde{k}_e - \frac{1}{2}\right)} > 0$. Since the denominator is also positive and the fraction is multiplied by -1, we get the probability of winning is decreasing when σ^* is interior. The results extends easily to the case where $\sigma^* = 0$ (refer to Proposition 1 and 2)

□

Proposition 5

Proof Let $F(\sigma_c)$ denote the expected social welfare (i.e., expression (2)). It can be shown that

$$F(\sigma_c) = \begin{cases} p(\sigma_c)\mathbb{E}(w_r) - \frac{\theta}{2}\left(\sigma_c^2 - \sigma_c + \frac{1}{2}\right) - \frac{\theta}{8} & \text{if } \sigma_c \le \frac{1}{2} \\ p(\sigma_c)\mathbb{E}(w_l) - \frac{\theta}{2}\left(\sigma_c^2 - \sigma_c + \frac{1}{2}\right) - \frac{\theta}{8} & \text{if } \sigma_c > \frac{1}{2} \end{cases}$$

where $\mathbb{E}(w_r)$ is the expectation of $w_r(\cdot)$ taken with respect to x (similarly for $\mathbb{E}(w_l)$). First, note that $F(\sigma_c)$ is not symmetric around $\frac{1}{2}$ unless $\mathbb{E}(w_r) = \mathbb{E}(w_l)$. It is the generically the case $\mathbb{E}(w_r) \neq \mathbb{E}(w_l)$. Now, clearly if $\mathbb{E}(w_r) > \mathbb{E}(w_l)$, then for any $\sigma_c > \frac{1}{2}$, $F(\sigma_c) < F(1 - \sigma_c)$ and therefore the ethical voting rule must be$\sigma_c^* \in [0, \frac{1}{2}]$. Similarly, if $\mathbb{E}(w_r) < \mathbb{E}(w_l)$, then it must be the case that $\sigma_c^* \in [\frac{1}{2}, 1]$. □

Proposition 6

Proof The ethical voting rule must minimize the aggregate opportunity cost of misaligned voting conditional on a certain probability of center-right victory. Hence, we first need to identify the opportunity cost of voting for the certain-right. If the voter would have abstained otherwise (i.e. $\max 1 - x, x - t \le 0$), the opportunity cost of voting for the center right is $t - x$. If the voter would have turnout and vote for the center-left (i.e. $1 - x - t > 0, x \le \frac{1}{2}$), the opportunity cost of switching his vote is $1 - 2x$. The aggregate opportunity cost will be minimized when $\max_{(x,t)\in R} t - x = \max_{(x,t)\in R} 1 - 2x$. Otherwise, the same probability of victory can be achieved at a lower cost by making abstainers head to the poll instead of center-left voters switching their votes to the center-right or viceversa.

Say that ethical voting rule $\sigma_c^* \in [0, \frac{1}{2}]$ requires all voters with $x \ge \sigma_c^*$ who are already casting a ballot for the center-left to switch their votes to the center-right. This means that $\max_{(x,t)\in R} 1 - 2x = 1 - 2\sigma_c^*$. As $\max_{(x,t)\in R} t - x = \max_{(x,t)\in R} 1 - 2x$, $\max_{(x,t)\in R} t - x = 1 - 2\sigma_c^*$. Hence, ethical voting rule σ_c^* requires abstainers with $t \le 1 - 2\sigma_c^* + x$ to vote for the center-right.

Finally, notice that $1-2\sigma_c^*+x=1-x$ when $x=\sigma_c^*$. This means that the restriction that the ethical voting rule imposes on center-left voter and that on abstainers intersect at the boundary that separates center-left voters from abstainers (i.e. $1-x-t=0$). Hence, we obtain that, under ethical voting rule $\sigma_c^* \in [0, \frac{1}{2}]$, voters of type (x,t)

1. vote for the center-right candidate if $\sigma_c^* \le x$ and $t \le 1-2\sigma_c^*+x$.
2. vote for the center-left if $x < \sigma_c^*$ and $t < 1-x$.
3. abstain otherwise.

□

Proposition 7

Proof First, we obtain the expression for the probability of victory for the centrist as a function of the ethical voting rule i.e., $P(\sigma_c)$. Note that for $\tau \le \frac{1-k_e}{2}$, the probability of centrist victory is 1. On the other hand, if $\sigma_c > 1-\frac{k_e}{1-k_e}$ and $\tau > (1-\sigma_c)(1-k_e)$, then we have the probability of victory being 0. Thus, we have that:

If $\sigma_c \le 1-\frac{k_e}{1-k_e} = 1-\tilde{k}_e$,

$$
\begin{aligned}
P(\sigma_c) &= \frac{1-k_e}{2k_e} + \int_{\frac{1-k_e}{2}}^{k_e}\left(1-\frac{\frac{2x}{1-k_e}-1}{1-2\sigma_c}\right)\frac{1}{k_e}dx \\
&= \frac{1-k_e}{2k_e} + \left.\frac{2(1-\sigma_c)\tau-\frac{\tau^2}{1-k_e}}{(1-2\sigma_c)k_e}\right|_{\frac{1-k_e}{2}}^{k_e} \\
&= 1+\frac{1-\tilde{k}_e}{1-2\sigma_c}-\frac{1}{4(1-2\sigma_c)\tilde{k}_e}
\end{aligned}
$$

If $\sigma_c > 1-\tilde{k}_e$,

$$
\begin{aligned}
P(\sigma_c) &= \frac{1-k_e}{2k_e} + \int_{\frac{1-k_e}{2}}^{(1-\sigma_c)(1-k_e)}\left(1-\frac{\frac{2x}{1-k_e}-1}{1-2\sigma_c}\right)\frac{1}{k_e}dx \\
&= \frac{1-k_e}{2k_e} + \frac{(1-\sigma_c)^2(1-k_e)}{1-2\sigma_c} - \frac{\left(\frac{3}{4}-\sigma_c\right)(1-k_e)}{(1-2\sigma_c)k_e} \\
&= \frac{(1-\sigma_c)^2}{\left(1-2\sigma_c\right)\tilde{k}_e} - \frac{1}{4(1-2\sigma_c)\tilde{k}_e}
\end{aligned}
$$

Recall $F(\sigma_c)$ is the ethical voter's objective function as defined in Eq. 1. With some algebra, one can show that

$$F'(\sigma_c) = \begin{cases} -\frac{(1-2\tilde{k}_e)^2}{2(1-2\sigma_c)^2\tilde{k}_e} + \frac{\theta}{w}(1-2\sigma_c) & \forall \sigma_c \leq 1-\tilde{k}_e \\ -\frac{1}{2\tilde{k}_e} + \frac{\theta}{w}(1-2\sigma_c) & \forall \sigma_c > 1-\tilde{k}_e \end{cases}$$

It can be verified that the second order condition is satisfied and so the FOC is sufficient and necessary the optimal ethical voting rule. In addition, the cross partial with respect to σ_c and $\tilde{k}_e$ satisfies

$$\frac{\partial^2 F}{\partial \tilde{k}_e \partial \sigma_c} = \begin{cases} \frac{1-4\tilde{k}_e^2}{2\tilde{k}_e^2(1-2\sigma_c)^2} < 0 & \forall \sigma_c \leq 1-\tilde{k}_e \\ \frac{1}{2y^2} > 0 & \forall \sigma_c > 1-\tilde{k}_e \end{cases}$$

Therefore, the objective function F is submodular in σ_c and $\tilde{k}_e$ in the range $\sigma_c \leq 1-\tilde{k}_e$ and is supermodular otherwise.

Now that F' is continuous at $1-\tilde{k}_e$ and $F'\left(1-\tilde{k}_e\right)$ is increasing in $\tilde{k}_e$. Moreover, if $\frac{\theta}{w} \leq \frac{1}{2}$, then $F'\left(1-\tilde{k}_e\right) < 0$. In this case, the optimal ethical rule satisfies $\sigma_c^* \leq 1-\tilde{k}_e$. We know from above that the objective function is submodular for $\sigma_c \leq 1-\tilde{k}_e$. Therefore by monotone comparative statics, we get that σ_c^* is always decreasing in $\tilde{k}_e$, i.e., there is always more misaligned voting in response to greater extremist popularity. Now, if $\frac{\theta}{w} > \frac{1}{2}$, then there exists some threshold $\bar{k}$ defined by $F'(1-\bar{k}) = 0$ such that if $\tilde{k}_e > (<)\bar{k}$, then $\sigma_c^* > (<)1-\tilde{k}_e$. Given our observation about the super/submodularity of the objective function and monotone comparative statics, it follows that for $\tilde{k}_e \leq \bar{k}$, we have that σ_c^* is decreasing in $\tilde{k}_e$ but for $\tilde{k}_e > \bar{k}$, σ_c^* is increasing. That is we get non-monotonicity in misaligned voting. □

Proposition 8

Proof We will argue that to achieve the same probability of victory, it is less costly for the ethical voters to misalign their votes for the center-right candidate than for the center-left candidate. To show this, it is sufficient to demonstrate that for any realizations of the fraction of ethical voters, q_c, the aggregate expressive benefit is higher conditional on the center-right candidate to attain vote share $y > G\left(\frac{1}{2}\right)$ than conditional on the center-left candidate to attain the same vote share. Let $q_c > 0$ be arbitrary, and let $K_l(y)$ and $K_r(y)$ denote the expressive benefit to the ethical voter for achieving vote share $y > G\left(\frac{1}{2}\right)$ for the center-left and center-right candidates, respectively. Let c_l and c_r be cutoffs that solve, respectively,

$$y = q_c G(c_l) + (1-q_c)G\left(\frac{1}{2}\right)$$
$$y = q_c\left(1-G(c_r)\right) + (1-q_c)\left(1-G\left(\frac{1}{2}\right)\right)$$

In other words, c_l and c_r are the ethical voting rules that achieves vote share y for the center-left and center-right candidates respectively. Now, based on the formulation of $B_E(\sigma_c)$, we can compute the marginal change to expressive benefit of an increase in vote share for the center-left candidate, which is

$$\frac{dK_l}{dy} = \frac{dK_l}{dc_l} \cdot \frac{dc_l}{dy} = \theta(1-2c_l)g(c_l) \cdot \frac{1}{q_c g(c_l)} = \frac{\theta}{q_c}(1-2c_l)$$

Similar calculations for the center-right candidates gets that $\frac{dK_r}{dy} = \frac{\theta}{q_c}(2c_r - 1)$. Now, by the premise of the Proposition, it must that that $c_l > \frac{1}{2}, c_r < \frac{1}{2}$, and $c_l - \frac{1}{2} > \frac{1}{2} - c_r$. It follows that $\frac{dK_l}{dy} < \frac{dK_r}{dy}$. Since this is true for all y, taking integral over y we get that $K_l(y) < K_r(y)$. And since this relationship holds for all realizations of q_c, it follows that there is greater expressive benefit associated with the center-right candidate achieving a certain winning probability than with the center-left candidate achieving the same winning probability. □

Expressions for $P(\sigma_c)$, $C_E(\sigma_c)$ and C_S under Turnout Costs

- $P(\sigma_C)$

  Given the results in Proposition 6, the fraction of ethical voters that vote for the center-right given σ_c is $1 - \sigma_c - \sigma_c^2$. Also, note that among non-ethical centrists, $\frac{3}{8}$ will vote for the center-right, $\frac{3}{8}$ for the center-left, and $\frac{1}{4}$ abstains. Hence, the center-right candidate wins the election if:

$$(1-q_c)\frac{3}{8} + q_c(1-\sigma_c-\sigma_c^2) \geq \tilde{k}_e$$
$$q_c \geq \frac{\tilde{k}_e - \frac{3}{8}}{\frac{5}{8} - \sigma_c - \sigma_c^2}$$

  Hence,

$$P(\sigma_C) = \max\left\{0, 1 - \frac{\tilde{k}_e - \frac{3}{8}}{\frac{5}{8} - \sigma_c - \sigma_c^2}\right\}$$

- $C_E(\sigma_c)$

$$C_E(L,R) = \int_{(x,t)\in L} \theta(1-x-t)dxdt + \int_{(x,t)\in R} \theta(x-t)dxdt$$
$$C_E(\sigma_c) = \theta\int_0^{\sigma_c}\int_0^{1-x}(1-x-t)dtdx + \theta\int_{\sigma_c}^1\int_0^{x+\sigma_c}(x-t)dtdx$$
$$= \theta\int_0^{\sigma_c}\frac{(1-x)^2}{2}dx + \theta\int_{\sigma_c}^1\frac{x^2-(1-2\sigma_c)^2}{2}dx$$
$$= \frac{\theta}{2}\left(\sigma_c - \sigma_c^2 + \frac{\sigma_c^3}{3}\right) + \frac{\theta}{2}\left(\frac{1}{3} - (1-2\sigma_c)^2 - \frac{\sigma_c^3}{3} - (1-2\sigma_c)^2\sigma_c\right)$$

- C_S

  Notice that $C_S = C_E(\frac{1}{2})$. Hence

$$C_S = \frac{\theta}{2}\left(\frac{1}{2} - \left(\frac{1}{2}\right)^2 + \frac{\left(\frac{1}{2}\right)^3}{3}\right) + \frac{\theta}{2}\left(\frac{1}{3} - (1 - 2\frac{1}{2})^2 - \frac{\left(\frac{1}{2}\right)^3}{3} - (1 - 2\frac{1}{2})^2 \frac{1}{2}\right)$$

$$= \frac{\theta}{2}\left(\frac{7}{24}\right) + \frac{\theta}{2}\left(\frac{7}{24}\right) = \frac{7}{24}\theta$$

Proposition 9

Proof First, note that if $\sigma_r = \frac{1}{2}$ (i.e., all center-right voters vote for the center-right), the center-left candidate never wins the election. Hence, the probability of a centrist victory is given by $P(\sigma_l, \frac{1}{2}) = 1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - \sigma_l}$. This expression is the same as the one in our baseline framework. Hence, the maximization problem faced by the ethical center-left voters is the same as the one centrists faced in our baseline model, which means that σ_c^* is a best-response for center-left ethical voters.

Now, suppose that $\sigma_l = \sigma_c^*$. We want to show that $\sigma_r^* = \frac{1}{2}$ is a best response for the center-right ethical voters. First, note that center-right voters would never adopt $\frac{1}{2} < \sigma_r \leq 1 - \sigma_l^*$ as $P(\sigma_l^*, \frac{1}{2}) \geq P(\sigma_l, \sigma_r)$. This is because the center-left candidate never wins when $\sigma_r \leq 1 - \sigma_l$ but the expressive benefit is lower given that $\frac{1}{2} < \sigma_r \leq 1 - \sigma_l^*$. If center-right voters choose $\sigma_r > 1 - \sigma_l^*$, we have that $G_r(\sigma_l^*, \sigma_r) < G_l(1 - \sigma_r, \frac{1}{2})$ as σ_r induces the same expressive benefit to center-right voters as $\sigma_l = 1 - \sigma_r$ to center-left voters, but $P(\sigma_l^*, \sigma_r) = 1 - \frac{\tilde{k}_e - \frac{1}{2}}{\sigma_l^* - (1 - \sigma_r)} < P(1 - \sigma_r, \frac{1}{2}) = 1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - (1 - \sigma_r)}$. Since $\sigma_l = \sigma_c^*$ is the unique best response for center-left voters when $\sigma_r = \frac{1}{2}$, we have $G_l(1 - \sigma_r, \frac{1}{2}) < G_l(\sigma_c^*, \frac{1}{2})$. Finally, note that $G_l(\sigma_c^*, \frac{1}{2}) \leq G_r(\sigma_c^*, \frac{1}{2})$ since both the center-right and center-left voters face the same probability of a centrist victory but center-left voters have a weakly lower expressive benefit as they are misaligning their vote. The discussion above means that $G_r(\sigma_l^*, \sigma_r) < G_l(1 - \sigma_r, \frac{1}{2}) < G_l(\sigma_c^*, \frac{1}{2}) \leq G_r(\sigma_c^*, \frac{1}{2})$ for all $\sigma_r > \frac{1}{2}$. Hence, $\sigma_r^* = \frac{1}{2}$ is a best response for center-right ethical voters. □

Proposition 10

Proof Suppose that $\sigma_r = \frac{1}{2}$, then the center-left never wins and so the probability of a centrist victory is $P(\sigma_l, \sigma_r) = 1 - \frac{\tilde{k}_e - \frac{1}{2}}{\frac{1}{2} - \sigma_l}$. Note that this is the same expression we had for our baseline framework. Hence, the best response for the center-left ethical voters, σ_l^* is equal to σ_c^* as in our baseline framework.

Now, suppose that $\sigma_l = \sigma_c^* = \frac{1}{2}$ (i.e., when it is optimal for the center-left to vote sincerely), it is not optimal for center-right voters to deviate from $\sigma_r = \frac{1}{2}$. In that case, center-right voters face the same problem as center-left voters, so their optimal strategy is to vote sincerely as well (i.e., $\sigma_r^* = \frac{1}{2}$).

Next, we consider the case when $\sigma_l = \sigma_c^* = 0$. Before proceeding, lets compute the probabilities of a centrist victory for when $\sigma_l \in [0, 1-\tilde{k}_e]$. For the center-left candidate to win, it must be the case that $q_l \le \frac{\frac{1}{2}-\tilde{k}_e+q_r(\sigma_r-\frac{1}{2})}{\frac{1}{2}-\sigma_l}$. This is means that the probability of a center-left victory, $P_l(\sigma_l, \sigma_r)$ is given by:

$$P_l(\sigma_l, \sigma_r) = \int_{\min\left\{\frac{\tilde{k}_e-\frac{1}{2}}{\sigma_r-\frac{1}{2}},1\right\}}^{1} \min\left\{\frac{\frac{1}{2}-\tilde{k}_e+q_r(\sigma_r-\frac{1}{2})}{\frac{1}{2}-\sigma_l}, 1\right\} dq_r$$

and after some tedious algebra, we have that

$$P_l(\sigma_l, \sigma_r) = \begin{cases} 0 & \text{if } \sigma_r \le \tilde{k}_e \\ \frac{\sigma_r-\frac{1}{2}}{2(\frac{1}{2}-\sigma_l)} - \frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l} + \frac{(\tilde{k}_e-\frac{1}{2})^2}{2(\frac{1}{2}-\sigma_l)(\sigma_r-\frac{1}{2})} & \text{if } \tilde{k}_e < \sigma_r < \tilde{k}_e + \frac{1}{2} - \sigma_l \\ 1 - \frac{\tilde{k}_e-\frac{\sigma_l}{2}-\frac{1}{4}}{\sigma_r-\frac{1}{2}} & \text{if } \sigma_r \ge \min\left\{\tilde{k}_e + \frac{1}{2} - \sigma_l, 1\right\} \end{cases}$$

The center-right candidate wins if $q_l > \frac{\tilde{k}_e-\frac{1}{2}+q_r(\sigma_r-\frac{1}{2})}{\frac{1}{2}-\sigma_l}$. Hence, the probability that the center-right wins, $P_r(\sigma_l, \sigma_r)$ is:

$$P_r(\sigma_l, \sigma_r) = \int_0^{\min\left\{\frac{1-\tilde{k}_e-\sigma_l}{\sigma_r-\frac{1}{2}},1\right\}} 1 - \frac{\tilde{k}_e - \frac{1}{2} + q_r(\sigma_r-\frac{1}{2})}{\frac{1}{2}-\sigma_l} dq_r$$

and after some tedious algebra, we have that

$$P_r(\sigma_l, \sigma_r) = \begin{cases} 1 - \frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l} - \frac{\sigma_r-\frac{1}{2}}{2(\frac{1}{2}-\sigma_l)} & \text{if } \sigma_r \le 1 - \tilde{k}_e + \frac{1}{2} - \sigma_l \\ \frac{(1-\tilde{k}_e-\sigma_l)^2}{2(\sigma_r-\frac{1}{2})(\frac{1}{2}-\sigma_l)} & \text{if } \sigma_r > 1 - \tilde{k}_e + \frac{1}{2} - \sigma_l \end{cases}$$

Note that the above implies that, given $\sigma_l \in [0, 1-\tilde{k}_e]$, center-right voters have no incentive to deviate from $\sigma_r = \frac{1}{2}$ to $\sigma_r \le \max\left\{\tilde{k}_e, 1-\tilde{k}_e+\frac{1}{2}-\sigma_l\right\}$. In particular, the center-right voters will not deviate to $\sigma_r \le \tilde{k}_e$ is trivial as, in that case there is no increase in the probability of a center-left victory from increasing σ_r (and there is a cost in terms of a lower probability of a center-right victory). If $1-\tilde{k}_e+\frac{1}{2}-\sigma_l > \tilde{k}_e$ and $\sigma_r \le 1-\tilde{k}_e+\frac{1}{2}-\sigma_l$, the probability of a centrist victory is $P(\sigma_l, \sigma_r) = P_l(\sigma_l, \sigma_r) + P(\sigma_l, \sigma_r) = 1 - 2\frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l} + \frac{(\tilde{k}_e-\frac{1}{2})^2}{2(\frac{1}{2}-\sigma_l)(\sigma_r-\frac{1}{2})}$. This is lower than $P(\sigma_l, \frac{1}{2}) = P_l(\sigma_l, \frac{1}{2}) + P(\sigma_l, \frac{1}{2}) = 1 - \frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l}$ as $-\frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l} + \frac{(\tilde{k}_e-\frac{1}{2})^2}{2(\frac{1}{2}-\sigma_l)(\sigma_r-\frac{1}{2})} < -\frac{\tilde{k}_e-\frac{1}{2}}{\frac{1}{2}-\sigma_l} + \frac{(\tilde{k}_e-\frac{1}{2})^2}{2(\frac{1}{2}-\sigma_l)(\tilde{k}_e-\frac{1}{2})} = -\frac{\tilde{k}_e-\frac{1}{2}}{2(\frac{1}{2}-\sigma_l)} < 0, \forall \sigma_r > \tilde{k}_e$.

Now, suppose that $\sigma_l = \sigma_c^* = 0$. In this case,

$$P_l(\sigma_l, \sigma_r) = \sigma_r - \frac{1}{2} - 2\left(\tilde{k}_e - \frac{1}{2}\right) + \frac{(\tilde{k}_e - \frac{1}{2})^2}{(\sigma_r - \frac{1}{2})}$$

and

$$P_r(\sigma_l, \sigma_r) = \begin{cases} 1 - 2\left(\tilde{k}_e - \frac{1}{2}\right) - \left(\sigma_r - \frac{1}{2}\right) & \text{if } \sigma_r \leq 1 - \tilde{k}_e + \frac{1}{2} \\ \frac{(1-\tilde{k}_e)^2}{\sigma_r - \frac{1}{2}} & \text{if } \sigma_r > 1 - \tilde{k}_e + \frac{1}{2} \end{cases}$$

As argued above, the center-right voters will not deviate from $\sigma_r = \frac{1}{2}$ to

$$\sigma_r \leq \max\left\{\tilde{k}_e, 1 - \tilde{k}_e + \frac{1}{2} - \sigma_l\right\}.$$

For $\sigma_r \geq 1 - \tilde{k}_e + \frac{1}{2}$, we have that $P(0, \sigma_r) = (\sigma_r - \frac{1}{2}) - 2(\tilde{k}_e - \frac{1}{2}) + \frac{(\tilde{k}_e - \frac{1}{2})^2}{\sigma_r - \frac{1}{2}} + \frac{(1-\tilde{k}_e)^2}{\sigma_r - \frac{1}{2}}$ $< (1 - \frac{1}{2}) - 2(\tilde{k}_e - \frac{1}{2}) + \frac{(\tilde{k}_e - \frac{1}{2})^2}{(1-\tilde{k}_e+\frac{1}{2})-\frac{1}{2}} + \frac{(1-\tilde{k}_e)^2}{(1-\tilde{k}_e+\frac{1}{2})-\frac{1}{2}} = 1 - 2(\tilde{k}_e - \frac{1}{2}) = P(0, \frac{1}{2})$. Thus the center right ethical voters would not find such a deviation profitable. Therefore, we have that $\sigma_r^* = \frac{1}{2}$ is a best response to $\sigma_l = 0$.

Now, notice that $P(\sigma_l, \sigma_r)$ is continuous in its arguments. Hence, for all δ, there is ϵ such that for $\sigma_l^* = \sigma_c^* < \epsilon$, $|P(\sigma_l^*, \sigma_r) - P(0, \sigma_r)| < \delta$, $\forall \sigma_r \in [\frac{1}{2}, 1]$. Since the only profitable deviations imply a decrease of at least $\int_{\frac{1}{2}}^{\sigma_r} \theta(1-x)dx + \int_{\sigma_r}^{1} \theta x dx - \frac{3}{8}\theta > 0$ in $G_r(\sigma_l, \sigma_r)$, there is $\epsilon > 0$ for which $\sigma_r^* = \frac{1}{2}$ and $\sigma_l^* < \epsilon$ is a consistent rule profile.

□

References

Blais A, Nadeau R (1996) Measuring strategic voting: a two-step procedure. Elect Stud 15(1):39–52

Blais A, Young R, Lapp M (2000) The calculus of voting: an empirical test. Eur J Polit Res 37(2):181–201

Bouton L (2013) A theory of strategic voting in runoff elections. Am Econ Rev 103(4):1248–88

Bouton L, Ogden BG (2018) Group-based voting in multicandidate elections. Available at: https://www.nber.org/papers/w23898.pdf. Accessed 15 July 2018

Cain BE (1978) Strategic voting in Britain. Am J Polit Sci 22(3):639–655

Coate S, Conlin M (2004) A group rule–utilitarian approach to voter turnout: theory and evidence. Am Econ Rev 94(5):1476–1504

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

Cox GW (1994) Strategic voting equilibria under the single nontransferable vote. Am Polit Sci Rev 88(03):608–621

Cox GW (1997) Making votes count: strategic coordination in the world's electoral systems (Political Economy of Institutions and Decisions). Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139174954

Degan A, Merlo A (2009) Do voters vote ideologically? J Econ Theory 144(5):1868–1894

DellaVigna S, List JA, Malmendier U, Rao G (2016) Voting to tell others. Rev Econ Stud 84(1):143–181

Feddersen T, Sandroni A (2006a) A theory of participation in elections. Am Econ Rev 96(4):1271–1282

Feddersen T, Sandroni A (2006b) Ethical voters and costly information acquisition. Q J Polit Sci 1(3):287–311
Feddersen TJ (1992) A voting model implying Duverger's law and positive turnout. Am J Polit Sci 36(4):938–962
Fey M (1997) Stability and coordination in Duverger's law: a formal model of preelection polls and strategic voting. Am Polit Sci Rev 91(01):135–147
Fisher SD (2000) Intuition versus formal theory: tactical voting in England 1987–1997. Available at http://www.nuff.ox.ac.uk/users/fisher/FisherLSE.pdf. Accessed 10 Jan 2018
Fujiwara T (2011) A regression discontinuity test of strategic voting and Duverger's law. Q J Polit Sci 6(3–4):197–233
Funk P (2010) Social incentives and voter turnout: evidence from the Swiss mail ballot system. J Eur Econ Assoc 8(5):1077–1103
Hamlin A, Jennings C (2011) Expressive political behaviour: foundations, scope and implications. Br J Polit Sci 41(3):645–670
Hamlin A, Jennings C (2019) Expressive voting. In: The oxford handbook of public choice, vol 1. Oxford University Press
Harsanyi JC (1977) Morality and the theory of rational behavior. Soc Res 44(4):623–656
Harsanyi JC (1992) Game and decision theoretic models in ethics. Handb Game Theory Econ Appl 1:669–707
Kiewiet DR (2013) The ecology of tactical voting in britain. J Elect Pub Opin Part 23(1):86–110
Myatt DP (2007) On the theory of strategic voting. Rev Econ Stud 74(1):255–281
Myerson RB (2002) Comparison of Scoring rules in poisson voting games. J Econ Theory 103(1):219–251
Myerson RB, Weber RJ (1993) A Theory of voting equilibria. Am Polit Sci Rev 87(01):102–114
Palfrey T (1989) A mathematical proof of Duverger's law. In: Ordeshook PC (ed) Models of strategic choice in politics. University of Michigan Press, Ann Arbor, pp 69–91
Piketty T (2000) Voting as communicating. Rev Econ Stud 67(1):169–191
Piolatto A, Schuett F (2015) Media competition and electoral politics. J Pub Econ 130:80–93
Pons V, Tricaud C (2018) Expressive voting and its cost: evidence from runoffs with two or three candidates. Econometrica 86(5):1621–1649
Spenkuch JL (2018) Expressive vs. strategic voters: an empirical assessment. J Pub Econ 165:73–81