WORKING PAPERS

N° 122

Does universalization ethics justify participation in large elections?

Ingela Alger, Konrad Dierks and Jean-François Laslier

Institute for Advanced Study in Toulouse

Does universalization ethics justify participation in large elections? *

Ingela Alger[†] Konrad Dierks[‡] Jean-François Laslier[§]

April 15, 2025

Abstract

We analyze the turnout decisions of ethical voters, equipped with (semi-)Kantian preferences: a voter considers the election outcome that would arise if other voters behaved like him. The "others" can be limited to co-partisans ("partisan ethics") or not ("non-partisan ethics"). In a standard model with two candidates, a known underdog, a continuum of voters, and a continuous power-sharing rule, we introduce two novel elements: core constituent groups, and distinct election stakes for the two partisan groups. Under partisan ethics, when an equilibrium exists, turnout is positive for both sides if the election is not of the winner-take-all kind. Under non-partisan ethics an equilibrium always exists and turnout is positive for one side only. There sometimes exist equilibria where the underdog wins. Moreover, multiple equilibria sometimes arise, possibly with different winners. Voters further face a coordination problem under equilibrium multiplicity in the non-partisan case.

Keywords: voter turnout, ethical voter, universalization, *Homo moralis*, Kantian morality

*I.A. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 789111 - ERC EvolvingEconomics), as well as IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program). J.-F. L. thanks the EUR grant ANR-17-EURE-0001. K.D. thanks the EUR grant ANR-18-EURE-0021.

[†]CNRS, Toulouse School of Economics, Institute for Advanced Study in Toulouse. ingela.alger@tse-fr.eu

[‡]PhD candidate, Toulouse School of Economics, University of Toulouse Capitole. konrad.dierks@tse-fr.eu

[§]CNRS, Paris School of Economics. jean-francois.laslier@ens.fr

1 Introduction

1.1 The turn-out question

Election outcomes depend not only on voters' political preferences, but also on which voters actually vote. To wit, in the 20th century in the U.S.A., turnout rates in the three groups of registered voters (Democrats, Republicans, Independents) shifted over time (DeNardo, 1980; Nagel and McNulty, 1996). So turnout matters, and it has been found to vary not only over time but also across countries, and to be correlated with macro-economic factors, the type of election, and even the weather on the election day, to name just a few of the variables that have been examined (see, e.g., Blais and Daoust, 2020; Cancela and Geys, 2016; Frank and Martínez i Coma, 2023, and references therein).

Understanding these patterns requires understanding the individual turnout decision-making process. Perceived benefits and costs of voting at the individual level are at the center of any rational voter theory of turnout (Downs, 1957). Voters driven by purely instrumental concerns are expected to incur a cost to participate only if they can expect to be pivotal (Krishna and Morgan, 2015; Ledyard, 1984; Myerson, 2000; Palfrey and Rosenthal, 1985), but a variety of other benefits have been invoked to explain turnout, such as a desire to express allegiance to the political system, to participate in the democratic process, to express an opinion, to affirm loyalty to a party, to fulfill one's duty, or to comply with a social norm (Riker and Ordeshook, 1968; Fiorina, 1976; Morton, 1987; Schuessler, 2000; Feddersen and Sandroni, 2006; Coate and Conlin, 2004; Feddersen, Gailmard, and Sandroni, 2009; D. K. Levine and Mattozzi, 2020). Several of these have been shown to matter empirically (Blais, 2000; Blais and Achen, 2019; Gerber, Green, and Larimer, 2008; Rogers, Green, Ternovski, and Young, 2017).[1] We propose a theoretical model where turnout is driven by ethical concerns.

1.2 The ethics of (partial) universalization

Our inquiry is about the moral justification of turnout. The driving force is a form of ethical preferences—dubbed *Homo moralis*—which capture the principle of universalization: when contemplating a course of action, a *Homo moralis* evaluates what his material payoff would be if, hypothetically, a share κ of the population to which he belongs would follow the same course of action, where κ is the individual's *degree of universalization*. One obtains the standard materialistic *Homo oeoconomicus* for $\kappa = 0$ and the Kantian model of Laffont

[1] See also the survey by Dhillon and Peralta (2002) and the literature discussion in Coate and Conlin (2004), as well as the book by Aytaç and Stokes (2019).

(1975) for $\kappa = 1$, while values of κ between 0 and 1 trigger partial universalization (Alger and Weibull, 2013).

Universalization concerns have roots in traditional philosophy: the contemplation of counterfactual situations that *Homo moralis* engages in is reminiscent of Kant's categorical imperative (Kant, 1785), although philosophers warn that a Kantian "maxim" is not a course of action (Braham and van Hees, 2020). In any case, the logic of universalization spreads over moral theories (Gravel, Laslier, and Trannoy, 2000), and *Homo moralis* preferences in particular have been derived in an axiomatic model of universalization concerns (Salonia, 2024). Moreover, *Homo moralis* preferences have an evolutionary foundation (see Alger and Weibull, 2013 and Alger, Weibull, and Lehmann, 2020)[2], and experimental evidence supports the relevance of universalization concerns for decisions in social dilemmas (Capraro and Rand (2018), Tappin and Capraro (2018), Miettinen, Kosfeld, Fehr, and Weibull (2020), Levine, Kleiman-Weber, Schultz, and Cushman (2020), Van Leeuwen and Alger (2024).

In a society where power is shared based on election outcomes, an individual's decision to vote or not affects others. For the part of the electorate for which the cost of voting matters, the situation thus has the characteristics of a social dilemma. The consequences of *Homo moralis* preferences have been investigated in a host of other social dilemmas, such as two-player coordination, prisoner's dilemma, and hawk-dove games (Bergstrom, 1995; Alger and Weibull, 2013); and n-player public goods games (Alger and Weibull, 2017).[3] More closely related to our model are the studies that take interest in interactions between individuals in large populations. These studies have examined consumption with externalities (Laffont, 1975, Eichner and Pethig, 2022) as well as tax compliance (Muñoz-Sobrado, 2022). This research shows that *Homo moralis* preferences tend to help solve the social dilemmas in question, in the sense that they promote pro-social behavior. This is also the case in the two classical voting issues analyzed by Alger and Laslier (2022): *Homo moralis* preferences are seen there to help a divided majority to coordinate on one candidate, and to foster information aggregation in a large population (thus restoring the Condorcet Jury Theorem). In these models the turnout issue is, however, fully neglected. This is also true in the model by Eichner and Pethig (2021) where individuals with a variant of *Homo moralis* preferences vote over climate policies.

Hence the question which is at the heart of this paper: do *Homo moralis* preferences

[2]Our approach is thus close in spirit to Conley, Toossi, and Wooders (2006), who base their voters' motivation to participate in elections on evolutionary arguments. See also the book by Hatemi and McDermott (2011), which *inter alia* cites evidence of intriguing correlations between biological factors such as genes on the one hand, and political preferences and even turnout on the other hand.

[3]Issues such as contracting with teams (Sarkisian, 2017, 2021,Mohanty, Rao, and Roy, 2024), bilateral trade (Norman, 2020), and bargaining (Juan-Bartroli and Karagözoğlu, 2024) have also been analyzed.

motivate voters to participate in large elections? Our analysis reveals that this is not necessarily the case. Before summarizing our findings (see Subsection 1.6), we need to introduce some further elements.

1.3 The two-party model of costly participation

We evaluate the consequences of *Homo moralis* preferences in a standard political model with three key elements, two of which are novel in the theoretical literature on turnout. As in most models of turnout, there are two candidates (or parties, or referendum proposals), A and B. We take B to be the (known) underdog, in the sense that it has the weakest support in the whole electorate. Some voters always turn out to vote, perhaps because of a deep sense of duty, a long-held habit, a strong wish to signal support of democracy, etc. They are the core voters (DeNardo, 1980), or the candidate's *base* (this is the first novel element). Our analysis concerns the other voters, who do *not* systematically turn out to vote. Their cost of voting is uncertain at the individual level, although the cost distributions are known. Each such *cost-sensitive voter* decides on a threshold strategy: she votes if and only if the realized cost falls short of this threshold.

The distribution of expressed votes across the two candidates determines the political outcome. At this level, we keep the familiar zero-sum pattern of two-party electoral competition: the outcome is positive for one side and negative for the other. But it need not be a winner-take-all election. Our political outcome gain-loss function also encompasses institutional settings where the vote share itself matters and political power is shared so that "the loser gets some". This second key element captures a range of possible institutional settings, from the pure majoritarian case (obtained as a limit case of our model) to a kind of "random dictatorship", or "proportional two-party system".[4]

At the level of the voters we break the symmetry and introduce a parameter that we call *the stake of the election*. As a first interpretation, this parameter represents the importance of the political benefit obtained through the election as perceived by the underdog's supporters relative to the other group's supporters (this is the third key element of the political model, and the second novel element). As an illustrating example, if the underdog tends to represent low-income households, the stake is expected to be higher, the stronger are the redistributive consequences of the election. As a second interpretation, the parameter captures the electorate's trust in policy-makers. For example, if the underdog represents a part of population that has lost trust in the government's willingness to implement policies

[4]Modifying this way the political benefit function in a two-party model in order to contrast proportional representation with winner-take-all is used, for instance, by Lizzeri and Persico (2001). Studying turnout, Herrera et al. (2016) similarly modify the outcome function.

that would favor them, the value of the stake parameter would be low. To summarize, the material consequences of the election for each individual depend on the political outcome, the stake, and the individual's cost (that may be incurred or not).

1.4 Partisan vs. non-partisan morality

By definition, universalization ethics implies a reference to a group, the population to which the individual belongs. Traditional discussion of morality takes a "universal" point of view, taking the whole human population as the reference group (Rousseau, 1762; Kant, 1785 ; Rawls, 1971; Harsanyi, 1977; see Gibbard, 1990 and Hausman and McPherson, 1996). But, as pointed out by Laffont (1975): "heterogeneity of people creates great difficulties for the definition of modes of Kantian behaviour" and Alger and Laslier (2022), in their analysis of vote splitting and the coordination on close candidates, are led to distinguish between two possibilities for a voter's reference group. Here we will also examine two settings: the *partisan* setting and the *non-partisan* one.

In the non-partisan setting, the reference group can be interpreted as the set of all (cost-sensitive) voters, while in the partisan setting there are two distinct populations, one for each candidate. More precisely, in the partisan setting, the voter applies the universalization argument to his set of co-partisans, by evaluating what the outcome of the election would be if—hypothetically—a share κ of his co-partisans were to choose the same threshold as the voter himself.[5] By contrast, in the non-partisan setting, the voter applies the universalization argument to the set of cost-sensitive voters, and chooses two cost thresholds rather than one, by taking into account the expected benefits and costs over the two possible preference realizations, "behind a veil of ignorance".

Our objective is to characterize rational behavior in these (one- or two-) population games, assuming that all voters have the same degree of universalization κ. The rich setting enables us to address a host of questions: is turnout positive in equilibrium, and how do turnout rates depend on the primitives (the way in which the relative margins affect the material benefits accruing to the parties, the candidates' bases, the stake of the election)? Are there equilibria in which the underdog wins the election? Finally, can there be failure of equilibrium existence, and can there be multiple equilibria?

[5] We refer the reader to Alger and Laslier, 2022 for arguments showing that this is a natural extension of the 2-player model of *Homo moralis* by Alger and Weibull, 2013 to a continuum population.

1.5 Related literature

Prior to summarizing our findings, we compare our formalization of ethically driven voters to existing ones. An early formalization (Harsanyi, 1980) of ethical voters posits that voters are *rule utilitarians*: in Harsanyi's words, such a voter does not look at the various issues from a partisan point of view but from the standpoint of an impartial but humane and sympathetic observer.[6] Furthermore, Harsanyi defines the moral behavior of rule-utilitarian individuals as "involving a firm commitment [...] to a specific moral strategy" (p.115 in Harsanyi, 1980), where the moral strategy maximizes the sum of individual utilities. Our non-partisan setting is in line with Harsanyi's view of voters as "impartial observers", but our formalization of ethically driven voters does not amount to altruistic utilitarianism. Instead, it amounts to a self-centered universalization thought experiment: a voter considers each course of action in the light of what *her* material well-being would be if some share of the others voters were to choose the same course of action.

Moreover, our formalization of ethical voters should not be confused with group-based voter participation models in which strategic decisions are made at the collective level and an ethical voter applies a decision rule – that is a cost threshold, like in our model – which maximizes the group's aggregate material well-being, given the other group's cost threshold (Coate and Conlin, 2004; Feddersen and Sandroni, 2006; Herrera, Morelli, and Nunnari, 2016). By adopting this dutiful behavior, such a voter receives a constant payoff $D > 0$. For each group, the equilibrium cost threshold optimally trades off the probability of winning against the group-aggregate expected cost of voting, given the other group's threshold.

By contrast, in our model a voter's utility depends directly on the material benefit she would enjoy if some share of the others also applied the same cost threshold, and she considers only her own (expected) cost when evaluating cost thresholds. In general, the *Homo moralis* model departs from standard material rationality but is rational in the sense that:

- Individual behavior is obtained by the maximization of a well-defined personal objective function. (This contrasts with group-based models.)[7]
- This objective is individualistic; the individual only cares about her personal payoff. (This contrasts with altruistic motivations.)[8]

[6]"In any social situation, each participant will tend to look at the various issues from his own, self-centered, partisan point of view. In contrast, if anybody wants to assert the situation from a *moral* point of view in terms of some standard of justice and equity, this will essentially amount to looking at it from the the standpoint of an impartial but humane and sympathetic observer." (p. 623 in Harsanyi, 1977.)

[7]We refer to subsection 3.5 for a more detailed comparison with group-based models.

[8]Alger and Weibull (2017) and Laslier (2023) study the relation between (partial) universalization ethics and (partial) Beckerian altruism.

- The beliefs about the other players are correct at equilibrium. (This contrasts with magical thinking.)[9]

To sum up, in our model each voter will be seen to maximize her own well-defined utility function, and, at equilibrium, the beliefs about the other voters' behaviors are correct. No voter believes that she is pivotal and voting is costly.

1.6 Overview of results

The universalization thought experiment entailed by *Homo moralis* preferences makes each voter act *as if* his decision had a real weight on the vote shares, and thus a real weight on the associated benefit. When considering an increase in his threshold strategy, a voter thus compares this hypothetical marginal benefit with the increase in the expected cost. Since we posit a cost distribution which implies an infinitesimal increase in the expected cost when the threshold is close to zero, this generates a willingness to incur a cost to participate, as long as the marginal benefit is positive. This willingness to participate is, however, not sufficient to guarantee costly *equilibrium* turnout.

In this respect, a striking difference appears between the majoritarian, winner-take-all setting, and the settings with some power sharing. In the winner-take-all setting, if an equilibrium exists, it is in most cases such that participation is nil in both groups. Such an equilibrium can be sustained when the cost necessary to obtain an increase in the benefit (from loss to gain for the supporters with the smallest base) is too large; conversely, an equilibrium with positive turnouts by both groups cannot be sustained in the winner-take-all case, because at such an equilibrium there must be a tie, which implies that the additional cost necessary to obtain a jump in the benefit (from tie to gain) is close to zero. The only winner-take-all setting where there exists an equilibrium with positive turnout is the non-partisan case, where only one group turns out: this group internalizes the externality imposed on the other group, and hence an equilibrium turnout yielding a tie can be sustained.

By contrast, under power sharing, we find that for any positive degree of universalization κ, in any equilibrium aggregate turnout is strictly positive (except in one knife-edge case in the nonpartisan setting). This is because the strictly increasing power-sharing rule makes any change in the vote shares consequential. A qualitative difference appears between the partisan and the non-partisan setting, however: in the latter, the voters take into account the effects of both cost thresholds on the expected utility, thereby internalizing the externalities they generate across the two groups of cost-sensitive voters. We show that as a result the

[9]See Daley and Sadowski (2017), who propose a model in which individuals falsely believe that their actions influence others' actions. Also, unlike Roemer (2010) we do not need to adopt an equilibrium notion which differs from standard Nash equilibrium.

only candidate that obtains a turnout among its cost-sensitive supporters is that with the highest expected net benefit from voting. By contrast, in the partisan setting there is no internalization of externalities, and the cost-sensitive voters of both groups incur a positive expected voting cost in any equilibrium.

In the power-sharing settings, the core constituents, or bases, as well as the stake of the election, are key determinants of the set of equilibrium outcomes. In particular, they matter for whether the underdog can win the election (in the sense that he gets more power than the leader). First, if the underdog's base exceeds that of the leader, there may exist equilibria in which the underdog wins. This result is due to the cost advantage that a large base confers on the cost-sensitive voters: the base enables them to reach higher turnout levels at a lower cost. This contrasts sharply with the results in other models with known underdogs (Feddersen and Sandroni, 2006; Herrera et al., 2016), where the underdog gets the smallest expected vote share in the unique equilibrium. Second, if the stake of the election outcome is higher for underdog than for leader supporters, the underdog supporters can be motivated to turn out at a high enough rate to win the election, even if their base falls short of the leader's base, and even if the leader supporters enjoy a favorable cost distribution compared to the underdog supporters.

By contrast to most of the literature, we do not impose assumptions guaranteeing equilibrium existence and uniqueness. In some games where players have *Homo moralis* preferences, equilibrium existence issues arise (Bomze, Schachinger, and Weibull, 2021). In the present case, we show existence in the non-partisan setting, and we observe with examples that existence sometimes fails in the partisan setting. On the other hand, multiple equilibria are not uncommon, in both settings.[10]

An important consequence of multiple equilibria in the non-partisan setting, where only one group incurs a cost to participate, is that it represents a coordination problem for the voters, who typically prefer one equilibrium over the other(s). Such coordination problems imply that voters are *more* likely to incur a cost to vote, the *higher* is the share of co-partisans that they believe will turn out to vote. The reason is the same as the one evoked above: a large turnout among the other cost-sensitive co-partisans implies that any individual voter can envisage reaching higher (counterfactual) turnout levels at a lower cost.

Interestingly, when there are multiple equilibria, these sometimes give rise to different election winners. This further points to the inability of the voters to easily coordinate their turnout decisions (a coordination problem which is solved by assumption in the group-based

[10]To gauge how common non-existence and multiplicity is, we provide illustrating examples, and an online tool that enables the reader to explore other parameter sets (available at: https://KonradEcon.github.io/homo-moralis-turnout-appendix/). There is also an appendix, where we provide further examples and also derive sufficient conditions for existence and uniqueness.

voter participation models, referred to above).

In the next section we describe the political model, and in the following two sections we analyze the partisan and the non-partisan settings, assuming that voters have *Homo moralis* preferences. A final section provides a discussion.

2 The political model

2.1 Institutional setting and political outcome

An election is taking place with two candidates, A and B.The electorate is formalized as a continuum, divided in two groups: a group of size $\bar{a}$ supporting A, and a group of size $\bar{b} < \bar{a}$ supporting B. Since the B-supporters are less numerous, we will refer to their group as the *underdog supporters*, and the group of A-supporters as the *leader supporters*.

Each voter either votes for their preferred candidate or abstains, and candidates A and B receive $a \leq \bar{a}$ and $b \leq \bar{b}$ votes, respectively, generating the relative margins α and β:

$$\alpha = \frac{a-b}{a+b} \qquad \text{and} \qquad \beta = -\alpha = \frac{b-a}{a+b}. \tag{1}$$

The election outcome generates some material (instrumental) benefit to the voters. This benefit is given by a strictly increasing and twice differentiable function $h : [-1, 1] \to \mathbb{R}$ of the relative margin of one's candidate. We assume that h is symmetric around 0, i.e., $h(-x) = -h(x)$ and $h'(x) = h'(-x)$, and that $h''(x) \leq 0$ for all $x > 0$. Moreover, we normalize $h(1) = -h(-1) = 1$. The competition between the two candidates is thus zero-sum, and the marginal impact on the material benefit is the largest at $x = 0$, the threshold value of x above which the candidate wins the election by securing a greater total turnout than the other candidate. In particular, our setting includes functions for which the slope of h at 0 is arbitrarily large and close to 0 elsewhere; this limit case of our model approximates the classical winner-take-all setting. However, by including h-functions such that the slope is sizeable everywhere, our model also encompasses situations where voters care about the margin of victory: this assumption is natural for parliamentary elections, where margins determine the number of seats obtained. The other limit case, opposed to the winner-take-all setting, consists in taking h to be linear. A possible interpretation is that decisions will be taken by one side or the other, yielding outcomes +1 and -1, with a probability equal to the proportion of votes obtained by A and B. This has a flavor of proportional representation and could be called "random dictatorship among participants".

We further assume that there is a parameter $\rho > 0$, which we call the *stake of the election*: the material benefit is $h(\alpha)$ for the A-supporters and $\rho h(\beta)$ for the B-supporters.

The election is more important for the underdog supporters than for the leader supporters if $\rho > 1$, while the opposite is true if $\rho < 1$. This parameter can also be interpreted as the voters' trust: for example, $\rho < 1$ could mean that the B-supporters do not fully trust that the power obtained by B will be used to implement the promised policies. We will say that the stake is neutral if $\rho = 1$.

In the numerical examples we will use the following specification for h:

$$h(x) = \frac{\arctan(mx)}{\arctan(m)}. \tag{2}$$

The parameter $m \in \mathbb{R}_{>0}$ changes the slope of h: the larger is m, the larger is the marginal benefit for small margins and the smaller is the marginal benefit for large margins. The linear case is obtained for $m \to 0$ and the step function for $m \to \infty$. See Figure 1.[11]

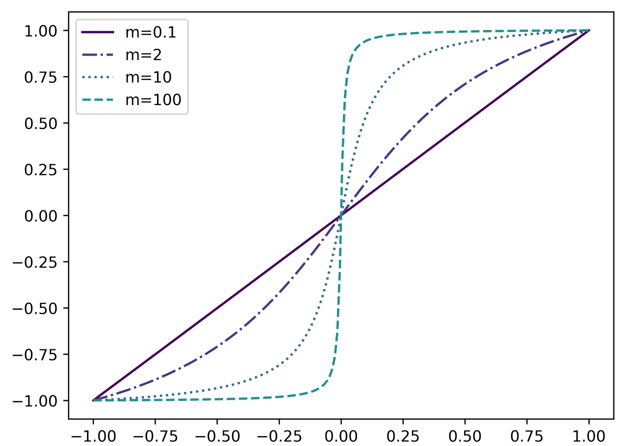


Figure 1: h as defined in equation (2) for different values of m

2.2 Voting costs and strategies

Some voters always turn out to vote—they may be driven by a strong sense of civic duty, a strong social pressure, a habit, or any other motivation outside of this model. There is a mass $0 < a_0 < \bar{b} < \bar{a}$ of such voters who vote for A, and a mass $0 < b_0 < \bar{b}$ of such voters who vote for B. We will refer to a_0 as *A's base* and to b_0 as *B's base*. The model examines the behavior of the remaining voters, a mass $a_v = \bar{a} - a_0$ of which are A-supporters, and a mass $b_v = \bar{b} - b_0$ of which are B-supporters. These voters are cost-sensitive: each of them faces an almost surely positive random cost of voting, and their turnout decision depends on their realized cost. Formally, the function $f_A : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ maps each cost to the probability density for

[11] This gives m a similar role as the power-sharing parameter γ in the benefit term in the model of Herrera et al. (2016). In Appendix B we check that their political outcome function is a particular case of ours.

a cost-sensitive A-supporter to have that voting cost. We assume that the support of f_A, i.e. $\{x \in \mathbb{R}_{\geq 0} : f_A(x) > 0\}$, is either an interval $[0, \bar{c}]$, for some $\bar{c} \in \mathbb{R}_{>0}$, or $\mathbb{R}_{\geq 0}$; and f_A is continuous on its support. We denote by $F_A(c)$ the proportion of cost-sensitive A-supporters whose cost realization falls short of c:

$$F_A(c) = \int_0^c f_A(t)\, dt, \tag{3}$$

so that $1 = F(+\infty)$. The same assumptions apply to group B, with notation f_B, and F_B.

We restrict attention to threshold strategies. A cost-sensitive A-supporter i picks a threshold $s_A^i \in \mathbb{R}_{\geq 0} \cup \{\infty\}$, and votes (for A) if the cost realization c_A^i does not exceed s_A^i and abstains otherwise. Likewise, a cost-sensitive B-supporter j picks a threshold $s_B^j \in \mathbb{R}_{\geq 0} \cup \{\infty\}$. Voters have correct beliefs about the voting cost distributions.

We also restrict attention to type-homogenous strategy profiles, in which all voters with the same preference over the candidates choose the same strategy. At a type-homogenous strategy profile $s = (s_A, s_B)$, the realized turnouts are

$$a(s_A) = a_0 + a_v F_A(s_A) \qquad \text{and} \qquad b(s_B) = b_0 + b_v F_B(s_B), \tag{4}$$

and the following relative vote margins obtain:

$$\alpha(s) = \frac{a(s_A) - b(s_B)}{a(s_A) + b(s_B)} \qquad \text{and} \qquad \beta(s) = -\alpha(s). \tag{5}$$

3 Partisan ethics (the *ex post* setting)

Under partisan ethics, the reference group is taken to be the other cost-sensitive voters who have the same preferences over the two candidates; the participation strategy is decided *ex post*, once the voter's affiliation is known. Thus, each voter i in group A (resp. each voter j in group B) evaluates any strategy s_A^i (resp. s_B^j) in the light of the material benefit that would realize if—hypothetically—a fraction $\kappa \in [0, 1]$ of the other cost-sensitive A-supporters (resp. B-supporters) were also to play s_A^i (resp. s_B^j) instead of the strategies they are actually using.

3.1 Payoff computations

At a type-homogenous strategy profile $s = (s_A, s_B)$, each A-supporter, respectively B-supporter, obtains expected net material benefit

$$EU_A(s) = h(\alpha(s)) - \int_{c=0}^{s_A} c f_A(c)\, dc \tag{6}$$

$$EU_B(s) = \rho h(\beta(s)) - \int_{c=0}^{s_B} c f_B(c)\; dc. \tag{7}$$

Homo moralis preferences induce A-supporter i to consider the expected net material benefit he would obtain if, hypothetically, a share κ of the other A-supporters were to use the same strategy as his own. He thus considers the hypothetical number of votes

$$a^\kappa(s_A, s_A^i) = a_0 + (1-\kappa)a_v F_A(s_A) + \kappa a_v F_A(s_A^i) \tag{8}$$

in favor of A, with the corresponding hypothetical relative vote margin

$$\alpha^\kappa(s, s_A^i) = \frac{a^\kappa(s_A, s_A^i) - b(s_B)}{a^\kappa(s_A, s_A^i) + b(s_B)}, \tag{9}$$

and this defines the voter's expected utility

$$EU_A^\kappa(s, s_A^i) = h(\alpha^\kappa(s, s_A^i)) - \int_{c=0}^{s_A^i} c f_A(c)\, dc. \tag{10}$$

Likewise, each B-supporter j considers the hypothetical number of votes

$$b^\kappa(s_B, s_B^j) = b_0 + (1-\kappa)b_v F_B(s_B) + \kappa b_v F_B(s_B^j) \tag{11}$$

in favor of B, with the corresponding relative vote margin

$$\beta^\kappa(s, s_B^j) = \frac{b^\kappa(s_B, s_B^j) - a(s_A)}{a(s_A) + b^\kappa(s_B, s_B^j)}, \tag{12}$$

and obtains expected utility

$$EU_B^\kappa(s, s_B^j) = \rho h(\beta^\kappa(s, s_B^j)) - \int_{c=0}^{s_B^j} c f_B(c)\, dc. \tag{13}$$

These equations reveal the main driver of the behavior of voters under partisan ethics in our model. Consider Equation 10. When $\kappa > 0$ an increase in s_A^i gives the voter a utility kick from knowing that if a share κ of his co-partisans were to adopt the strategy s_A^i as well, the vote share in favor of A, and thereby also his own benefit h, would increase.

Before going further, we proceed to a change of variables that simplifies the analysis. Seeing from (4) that the threshold s_A that yields turnout a is $F_A^{-1}\left(\frac{a-a_0}{a_v}\right) \in [0, \infty]$ and the threshold s_B that yields turnout b is $F_B^{-1}\left(\frac{b-b_0}{b_v}\right) \in [0, \infty]$, we write the expected utilities in

(10) and (13) as follows:

$$EU_A\left(a,b,a^i\right)=h\left(\alpha^\kappa\left(a,b,a^i\right)\right)-C_A\left(a^i\right) \tag{14}$$

$$EU_B\left(a,b,b^j\right)=\rho h\left(\beta^\kappa\left(a,b,b^j\right)\right)-C_B\left(b^j\right), \tag{15}$$

where

$$\alpha^\kappa\left(a,b,a^i\right)=\frac{(1-\kappa)\,a+\kappa a^i-b}{(1-\kappa)\,a+\kappa a^i+b} \qquad \text{and} \qquad \beta^\kappa\left(a,b,b^j\right)=\frac{(1-\kappa)\,b+\kappa b^j-a}{(1-\kappa)\,b+\kappa b^i+a} \tag{16}$$

$$C_A\left(a^i\right)=\int_0^{F_A^{-1}\left(\frac{a^i-a_0}{a_v}\right)} cf_A\left(c\right)dc \qquad \text{and} \qquad C_B\left(b^j\right)=\int_0^{F_B^{-1}\left(\frac{b^j-b_0}{b_v}\right)} cf_B\left(c\right)dc. \tag{17}$$

Henceforth, the strategy of A-supporter i is thus a "turnout" $a^i\in[a_0,\bar{a}]$, and that of B-supporter j a "turnout" $b^j\in[b_0,\bar{b}]$, although it should be clear to the reader that what these voters are really choosing are the cost thresholds that would yield these turnout levels. This change of variables facilitates analysis because the functions C_A and C_B are strictly convex, for any cost distributions F_A and F_B satisfying our assumptions.

Lemma 1. *Both C_A and C_B are strictly convex and strictly increasing.*

Proof. By a substitution $z=F_A\left(c\right)$

$$C_A\left(a\right)=\int_0^{\frac{a-a_0}{a_v}} F_A^{-1}\left(z\right)dz. \tag{18}$$

Then,

$$C_A'\left(a\right)=\frac{1}{a_v}F_A^{-1}\left(\frac{a-a_0}{a_v}\right)>0 \tag{19}$$

for $a>a_0$, with $C_A'\left(a_0\right)=0$, and

$$C_A''\left(a\right)=\frac{1}{a_v^2}\frac{1}{f_A\left(F_A^{-1}\left(\frac{a-a_0}{a_v}\right)\right)}>0 \tag{20}$$

for $a\geq a_0$. The same argument applies to C_B. □

In the numerical examples we use the functions

$$C_A\left(a\right)=\tfrac{\theta_A}{2}\left(\tfrac{a-a_0}{a_v}\right)^2 \qquad \text{and} \qquad C_B\left(b\right)=\tfrac{\theta_B}{2}\left(\tfrac{b-b_0}{b_v}\right)^2, \tag{21}$$

which correspond to a uniformly distributed cost on $[0, \theta_A]$ for the A-supporters and a uniformly distributed cost on $[0, \theta_B]$ for the B-supporters.

With this notation, at a type-homogenous Nash equilibrium a first requirement is that each A-supporter i best-responds not only to the B-supporters' strategy but also to the other A-supporters' strategy. Any such A-*consistent strategy* satisfies the fixed-point equation

$$a \in \arg \max_{a^i \in [a_0, \bar{a}]} \; h\left(\frac{(1-\kappa)\, a + \kappa a^i - b}{(1-\kappa)\, a + \kappa a^i + b}\right) - C_A\left(a^i\right). \tag{22}$$

Likewise, any B-supporter j must use a B-*consistent strategy*, defined by a similar equation, so that a strategy profile (a^*, b^*) is a type-homogenous Nash equilibrium if and only if a^* is A-consistent and b^* is B-consistent:

$$\begin{cases} a^* \in \arg\max_{a^i \in [a_0, \bar{a}]} \; h\left(\frac{(1-\kappa)a^* + \kappa a^i - b^*}{(1-\kappa)a^* + \kappa a^i + b^*}\right) - C_A\left(a^i\right) \\ b^* \in \arg\max_{b^j \in [b_0, \bar{b}]} \; \rho h\left(\frac{(1-\kappa)b^* + \kappa b^j - a^*}{a^* + (1-\kappa)b^* + \kappa b^j}\right) - C_B\left(b^j\right). \end{cases} \tag{23}$$

To better understand how an individual with *Homo moralis* preferences is expected to behave, we show in Figure 2 the material benefit and expected utility that an A-supporter can achieve for some given a, b, and depending on κ. In the two graphs on the left, the expected utility is plotted as a function of the hypothetical number of votes a^κ (see (8)) for four different values of κ. The line connected by black arrows shows the individual best response as κ varies (the direction of the arrow indicating how the best response changes as κ increases). In the two graphs on the right, the line isolates the benefit term $h(\alpha^\kappa(\cdot))$, still as a function of a^κ: the entire line shows the range of benefits that a *Homo moralis* with $\kappa = 1$ can enjoy, while the colored parts show the corresponding ranges for four other values of $\kappa < 1$: the range shrinks as κ decreases. The parameter values are $m = 5$, $\theta_A = \theta_B = 2$, $a_0 = 0.5$, $a_v = 1.25$, $b_0 = 0.1$, and $b_v = 1.25$. In all the graphs, $b = 1.2$, while $a = 1.2$ in the two graphs at the top but $a = 0.9$ in the two graphs at the bottom. Interestingly, the two graphs on the left show that the expected utility associated to the best response is neither generally increasing nor decreasing in κ.

Figure 3 shows the same objects as in Figure 2, but for $m = 20$. Comparing the two graphs on the right, we see that with this higher value of the curvature parameter, the benefit increase a *Homo moralis* with $\kappa < 1$ can enjoy depends heavily on what the co-partisans do. For example, if $\kappa = 0.2$, with $a = 1.2$ the range of attainable benefits is roughly between -0.6 and 0.5, while with $a = 0.9$ it is roughly between -0.85 and -0.55. In the graphs on the left we see that when $a < b$ (the bottom graph), the individual best response for an A-supporter may jump upwards as κ increases. This discontinuous pattern of the individual

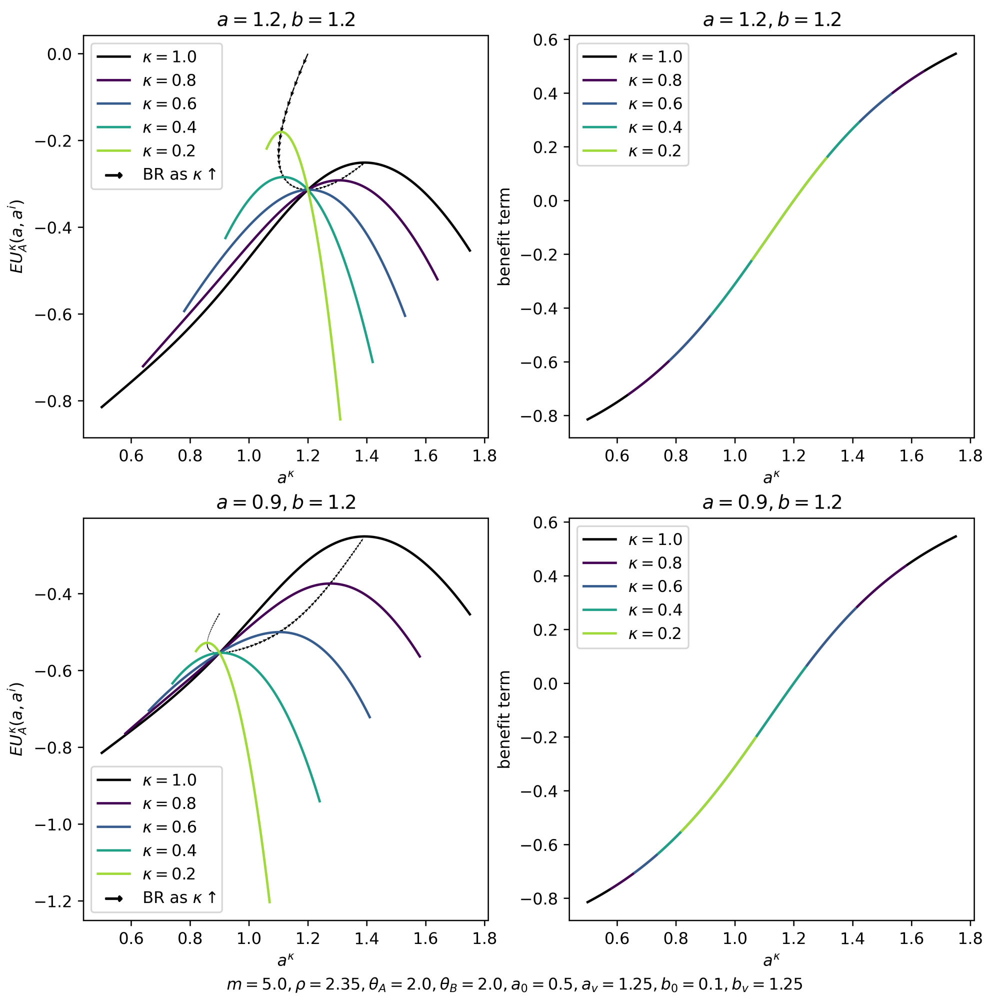


Figure 2: Utilities and benefits that an individual A-supporter can reach, for given turnouts a and b among the other voters, as κ varies; low m

best responses can also be seen in the A-consistent strategies (see (22)), shown in Figure 4 for four different values of b, and as a function of the curvature parameter m. We see that for $b = 1.2$ and $b = 1.5$, there are multiple A-consistent strategies when m is high enough. In this figure, $\kappa = 0.3$, $\theta_A = \theta_B = 2$, $a_0 = 0.5$, $a_v = 1.25$, $b_0 = 0.1$ and $b_v = 1.25$.

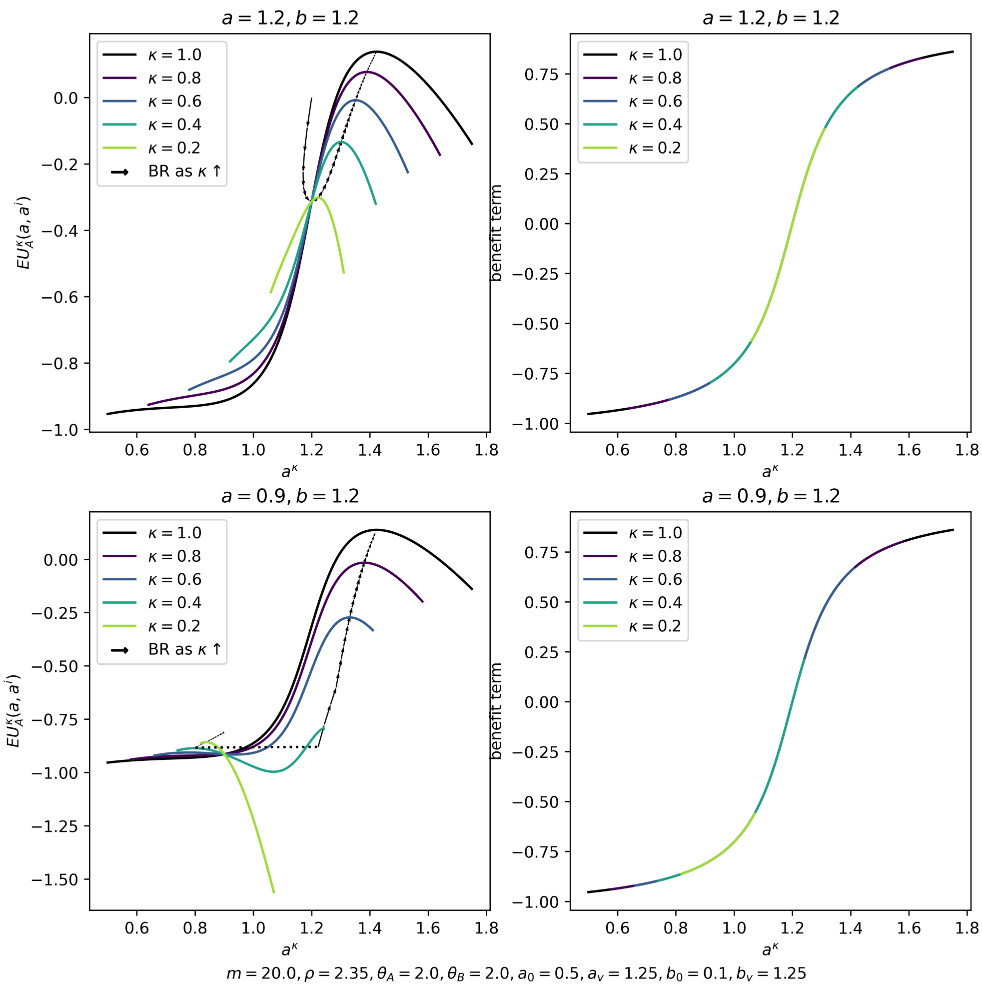


Figure 3: Utilities and benefits that an individual A-supporter can reach, for given turnouts a and b among the other voters, as κ varies; high m

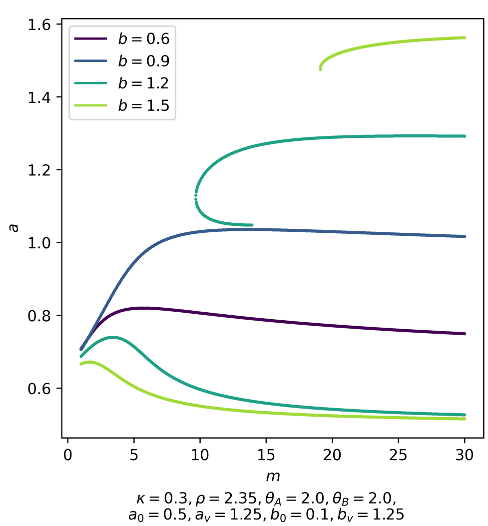


Figure 4: A-consistent strategies as m varies for some b

3.2 Power sharing: a never-a-best-response result

We now state our first main result: under partisan ethics, for any $\kappa > 0$ costly participation is positive on both sides, and this is a best response property, which holds even out of equilibrium. This could have been anticipated already by recalling that the benefit terms in the expected utilities in (10) and (13) are strictly increasing in s_A^i respectively s_B^i, while the marginal effect on the expected paid cost of an increase in s_A^i at $s_A^i = 0$ is $0 \times f_A(0) = 0$.

Proposition 1. *Under partisan ethics and power sharing,*

- *if $\kappa = 0$, $a = a_0$ (resp. $b = b_0$) is the dominant strategy for A-supporters (resp. B-supporters);*
- *if $\kappa \in (0, 1]$, $a = a_0$ (resp. $b = b_0$) is never a best response for a voter in group A (resp. B), and thus any equilibrium (a^*, b^*) is such that $\bar{a} \geq a^* > a_0$ and $\bar{b} \geq b^* > b_0$.*

Proof. For any turnout levels (a, b) from the other voters, A-supporter i's marginal utility of a^i is

$$\frac{\partial}{\partial a^i} U_A^\kappa \left(a, b, a^i\right) = h'(\alpha^\kappa(a, b, a^i)) \frac{2\kappa b}{\left[(1-\kappa)\, a + \kappa a^i + b\right]^2} - C_A' \left(a^i\right), \tag{24}$$

where $h' > 0$. Recalling, from the proof of Lemma 1, that

$$C_A\left(x\right) = \int_0^{F_A^{-1}\left(\frac{x - a_0}{a_v}\right)} c f_A\left(c\right) dc \quad \text{and} \quad C_A'\left(x\right) = \frac{1}{a_v} F_A^{-1}\left(\frac{x - a_0}{a_v}\right), \tag{25}$$

we see that $C_A'(a_0) = 0$ and $C_A'(x) > 0$ for $x > a_0$ by our assumptions on f_A. It follows that if $\kappa = 0$ the unique best response to any (a, b) is $a^i = a_0$, while if $\kappa > 0$, $a^i = a_0$ is never a best response. Idem for any B-supporter. □

Without universalisation (that is for $\kappa = 0$), the cost-sensitive voters are not willing to incur any cost to vote, since with a continuum of voters each individual vote has a nil effect on the election outcome. The second part of the proposition shows that a willingness to incur some cost of voting arises as soon as the degree of universalization is strictly positive. This is because (i) for any $\kappa \in (0, 1]$, the individual voter derives utility from considering the benefit that would realize if some positive share of the voters voted, and (ii) the smallest cost realizations approach zero (as per our assumption on the voting cost distributions).

Before studying further equilibria under power sharing, we present the results for the winner-take-all setting, which contrast sharply with the result stated in Proposition 1.

3.3 Winner-take-all: equilibrium characterisation

The pure winner-take-all case corresponds to a discontinuous step-function sign($\cdot$). The continuous function $h(\cdot)$ is not suitable for handling this case. Therefore, we proceed by approximation as follows:

Definition 1. Let $(h_t)_{t=1,2...}$ be a sequence of functions that all satisfy the hypothesis of the model and such that, for any $x \in [-1,1]$, $\lim_{t\to\infty} h_t(x) = \text{sign(x)}$. Such a sequence $(h_t)_{t=1,2,...}$ will be called an *approximating sequence.*

Equation 2 is an example of such a sequence.

Definition 2. We say that a pair (a,b) is *sustained as a limit winner-take-all equilibrium* under partisan (resp. non partisan) ethics if there exists a sequence $(a_t, b_t, h_t)_{t\in\mathbb{N}}$ such that

- $(h_t)_{t=1,2...}$ is an approximating sequence,
- for all t, (a_t, b_t) is an equilibrium of the partisan (resp. non partisan) game when the political outcome function is h_t, and
- $\lim_{t\to\infty}(a_t, b_t) = (a,b)$.

To prepare the proofs, we establish some properties of $(h_t)_{t=1,2,...}$. Indeed, as both h_t and its argument vary with t, we will need to establish that h_t converges uniformly on closed intervals not containing 0. Moreover, in order to be able to argue using the first-order conditions, we also show that h_t' converges to 0 uniformly on closed intervals not containing 0. While the uniform convergence of h_t is a consequence of monotonicity, the uniform convergence of h_t' is a consequence of concavity (resp. convexity).

Lemma 2. *Let* $0 < \varepsilon < 1$. *Then,*

1. h_t *converges uniformly to* 1 *on* $[\varepsilon, 1]$ *and to* -1 *on* $[-1,-\varepsilon]$, *and*
2. h_t' *converges uniformly to* 0 *on* $[-1,-\varepsilon]\cup[\varepsilon,1]$.

Proof. For the first statement, note that h_t is increasing. Therefore, for any $x \in [\varepsilon, 1]$, $h_t(\varepsilon) \leq h_t(x) \leq 1$. Thus, $\sup_{x\in[\varepsilon,1]} |h_t(x) - 1| \leq 1 - h_t(\varepsilon)$ and the result follows by the pointwise convergence of h_t to 1 at ε. The convergence to -1 on $[-1,-\varepsilon]$ follows by symmetry. For the second statement, the concavity of h_t on positive numbers implies that for any $x > 0$,

$$\lim_{t\to\infty} h_t'(x) = 0.$$

To see this, note that on intervals $[y,x]$ and $[x,z]$, concavity of h_t implies $\frac{h_t(z)-h_t(x)}{z-x} < h_t'(x) < \frac{h_t(x)-h_t(y)}{x-y}$, and then apply the sandwich theorem. The same result is also obtained for $x<0$. Finally, since h_t' is decreasing for $x>0$, h_t' converges uniformly to 0 on $[\varepsilon,1]$, and likewise on $[-1,-\varepsilon]$. □

We are now in position to state our second main result: under partisan ethics, winner-take-all, unlike power-sharing, leads to equilibrium costly participation being (in the limit) nil for both supporter groups, leaving the bases to determine the result of the election. The reasoning that applies in the power sharing case indeed breaks down in the limit case, because the political outcome function h is then locally flat.

Proposition 2 (Partisan ethics)**.** *Let $a_0 \neq b_0$ and let (a,b) be sustained as a limit equilibrium in the winner-take-all case. Then it must be that $(a,b)=(a_0,b_0)$, and the underdog wins if $b_0>a_0$ while the leader wins if $a_0>b_0$. If $a_0=b_0$ and $\kappa\in(0,1]$, no limit equilibrium exists. If $\kappa=0$, (a_0,b_0) is the unique winner-take-all limit equilibrium.*

Proof of Proposition 2: Let (a,b) be sustained as a limit winner-take-all equilibrium under partisan ethics by a sequence (a_t,b_t,h_t). The continuity of the function α implies that, if $a\neq b$, then $\alpha(a_t,b_t)$ converges to $\alpha(a,b)\neq 0$, hence $h_t'(\alpha(a_t,b_t))$ tends to 0, using the uniform convergence established above. It follows from the equilibrium conditions (Equations 88 and 32) that $C_A'\left(\frac{a_t-a_0}{a_v}\right)$ tends to 0. By continuity of the function C_A' this implies that a_t tends to a_0. The same argument holds for B.

We now assume that $a=b$ with (a,b) such that $a_0\le a<a_0+a_v$ and $b_0\le b<b_0+b_v$. Notice that there must be infinitely many (a_t,b_t) such that $b_t\le a_t$ or infinitely many (a_t,b_t) such that $a_t\le b_t$ (both can be the case). We will assume that there are infinitely many $b_t\le a_t$ since the argument below works analogously with infinitely many $a_t\le b_t$. By extracting a subsequence, we will assume that the whole sequence is such that $b_t\le a_t$.

Let $\delta>0$ and consider a deviation $\hat{b}_t$ such that

$$\kappa\hat{b}_t+(1-\kappa)\,b_t=a+\delta, \tag{26}$$

i.e.

$$\hat{b}_t=\frac{a+\delta-(1-\kappa)\,b_t}{\kappa}\to a+\frac{\delta}{\kappa}. \tag{27}$$

Since $b_t\to a$ and since we assumed that $a=b<b_0+b_v$, as long as δ is small enough, there exists a $T\in\mathbb{N}$ large enough such that $\hat{b}_t\in[b_0,b_0+b_v]$, i.e. such that $\hat{b}_t$ is a feasible deviation for all $t>T$.

Then, recalling $(a_t,b_t)\to(a,a)$, and by possibly increasing T, we can ensure that $\beta\,(a_t,a+\delta)\in[\varepsilon,1]$ for all $t>T$, for some $\varepsilon>0$ small enough.

We will now show that $\hat{b}_t$ is a profitable deviation for large enough t. Indeed, for $t > T$, by the uniform convergence of h_t towards 1 on $[\varepsilon, 1]$ and the continuity of C_B,

$$U^{\kappa}_{B,t}\left(a_t, b_t, \hat{b}_t\right) - U^{\kappa}_{B,t}\left(a_t, b_t, b_t\right) = \rho \underbrace{h_t\left(\beta\left(a_t, a+\delta\right)\right)}_{\to 1} - \rho \underbrace{h_t\left(\beta\left(a_t, b_t\right)\right)}_{\leq 0} + \underbrace{C_B\left(b_t\right) - C_B\left(\hat{b}_t\right)}_{\to C_B(a) - C_B\left(a+\frac{\delta}{\kappa}\right)} \tag{28}$$

is strictly positive for t large enough as well as δ small enough. The continuity of C_B is used twice in the argument: first to establish the convergence of the cost terms, and second to argue that the cost difference is arbitrarily small for small enough δ.

Having shown that there exist profitable deviations in the approximating sequence of equilibria for the individuals supporting at least one of the parties, we have reached a contradiction, i.e. there cannot be a limit winner-take-all equilibrium (a, a) with $a_0 \leq a < a_0 + a_v$ and $b_0 \leq b < b_0 + b_v$.

Let us now assume that $\left(\bar{b}, \bar{b}\right)$ is sustained as a limit winner-take-all equilibrium. Recall $\bar{a} > \bar{b}$. Assume that $h_t\left(\alpha\left(a_t, b_t\right)\right)$ does not converge to 1. Then, there exists a δ such that there exists a subsequence of (a_t, b_t) which satisfies $h_t\left(\alpha\left(a_t, b_t\right)\right) < 1 - \delta$. Similarly to the sequence of eventually profitable deviations we constructed above for B, we can then construct a sequence of eventually profitable deviations for A so that we can conclude that $h_t\left(\beta\left(a_t, b_t\right)\right) \to 1$. This in turn implies that for t large enough, B-supporters benefit from deviating to a zero effort threshold, since the benefit of keeping a high turnout vanishes as $t \to \infty$.

Hence, there cannot be a limit winner-take-all equilibrium $\left(\bar{b}, \bar{b}\right)$ and the only possible winner-take-all limit equilibrium is (a_0, b_0), whenever $a_0 \neq b_0$. Otherwise, there is no winner-take-all limit equilibrium.

□

This proposition says that if there exists a limit equilibrium, then (a) $a_0 \neq b_0$, and (b) it must be such that no cost-sensitive voters turn out. The reason is simple. If, to the contrary, there was some costly turnout among either A- or B-supporters, then it must be such that their candidate wins. Indeed, if their candidate lost, they would have an incentive to eliminate their turnout; and if the two candidates obtained the exact same number of votes, they would have an incentive to slightly increase their turnout to secure a victory (this is also why there exists no equilibrium when $a_0 = b_0$). But given that the strategy is a continuous variable, there would then exist some $\varepsilon > 0$ such that they could reduce their turnout by ε without losing, thus reducing the cost while keeping the benefit of the victory.

Clearly, and like in the power sharing setting (see Proposition 1), if $\kappa = 0$ the dominant strategy is to not participate. Existence of the limit equilibrium $(a, b) = (a_0, b_0)$ is then not an issue. The following proposition shows that existence is in fact ensured as long as κ is not too large. For the purpose of this proposition, whenever $C_A(\bar{a}) \geq 2$, respectively $C_B(\bar{b}) \geq 2\rho$, let the threshold values $\tilde{\kappa}_A > 0$ and $\tilde{\kappa}_B > 0$ be implicitly defined by

$$C_A\left(\frac{b_0 - a_0}{\tilde{\kappa}_A} + a_0\right) = 2 \tag{29}$$

$$C_B\left(\frac{a_0 - b_0}{\tilde{\kappa}_B} + b_0\right) = 2\rho, \tag{30}$$

where 2 is the difference between the (limit) material benefit in case of a victory of the own candidate and that in case of their loss, that is $2 = \text{sign}(1) - \text{sign}(-1)$. The existence of $\tilde{\kappa}_A$ and $\tilde{\kappa}_B$ follows from $C_A(a_0) = C_B(b_0) = 0$ and the continuity of C_A and C_B. While $\tilde{\kappa}_j$ can exceed 1, κ in the proposition below remains within $[0, 1]$.

Proposition 3 (Partisan ethics)**.** *If $a_0 > b_0$, then $(a, b) = (a_0, b_0)$ is a limit equilibrium if and only if $\kappa \leq \tilde{\kappa}_B$, where $\tilde{\kappa}_B$ is defined in* (30) *if $C_B(\bar{b}) \geq 2\rho$, and $\tilde{\kappa}_B = 0$ otherwise. If $a_0 < b_0$, then $(a, b) = (a_0, b_0)$ is a limit equilibrium if and only if $\kappa \leq \tilde{\kappa}_A$, where $\tilde{\kappa}_A$ is defined in* (29) *if $C_A(\bar{a}) \geq 2$, and $\tilde{\kappa}_A = 0$ otherwise.*

Sketch of proof. (The full proof can be found in Appendix A.1.) Without loss of generality, let $a_0 > b_0$. Let $\kappa \leq \tilde{\kappa}_B$. Taking our example benefit functions from (2) (replacing m with t), one can find a sequence of solutions to the first-order conditions that are unique within $[a_0, a_0 + \varepsilon] \times [b_0, b_0 + \varepsilon]$. It is straightforward to see that there is no profitable deviation for A-supporters for high t. By distinguishing cases around $b_w = \frac{a_0 - b_0}{\tilde{\kappa}_B} + b_0$, the hypothetical deviation that would make B-supporters achieve a tie in the limit, one can equally exclude the existence of a profitable deviation for B-supporters. This is where the definition of $\tilde{\kappa}_B$ comes in: $\kappa \leq \tilde{\kappa}_B$ ensures that any such deviation is too costly. Conversely, if $\kappa > \tilde{\kappa}_B$, for t large enough, we can find a profitable deviation towards $b_w + \varepsilon$: the cost is then less than the benefit in the limit (the benefit converges to 2ρ) and for small enough ε. □

The intuition for this result is simple. Recall from Proposition 2 that if an equilibrium exists, the party with the smallest base loses the election. But then any voter in the losing group would deviate to a turnout which, if also applied by a share κ of the co-partisans, would guarantee a victory, if the utility kick from doing so exceeds the expected cost thereof. Clearly, a low enough κ is both necessary and sufficient for this deviation to not be utility-enhancing.

3.4 Power sharing: discussion

Propositions 2 and 3 show that in the winner-take-all limit, existence is not guaranteed, but when an equilibrium exists it is unique, and the candidate with the largest base wins. In this section we discuss whether these properties generalize to the power sharing system.

While a sufficient condition for an equilibrium to exist would be that both objective functions in (40) are quasi-concave in a^i respectively b^j, the strict convexity of h for negative relative margins implies that quasi-concavity is not guaranteed. Below we will see examples of non-existence.[12] But we will also see examples with multiple equilibria. We begin by deriving a general result. To prepare the ground for this, by Proposition 1 for any $\kappa > 0$ any equilibrium (a^*, b^*) satisfies the first-order conditions:

$$\frac{\partial}{\partial a^i} U_A^\kappa\left(a, b, a^i\right)|_{a^i=a^*, b=b^*} \quad = \quad \frac{2\kappa b^* h'(\alpha(a^*, b^*))}{(a^*+b^*)^2} - C_A'\left(a^*\right) \begin{cases} = 0 \text{ if } a^* \in (a_0, \bar{a}) \\ \geq 0 \text{ if } a^* = \bar{a} \end{cases} \tag{31}$$

$$\frac{\partial}{\partial b^i} U_B^\kappa\left(a, b, b^j\right)|_{a=a^*, b^j=b^*} \quad = \quad \frac{2\kappa a^* \rho h'(\beta(a^*, b^*))}{(a^*+b^*)^2} - C_B'\left(b^*\right) \begin{cases} = 0 \text{ if } b^* \in (b_0, \bar{b}) \\ \geq 0 \text{ if } b^* = \bar{b}. \end{cases} \tag{32}$$

Notice that, for any interior equilibrium $(a^*, b^*) \in (a_0, \bar{a}) \times (b_0, \bar{b})$, the two equations and the fact that $\alpha(a, b) = -\beta(a, b)$ and $h'(x) = h'(-x)$ together imply:

$$\frac{b^*}{\rho a^*} = \frac{C_A'(a^*)}{C_B'(b^*)}. \tag{33}$$

The result concerns settings where the leader supporters have an absolute cost advantage over the underdog supporters: this occurs when their base a_0 and their mass a_v of cost-sensitive voters is at least as large, and its cost distribution F_A is more favorable, compared to the corresponding features of the underdog supporters. One might then expect the leader to always win. The next proposition identifies conditions under which this intuition is confirmed. Following the proposition, however, we will show settings where the underdog wins, in spite of the cost advantage enjoyed by the leader supporters.

Proposition 4 (Partisan ethics)**.** *Suppose that the leader supporters enjoy a cost advantage over the underdog supporters: their base is at least as large ($a_0 \geq b_0$), their mass of cost-sensitive voters is at least as large ($a_v \geq b_v$), and the cost distributions favor them ($F_A(c) \geq F_B(c)$ for all $c \in \mathbb{R}_{>0}$), with at least one of the inequalities holding strictly. Then, all other*

[12] See also Appendix C.2 for a numerical example with non-existence. Moreover, in Appendix C.3 we derive sufficient conditions for existence and uniqueness; these are similar to those adopted by Herrera et al. (2016).

parameters being fixed, there exists $\varepsilon > 0$ such that for any stake $\rho < 1 + \varepsilon$, the leader wins in any equilibrium ($\alpha(a^*, b^*) > 0$).

Proof. Suppose, by contradiction, that $b^* \geq a^*$, in which case either both a^* and b^* are interior, or $b^* = \bar{b}$ and a^* is interior. Clearly, if $a_0 \geq b_0$, $a_v \geq b_v$, and $F_A(c) \geq F_B(c)$ for all $c \in \mathbb{R}_{>0}$, with at least one of the inequalities holding strictly, then $C_A'(x) < C_B'(x)$ for all x. Hence, $b^* \geq a^*$ implies $C_A'(a^*) < C_B'(b^*)$, and $b^* C_B'(b^*) > a^* C_A'(a^*)$. If b^* is interior, or if $b^* = \bar{b}$ and the first-order condition (32) holds as an equality, the inequality $b^* C_B'(b^*) > a^* C_A'(a^*)$ contradicts (33) if $\rho = 1$. If $b^* = \bar{b}$ and the first-order condition holds as a strict inequality, i.e., $h'(\beta(a^*, b^*)) \frac{2\kappa a^* \rho}{(a^*+b^*)^2} - C_B'(b^*) > 0$, then this inequality and (88) together imply $b^* C_B'(b^*) < \rho a^* C_A'(a^*)$. A contradiction with the inequality $b^* C_B'(b^*) > a^* C_A'(a^*)$ is reached if $\rho \leq 1$. By continuity, there exists $\bar{\varepsilon} > 0$ such that the contradiction also obtains for $\rho = 1 + \varepsilon$ for any $\varepsilon \in (0, \bar{\varepsilon})$. □

We next lift two of the conditions identified in the proposition, and show that the underdog supporters can secure a power advantage of their candidate over the other candidate, either if they perceive a high enough stake (Example 1), or if their base is large compared to the leader's (Example 2). In the accompanying figures, we display the curve showing all the A-consistent turnouts (recall (22)) and the curve showing all the B-consistent turnouts. A type-homogenous equilibrium (a^*, b^*) is such that a^* is A-consistent given b^*, and b^* is B-consistent given a^* (recall (23)). Any (a^*, b^*) where the two curves intersect is an equilibrium.

Example 1 (The underdog wins thanks to a high stake ρ)**.** In Figure 5, the A-supporters enjoy a strictly larger base ($a_0 = 0.5 > 0.4 = b_0$), a strictly larger mass of cost-sensitive voters ($a_v = 1.4 > 1.1 = b_v$), and a strictly more favorable cost distribution ($\theta_A = 1.1 < 2.0 = \theta_B$), but the stake for the B-supporters is high ($\rho = 3.6$). We see that there exists a unique equilibrium, since the curves intersect only once. Moreover, since (a^*, b^*) is above the dashed line, along which the turnouts are equal, the underdog wins, even though its supporters represent only $1.5/3.4 \approx 44\%$ of the electorate. Note further that approximately $1.313/(0.5+1.4) \approx 69\%$ of the A-supporters and $1.359/(0.4+1.1) \approx 91\%$ of the B-supporters participate in the election, although the degree of universalization is only $\kappa = 0.5$. In other words, full universalization is not necessary for high rates of participation to obtain.

Example 2 (The underdog wins thanks to a large base b_0)**.** In Figure 6, the A-supporters enjoy a strictly larger mass of cost-sensitive voters ($a_v = 1.25 > 1.0 = b_v$), while the B-supporters enjoy a strictly larger base ($b_0 = 0.7 > 0.5 = a_0$). The cost distribution parameters are the same ($\theta_A = \theta_B = 2.0$). These assumptions together imply that the

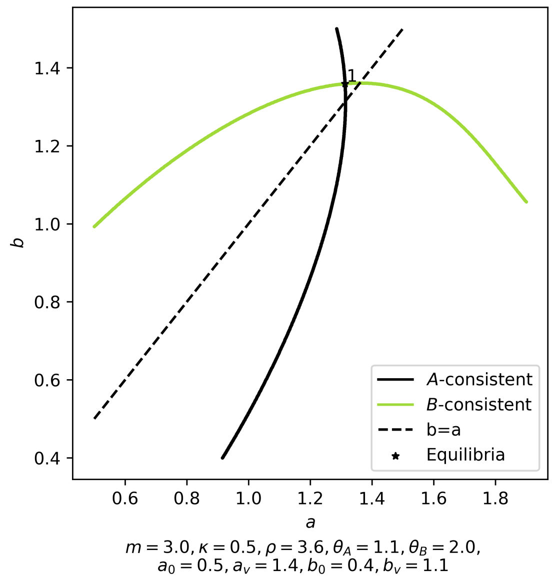


Figure 5: An example where the underdog wins in spite of the leader supporters's cost advantage ($\theta_A < \theta_B$, $a_v > b_v$, and $a_0 > b_0$), thanks to a high enough stake ρ.

underdog supporters enjoy a cost advantage over the leader supporters for turnouts close enough to the underdog's base b_0. The figure shows that although the stake for the B-supporters is neutral ($\rho = 1$), the underdog wins at the unique equilibrium, since the unique intersection of the curves showing the A-consistent and the B-consistent strategies intersect once, at a point above the dashed line, along which the turnouts are equal. In this case the underdog supporters represent $1.7/3.45 \approx 49\%$ of the electorate.

The preceding examples both feature a unique equilibrium. However, equilibria may fail to exist, or there may exist multiple equilibria. Interestingly, when there are multiple equilibria, they may fail to generate the same winner. An illustration is provided in Figure 7. The figures in the first column show the number of type-homogneous Nash equilibria (a number that varies between 0 and 5). The figures in the second column then show, for the same parameter configurations, the number of equilibria in which the underdog wins the election. In each graph, we vary two parameters while the others are as listed at the bottom of the figure; these parameter values are the result of a search of the parameter space for examples with many equilibria. In all the graphs, it is the universalization parameter (κ) that varies along the horizontal axis. On the vertical axis it is the curvature parameter of the benefit function (m) that varies in the first line of figures, while it is the underdog's base (b_0) in the second line, and the stake (ρ) in the third line. Moreover, in the first two lines, where the stake ρ is fixed, it is high enough to generate equilibria where the underdog wins ($\rho = 5$).

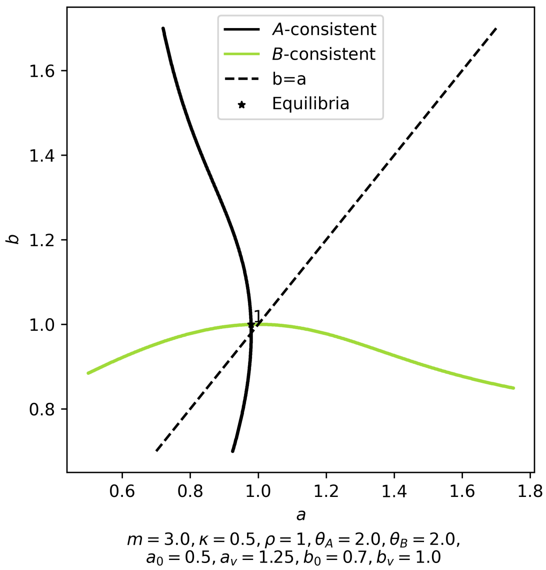


Figure 6: An example where the underdog wins thanks to a larger base ($b_0 > a_0$), in spite of a neutral stake ($\rho = 1$) and identical cost distributions ($\theta_A = \theta_B$).

The following patterns appear. First, equilibrium multiplicity (respectively non-existence) appears only for sufficiently low (respectively high) values of κ. Second, the first line shows that when there exists a unique equilibrium, the leader wins when the value of the curvature parameter m is either low enough or high enough, while the underdog wins for the set of values of m in between. We further see that as κ increases, the interval of m-values for which the underdog wins gets closer to 0. Third, the size of the base b_0 has an unambiguously positive effect on the underdog's prospects of winning, as seen in the second line. Fourth, in the third line we see that, at least when the base of the leader exceeds that of the underdog, an increase in the stake ρ does not necessarily lead to an increase in the number of equilibria with a victory for the underdog.

To get a sense of how large a victory the underdog may obtain due to a higher stake, a general result is obtained when the relative frequency of cost-sensitive voters is the same for groups A and B ($a_0/\bar{a} = b_0/\bar{b}$ or, equivalently, $a_0/b_0 = a_v/b_v = \bar{a}/\bar{b}$), and all the cost-sensitive voters face the same cost distribution ($F_A = F_B = F$). This set of assumptions can be interpreted as an assumption of independence between costs and party affiliation (note that the parameter values used in Figure 7 correspond to this case, since $\frac{a_0}{a_v} = \frac{b_0}{b_v} \approx 0.94$). The next proposition shows that, for interior equilibria, the turnout rates ($\frac{a^*}{\bar{a}}$ and $\frac{b^*}{\bar{b}}$) are the same for both groups unless $\rho \neq 1$, in which case the group with the highest stake turns out at a greater rate. The proposition further provides bounds on how different the equilibrium turnout rates can then be. These bounds depend only on the stake ρ and are universal with respect to the form of the function h and to the value of the parameter κ. They are therefore

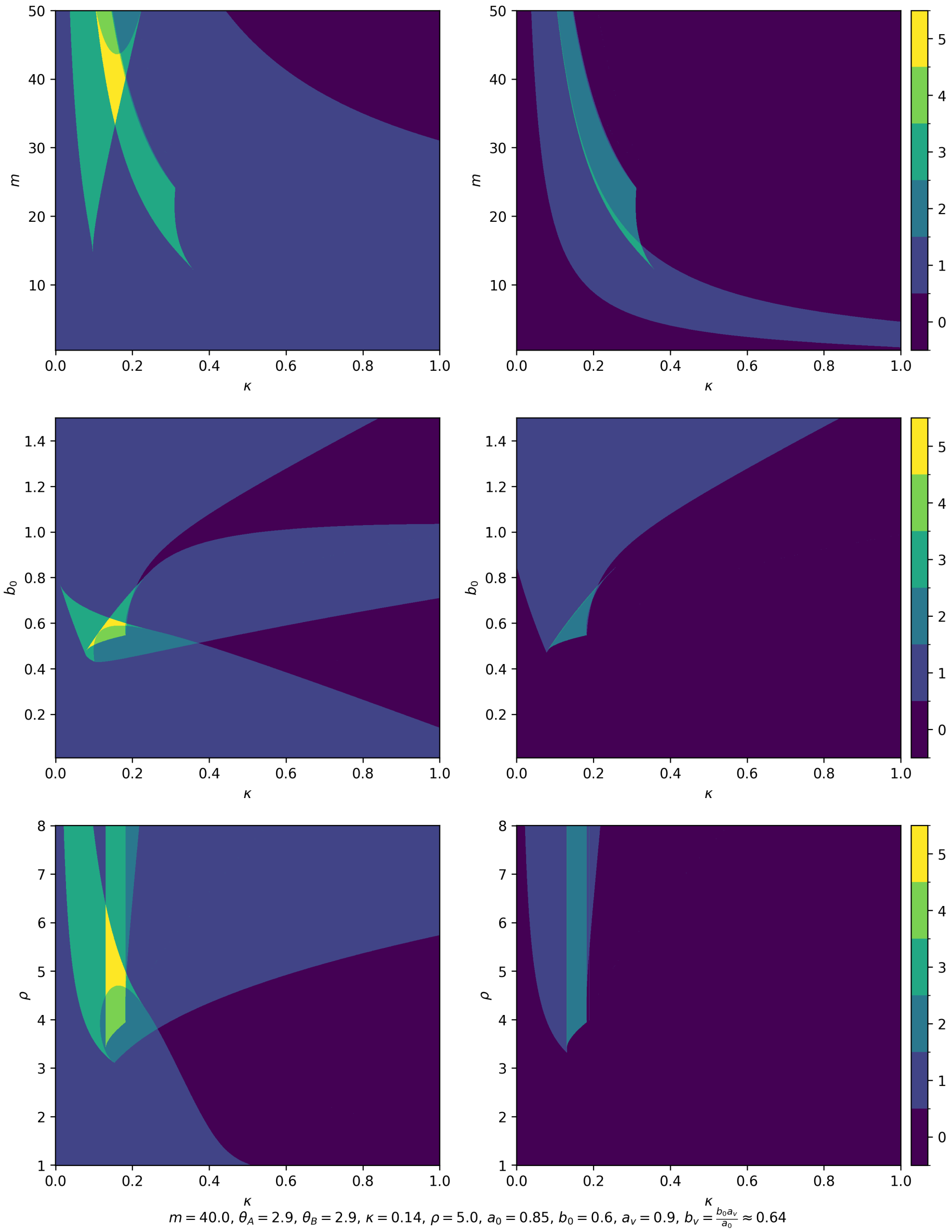


Figure 7: Multiplicity of equilibria (left column) and multiplicity of equilibria where the underdog wins (right column), depending on (κ, m) (first line), (κ, ρ) (second line), (κ, b_0) (third line)

valid whether the election is close to winner-take all, a random dictatorship, or anything in between, and for any degree of morality $\kappa \in (0, 1]$.

Proposition 5 (Partisan ethics)**.** *Suppose that* $\frac{a_0}{\bar{a}} = \frac{b_0}{\bar{b}}$ *and that* $F_A(c) = F_B(c) = F(c)$ *for all* $c \in \mathbb{R}_{>0}$. *Then, at any interior equilibrium* $(a^*, b^*) \in (a_0, \bar{a}) \times (b_0, \bar{b})$*:*

1. if $\rho = 1$:

$$\frac{a^*}{\bar{a}} = \frac{b^*}{\bar{b}}; \tag{34}$$

2. if $\rho > 1$:

$$\frac{a^*}{\bar{a}} < \frac{b^*}{\bar{b}} < \rho \cdot \frac{a^*}{\bar{a}}; \tag{35}$$

3. if $\rho < 1$:

$$\frac{a^*}{\bar{a}} > \frac{b^*}{\bar{b}} > \rho \cdot \frac{a^*}{\bar{a}}. \tag{36}$$

Proof. Take $\rho > 1$. Given that a^* and b^* are interior, equation (33) applies, and writes:

$$\frac{b^*}{b_v} F^{-1}\left(\frac{b^*}{b_v} - r\right) = \rho \cdot \frac{a^*}{a_v} F^{-1}\left(\frac{a^*}{a_v} - r\right) \tag{37}$$

for $r = \frac{a_0}{a_v} = \frac{b_0}{b_v}$. First, suppose that $\frac{b^*}{b_v} \leq \frac{a^*}{a_v}$, then: $F^{-1}\left(\frac{b^*}{b_v} - r\right) \geq \rho \cdot F^{-1}\left(\frac{a^*}{a_v} - r\right)$ and since $\rho > 1$, $F^{-1}\left(\frac{b^*}{b_v} - r\right) > F^{-1}\left(\frac{a^*}{a_v} - r\right)$. Because F is strictly increasing by our assumption on the support of f, this implies $\frac{b^*}{b_v} > \frac{a^*}{a_v}$, a contradiction. Next, suppose that $\frac{b^*}{b_v} \geq \rho \frac{a^*}{a_v}$. Then, similar reasoning yields $F^{-1}\left(\frac{b^*}{b_v} - r\right) \leq F^{-1}\left(\frac{a^*}{a_v} - r\right)$ and $\frac{b^*}{b_v} \leq \frac{a^*}{a_v} < \rho \frac{a^*}{a_v}$, a contradiction. This proves the result for $\rho > 1$. The reasoning is identical for $\rho < 1$.

For the case $\rho = 1$, suppose for a contradiction $\frac{b^*}{b_v} < \frac{a^*}{a_v}$. Then, (37) implies $F^{-1}\left(\frac{b^*}{b_v} - r\right) > F^{-1}\left(\frac{a^*}{a_v} - r\right)$, which in turn implies $\frac{b^*}{b_v} > \frac{a^*}{a_v}$ because F is strictly increasing, a contradiction. $\frac{b^*}{b_v} > \frac{a^*}{a_v}$ yields a contradiction by the same reasoning, implying $\frac{b^*}{b_v} = \frac{a^*}{a_v}$. □

The proposition describes the effect of the stake on turnout, when costs are not correlated with party affiliation. The participation rate is higher for the side with the higher stake, but less than proportionally: if the stake is x times higher for one side, the participation rate is, on that side, higher than on the other side, but less than x times higher.

3.5 Comparison with group-based models

We here compare our formalization of ethical voters with that adopted in group-based models of Coate and Conlin, 2004 and Feddersen and Sandroni, 2006.[13] In these models an ethical voter gets a "duty payoff" D from "doing their part" (Feddersen and Sandroni, 2006), where D exceeds the highest possible cost realization. A decision based on a "duty to vote" would thus lead all the ethical voters to vote. Dependence of an ethical voter's turnout decision on their cost realization obtains by positing that an ethical voter adopts a cost threshold

[13] See also the group-based models by Morton (1991) and Bierbrauer, Tsyvinski, and Werquin (2022), which have both endogenous turnout and endogenous party platforms.

in order to reduce the aggregate cost; this cost reduction is traded off against the loss with the associated reduced probability of winning. In other words, “doing their part” entails abstaining when the cost realization exceeds the threshold. The predicted turnout rates are obtained as “an equilibrium between two party planners”, each of which “looks at the total electoral benefit” for their preferred candidate “net of the total cost incurred by his supporters” (Herrera et al., 2016, p. 612). A counter-intuitive feature of these models is that, at the individual level, each ethical voter would be perfectly happy to incur any positive voting cost, since D is assumed to exceed the largest possible cost realization. The duty that such a voter feels obliged to fulfill thus consists of reducing the aggregate cost of voting, by abstaining from voting when the realized cost is above the equilibrium threshold: some voters “receive a [duty] payoff for not voting” (Feddersen and Sandroni, 2006, p. 1272).

By contrast, in our model each voter simply maximizes his own expected utility, and there is no constant duty payoff. Such utility maximization imposes fewer demands on the information the voter needs to select an ethical behavior, compared to the group-based models, in which an ethical voter needs to place himself in a social planner’s shoes to understand which cost threshold he or she should adopt to obtain the constant duty payoff D. With *Homo moralis* preferences, a voter instead evaluates each possible strategy applying a simple universalization calculus to the benefit, while the expected cost of the deviation is the individual’s own true expected cost (recall (14) and (15)). Indeed, *Homo moralis* preferences make a voter evaluate a strategy in the light of the expected material utility that would obtain if a share κ of the others also adopted the same strategy; whether or not others adopted a different threshold than they do would be irrelevant for this voter’s expected cost of voting.

To further understand this crucial conceptual difference between our model and group-based models, it is instructive to consider the special case of our model where all ethical voters have degree of universalization $\kappa = 1$. The expected utilities in (14) and (15) then boil down to

$$EU_A\left(a, b, a^i\right) = h\left(\frac{a^i - b}{a^i + b}\right) - C_A\left(a^i\right) \tag{38}$$

$$EU_B\left(a, b, b^j\right) = \rho h\left(\frac{b^j - a}{a + b^j}\right) - C_B\left(b^j\right). \tag{39}$$

The strategy profile (a^*, b^*) is a type-homogenous Nash equilibrium if and only if

$$\begin{cases} a^* \in \arg\max_{a^i \in [a_0, \bar{a}]} \; h\left(\frac{a^i - b^*}{a^i + b^*}\right) - C_A\left(a^i\right) \\ b^* \in \arg\max_{b^j \in [b_0, \bar{b}]} \; \rho h\left(\frac{b^j - a^*}{a^* + b^j}\right) - C_B\left(b^j\right). \end{cases} \tag{40}$$

There is a mathematical similarity of this special case with group-based models, since any (a^*, b^*) satisfying (40) could alternatively be interpreted as representing a Nash equilibrium of a game played by "two party planners" (Herrera et al., 2016, p. 612); likewise, see Definition 1 in Feddersen and Sandroni (2006) and the description of equilibrium on p.1481 in Coate and Conlin (2004). This leads to the following observation:

Remark 1. *If all cost-sensitive voters have* Homo moralis *preferences with degree of universalization $\kappa = 1$, any pair of threholds implemented at a type-homogenous Nash equilibrium of the two-population game would also be implemented at a Nash equilibrium of the two-player game between two party planners, each of whom seeks to maximize the aggregate material payoffs of their respective constituent groups. The type-homogenous Nash equilibrium implements the cost thresholds in a decentralized manner (in the sense that each voter simply maximizes her own expected utility).*

In sum, the key conceptual difference with group-based models is that in our model each voter's decision is individually rational. Furthermore, mathematically, our model is different as soon as $\kappa < 1$.

There are also other differences between our model and those by Coate and Conlin (2004), Feddersen and Sandroni (2006), and Herrera et al. (2016). In Coate and Conlin (2004) there is *ex ante* uncertainty about the distribution of voters into A- and B-supporters: voting then also entails information aggregation. The objective function of the "social planner" representing a group is the expected aggregate welfare of the members of this group. In Feddersen and Sandroni (2006) and Herrera et al. (2016) there is no *ex ante* uncertainty about the distribution of party preferences; the objective function of the "social planner" representing a group is the *per capita* benefit implied by the preferred candidate's net margin, net of the aggregate voting cost for the group members in Herrera et al. (2016), and net of the aggregate voting cost for all the ethical voters (in both groups) in Feddersen and Sandroni (2006). The latter model is thus reminiscent of a mix of the partisan setting considered above, where the co-partisans constitute the reference group for ethical voters, and the non-partisan setting considered in the next section, where the full set of cost-sensitive voters constitute their reference group. Such a non-partisan setting has not yet been formalized in the literature, although it appears to be natural in the context of ethical voters.

4 Non-partisan ethics (the *ex ante* setting)

We here follow Harsanyi's view by considering non-partisan ethics. In the non-partisan setting a cost-sensitive voter i chooses a strategy which is a pair of thresholds $s^i = (s^i_A, s^i_B) \in [0, \infty]^2$: the voter abstains when her cost for voting is larger than s^i_A if she prefers candidate

A, and when her cost for voting is larger than s^i_B if she prefers candidate B ($s^i_A = \infty$ respectively $s^i_B = \infty$ means that she votes independently of the realized cost). Two interpretations are possible. In the first, there is *ex ante* uncertainty regarding the candidate i prefers, and she selects the strategy behind the veil of ignorance, before this uncertainty is resolved. This may well describe how independent voters reason. In the second interpretation, there is no such uncertainty, but due to her ethical concern the individual adopts the viewpoint of Harsanyi's impartial observer, by inserting a veil of ignorance in her reasoning. Whatever interpretation is chosen, i selects the strategy $s^i = (s^i_A, s^i_B)$ before knowing her actual cost of voting. Voters have correct beliefs about the voting cost distributions, described in Section 2. Each individual votes for A or B, or abstains. Individual i in group A with realized cost c^i_A votes for A if and only if $c^i_A \leq s^i_A$. The corresponding assumptions are made for group B.

4.1 Payoff computations

As in the partisan setting, we are looking for homogenous equilibria, which here means that all the cost-sensitive voters choose the same strategy. At a homogenous equilibrium $s = (s_A, s_B)$, the realized turnouts are $a(s) = a_0 + a_v F_A(s_A)$ and $b(s) = b_0 + b_v F_B(s_B)$, respectively, for the two candidates, and the following relative vote margins obtain:

$$\alpha(s) = \frac{a(s_A) - b(s_B)}{a(s_A) + b(s_B)} \qquad \text{and} \qquad \beta(s) = -\alpha(s), \tag{41}$$

so that each voter obtains expected net material benefit

$$EU(s) = \frac{1}{a_v + b_v}\left[a_v h(\alpha(s)) + b_v \rho h(\beta(s)) - a_v \int_{c=0}^{s_A} c f_A(c)\, dc - b_v \int_{c=0}^{s_B} c f_B(c)\, dc\right]. \tag{42}$$

With a non-partisan ethic, the voter takes into account both the expected benefits and the expected costs that the thresholds s_A and s_B entail; the benefits are weighted by the relative population shares of the groups, to reflect the *ex ante* perspective that the voter adopts. Henceforth, without loss of generality, we will drop the constant positive factor $1/(a_v + b_v)$. Since $h(-x) = -h(x)$, one obtains the following expression for the expected utility of a voter i with *Homo moralis* preferences, given that all the other voters use strategy s:

$$EU^\kappa(s, s^i) = (a_v - \rho b_v)\, h(\alpha^\kappa(s, s^i)) - a_v \int_{c=0}^{s^i_A} c f_A(c)\, dc - b_v \int_{c=0}^{s^i_B} c f_B(c)\, dc, \tag{43}$$

where $\alpha^\kappa(s, s^i)$ is the hypothetical relative margin,

$$\alpha^\kappa(s, s^i) = \frac{(1-\kappa)[a(s_A) - b(s_B)] + \kappa[a(s_A^i) - b(s_B^i)]}{(1-\kappa)[a(s_A) + b(s_B)] + \kappa[a(s_A^i) + b(s_B^i)]}. \tag{44}$$

Applying the same change of variables as under partisan ethics, we henceforth assume that an individual i's strategy is a pair $(a^i, b^i) \in [a_0, \bar{a}] \times [b_0, \bar{b}]$, and we write (a, b) for the strategy used by the other voters at a homogenous equilibrium, so that the expected utility of i is

$$EU^\kappa(a, b, a^i, b^i) = (a_v - \rho b_v) \cdot h(\alpha^\kappa\left(a, b, a^i, b^i\right)) - a_v C_A\left(a^i\right) - b_v C_B\left(b^i\right), \tag{45}$$

where the hypothetical relative margin is expressed as

$$\alpha^\kappa\left(a, b, a^i, b^i\right) = \frac{(1-\kappa)\,a + \kappa a^i - (1-\kappa)b - \kappa b^i}{(1-\kappa)\,a + \kappa a^i + (1-\kappa)\,b + \kappa b^i}. \tag{46}$$

Under our assumptions, this is a continuously differentiable function of (a^i, b^i).

4.2 Power sharing: equilibrium existence and key property

By contrast to the partisan setting, here we can establish general equilibrium existence, thanks to the aforementioned change of variables together with Lemma 1, and state our third main result: under non-partisan ethics, at equilibrium, exactly one group incurs positive voting costs (except in two knife-edge cases, in which nobody votes).

Proposition 6 (Non-partisan ethics)**.** *Under power sharing, an equilibrium always exists. If $\kappa = 0$, then $(a^*, b^*) = (a_0, b_0)$ is the unique equilibrium, while if $\kappa \in (0, 1]$ then:*

- *if $a_v > \rho b_v$, any equilibrium is such that $a^* > a_0$ and $b^* = b_0$;*
- *if $\rho b_v > a_v$ any equilibrium is such that $a^* = a_0$ and $b^* > b_0$;*
- *if $a_v = \rho b_v$, then $(a^*, b^*) = (a_0, b_0)$ is the unique equilibrium.*

Proof. We begin by proving existence. Consider the auxiliary function

$$\Phi(a, b) = \kappa\lambda h(\alpha(a, b)) - a_v C_A\left(a\right) - b_v C_B\left(b\right) \tag{47}$$

where $\lambda = a_v - \rho b_v$. It takes values in $\mathbb{R} \cup \{-\infty\}$. We begin by showing that (a^*, b^*) is an equilibrium if it is a global maximum of Φ. Thus, let (a^*, b^*) be a point where Φ reaches its maximum, and suppose, by contradiction, that there exists (a', b') such that

$EU^{\kappa}(a^*, b^*, a', b') > EU(a^*, b^*)$, that is:

$$a_v C_A(a^*) + b_v C_B(b^*) - a_v C_A(a') - b_v C_B(b') > \lambda\left[h(\alpha(a^*, b^*)) - h(\alpha(a^{\kappa}, b^{\kappa}))\right] \quad (48)$$

for $a^{\kappa} = (1-\kappa)a^* + \kappa a'$ and $b^{\kappa} = (1-\kappa)b^* + \kappa b'$. Since (a^*, b^*) maximizes Φ, we have $\Phi(a^{\kappa}, b^{\kappa}) \leq \Phi(a^*, b^*)$, which writes:

$$a_v C_A(a^*) + b_v C_B(b^*) - a_v C_A(a^{\kappa}) - b_v C_B(b^{\kappa}) \leq \kappa\lambda\left[h(\alpha(a^*, b^*)) - h(\alpha(a^{\kappa}, b^{\kappa}))\right]. \quad (49)$$

Combining the two previous equations we find, upon rearranging the terms,

$$(1-\kappa)a_v C_A(a^*) + \kappa a_v C_A(a') + (1-\kappa)b_v C_B(b^*) + \kappa b_v C_B(b') < a_v C_A(a^{\kappa}) + b_v C_B(b^{\kappa}). \quad (50)$$

This contradicts the convexity of the functions C_A, C_B (see Lemma 1).

The second part of the proof consists in showing that Φ admits a maximum, which is a sufficient condition for an equilibrium to exist, given the first part of the proof. To see this, first note that $h(\alpha(a,b))$ is continuous on $[a_0, \bar{a}] \times [b_0, \bar{b}]$ and takes values in $\mathbb{R}$. Moreover, in the case where C_A, C_B are continuous functions on $[a_0, \bar{a}] \times [b_0, \bar{b}]$ taking values in $\mathbb{R}$, one can conclude by the extreme value theorem, observing that Φ is continuous and takes values in $\mathbb{R}$. In the remaining case, where $C_A(\bar{a}) = \infty$ or $C_B(\bar{b}) = \infty$, we have that $C_A(a) \to \infty$ for $a \to \bar{a}$ or $C_B(b) \to \infty$ for $b \to \bar{b}$. Then, since h is bounded on $[a_0, \bar{a}] \times [b_0, \bar{b}]$, one can find some $\varepsilon_A, \varepsilon_B \geq 0$ such that Φ takes real values on $[a_0, \bar{a} - \varepsilon_A] \times [b_0, \bar{b} - \varepsilon_B]$ and such that $\Phi(a,b) \leq \Phi\left(\min(a, \bar{a} - \varepsilon_A), \min\left(b, \bar{b} - \varepsilon_B\right)\right)$, which allows to conclude using the extreme value theorem on $[a_0, \bar{a} - \varepsilon_A] \times [b_0, \bar{b} - \varepsilon_B]$.

We now turn to proving the properties of the equilibrium. Given that all other voters use strategy (a, b), i's expected marginal utility from a^i is

$$\begin{aligned} \frac{\partial}{\partial a^i} EU^{\kappa}\left(a, b, a^i, b^i\right) &= (a_v - \rho b_v) h'(\alpha^{\kappa}(a, b, a^i, b^i)) \frac{\partial \alpha^{\kappa}(a, b, a^i, b^i)}{\partial a^i} - a_v C_A'(a^i) \\ = (a_v - \rho b_v)\, h'\left(\alpha^{\kappa}(a, b, a^i, b^i)\right) & \frac{2\kappa[(1-\kappa)b + \kappa b^i]}{[(1-\kappa)a + \kappa a^i + (1-\kappa)b + \kappa b^i]^2} - a_v C_A'\left(a^i\right) \end{aligned} \quad (51)$$

and the expected marginal utility from b^i is:

$$\begin{aligned} \frac{\partial}{\partial b^i} EU^{\kappa}\left(a, b, a^i, b^i\right) &= (\rho b_v - a_v) h'(\beta^{\kappa}(a, b, a^i, b^i)) \frac{\partial \beta^{\kappa}(a, b, a^i, b^i)}{\partial b^i} - b_v C_B'(b^i) \\ = (\rho b_v - a_v)\, h'\left(\beta^{\kappa}(a, b, a^i, b^i)\right) & \frac{2\kappa[(1-\kappa)a + \kappa a^i]}{[(1-\kappa)a + \kappa a^i + (1-\kappa)b + \kappa b^i]^2} - b_v C_B'\left(b^i\right). \end{aligned} \quad (52)$$

Since $C_A'(a_0) = C_B'(b_0) = 0$, and $C_A(a^i), C_B(b^i) > 0$ for $a^i > 0$ or $b^i > 0$ (see (19)), a best response (a^i, b^i) to (a, b) has $a^i > a_0$ if and only if the first of the two terms in (51) is strictly positive. Since h is strictly increasing, and $b \geq b_0 > 0$, this is true if and only if $\kappa(a_v - \rho b_v) > 0$; otherwise, any best response has $a^i = a_0$. Likewise, a best response (a^i, b^i) to (a, b) has $b^i > b_0$ if and only if $\kappa(\rho b_v - a_v) > 0$; otherwise, any best response has $b^i = b_0$. These arguments prove the proposition 6. □

This proposition shows that both the degree of universalization (κ) and the stake of the election for the underdog supporters (ρ) matter for the qualitative nature of the set of equilibria. First, if voters have no universalization concerns ($\kappa = 0$) or if the expected benefit that one group obtains from a positive margin of its candidate exactly outweighs the expected cost that the other group garners from this margin ($\rho = a_v/b_v$), then turnout is confined to the bases a_0 and b_0, in which case the underdog wins if and only if it has a larger base than the leader ($b_0 > a_0$). Second, whenever $a_v \neq \rho b_v$, any positive degree of universalization $\kappa > 0$ triggers participation of a positive mass of cost-sensitive voters. The reason is clear: a $\kappa > 0$ triggers in the individual voter a utility kick from contemplating the margin that her preferred candidate would obtain if all the other cost-sensitive voters selected the same strategy as herself; the voter is willing to incur a positive voting cost to obtain this utility kick. Third, and in stark contrast with the partisan setting, here a voter internalizes the negative externality that voting for one candidate has on the group supporting the other candidate. Hence, she votes only if she belongs to the group that obtains the highest expected benefit from its candidate's margin. The fact that voters in one of the groups do not vote facilitates equilibrium existence, since the A-supporters and B-supporters will thus not seek to outbid each other. Equilibrium characterization is also facilitated, since it now suffices to identify turnouts such that the voters who do vote, best-respond only to their co-partisans.

4.3 Winner-take-all: equilibrium characterization

We adopt the same approximation as in the partisan setting to analyze the winner-take-all limit and obtain the fourth main result, which states that under non-partisan ethics winner-take-all elections may result in two situations only: either cost-sensitive voters incur no cost and the side with the largest base wins (outcome (a_0, b_0)), or the side with the largest base pays no cost and the side with the smallest base pays to match the other base (outcome (a_0, a_0) or (b_0, b_0)).

The next proposition states the result formally and details the conditions for these outcomes to hold. For the purpose of this proposition, if $\rho b_v - a_v > 0$ let $\tilde{\kappa}_B$ be implicitly defined by

$$2\left(\rho b_v - a_v\right) = b_v C_B \left(\frac{a_0 - b_0}{\tilde{\kappa}_B} + b_0\right), \tag{53}$$

while if $a_v - \rho b_v > 0$ let $\tilde{\kappa}_A$ be implicitly defined by

$$2\left(a_v - \rho b_v\right) = a_v C_A\left(\frac{b_0 - a_0}{\tilde{\kappa}_A} + a_0\right). \tag{54}$$

Proposition 7 (Non-partisan ethics)**.** *In the limit "winner-take-all" case, let (a, b) be sustained as limit equilibrium.*

1. *If $\rho b_v - a_v > 0$, then $(a, b) \in \{(a_0, b_0), (a_0, a_0)\}$. Moreover:*
   (a) *If $b_0 > a_0$, (a_0, b_0) is the unique limit equilibrium.*
   (b) *If $b_0 < a_0$, (a_0, b_0) is a limit equilibrium if and only if $\kappa \le \tilde{\kappa}_B$, where $\tilde{\kappa}_B$ is defined in (53) if $C_B\left(\bar{b}\right) \ge 2\left(\rho b_v - a_v\right)$, and $\tilde{\kappa}_B = 0$ otherwise. Moreover, (a_0, a_0) is a limit equilibrium if and only if $2\left(\rho b_v - a_v\right) > b_v C_B\left(a_0\right)$ and $\kappa > 0$ and $a_0 \le \bar{b}$.*
2. *If $a_v - \rho b_v > 0$, then $(a, b) \in \{(a_0, b_0), (b_0, b_0)\}$. Moreover:*
   (a) *If $a_0 > b_0$, (a_0, b_0) is the unique limit equilibrium.*
   (b) *If $a_0 < b_0$, (a_0, b_0) is a limit equilibrium if and only if $\kappa \le \tilde{\kappa}_A$, where $\tilde{\kappa}_A$ is defined in (54) if $C_A\left(\bar{a}\right) \ge 2\left(a_v - \rho b_v\right)$, and $\tilde{\kappa}_A = 0$ otherwise. Moreover, (b_0, b_0) is a limit equilibrium if and only if $2\left(a_v - \rho b_v\right) > a_v C_A\left(b_0\right)$ and $\kappa > 0$ and $b_0 \le \bar{a}$.*

Sketch of proof. (The full proof can be found in Appendix A.2.) Without loss of generality, we can assume $\rho b_v - a_v > 0$. Recall from Proposition 6 that for any sequence (h_t), for all t, an equilibrium exists and is such that $a_t = a_0$.

We first show that $(a, b) \in \{(a_0, b_0), (a_0, a_0)\}$. This follows from the first-order condition: indeed, due to the uniform convergence of h_t' towards zero, the benefit term vanishes in the limit, except around 0, which, if attainable, corresponds to $b_0 = a_0$. Therefore, in the limit, the solution of the first-order condition either features no cost ($b = b_0$) or converges to a_0.

In the case $b_0 > a_0$, we use the concavity of h_t on $[0, 1]$ in order to show uniqueness of the equilibrium for all t. By the same argument as above, it follows that $b_t \to b_0$.

In the case $b_0 < a_0$, the proofs around $\tilde{\kappa}_B$ closely mirror the proof of Proposition 3. For the case $2\left(\rho b_v - a_v\right) > b_v C_B\left(a_0\right)$, we first prove that the necessary condition for the equilibrium admits a unique solution on $[a_0, \bar{b}]$ for t large enough. Moreover, using the uniform convergence of h_t' on closed intervals that do not contain 0, we show that the solution of the necessary condition gets arbitrarily close to a_0. In order to finish the proof of existence of an equilibrium, we must exclude profitable deviations. First, we argue that due to the concavity of the objective on that area, there cannot be profitable deviations b_t^i such that $b_t^\kappa \in \left[a_0, \bar{b}\right]$. This also implies that $h_t'\left(\beta\left(a_0, b_t\right)\right) \to 1$, as otherwise one could construct a profitable upward deviation for large enough t. Finally, we distinguish deviations such that $a_0 > b_t^\kappa > a_0 - \varepsilon$ and those such that $b_t^\kappa \leq a_0 - \varepsilon$. For deviations such that $a_0 > b_t^\kappa > a_0 - \varepsilon$, there is a minimum loss in the benefit term that does not depend on ε, whereas the cost savings from the deviation can be made arbitrarily small for large t by choosing a smaller ε. Then, we can exclude that deviations such that $b_t^\kappa \in [b_0, a_0 - \varepsilon]$ are profitable: the benefit of any such deviation, by the uniform convergence of h_t, converges to $2(\rho b_v - a_v)$, whereas the cost can be reduced by at most $b_v C_B(a_0)$. In the case $2(\rho b_v - a_v) \geq b_v C_B(a_0)$, we assume for a contradiction that a limit equilibrium exist and show that deviating to b_0 is profitable for any t as the cost savings exceed the losses in the benefit term. □

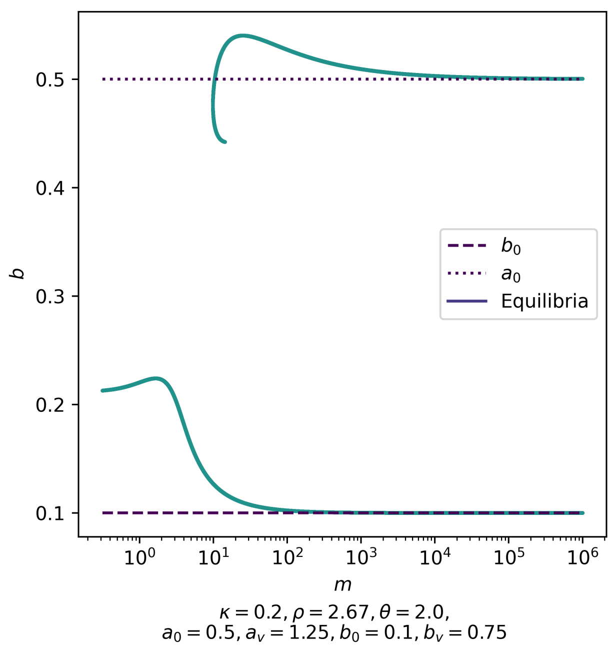


Figure 8: Possible equilibrium b for varying m in the case $\rho b_v > a_v$

Example 3. The two equilibrium types can co-exist. An illustration is provided in Figure 8, for the case $\rho b_v > a_v$. Each solid line shows equilibrium turnouts among the B-supporters, as a function of the curvature parameter m of the benefit function h. For m small enough, there is a unique equilibrium, with a rather low turnout. When m reaches a value around 10, a second equilibrium appears, with a higher turnout. The coexistence of an equilibrium with a participation rate close to 0 and one with a participation rate that matches A's base a_0, that is, those that can be supported in the limit case, appears for values of m around 10^5 and above.

4.4 Power sharing: discussion

Returning to the power-sharing setting, we examine only non-trivial settings where $\kappa(a_v - \rho b_v) \neq 0$. To begin, consider the case $a_v > \rho b_v$. Then, any equilibrium (a^*, b^*) is such that only the base turns out among the B-supporters ($b^* = b_0$). The marginal utility at strategy $a^i = a$ for any A-supporter i, given that the other A-supporters select strategy a, thus writes:

$$A(a) \equiv \frac{\partial}{\partial a^i} EU^\kappa \left(a, b, a^i, b^i\right)|_{a^i=a, b=b^i=b_0} = (a_v - \rho b_v) h'(\alpha(a, b_0)) \frac{2\kappa b_0}{(a + b_0)^2} - a_v C_A'(a) . \tag{55}$$

The necessary first-order condition for a^* to be an equilibrium is $A(a^*) \geq 0$, which must hold as an equality if a^* is interior, $a^* \in (a_0, \bar{a})$. The necessary second-order condition for such an interior solution is:

$$\begin{aligned} \frac{\partial^2}{\partial (a^i)^2} EU^\kappa \left(a, b, a^i, b^i\right)|_{a^i=a^*, b=b^i=b_0} &= (a_v - \rho b_v) h''(\alpha(a^*, b_0)) \frac{4\kappa^2 b_0^2}{(a^* + b_0)^4} \\ &- (a_v - \rho b_v) h'(\alpha(a^*, b_0)) \frac{4\kappa^2 b_0}{(a^* + b_0)^3} \\ &- a_v C_A''(a^*) \leq 0. \end{aligned} \tag{56}$$

Likewise, for $\rho b_v > a_v$, define

$$B(b) \equiv \frac{\partial}{\partial b^i} EU^\kappa \left(a, b, a^i, b^i\right)|_{a=a^i=a_0, b^i=b} = (\rho b_v - a_v) h' \left(\beta(a_0, b)\right) \frac{2\kappa a_0}{(a_0 + b)^2} - b_v C_B'(b) , \tag{57}$$

so that the necessary first-order condition for any equilibrium b^* is $B(b^*) \geq 0$, which must hold as an equality if b^* lies in the interior $(b_0, \bar{b})$. The necessary second-order condition for such an interior solution is:

$$\begin{aligned} \frac{\partial^2}{\partial (b^i)^2} EU^\kappa \left(a, b, a^i, b^i\right)|_{a=a^i=a_0, b^i=b^*} &= (\rho b_v - a_v) h''(\beta(a_0, b^*)) \frac{4\kappa^2 a_0^2}{(a_0 + b^*)^4} \\ &- (\rho b_v - a_v) h'(\beta(a_0, b^*)) \frac{4\kappa^2 a_0}{(a_0 + b^*)^3} - b_v C_B''(b^*) \leq 0. \end{aligned} \tag{58}$$

If there is a unique a^* (respectively b^*) satisfying the first-order condition, then it is the unique equilibrium (by virtue of Proposition 6). We begin by identifying a sufficient condition for this to obtain (for completeness, in the proof we also show that the second-order condition holds as a strict inequality).

Proposition 8 (Non-partisan ethics)**.** *Suppose that $\kappa \in (0, 1]$. If $a_v > \rho b_v$ and $a_0 \geq b_0$, there is a unique equilibrium (a^*, b^*). At this equilibrium, the underdog supporters pay no cost: $b^* = b_0$ and the leader wins: $\alpha(a^*, b_0) > 0$. Likewise, if $\rho b_v > a_v$ and $b_0 \geq a_0$, there is a unique equilibrium (a_0, b^*). At this equilibrium, the leader supporters pay no cost: $a^* = a_0$ and the underdog wins: $\beta(a_0, b^*) > 0$.*

Proof. It is sufficient to prove the result for one of the cases, say $\rho b_v > a_v$. We begin by proving the following claim: if $h''(\beta(a_0, b)) \leq 0$ for all $b \in [b_0, \bar{b}]$ there either exists a unique $b^* \in (b_0, \bar{b})$ satisfying $B(b^*) = 0$ and such that (58) holds strictly, or $B(b) > 0$ for all $b \in \left(b_0, \bar{b}\right)$. To see this, note first that if $h''(\beta(a_0, b)) \leq 0$ for all $b \in [b_0, \bar{b}]$, then the first term in $B(b^*)$ is non-increasing in b^*; this term is also strictly positive for $b^* = b_0$ (since $h' > 0$). Since the second term equals 0 for $b^* = b_0$ and is strictly increasing in b^*, the claim follows.

If $b_0 \geq a_0$, the two statements in the proposition then follow immediately from the fact that $\rho b_v > a_v$ implies $b^* > b_0$ and $a^* = a_0$. Indeed, $\beta(a_0, b)$ is thus strictly positive for any $b \in (b_0, \bar{b})$, and our assumptions on h then imply that $h''(\beta(a_0, b)) \leq 0$ for any $b \in (b_0, \bar{b}]$. □

Just like in the partisan setting, the bases a_0 and b_0 play a crucial role. If the cost-sensitive voters who do participate in the election can rely on a larger base than that of the other candidate, then they win the election independent of their turnout rate. This implies a decreasing marginal benefit and an increasing marginal cost of increases in the turnout, and hence equilibrium uniqueness[14]

Next, we examine what happens if the group that votes (i.e., the leader supporters if $a_v > \rho b_v$ and the underdog supporters if $\rho b_v > a_v$) has a smaller base than the other group. The marginal benefit is then increasing for turnout rates close enough to the base, so that there may be multiple candidates a^* (respectively b^*) satisfying the first- and second-order conditions. Each such candidate is an equilibrium if there do not exist utility-enhancing global deviations. The main question we investigate is whether equilibrium uniqueness obtains. Examination of the special case of full universalization provides some initial insights.

Proposition 9 (Non-partisan ethics)**.** *Suppose that $\kappa = 1$. If there exist multiple equilibria, they all generate the same expected utility to the cost-sensitive voters.*

Proof. It is sufficient to prove the result for one of the cases, say $\rho b_v > a_v$. Plugging in $a = a_0$ and $\kappa = 1$ into the expected utility (45), the expected utility becomes independent of b, the turnout rate among the other voters, and thus a function of b^i only:

$$EU^\kappa(a_0, b, a_0, b^i) = (\rho b_v - a_v) h\left(\frac{b^i - a_0}{a_0 + b^i}\right) - C_B(b^i). \tag{59}$$

Hence, each individual voter simply chooses some value of b^i that maximizes this expression. If there are multiple solutions, they must yield the same expected utility. □

Clearly, if $\kappa = 1$, each voter's objective function coincides with that of a utilitarian planner, and hence, if there are multiple equilibrium turnout rates, they must give rise to

[14]See Appendix D.3 for a more general result on equilibrium uniqueness, which relies on the same logic.

the same expected utility. The following example showcases such a situation, for the case $\rho b_v > a_v$. This example is interesting also for two other reasons: first, it shows that full universalization does not guarantee a high turnout rate; second, and similar to the partisan setting, the different equilibria do not necessarily lead to the same winner.

Example 4 (Multiple equilibria under full universalization ($\kappa = 1$))**.** Each curve in Figure 9 shows the expected utility for a B-supporter i as a function of their strategy b^i, for three different values for the mass of cost-sensitive B-supporters, b_v. For one of these values, namely $b_v \approx 0.609$, and the other parameter values are as specified in the figure legend, there are two equilibria, shown as stars: the underdog wins at one of them ($b^* \approx 0.6 > a_0$) but loses at the other (b^* is close to $b_0 = 0.2$). The figure further shows that small variations in the parameter values may induce discrete jumps in b^*: for b_v slightly above 0.609, there is a unique equilibrium turnout, at which the underdog wins, while for b_v slightly below 0.609, there is a unique equilibrium turnout, close to b_0.

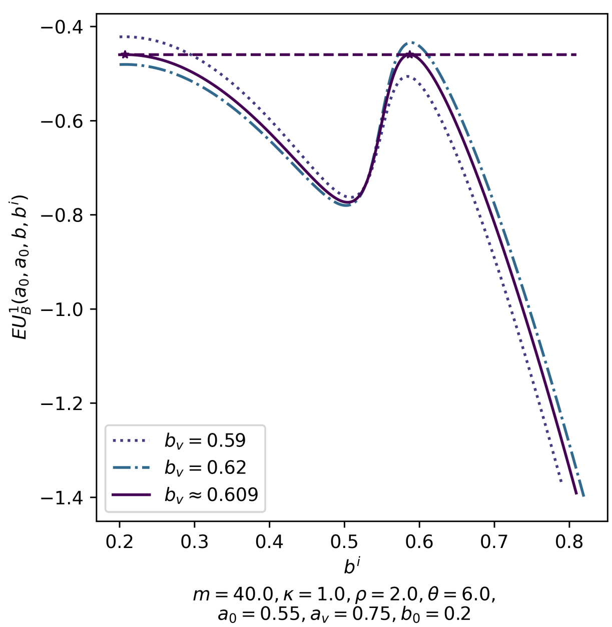


Figure 9: Existence of two equilibria when $\kappa = 1$

While multiplicity of equilibria appears only in knife-edge cases under full universalization ($\kappa = 1$), it is not an uncommon phenomenon under partial universalization ($\kappa \in (0, 1)$). By contrast to the full universalization case, these equilibria do not necessarily give rise to the same utility for those who turn out, as shown in the following example.

Example 5 (Multiple equilibria under partial universalization ($\kappa \in (0, 1)$))**.** Considering still the case $\rho b_v > a_v$, Figure 10 shows, for $\kappa = 0.4$, an example with two equilibrium strategies

b^*, indicated by stars. The curve going through each of these points shows the expected utility of some B-supporter i, given that the other B-supporters turn out in accordance with the indicated equilibrium. Each curve shows that this voter does not wish to deviate, since the utility is lower for all other values of b^i than b^*. Like in the example under full universalization in Figure 9, here one equilibrium turnout is close to the base, $b^* \approx 0.12$, while the other makes the underdog win, $b^* \approx 0.49 > a_0 = 0.45$. This victory obtains despite the underdog's base being weak compared to that of the leader (compare $b_0 = 0.1$ to $a_0 = 0.45$). We further see in the figure that the high-turnout equilibrium gives a substantially higher expected utility than the low-turnout one.

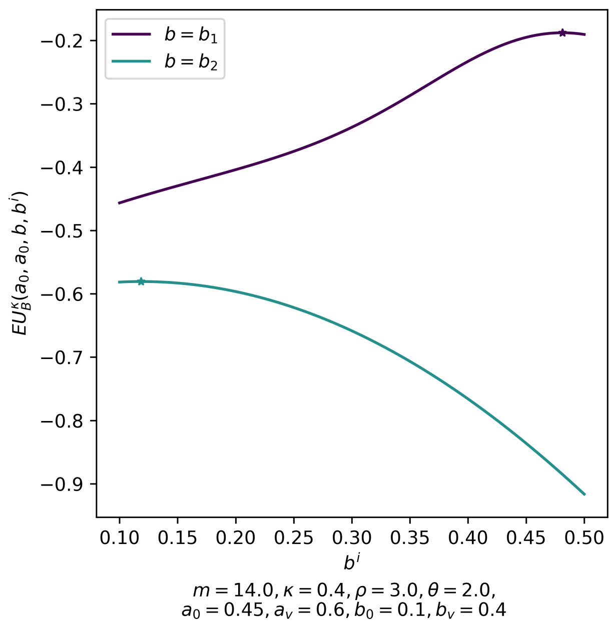


Figure 10: Utility from deviating from each equilibrium: $b_1 \approx 0.49$, $b_2 \approx 0.12$

In the preceding example the underdog supporters face a coordination problem, and they would prefer to coordinate on the high-turnout equilibrium. A question of interest is whether such coordination problems — i.e., co-existence of equilibria with substantially different turnouts, where one equilibrium is preferred to the other(s) — are common. We here examine the necessary conditions for global deviations not to exist. This will provide some insights about parameter values that might give rise to such multiplicity of equilibria. We do this for the case $\rho b_v > a_v$.

Thus, consider some turnout rate $b \in (b_0, \bar{b})$. For b to be an equilibrium, an individual voter must not wish to deviate to any $b' \neq b$. Considering first downward deviations $b' < b$, the following condition must hold:

$$(\rho b_v - a_v)\left[h(\beta(a_0, b)) - h(\beta^\kappa(a_0, b, a_0, b'))\right] \geq C_B(b) - C_B(b') \quad \forall b' \in [b_0, b), \tag{60}$$

where

$$\beta\left(a_0, b\right) = \frac{b - a_0}{a_0 + b} \quad \text{and} \quad \beta^\kappa\left(a_0, b, a_0, b'\right) = \frac{(1-\kappa)b + \kappa b' - a_0}{a_0 + (1-\kappa)\,b + \kappa b'}. \tag{61}$$

Any downward deviation $b' < b$ reduces the cost, i.e., the right-hand side of (60) is strictly positive. For any $b' < b$ the left-hand side is equal to zero if $\kappa = 0$, and it is increasing in κ: the utility loss that the voter incurs from a decline in its preferred candidate's margin gets larger as her degree of universalization gets larger. Hence, any value of $\kappa > 0$ imposes an upper bound on the turnout rate that can be sustained in equilibrium. In particular, the voter must not be tempted by abstention ($b' = b_0$), the deviation that would maximize the cost saving $C_B(b) - C_B(b')$, and we note that the deviation to abstention generates a cost saving that is larger the smaller is the base b_0. Taken together, these observations suggest that the underdog supporters can achieve a victory only if κ is large enough, and that this constraint on κ is stronger the weaker is the base b_0. Noting further that C_B is decreasing in the size of the cost-sensitive electorate b_v (see (17)), *ceteris paribus* the constraint on κ is also stronger the smaller is b_v. Finally, (60) clearly implies that the stake for the B-supporters (ρ) must be large for an equilibrium with a higher turnout to be sustained.

Considering now upward deviations $b' > b$, the following condition must hold for an individual not to wish to deviate:

$$C_B(b') - C_B(b) \geq (\rho b_v - a_v)\left[h(\beta^\kappa(a_0, b, a_0, b')) - h(\beta(a_0, b)\right] \quad \forall b' \in (b, \bar{b}]. \tag{62}$$

Any upward deviation $b' > b$ raises the cost, i.e., the left-hand side is strictly positive. But it also raises the utility gain that the voter obtains from an increase in its preferred candidate's margin, as long as her degree of universalization is strictly positive: the right-hand side equals zero if $\kappa = 0$ and is increasing in κ. Hence, for any $\kappa > 0$ there is a lower bound on the turnout rate that can be sustained in equilibrium. In particular, if the underdog has a small base b_0 and κ is close enough to 1 — so that the right-hand side of (62) is large — we should expect existence of equilibria with a turnout rate close to b_0 only if voting costs are high enough. Since C_B is decreasing in the size of the cost-sensitive electorate (b_v), *ceteris paribus* low turnout equilibria also require b_v to be small enough. Finally, (62) clearly implies that low turnout equilibria are more likely to be sustained the smaller is the stake ρ.

Taken together, the preceding arguments suggest that the aforementioned coordination problem should be expected only if κ is neither too large nor too small, $a_0 - b_0$ is large enough, and $\rho b_v - a_v$ is neither too large nor too small.

Recalling that the same arguments apply to the case $a_v > \rho b_v$, the leader's supporters may also face a coordination problem: if the leader's base a_0 is small enough compared to that of the underdog b_0, and κ is moderate, then there may exist two equilibria, one with a high turnout and with a low turnout, and the leader may suffer a sizeable loss in the latter.

Further examples are provided in Appendix C.2, and Appendix C.3 provides sufficient conditions for there to be a unique equilibrium.

5 Discussion and conclusions

In this paper we have tried to understand what follows if some people base their turnout decisions on an argument of the form "Voting is the right thing to do because there would be bad consequences for me if too many people abstained," formalized through the *Homo moralis* preferences (Alger and Weibull, 2013), which capture well such partial universalization. The point is particularly relevant in circumstances where voting is costly and each single vote has a negligible effect on the relative number of votes obtained by the candidates, a feature that we capture by modeling the electorate as a continuum.

We find, first, that any extent of *Homo moralis* universalization ethics, i.e., any positive κ in the model, justifies participation in large electorates in most cases. It is generally true as long as voters perceive some benefit to any increase in the favorite candidate's margin. This corresponds to the power sharing setting of our model, where both the winning and the losing side stand to gain from further increasing their share of the expressed votes. By contrast, *Homo moralis* universalization ethics do not generally justify participation in large electorates in the winner-take-all limit case of our model, where there is a second reason for why votes have a negligible impact: the marginal benefit from further increasing the share of expressed votes is nil except at the point where the two candidates tie precisely.

Second, our analysis reveals why it is important for a candidate to have a large base, that is, a large share of voters who always turn out to vote for them. Such a base motivates the cost-sensitive voters to vote, by reducing the cost of garnering the utility benefit from envisaging the positive consequences of participation. The base is then a complement, and not a substitute, to the turnout of cost-sensitive voters. We show that a large enough base for the underdog compared to that of the topdog can even trigger a large enough turnout among the cost-sensitive voters for the underdog to win the election. This can occur even if the underdog's supporters do not perceive a particularly high stake in the election (i.e., even if the stake parameter $\rho = 1$), and the effect is strengthened if the stake is not neutral ($\rho > 1$). That being said, if the base is too large, it becomes a substitute for the cost-sensitive voters' turnout, since it reduces the marginal benefit of higher turnout rates.

A third pattern is that high values of κ do not necessarily guarantee high turnout rates, because voters may face coordination problems. Indeed, similar to the base of one's group, an increase in the share of other cost-sensitive voters who are expected to vote dampens the cost that an individual cost-sensitive voter needs to incur to reach a certain benefit. While this explains why there may exist equilibria with high turnout rates, it also explains why such equilibria can sometimes co-exist with equilibria with very low turnout rates.

Many questions remain. In particular, it would be interesting to extend the setting to heterogeneity in the degrees of universalization, to more than two parties, and a finite electorate. On the empirical side, while several motivations behind turnout decisions have already been documented and studied (see, e.g., Aytaç and Stokes, 2019; Blais, 2000; Blais and Daoust, 2020; Carlsson and Johansson-Stenman, 2010; Downs, 1957; Gerber et al., 2008; Hatemi and McDermott, 2011), it appears that no study has sought to detect universalization ethics as a driver of turnout decisions. Relatedly, it might be fruitful to examine whether *Homo moralis* preferences can help explain some of the patterns in the experimental literature on costly turnout, in particular, turnout rates above those predicted by the pivotal-voter model, higher turnout rates in close elections, and the mixed evidence on "underdog effects", whereby individuals in the minority tend to vote at higher rates than those in the majority, and "bandwagon effects", where the opposite is true (see, e.g., Agranov, Goeree, Romero, and Yariv, 2018; Blais, Pilet, Van der Straeten, Laslier, and Héroux-Legault, 2014; Duffy and Tavits, 2008; Faravelli, Kalayci, and Pimienta, 2020; Großer and Schram, 2010; Hizen, Kikuchi, Koriyama, and Masuda, 2025; D. K. Levine and Palfrey, 2007).

References

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

Alger, I., & Laslier, J.-F. (2022). *Homo moralis* goes to the voting booth: Coordination and information aggregation. *Journal of Theoretical Politics*, *34*(2), 280–312.

Alger, I., & Weibull, J. W. (2013). *Homo moralis*—preference evolution under incomplete information and assortative matching. *Econometrica*, *81*(6), 2269–2302.

Alger, I., & Weibull, J. W. (2017). Strategic behavior of moralists and altruists. *Games*, *8*(3).

Alger, I., Weibull, J. W., & Lehmann, L. (2020). Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory*, *185*, 104951.

Aytaç, S. E., & Stokes, S. C. (2019). *Why Bother? Rethinking Participation in Elections and Protests*. Cambridge: Cambridge University Press.

Bergstrom, T. C. (1995). On the evolution of altruistic ethical rules for siblings. *American Economic Review*, *85*(1), 58–81.

Bierbrauer, F., Tsyvinski, A., & Werquin, N. (2022). Taxes and turnout: When the decisive voter stays at home. *American Economic Review*, *112*(2), 689–719.

Blais, A. (2000). *To Vote or Not to Vote? the Merits and Limits of Rational Choice Theory.* Pittsburgh: University of Pittsburgh Press.

Blais, A., & Achen, C. H. (2019). Civic duty and voter turnout. *Political Behavior*, *41*, 473–497.

Blais, A., & Daoust, J.-F. (2020). *The Motivation to Vote: Explaining Electoral Participation.* Vancouver: University of British Columbia Press.

Blais, A., Pilet, J.-B., Van der Straeten, K., Laslier, J.-F., & Héroux-Legault, M. (2014). To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections. *Electoral Studies*, *36*, 39–50.

Bomze, I., Schachinger, W., & Weibull, J. (2021). Does moral play equilibrate? *Economic Theory*, *71*(1), 305–315. doi:10.1007/s00199-020-01246-4

Braham, M., & van Hees, M. (2020). *Kantian* Kantian optimization. *Erasmus Journal for Philosophy and Economics*, *13*, 30–42.

Cancela, J., & Geys, B. (2016). Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies*, *42*, 264–275.

Capraro, V., & Rand, D. G. (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*, *13*, 99–111.

Carlsson, F., & Johansson-Stenman, O. (2010). Why Do You Vote and Vote as You Do? *Kyklos*, *63*(4), 495–516.

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

Conley, J. P., Toossi, A., & Wooders, M. (2006). Memetics and voting: How nature may make us public spirited. *International Journal of Game Theory*, *35*, 71–90.

Daley, B., & Sadowski, P. (2017). Magical thinking: A representation result. *Theoretical Economics*, *12*(2), 909–956.

DeNardo, J. (1980). Turnout and the vote: The joke's on the Democrats. *American Political Science Review*, *74*(2), 406–420.

Dhillon, A., & Peralta, S. (2002). Economic theories of voter turnout. *Economic Journal*, *112*(480), F332–F352.

Downs, A. (1957). *An Economic Theory of Democracy.* New York: Harper and Row.

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

Eichner, T., & Pethig, R. (2021). Climate policy and moral consumers. *Scandinavian Journal of Economics*, *123*(4), 1190–1226.

Eichner, T., & Pethig, R. (2022). Kantians defy the economists' mantra of uniform Pigovian emissions taxes. *Ecological Economics*, *200*, 107514.

Faravelli, M., Kalayci, K., & Pimienta, C. (2020). Costly voting: a large-scale real effort experiment. *Experimental Economics*, *23*(2), 468–492.

Feddersen, T., Gailmard, S., & Sandroni, A. (2009). Moral bias in large elections: Theory and experimental evidence. *American Political Science Review*, *103*(2), 175–192.

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

Fiorina, M. P. (1976). The voting decision: Instrumental and expressive aspects. *Journal of Politics*, *38*(2), 390–413.

Frank, R. W., & Martínez i Coma, F. (2023). Correlates of voter turnout. *Political Behavior*, *45*(2), 607–633.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, *102*(1), 33–48.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, Mass.: Harvard University Press.

Gravel, N., Laslier, J.-F., & Trannoy, A. (2000). Consistency between tastes and values: A universalisation approach. *Social Choice and Welfare*, *17*, 129–137.

Großer, J., & Schram, A. (2010). Public opinion polls, voter turnout, and welfare: An experimental study. *American Journal of Political Science*, *54*(3), 700–717.

Harsanyi, J. C. (1977). Morality and the theory of rational behavior. *Social Research: An International Quarterly*, *44*(4), 623–656.

Harsanyi, J. C. (1980). Rule utilitarianism, rights, obligations and the theory of rational behavior. *Theory and Decision*, *12*(2), 115–133.

Hatemi, P. K., & McDermott, R. (2011). *Man is by Nature a Political Animal: Evolution, Biology, and Politics*. Chicago: University of Chicago Press.

Hausman, D., & McPherson, M. (1996). *Economic Analysis and Moral Philosophy*. Cambridge, Mass.: Cambridge University Press.

Herrera, H., Morelli, M., & Nunnari, S. (2016). Turnout across democracies. *American Journal of Political Science*, *60*(3), 607–624.

Hizen, Y., Kikuchi, K., Koriyama, Y., & Masuda, T. (2025). Jumping on the bandwagon and off the Titanic: An experimental study of turnout in two-tier voting. *European Journal of Political Economy*, *86*, 102646.

Juan-Bartroli, P., & Karagözoğlu, E. (2024). Moral preferences in bargaining. *Economic Theory*. Retrieved from https://doi.org/10.1007/s00199-023-01544-7

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten. [in English: Groundwork of the Metaphysics of Morals (1964).* New York: Harper Torch Books].

Krishna, V., & Morgan, J. (2015). Majority rule and utilitarian welfare. *American Economic Journal: Microeconomics*, *7*(4), 339–375.

Laffont, J.-J. (1975). Macroeconomic constraints, economic efficiency and ethics: An introduction to Kantian economics. *Economica*, *42*(168), 430–437.

Laslier, J.-F. (2023). Universalization and altruism. *Social Choice and Welfare*, *60*, 579–594.

Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public Choice*, *44*(1), 7–41.

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

Levine, D. K., & Palfrey, T. R. (2007). The Paradox of Voter Participation? A Laboratory Study. *American Political Science Review*, *101*(1), 143–158.

Levine, S., Kleiman-Weber, M., Schultz, L., & Cushman, F. (2020). The logic of universalization guides moral judgment. *PNAS*, *117*, 26158–26169.

Lizzeri, A., & Persico, N. (2001). The provision of public goods under alternative electoral incentives. *American Economic Review*, *91*(1), 225–239.

Miettinen, T., Kosfeld, M., Fehr, E., & Weibull, J. W. (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization*, *173*, 1–25.

Mohanty, S., Rao, K. M., & Roy, J. (2024). Kantian imperatives in public goods networks. *Journal of Economic Behavior & Organization*, *224*, 194–214.

Morton, R. B. (1987). A group majority voting model of public good provision. *Social Choice and Welfare*, *4*, 117–131.

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

Muñoz-Sobrado, E. (2022). *Taxing moral agents*. CESifo Working Paper No. 9867.

Myerson, R. B. (2000). Large Poisson games. *Journal of Economic Theory*, *94*(1), 7–45.

Nagel, J. H., & McNulty, J. E. (1996). Partisan effects off voter turnout in senatorial and gubernatorial elections. *American Political Science Review*, *90*(4), 780–793.

Norman, T. W. (2020). The evolution of monetary equilibrium. *Games and Economic Behavior*, *122*, 233–239.

Palfrey, T. R., & Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review*, *79*(1), 62–78.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Cambridge University Press.

Riker, W. H., & Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review*, *62*(1), 25–42.

Roemer, J. E. (2010). Kantian equilibrium. *The Scandinavian Journal of Economics*, *112*(1), 1–24.

Rogers, T., Green, D. P., Ternovski, J., & Young, C. F. (2017). Social pressure and voting: A field experiment conducted in a high-salience election. *Electoral Studies*, *46*, 87–100.

Rousseau, J.-J. (1762). *Du contrat social*. English trans. by V. Gourevitch: The Social Contract and Other Later Political Writings. Cambridge: Cambridge University Press.

Salonia, E. M. (2024). *A foundation for universalisation in games*. Toulouse School of Economics Working Paper No. 1586.

Sarkisian, R. (2017). Team incentives under moral and altruistic preferences: Which team to choose? *Games*, *8*(3), 37.

Sarkisian, R. (2021). Screening teams of moral and altruistic agents. *Games*, *12*(4).

Schuessler, A. A. (2000). Expressive voting. *Rationality and Society*, *12*(1), 87–119.

Tappin, B. M., & Capraro, V. (2018). Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. *Journal of Experimental Social Psychology*, *79*, 64–70.

Van Leeuwen, B., & Alger, I. (2024). Estimating social preferences and Kantian morality in strategic interactions. *Journal of Political Economy Microeconomics*, *2*, 665–706.

A Proofs of Propositions 3 and 7

A.1 Proof of Proposition 3

Without loss of generality, suppose that $a_0 > b_0$.

Part 1: If $\kappa \leq \tilde{\kappa}_B$, then (a_0, b_0) can be sustained as a limit winner-take-all equilibrium.

Let $\kappa \leq \tilde{\kappa}_B$. In order to show that (a_0, b_0) can be sustained as a limit winner-take-all equilibrium, we need to find an approximating sequence. For the benefit functions, we take the sequence specified in (2) of the main text:

$$h_t(x) = \frac{1}{\arctan(t)} \arctan(tx). \tag{63}$$

First, we will establish the existence of a sequence (a_t, b_t) of solutions to the first-order conditions such that $(a_t, b_t) \to (a_0, b_0)$. Using $h'(x) = h'(-x)$, the first-order conditions write

$$h_t'\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{2\kappa b_t}{(a_t + b_t)^2} = C_A'(a_t)$$
$$h_t'\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{2\kappa\rho a_t}{(a_t + b_t)^2} = C_B'(b_t)$$

Let $\varepsilon > 0$ be such that $b_0 + \varepsilon < a_0$. As $a \geq a_0 > 0$ and $b \geq b_0 > 0$, and by the uniform convergence of h_t' shown in Lemma 2, the left-hand sides of the first-order conditions converge to 0, as long as $(a_t, b_t) \in [b_0, b_0 + \varepsilon] \times [a_0, \bar{a}]$. Let $\delta = \min\{C_A'(a_0 + \varepsilon), C_B'(b_0 + \varepsilon)\}$. Then, there exists $T \in \mathbb{N}$ such that for all $t \geq T$, the left-hand sides are bounded by δ. For a fixed a_t, we can then find $b_t^* \in [b_0, b_0 + \varepsilon]$ which solves the first-order condition for b_t, and for a fixed b_t, we can find $a_t^* \in [a_0, a_0 + \varepsilon]$. We will now establish the uniqueness of such solutions. Consider

$$h_t'\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{2\kappa b_t}{(a_t + b_t)^2} - C_A'(a_t) = 0$$
$$h_t'\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{2\kappa\rho a_t}{(a_t + b_t)^2} - C_B'(b_t) = 0$$

The first-order condition for a_t is then strictly decreasing by the convexity of C_A and our assumptions on h_t. Therefore, there exists a unique solution a_t^* to the first order condition that depends continuously on b_t (because of the uniqueness and $[b_0, b_0 + \varepsilon]$ being compact).

Differentiating the first-order condition for b_t further,

$$-h_t''\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{4\kappa\rho a_t^2}{(a_t + b_t)^4} - h_t'\left(\frac{a_t - b_t}{a_t + b_t}\right)\frac{4\kappa\rho a_t}{(a_t + b_t)^3} - C_B''(b_t)$$

Since f_B is continuous on its support, $C_B''(b_t) > c_0 > 0$ for some $c_0 > 0$ (see the derivative of C in the proof of Lemma 1. While we cannot show that for any $\varepsilon > 0$, the second derivatives of any sequence of benefit functions satisfying our conditions converge to 0 uniformly on $[-1,-\varepsilon]$ and $[\varepsilon,1]$, it is straightforward to verify for the particular sequence $(h_t)_{n\in\mathbb{N}}$ chosen above. Therefore, the terms including the derivatives of h_t are bounded by c_0 on for $b_t \in [b_0, b_0+\varepsilon]$ for t large enough. Therefore, the first-order condition for b_t is strictly decreasing for t large enough, implying that there exists a unique b_t^* that solves the first order condition and depends continuously on a_t.

The continuous solutions to the individual first-order conditions give rise to a continuous mapping $F_t : [a_0, a_0+\varepsilon] \times [b_0, b_0+\varepsilon] \to [a_0, a_0+\varepsilon] \times [b_0, b_0+\varepsilon]$. A fixed point (a_t^*, b_t^*) of F_t exists by Brouwer's fixed point theorem and solves both first-order conditions.

As the left-hand sides of the first-order conditions converge to 0 uniformly, the left-hand sides converge to 0 when evaluated at (a_t^*, b_t^*). By the continuity of C_B', we may conclude that (a_t^*, b_t^*) converges to (a_0, b_0).

Having shown that there is a sequence of solutions to the first-order conditions that converges to (a_0, b_0), we need to show that for t large enough, there cannot be a profitable deviation. This is straightforward for A-supporters: the utility gain from any upward deviation converges to zero, therefore so does the cost. Downward deviations can be excluded (for t large enough) by adapting the above argument using the second derivative to the objective as a function of a_t^i. For B supporters, we consider any deviation $b_t^i \in \left[b_0, \bar{b}\right]$. We write $b_t^\kappa = \kappa b_t^i + (1-\kappa)\, b_t$. Moreover, we define

$$b_w = \frac{a_0 - b_0}{\kappa} + b_0, \tag{64}$$

the deviation that leads to a tie in the limit.

The B supporter would like to deviate or is indifferent if

$$C_B\left(b_t^i\right) - \underbrace{C_B\left(b_t\right)}_{\to 0} \le \rho h_t\left(\frac{b_t^\kappa - a_t}{b_t^\kappa + a_t}\right) - \rho \underbrace{h_t\left(\frac{b_t - a_t}{a_t + b_t}\right)}_{\to -1}. \tag{65}$$

We can distinguish three kinds of sequences of deviations (and any sequence can be decomposed into three subsequences of this kind): those that are bounded away from b_w downwards, those that are bounded away from b_w upwards and those that converge to b_w.

For the deviations that are bounded away from b_w downwards, the right-hand side converges to 0 by the uniform convergence of h_t on intervals bounded away from 0. For any deviation to be profitable for all large t, the left-hand side needs to converge to 0, implying that $b_t^i \to b_0$. But then b_t^i and b_t get arbitrarily close to each other for large t. Adapting the above argument using the second derivative to the objective as a function of b_t^i, we find that both b_t and b_t^i are the unique (for large enough t) solution to the same first order condition.

For the deviations that are bounded away from b_w upwards, the right hand side converges to 2ρ. The left-hand side, on the other hand, converges to a number strictly larger than $C_B\left(\frac{a_0-b_0}{\tilde{\kappa}_B} + b_0\right)$, implying that the deviation is not profitable by the definition of $\tilde{\kappa}_B$.

Finally, for any deviation that converges to b_w, we observe that the right-hand side is bounded by 2ρ, while the left-hand side converges to $C_B\left(\frac{a_0-b_0}{\tilde{\kappa}_B} + b_0\right)$. Notice that $a_t > a_0$ and that $b_t \ge a_t$, as otherwise the right-hand side is bounded by ρ. But then the right-hand side is bounded by 2ρ from above, while the left-hand side is strictly greater than $C_B\left(\frac{a_0-b_0}{\tilde{\kappa}_B}+b_0\right) \geq 2\rho$.

In sum, there cannot be a sequence of profitable deviations, showing that (a_0, b_0) is indeed a limit equilibrium.

Part 2: If $\kappa > \tilde{\kappa}_B$, then (a_0, b_0) cannot be supported as a limit winner-take-all equilibrium.

Let $\kappa > \tilde{\kappa}_B$. We proceed by contradiction and assume that (a_0, b_0) is supported as a limit equilibrium by (a_t, b_t, h_t). For a given $t \in \mathbb{N}$ (assumed to be large enough to have $a_s > b_s$ for all $s \geq t$), we consider the following deviation for a B-supporter for some $\varepsilon > 0$:

$$(1-\kappa)\, b_t + \kappa b_t^i = a_t + \varepsilon. \tag{66}$$

Notice that such a $b_t^i \in [b_0, \bar{b}]$ is indeed well-defined for t large enough and ε small enough. Solving for b_t^i,

$$b_t^i = \frac{a_t + \varepsilon - b_t}{\kappa} + b_t. \tag{67}$$

Hence, the deviation entails a cost $C_B\left(\frac{a_t+\varepsilon-b_t}{\kappa}+b_t\right)$. Therefore, the B supporter would like to deviate or is indifferent if and only if

$$C_B\left(\frac{a_t+\varepsilon-b_t}{\kappa}+b_t\right) \leq \rho h_t\left(\frac{b_t^i-a_t}{b_t^i+a_t}\right) - \rho h_t\left(\frac{b_0-a_0}{a_0+b_0}\right). \tag{68}$$

By the uniform convergence established in Lemma 2, the right-hand side converges to 2ρ, whereas the left-hand side converges to $C_B\left(\frac{a_0+\varepsilon-b_0}{\kappa}+b_0\right)$. This expression is decreasing in κ and increasing in ε. Therefore, if $\kappa > \tilde{\kappa}_B$, for ε small enough and t large enough, the B-supporter would like to deviate to b_t^i, therefore showing that (a_0, b_0) was not supported as a limit equilibrium.

A.2 Proof of Proposition 7

Without loss of generality, we assume $\rho b_v - a_v > 0$.

Part 1: Proof that $(a, b) \in \{(a_0, b_0), (a_0, a_0)\}$.

In light of Proposition 6, for $\kappa = 0$, (a_0, b_0) is the unique equilibrium for any t for any sequence of benefit functions, thus showing $(a, b) = (a_0, b_0)$.

Let $\kappa > 0$ and let (a_t, b_t) be a sequence of turnouts sustained at equilibrium for h_t.

In view of Proposition 6, for all t, $b_t > b_0$ and $a_t = a_0$, and b_t satisfies the first-order condition (see Section 4.4 in the main text):

$$0 \leq (\rho b_v - a_v) h_t'\left(\frac{b_t-a_0}{b_t+a_0}\right)\frac{2\kappa a_0}{b_t+a_0} - b_v C_B'(b_t), \tag{69}$$

where the inequality may hold strictly only for $b_t = \bar{b}$.

Let $\varepsilon > 0$. Assume, for a contradiction, that the sequence $(b_t)_{t\in\mathbb{N}}$ has an infinite number of points within $(b_0+\varepsilon, \bar{b}]$ (in the case $b_0 \geq a_0$) or within $(b_0+\varepsilon, a_0-\varepsilon) \cup (a_0+\varepsilon, \bar{b}]$ (in the case $a_0 > b_0$). We can then extract a subsequence (b_τ) that converges to $\tilde{b} \in \left[b_0+\varepsilon, \bar{b}\right]$ (in the case $b_0 \geq a_0$) or to $\tilde{b} \in [b_0+\varepsilon, a_0-\varepsilon] \cup [a_0+\varepsilon, \bar{b}]$ (in the case $a_0 > b_0$). In any case, $\frac{b_\tau - a_0}{b_\tau + a_0}$ tends to $\frac{\tilde{b}-a_0}{\tilde{b}+a_0} \neq 0$. By the uniform convergence established above, $h_t'\left(\frac{b_\tau - a_0}{b_\tau + a_0}\right)$ tends to 0, whereas $C_B'(b_\tau)$ tends to $C_B'(\tilde{b}) > 0$, implying that for τ large enough, b_τ fails to satisfy the above first-order condition. It follows that $\lim_{t\to\infty} b_t \in \{a_0, b_0\}$.

Part 2: If $b_0 > a_0$, then (a_0, b_0) is the unique limit equilibrium. Let h_t be a sequence satisfying our assumptions and converging to sign pointwise on $[-1,1]$. Recall that by Proposition 6, an equilibrium exists and is such that $a_t = a_0$. If $b_0 \geq a_0$, then $\beta(b, a_0) \geq 0$ for all $b \in [b_0, \bar{b}]$. By our assumptions, h is strictly concave on $[0,1]$. It is straightforward to verify that $b \mapsto \beta(b, a_0)$ is concave, so $b \mapsto h(\beta(b, a_0))$ is concave. As the cost is convex, both the utility function and the auxiliary function (see the proof of Proposition 6) are concave, implying the uniqueness of the equilibrium (b_t, a_0) for all t. Consider the first-order condition,

$$(\rho b_v - a_v)\underbrace{h_t'\left(\frac{b_t - a_0}{b_t + a_0}\right)}_{\to 0}\frac{2a_0}{(b_t+a_0)^2} - b_v C_B'(b_t) \geq 0. \tag{70}$$

By the uniform convergence of h_t', the benefit term converges to zero (recall $b_t \geq b_0 > a_0$). Therefore, the cost term needs to converge to zero as well, and by continuity of C_B' we may conclude that b_t indeed converges to b_0, proving both uniqueness and existence of (a_0, b_0) as a limit equilibrium.

Part 3.1.1: The case $b_0 < a_0$ and $\kappa \leq \tilde{\kappa}_B$.

In order to establish existence of (a_0, b_0) as an equilibrium for $\kappa \leq \tilde{\kappa}_B$, we consider the usual sequence, see also (63). As in the proof of Proposition 3 , we can show that for any $\varepsilon > 0$, for large enough t there exists a solution of the first-order condition that is unique in $[b_0, b_0+\varepsilon]$. Again by the same arguments as in the proof of Proposition 3 , there is no profitable deviation, therefore proving that (a_0, b_0) is a limit winner-take-all equilibrium.

Part 3.1.2: The case $b_0 < a_0$ and $\kappa > \tilde{\kappa}_B$.

For $\kappa > \tilde{\kappa}_B$, one can assume the existence of an approximating sequence for a contradiction and show that there is a profitable deviation for t high enough, by the same arguments as in the proof of Proposition 3 .

Part 3.2.1: The case $b_0 < a_0$ and $2(\rho b_v - a_v) > b_v C_B(a_0)$.

We will first establish that there exists a unique solution b_t^* to the first-order condition

$$(\rho b_v - a_v)\, h_t'\left(\frac{b_t - a_0}{b_t + a_0}\right)\frac{2\kappa a_0}{(b_t+a_0)^2} = b_v C_B'(b_t) \tag{71}$$

such that $b_t^* > a_0$ for t large enough. To that extent, notice that $C_B'(b_t)$ is bounded and positive on $[a_0, \bar{b}]$, whereas $h_t'(0)$ takes arbitrarily high values and $h_t'\left(\frac{\bar{b}-a_0}{\bar{b}+a_0}\right)$ converges to zero. The existence of a solution of the first-order condition such that $b_t^* > a_0$ thus follows from the intermediate value theorem. The uniqueness follows from $h_t'\left(\frac{b_t-a_0}{b_t+a_0}\right)\frac{2\kappa a_0}{(b_t+a_0)^2}$ being decreasing for $b_t \in [a_0, \bar{b}]$ (as a consequence of $h_t''(x) < 0$ for $x > 0$) and $C_B'(b_t)$ being increasing (as a consequence of Lemma 1). Moreover, we can show that for any $\varepsilon > 0$, there exists T such that for $t \geq T$, $b_t^* \in [a_0, a_0+\varepsilon]$: this is a consequence of the uniform convergence of h_t' towards 0 on closed intervals that do not include 0. As a result, $b_t^* \to a_0$ as

$t \to \infty$. In order to show that such a sequence (b_t^*) of solutions to the first-order condition is indeed an equilibrium for all t, we need to exclude profitable deviations (at least for large enough t, shifting the start of the sequence). For deviations b_t^i such that $b_t^\kappa \in [a_0, \bar{b}]$, this is straightforward: the objective is concave in that area by the usual arguments. This also implies that $h_t\left(\frac{b_t^*-a_0}{b_t^*+a_0}\right) \to 1$: indeed, if this was not the case, one could find a profitable upward deviation for large enough t, a contradiction.

To complete this part of the proof, consider any sequence of deviations (b_t^i) such that $b_t^\kappa \in [b_0, a_0)$ for all $t \in \mathbb{N}$. A deviation b_t^i is profitable only if

$$(\rho b_v - a_v)\left(h_t\left(\frac{b_t^* - a_0}{b_t^* + a_0}\right) - h_t\left(\frac{b_t^\kappa - a_0}{b_t^\kappa + a_0}\right)\right) \leq b_v C_B\left(b_t^*\right) - b_v C_B\left(b_t^i\right). \tag{72}$$

Given that $b_t^* \to a_0$, for all $\delta > 0$, there exists $\varepsilon > 0$ and T such that for all $t \geq T$, if $b_t^\kappa > a_0 - \varepsilon$, then $b_t^i = \frac{b_t^\kappa - (1-\kappa)b_t^*}{\kappa} \geq a_0 - \delta$.

Examining the criterion for the deviation to be profitable for a deviation such that $b_t^\kappa > a_0 - \varepsilon$, for $t \geq T$:

$$(\rho b_v - a_v)\left(\underbrace{h_t\left(\frac{b_t^* - a_0}{b_t^* + a_0}\right)}_{\to 1} - \underbrace{h_t\left(\frac{b_t^\kappa - a_0}{b_t^\kappa + a_0}\right)}_{<0}\right) \leq b_v \underbrace{C_B\left(b_t^*\right)}_{\to C_B(a_0)} - b_v \underbrace{C_B\left(b_t^i\right)}_{\geq C_B(a_0-\delta)}. \tag{73}$$

By the continuity of C_B, δ can be chosen small enough for the deviations not to be profitable for large enough t.

On the other hand, consider those deviations that satisfy $b_t^\kappa \leq a_0 + \varepsilon$. Then, by the uniform convergence of (h_t) on closed intervals that do not contain 0,

$$(\rho b_v - a_v)\left(\underbrace{h_t\left(\frac{b_t^* - a_0}{b_t^* + a_0}\right)}_{\to 1} - \underbrace{h_t\left(\frac{b_t^\kappa - a_0}{b_t^\kappa + a_0}\right)}_{\to -1}\right) \leq b_v \underbrace{C_B\left(b_t^*\right)}_{\to C_B(a_0)} - b_v \underbrace{C_B\left(b_t^i\right)}_{\geq 0}. \tag{74}$$

This condition cannot hold for large enough t as we assumed $2(\rho b_v - a_v) > b_v C_B(a_0)$.

Part 3.2.2: The case $b_0 < a_0$ and $2(\rho b_v - a_v) \leq b_v C_B(a_0)$.

Assume for a contradiction that $(a_0, b_t, h_t)_{t \in \mathbb{N}}$ is an approximating sequence for (a_0, a_0) and $2(\rho b_v - a_v) \leq b_v C(a_0)$. We consider the deviation for B supporters towards b_0. Let $b_t^\kappa = \kappa b_0 + (1-\kappa) b_t$. A B-supporter prefers to deviate or is indifferent if

$$(\rho b_v - a_v)\underbrace{\left(h_t\left(\frac{b_t - a_0}{b_t + a_0}\right) - h_t\left(\frac{b_t^\kappa - a_0}{b_t^\kappa + a_0}\right)\right)}_{<2} \leq b_v \underbrace{C_B\left(b_t\right)}_{>C_B(a_0)} - b_v \underbrace{C_B\left(b_0\right)}_{=0} \tag{75}$$

Therefore, if $2(\rho b_v - a_v) \leq b_v C(a_0)$, then b_0 is indeed a profitable deviation for any t.

B A comment on the benefit function used by Herrera, Morelli, and Nunnari (2016)

Using our notation, Herrera et al. (2016) posit the material benefit function $a^\gamma/(a^\gamma+b^\gamma)$ for the A-supporters and $b^\gamma/(a^\gamma+b^\gamma)$ for the B-supporters, where the parameter $\gamma \in [1,+\infty)$ captures the power sharing rule. By setting

$$h(\alpha)=\frac{(1+\alpha)^\gamma-(1-\alpha)^\gamma}{(1+\alpha)^\gamma+(1-\alpha)^\gamma}, \tag{76}$$

it is straightforward to show that

$$h(\alpha)=-1+2\frac{a^\gamma}{a^\gamma+b^\gamma}, \tag{77}$$

and similarly for β. Indeed,

$$\frac{(1+\alpha)^\gamma-(1-\alpha)^\gamma}{(1+\alpha)^\gamma+(1-\alpha)^\gamma}=\frac{\frac{2a}{a+b}^\gamma-\frac{2b}{a+b}^\gamma}{\frac{2a}{a+b}^\gamma+\frac{2b}{a+b}^\gamma}=-1+2\frac{a^\gamma}{a^\gamma+b^\gamma}. \tag{78}$$

It remains to show that the function in (76) fits the assumptions of our model. Clearly, it takes values between -1 and 1 and is symmetric around $\alpha=0$. Moreover, it is continuous and differentiable. Indeed, the first derivative is given by

$$h'(\alpha)=\frac{4\gamma\left((1-\alpha^2)^{\gamma-1}\right)}{((1+\alpha)^\gamma+(1-\alpha)^\gamma)^2}>0, \tag{79}$$

and therefore, the second derivative is given by

$$\begin{aligned}h''(\alpha)&=4\gamma\frac{-2\alpha(\gamma-1)(1-\alpha^2)^{\gamma-2}((1+\alpha)^\gamma+(1-\alpha)^\gamma)^2}{((1+\alpha)^\gamma+(1-\alpha)^\gamma)^4}\\&-4\gamma\frac{2\left((1-\alpha^2)^{\gamma-1}\right)((1+\alpha)^\gamma+(1-\alpha)^\gamma)\gamma\left((1+\alpha)^{\gamma-1}-(1-\alpha)^{\gamma-1}\right)}{((1+\alpha)^\gamma+(1-\alpha)^\gamma)^4}\\&=-8\gamma(\gamma-1)\alpha\frac{(1-\alpha^2)^{\gamma-2}}{((1+\alpha)^\gamma+(1-\alpha)^\gamma)^2}\\&-8\gamma^2\frac{(1-\alpha^2)^{\gamma-1}\left((1+\alpha)^{\gamma-1}-(1-\alpha)^{\gamma-1}\right)}{((1+\alpha)^\gamma+(1-\alpha)^\gamma)^3}.\end{aligned} \tag{80}$$

Clearly, this h satisfies our assumptions on the derivatives for any $\gamma\in[1,+\infty)$. Re-scaling the cost accordingly, this shows that our benefit term is more general.

An interactive tool that enables comparison of our benefit function and that used by Herrera et al. (2016) can be found here: https://KonradEcon.github.io/homo-moralis-turnout-appendix/.

C Partisan ethics

C.1 Computing equilibria with the arctan benefit function and uniform cost

We first describe how to compute A-consistent strategies; B-consistent strategies can be computed analogously. We look for an A-consistent strategy a given a strategy b played by B-supporters. The first-order condition for a, $0 = \frac{\partial}{\partial a^i} EU_A^\kappa \left(a, b, a^i\right)\big|_{a^i=a}$ is given by

$$0 = \frac{m}{\arctan(m)} \frac{1}{1 + \left(m\frac{a-b}{b+a}\right)^2} \frac{2\kappa b}{(b+a)^2} - \theta_A \frac{a - a_0}{a_v^2}. \tag{81}$$

After some algebra, we obtain a polynomial:

$$0 = a^3 + \left(2\frac{1-m^2}{1+m^2}b - a_0\right)a^2 + \left(b^2 - 2\frac{1-m^2}{1+m^2}a_0 b\right)a - b^2 a_0 - \frac{2m\kappa b a_v^2}{(1+m^2)\,\theta_A \arctan(m)} \tag{82}$$

Solving the polynomial gives candidate A-consistent strategies, to which we have to add the corner solution $\bar{a}$ (the other corner solution a_0 cannot be A-consistent, given Proposition 1).

Then, in order to rule out a profitable deviation, it is sufficient to compare the utility level associated to such a candidate a with the utility at any solution of $\frac{\partial}{\partial a^i} U^\kappa \left(a, b^*, a^i\right) = 0$ and with the utility at $\bar{a}$. The equation $\frac{\partial}{\partial a^i} U^\kappa \left(a, b^*, a^i\right) = 0$ can be rewritten as another degree three polynomial equation.

In order to find equilibria, we first find pairs (a, b) that simultaneously solve the first order conditions, i.e. $0 = \frac{\partial}{\partial a^i} EU_A^\kappa \left(a, b, a^i\right)\big|_{a^i=a}$ and $0 = \frac{\partial}{\partial b^i} EU_B^\kappa \left(a, b, a^i\right)\big|_{b^i=b}$. As above, these can both be rewritten as polynomials in (a, b). We can solve the system of polynomials numerically using the resultant method, where we use SymPy to compute the resultant and NumPy to compute roots.

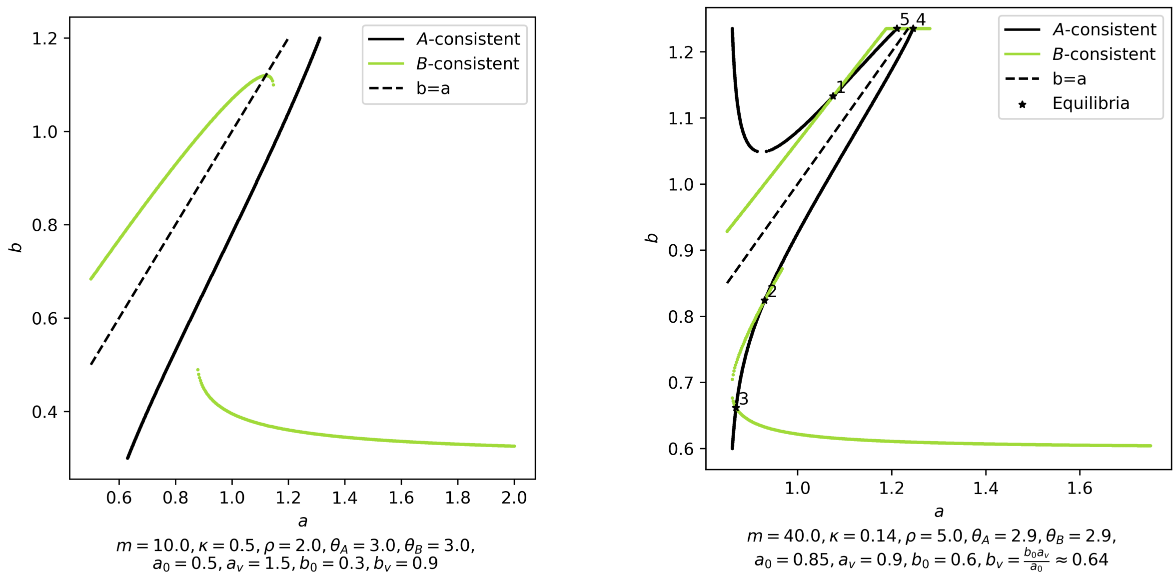
Other candidate equilibria are $\left(\bar{a}, \bar{b}\right)$, $\left(a, \bar{b}\right)$ where a is a candidate A-consistent strategy given $\bar{b}$, and $(\bar{a}, b)$ where b a candidate B-consistent strategy given $\bar{a}$. For each candidate equilibrium, we check for profitable deviations as described above.

C.2 Examples

Example 6 (a setting with no equilibrium)**.** Equilibria may fail to exist. This is illustrated in Figure 11a, where the curve showing the A-consistent turnouts does not intersect with the curve showing the B-consistent turnouts.

Example 7 (a setting with multiple equilibria)**.** There may exist more than one equilibrium. In Figure 11b we show an example with five equilibria: two of them exhibit low turnouts and a victory by the leader, while the other three exhibit high turnouts and a victory by the underdog.

Readers interested in exploring how existence and/or mulitplicity issues depend on the parameter sets can use an online tool available at: https://KonradEcon.github.io/homo-moralis-turnout-appendix/.



(a) An example of failure of equilibrium existence

(b) An example with five equilibria

Figure 11: Two extreme cases on multiplicity of equilibria

C.3 Sufficient conditions for equilibrium existence and uniqueness

Here we identify sufficient conditions for there to exist a unique equilibrium (a^*, b^*). To prepare the ground for the statements and proofs, we define the auxiliary functions

$$\Phi_A(a,b) = \kappa h(\alpha(a,b)) - C_A(a), \quad (83)$$
$$\Phi_B(a,b) = \rho\kappa h(-\alpha(a,b)) - C_B(b). \quad (84)$$

We also define the *auxiliary game*: this is a simultaneous-move game between two players, call them Alice and Bob, who have strategy sets $[a_0, \bar{a}]$ and $[b_0, \bar{b}]$, and payoff functions Φ_A and Φ_B, respectively. We will use the term *population game* to refer to the game in the main text.

In the following statements, by single-peaked we mean that a function defined on $[a_0, \bar{a}]$ (or $[b_0, \bar{b}]$) is strictly increasing up to some $a \in (a_0, \bar{a})$ (or $b \in (b_0, \bar{b})$), and strictly decreasing thereafter. In case it is differentiable, this amounts to its first-derivative being first strictly positive, crossing zero once from above, and then being strictly negative.

Assumption 1. *Assume that for all (a,b), $a^i \mapsto EU_A^\kappa(a,b,a^i)$ and $b^i \mapsto EU_B^\kappa(a,b,b^i)$ are single-peaked. Moreover, assume that for all b, $a \mapsto \Phi_A(a,b)$ is single-peaked, and for all a, $b \mapsto \Phi_B(a,b)$ is single-peaked.*

Lemma 3. *Under Assumption 1, (a^*, b^*) is an equilibrium of the population game if, and only if, it is a Nash equilibrium of the auxiliary game.*

Proof. Claim: for any b, there exists a unique A-consistent strategy. To prove this, let us define $\phi : [a_0, \bar{a}] \to [a_0, \bar{a}]$ by

$$\phi(a) = \arg\max_{a^i} EU_A^\kappa\left(a, b, a^i\right). \tag{85}$$

This is indeed well-defined, since the argmax exists and is unique due to the single-peakedness assumption on EU_A^κ. By Berge's maximum theorem, ϕ is continuous. Therefore, Brouwer's fixed point theorem applies and there exists at least one fixed point, i.e. at least one A-consistent strategy. By the assumption that for all a, b, the unique maximum of $a^i \mapsto EU_A^\kappa(a, b, a^i)$ lies in $(a_0, \bar{a})$, we can conclude that any such A-consistent strategy lies in $(a_0, \bar{a})$ (is interior). Hence, it satisfies the first-order condition

$$0 = \frac{\partial}{\partial a^i} EU_A^\kappa\left(a, b, a^i\right)\bigg|_{a^i = a}. \tag{86}$$

Since

$$\frac{\partial}{\partial a}\Phi_A(a, b) = \frac{2\kappa b h'(\alpha(a, b))}{(a+b)^2} - C_A'(a), \tag{87}$$

and recalling the first-order condition

$$\frac{\partial}{\partial a^i} U_A^\kappa\left(a, b, a^i\right)|_{a^i = a^*, b = b^*} = \frac{2\kappa b^* h'(\alpha(a^*, b^*))}{(a^* + b^*)^2} - C_A'(a^*) \begin{cases} = 0 \text{ if } a^* \in (a_0, \bar{a}) \\ \geq 0 \text{ if } a^* = \bar{a}, \end{cases} \tag{88}$$

we conclude that any fixed point of (85) satisfies

$$0 = \frac{\partial}{\partial a^i} EU_A^\kappa\left(a, b, a^i\right)\bigg|_{a^i = a} = \frac{\partial}{\partial a}\Phi_A(a, b). \tag{89}$$

Since the single-peakedness assumption on Φ_A implies that there is a unique a that maximizes $\Phi_A(a, b)$, this completes the proof of the claim.

By repeating the same argument for B-consistent strategies, we conclude that (a^*, b^*) is an equilibrium of the population game if and only if

1. a^* maximizes $a \mapsto \Phi_A(a, b^*)$, and
2. b^* maximizes $b \mapsto \Phi_B(a^*, b)$.

□

By reducing the analysis to that of a standard two-player game, Lemma 3 facilitates identification of sufficient conditions for there to exist a unique equilibrium of the population game.

Assumption 2. *Let* $h(x) = \frac{\arctan(mx)}{\arctan(m)}$, *and assume that:*

1. $\kappa \in (0, 1]$,
2. $\frac{f_k(c)c}{F_k(c)}$, $k = A, B$, *is decreasing,*

3. $m \leq 1$ *or for some* $r < \frac{2m}{(m-1)^2}$,

$$\lim_{c \to 0} \frac{F_A(c)}{c^r} > 0, \tag{90}$$

4. $m \leq 1$ *or for some* $r < \frac{2m}{\rho(m-1)^2}$,

$$\lim_{c \to 0} \frac{F_B(c)}{c^r} > 0, \tag{91}$$

5. $\bar{s}_B \geq \frac{\rho \bar{a}(m^2+1)}{2\kappa b_v m \arctan(m)}$, *and*

6. $\bar{s}_A \geq \frac{\bar{b}(m^2+1)}{2\kappa a_v m \arctan(m)}$.

Proposition 10. *Under Assumption 1, there exists a unique equilibrium of the population game. Moreover, Assumption 1 holds under Assumption 2.*

Proof. Proving first the first claim, let us define a function $\psi : [a_0, \bar{a}] \times \left[b_0, \bar{b}\right] \to [a_0, \bar{a}] \times \left[b_0, \bar{b}\right]$ by

$$\psi(a, b) = \begin{pmatrix} \arg\max_{\tilde{a}} \Phi_A(\tilde{a}, b) \\ \arg\max_{\tilde{b}} \Phi_B\left(a, \tilde{b}\right) \end{pmatrix}, \tag{92}$$

which is well-defined by the assumptions on Φ_A and Φ_B. Applying Berge's maximum theorem, we conclude that ψ is continuous, allowing us to apply Brouwer's theorem. We deduce that ψ has at least one fixed point, proving equilibrium existence.

In order to prove uniqueness, we again rely on the auxiliary two-player game. Assume, for a contradiction, that (a_1, b_1) and (a_2, b_2) are both equilibria of that game. Then,

$$\begin{aligned} \Phi_A(a_1, b_1) &\geq \Phi_A(a_2, b_1), & (93) \\ \Phi_A(a_2, b_2) &\geq \Phi_A(a_1, b_2), & (94) \\ \Phi_B(a_1, b_1) &\geq \Phi_B(a_1, b_2), & (95) \\ \Phi_B(a_2, b_2) &\geq \Phi_B(a_2, b_1). & (96) \end{aligned}$$

Writing these expressions out,

$$\begin{aligned} \kappa h(\alpha(a_1, b_1)) - C_A(a_1) &\geq \kappa h(\alpha(a_2, b_1)) - C_A(a_2), & (97) \\ \kappa h(\alpha(a_2, b_2)) - C_A(a_2) &\geq \kappa h(\alpha(a_1, b_2)) - C_A(a_1), & (98) \\ -\rho\kappa h(\alpha(a_1, b_1)) - C_B(b_1) &\geq -\rho\kappa h(\alpha(a_1, b_2)) - C_B(b_2), & (99) \\ -\rho\kappa h(\alpha(a_2, b_2)) - C_B(b_2) &\geq -\rho\kappa h(\alpha(a_2, b_1)) - C_B(b_1). & (100) \end{aligned}$$

Rewriting,

$$
\begin{aligned}
\kappa h\left(\alpha\left(a_1,b_1\right)\right)-\kappa h\left(\alpha\left(a_2,b_1\right)\right) &\geq C_A\left(a_1\right)-C_A\left(a_2\right), &(101)\\
C_A\left(a_1\right)-C_A\left(a_2\right) &\geq \kappa h\left(\alpha\left(a_1,b_2\right)\right)-\kappa h\left(\alpha\left(a_2,b_2\right)\right), &(102)\\
\rho\kappa h\left(\alpha\left(a_1,b_2\right)\right)-\rho\kappa h\left(\alpha\left(a_1,b_1\right)\right) &\geq C_B\left(b_1\right)-C_B\left(b_2\right), &(103)\\
C_B\left(b_1\right)-C_B\left(b_2\right) &\geq \rho\kappa h\left(\alpha\left(a_2,b_2\right)\right)-\rho\kappa h\left(\alpha\left(a_2,b_1\right)\right). &(104)
\end{aligned}
$$

Combining and eliminating constant positive factors,

$$
\begin{aligned}
h\left(\alpha\left(a_1,b_1\right)\right)-h\left(\alpha\left(a_2,b_1\right)\right) &\geq h\left(\alpha\left(a_1,b_2\right)\right)-h\left(\alpha\left(a_2,b_2\right)\right), &(105)\\
h\left(\alpha\left(a_1,b_2\right)\right)-h\left(\alpha\left(a_1,b_1\right)\right) &\geq h\left(\alpha\left(a_2,b_2\right)\right)-h\left(\alpha\left(a_2,b_1\right)\right). &(106)
\end{aligned}
$$

Rewriting once more,

$$
\begin{aligned}
h\left(\alpha\left(a_1,b_1\right)\right)+h\left(\alpha\left(a_2,b_2\right)\right) &\geq h\left(\alpha\left(a_1,b_2\right)\right)+h\left(\alpha\left(a_2,b_1\right)\right), &(107)\\
h\left(\alpha\left(a_1,b_2\right)\right)+h\left(\alpha\left(a_2,b_1\right)\right) &\geq h\left(\alpha\left(a_2,b_2\right)\right)+h\left(\alpha\left(a_1,b_1\right)\right). &(108)
\end{aligned}
$$

Combining the two inequalities, we have equality throughout:

$$h\left(\alpha\left(a_1,b_1\right)\right)+h\left(\alpha\left(a_2,b_2\right)\right)=h\left(\alpha\left(a_1,b_2\right)\right)+h\left(\alpha\left(a_2,b_1\right)\right). \tag{109}$$

Multiplying with κ and subtracting $C_A\left(a_1\right)$ as well as $C_A\left(a_2\right)$ on both sides,

$$\Phi_A\left(a_1,b_1\right)+\Phi_A\left(a_2,b_2\right)=\Phi_A\left(a_1,b_2\right)+\Phi_A\left(a_2,b_1\right). \tag{110}$$

Since $\Phi_A\left(a_1,b_1\right)\geq\Phi_A\left(a_2,b_1\right)$ and $\Phi_A\left(a_2,b_2\right)\geq\Phi_A\left(a_1,b_2\right)$, we have

$$\Phi_A\left(a_1,b_1\right)=\Phi_A\left(a_2,b_1\right) \text{ and } \Phi_A\left(a_2,b_2\right)=\Phi_A\left(a_1,b_2\right). \tag{111}$$

Since we assumed that $a\mapsto\Phi_A\left(a,b\right)$ is single-peaked for any b, we deduce $a_1=a_2$. By repeating the same argument on Φ_B, we deduce $b_1=b_2$, thus proving equilibrium uniqueness of the auxiliary game. Lemma 3 then implies that this is also the unique equilibrium of the population game. This completes the proof of the first claim of the proposition.

We turn now to the second claim of the proposition. To begin, for any turnout levels (a,b) let us write the expected utility of B-supporter i as a function of the cutoff strategy s_B^i:

$$EU_B^\kappa\left(a,b,s_B^i\right)=\rho h\left(\beta^\kappa\left(a,b,s_B^i\right)\right)-\int_0^{s_B^i}cf\left(c\right)dc, \tag{112}$$

where

$$\beta^\kappa\left(a,b,s_B^i\right)=\frac{\left(1-\kappa\right)b+\kappa\left(b_vF_B\left(s_B^i\right)+b_0\right)-a}{\left(1-\kappa\right)b+\kappa\left(b_vF_B\left(s_B^i\right)+b_0\right)+a}. \tag{113}$$

Likewise, write the associated auxiliary function as a function of s_B:

$$\Phi_B\left(a,s_B\right)=\rho\kappa h\left(\beta\left(a,s_B\right)\right)-\int_0^{s_B}cf\left(c\right)dc. \tag{114}$$

Clearly, since F_B is strictly increasing, single-peakedness of $EU_B^{\kappa}(a, b, s_B^i)$ in s_B^i holds if and only if single-peakedness of $EU_B^{\kappa}(a, b, b^i)$ in b^i holds. Likewise, single-peakedness of $\Phi(a, s_B)$ in s_B holds if and only if single-peakedness of $\Phi_B(a, b)$ in b holds.

We will show that for all (a, b), $s_B^i \mapsto EU_B^{\kappa}(a, b, s_B^i)$ and $s_B \mapsto \Phi_B(a, s_B)$ are single-peaked, as the proof goes analogously for EU_A^{κ} and Φ_A (with $\rho = 1$).

In order to ease notation, let

$$\begin{aligned} M^{\kappa}\left(s_B^i\right) &= (1-\kappa)\, b + \kappa\left(b_v F_B\left(s_B^i\right) + b_0\right) - a, \text{ and} \\ T^{\kappa}\left(s_B^i\right) &= (1-\kappa)\, b + \kappa\left(b_v F_B\left(s_B^i\right) + b_0\right) + a. \end{aligned} \tag{115}$$

We then have

$$\frac{\partial}{\partial s_B^i}\beta^{\kappa}\left(a, b, s_B^i\right) = \frac{2\kappa a b_v f_B\left(s_B^i\right)}{\left(T^{\kappa}\left(s_B^i\right)\right)^2} \tag{116}$$

so that

$$\frac{\partial}{\partial s_B^i} EU_B^{\kappa}\left(a, b, s_B^i\right) = \rho h'\left(\beta^{\kappa}\left(a, b, s_B^i\right)\right)\frac{2\kappa a b_v f_B\left(s_B^i\right)}{\left(T^{\kappa}\left(s_B^i\right)\right)^2} - s_B^i f_B\left(s_B^i\right). \tag{117}$$

Hence, for $h(x) = \frac{\arctan(mx)}{\arctan(m)}$,

$$\begin{aligned} \frac{\partial}{\partial s_B^i} EU_B^{\kappa}\left(a, b, s_B^i\right) &= \rho\frac{2m\kappa a b_v f_B\left(s_B^i\right)}{\arctan(m)}\frac{1}{\left(T^{\kappa}\left(s_B^i\right)\right)^2}\frac{1}{1 + m^2\frac{\left(M^{\kappa}\left(s_B^i\right)\right)^2}{\left(T^{\kappa}\left(s_B^i\right)\right)^2}} - s_B^i f_B\left(s_B^i\right) \\ &= f_B\left(s_B^i\right)\left(\frac{2m\kappa\rho a b_v}{\arctan(m)}\frac{1}{\left(T^{\kappa}\left(s_B^i\right)\right)^2 + m^2\left(M^{\kappa}\left(s_B^i\right)\right)^2} - s_B^i\right) \\ &= \underbrace{\frac{f_B\left(s_B^i\right)}{\left(T^{\kappa}\left(s_B^i\right)\right)^2 + m^2\left(M^{\kappa}\left(s_B^i\right)\right)^2}}_{>0}\left(\frac{2m\kappa\rho a b_v}{\arctan(m)} - \left(\left(T^{\kappa}\left(s_B^i\right)\right)^2 + m^2\left(M^{\kappa}\left(s_B^i\right)\right)^2\right)s_B^i\right). \end{aligned} \tag{118}$$

Consider now the auxiliary function in (114). Since

$$\frac{\partial}{\partial s_B}\beta(a, s_B) = \frac{2ab_v f_B(s_B)}{\left(T^1(s_B)\right)^2}, \tag{119}$$

for $h(x) = \frac{\arctan(mx)}{\arctan(m)}$ we obtain

$$\begin{aligned} \frac{\partial}{\partial s_B}\Phi_B(a, s_B) &= \rho\kappa h'(\beta(a, s_B))\frac{2\kappa a b_v f_B(s_B)}{\left(T^1(s_B)\right)^2} - s_B f_B(s_B) \\ &= \underbrace{\frac{f_B(s_B)}{\left(T^1(s_B)\right)^2 + m^2\left(M^1(s_B)\right)^2}}_{>0}\left(\frac{2m\kappa\rho a b_v}{\arctan(m)} - \left(\left(T^1(s_B)\right)^2 + m^2\left(M^1(s_B)\right)^2\right)s_B\right). \end{aligned} \tag{120}$$

Therefore, to show the single-peakedness of $EU_B^{\kappa}(a, b, s_A^i)$ in s_B^i and of $\Phi_B(a, s_B)$ in s_B, it is sufficient that $\phi_B^{\tilde{\kappa}}(s_B) \overset{def}{=} \left(\left(T^{\tilde{\kappa}}\right)^2 + m^2\left(M^{\tilde{\kappa}}\right)^2\right) s_B$ is, for both $\tilde{\kappa} = \kappa$ and $\tilde{\kappa} = 1$,

(a) strictly increasing, and

(b) eventually greater than $\frac{2m\kappa\rho a b_v}{\arctan(m)}$.

To prove (a) it is sufficient to prove that prove that $\left(\phi_B^{\tilde{\kappa}}\right)'(s_B) > 0$, where

$$\left(\phi_B^{\tilde{\kappa}}\right)'(s_B) = \left(T^{\tilde{\kappa}}\right)^2 + m^2\left(M^{\tilde{\kappa}}\right)^2 + 2s_B\tilde{\kappa}b_v f_B(s_B)\left(T^{\tilde{\kappa}} + m^2 M^{\tilde{\kappa}}\right). \tag{121}$$

It is straightforward to see that $\left(\phi_B^{\tilde{\kappa}}\right)'(s_B) > 0$ holds for $m \leq 1$. For $m > 1$, writing $\tilde{b} = b - b_0$, and minimizing the expression over $a, \tilde{b}$ and b_0 using SymPy (the code is included in Section 5 below), we show that $\left(\phi_B^{\tilde{\kappa}}\right)'(s_B) > 0$ if

$$s_B < \frac{2F_B(s_B)\,m}{\rho f_B(s_B)(m-1)^2} \tag{122}$$

or, equivalently,

$$\frac{f_B(s_B)\,s_B}{F_B(s_B)} < \frac{2m}{\rho(m-1)^2}. \tag{123}$$

Since $f_B(c)c/F_B(c)$ is decreasing (by assumption 2), it is sufficient to have

$$\lim_{c\to 0}\frac{f_B(c)\,c}{F_B(c)} < \frac{2m}{\rho(m-1)^2}. \tag{124}$$

For this, in turn, it is sufficient to have, for some $r < \frac{2m}{\rho(m-1)^2}$,

$$\lim_{c\to 0}\frac{F_B(c)}{c^r} > 0. \tag{125}$$

Let us now turn to (b). It is sufficient to give a condition on $\bar{s}_B$ such that

$$\bar{s}_B\left(\left(T^{\tilde{\kappa}}\right)^2 + m^2\left(M^{\tilde{\kappa}}\right)^2\right) > \frac{2m\tilde{\kappa}\rho a b_v}{\arctan(m)}, \tag{126}$$

where M^{κ} and T^{κ} are evaluated at $s_B^i = \bar{s}_B$, for both $\tilde{\kappa} = \kappa$ and $\tilde{\kappa} = 1$.

Using SymPy, we minimize $\psi = \left(\left(T^{\tilde{\kappa}}\right)^2 + m^2\left(M^{\tilde{\kappa}}\right)^2\right)$ over a, $\tilde{b}$ and b_0; we also maximize the right-hand side by plugging in $a = \bar{a}$. We find that (126) holds if

$$\bar{s}_B\frac{4\tilde{\kappa}^2 b_v^2 m^2}{m^2+1} \geq \frac{2m\tilde{\kappa}\rho\bar{a}b_v}{\arctan m}, \tag{127}$$

or equivalently,

$$\bar{s}_B \geq \frac{\rho\bar{a}\left(m^2+1\right)}{2\tilde{\kappa}b_v m\arctan(m)}. \tag{128}$$

We also observe that this condition holds for both $\tilde{\kappa} \in \{\kappa, 1\}$ if

$$\bar{s}_B \geq \frac{\rho \bar{a}\left(m^2+1\right)}{2\kappa b_v m \arctan\left(m\right)}. \tag{129}$$

□

Conditions 2-6 of Assumption 2 are reminiscent of those that Herrera et al. (2016) adopt to ensure equilibrium existence and uniqueness in their model (see their "decreasing generalized reversed hazard rate (DGRHR) property" and their Condition 1). These conditions ensure that the density for low costs is large enough to avoid possible multiple peaks for low turnouts levels.

Observe that, for arbitrarily large m, part 3 and 4 of Assumption 2 require F_A and F_B to be arbitrarily steep at 0. Therefore, there cannot be a single continuous cost distribution on $[0, \infty)$ that satisfies Assumption 2 for all $m > 0$. A similar observation holds for arbitrarily large γ in Condition 1 of Herrera et al. (2016). We do, however, define a notion of limit equilibrium in the main text and can establish existence and uniqueness under some conditions.

D Non-partisan ethics

D.1 Computing equilibria with the arctan benefit function and uniform cost

We assume, without loss of generality, that $\rho b_v - a_v > 0$: indeed, we have established that in this case, $a = a_0$. Finding an equilibrium therefore amounts to finding $b > b_0$. In case the opposite inequality holds, one needs to find the equilibrium a for $b = b_0$.

Let us write out the first-order condition $0 = \frac{\partial}{\partial b^i} EU^\kappa\left(a, b, a^i, b^i\right)|_{b^i=b, a^i=a=a_0}$:

$$0 = \frac{m\left(\rho b_v - a_v\right)}{\arctan\left(m\right)} \frac{1}{1+\left(m\frac{b-a_0}{b+a_0}\right)^2} \frac{2\kappa b}{\left(b+a_0\right)^2} - \theta \frac{b-b_0}{b_v}. \tag{130}$$

It is then straightforward to rewrite this equation as a polynomial equation in b. Finding roots of the polynomial yields candidate equilibria.

Finally, for some candidate equilibrium b, one can write out $0 = \frac{\partial}{\partial b^i} EU^\kappa\left(a, b, a^i, b^i\right)|_{a^i=a=a_0}$ (observe that it can be rewritten as a polynomial equation in b^i, we use SymPy) and check if any solution or $\bar{b}$ is associated with a higher expected utility, in order to rule out profitable deviations.

D.2 Further examples

Example 8. Figure 12 shows the set of equilibrium turnouts for the leader supporters, as a function of a_0, and for three values of κ. For $\kappa = 1$, we see that there is a unique equilibrium for any value of a_0, while for $\kappa = 0.8$ and $\kappa = 0.5$, there are two equilibria for small values of a_0 and a unique equilibrium for large enough values of a_0. The leader wins at equilibria above the dashed line, which corresponds to $a = b_0$. The figure thus shows that if the base a_0 is small, the leader supporters face a coordination problem: they may win or lose. By contrast, a victory for the leader is guaranteed if the base a_0 is large enough.

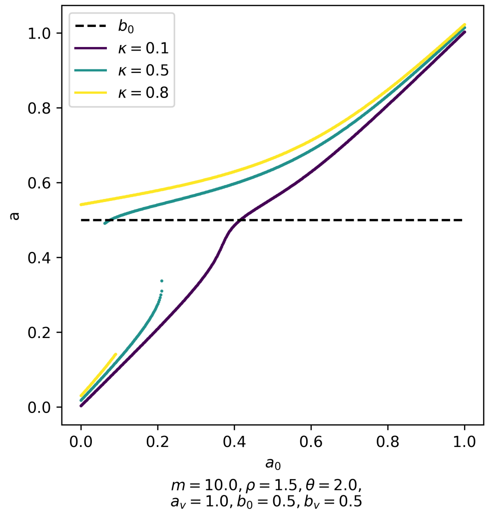


Figure 12: Set of equilibrium turnouts a for different values of a_0

We now examine whether there may be even more than two equilibria.

Example 9. Returning to settings where $\rho b_v - a_v > 0$, Figure 13 shows an example with three equilibria. Like for the example with two equilibria in the main text, here the expected utility is higher the higher is the equilibrium turnout.

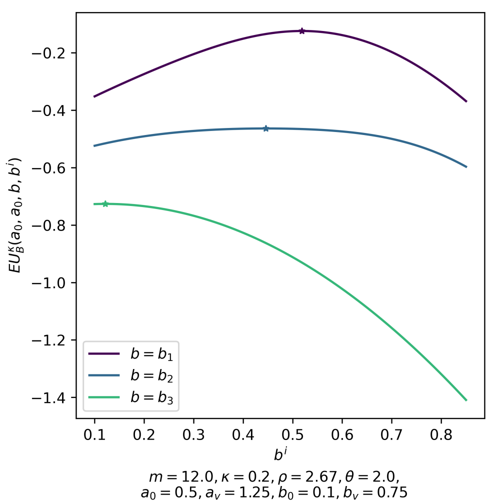


Figure 13: Utility from deviation for each equilibrium candidate

By contrast to the partisan setting, however, in our numerical examples we did not identify any parameter values for which there are more than three equilibria.

Example 10. In Figure 14, we still examine the case $\rho b_v > a_v$ and we vary two parameters at a time. Then, we plot the number of equilibria in panel (a) and the number of equilibria such that the underdog wins in panel (b).

The first line of figures shows how the set of equilibria varies with the degree of universalization κ and m, the curvature parameter for the h function. For low enough values of m, the expected utility is concave and equilibrium uniqueness obtains. Multiplicity of equilibria appears for a value of m around 5.

The second line of figures shows how the set of equilibria varies with the degree of universalization κ and b_0, the size of candidate B's base. A higher b_0 reduces the cost for B-supporters to match A's base. Thus, a high enough b_0 leads to concavity (or at least single-peakedness) of the expected utility and yields uniqueness of equilibria. For lower values of b_0, multiplicity occurs for intermediate values of κ.

The third line of figures shows how the set of equilibria varies with the degree of universalization κ and the stake ρ. For high (resp. low) enough values of ρ and κ there is a unique equilibrium, in which the underdog wins (resp. loses). The coordination problem appears either if κ is not very high but ρ is, or the reverse, and it is the combination of a modest κ and a high ρ that favors the appearance of more than two equilibria.

Note that the figures confirm one of the conclusions from our analysis of global deviations above: multiplicity of equilibria appears only for values of κ neither too close to 0, nor too close to 1. One exception appears in the third line, however, where a value of ρ slightly below 2 corresponds to a knife-edge case with two equilibria for $\kappa = 1$.

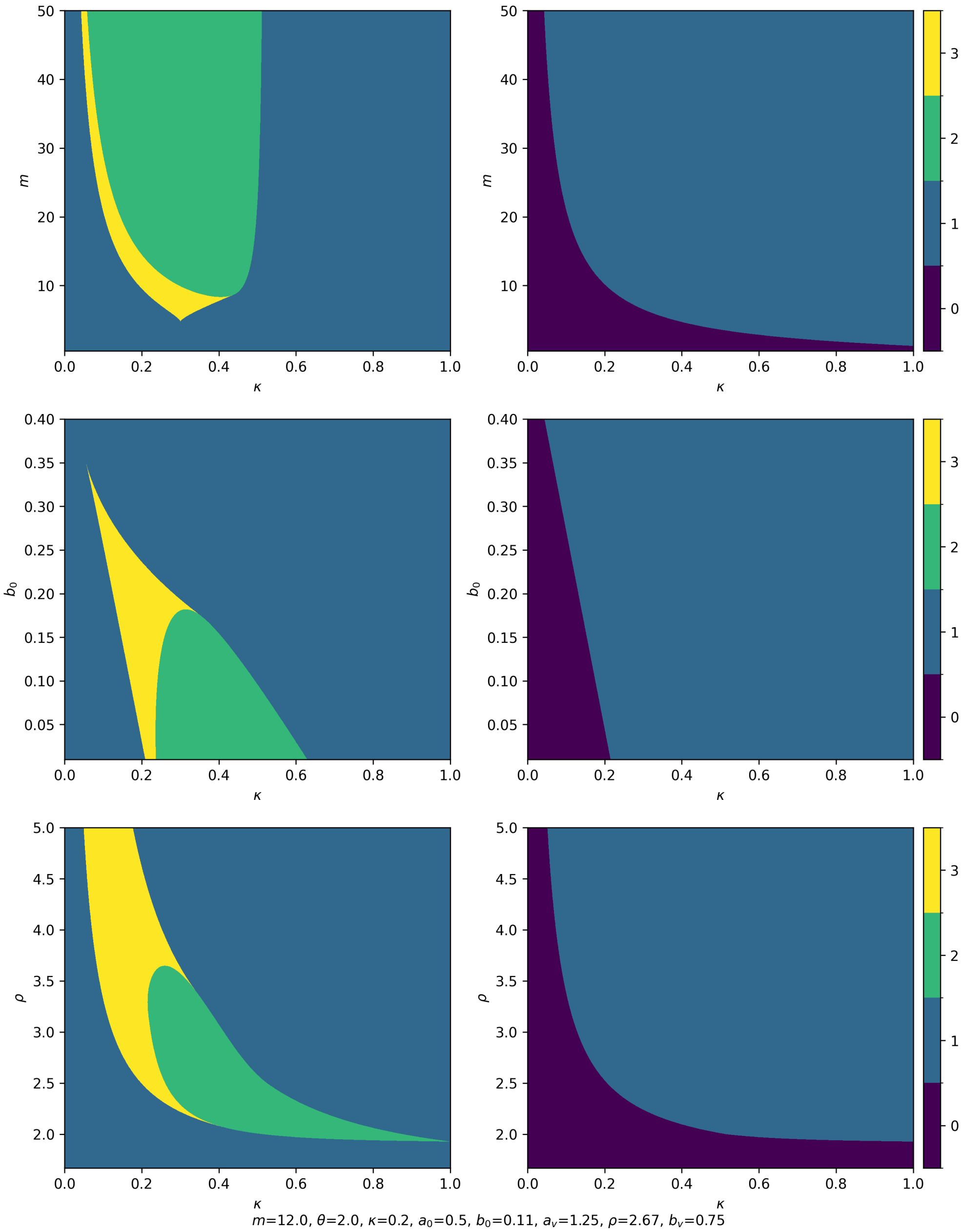


Figure 14: Multiplicity of equilibria (left column) and multiplicity of equilibria where the underdog wins (right column), depending on (κ, m) (first line), (κ, ρ) (second line), (κ, b_0) (third line)

D.3 Sufficient conditions for equilibrium uniqueness

Sufficient conditions for uniqueness are analogous to the partisan case described in Appendix C.3. We will show that uniqueness holds if the auxiliary function from the proof of

Proposition 6 is single-peaked in a suitable sense. Recall

$$\begin{aligned}\Phi(a,b) &= \kappa a_v h(\alpha(a,b)) + \rho\kappa b_v h(\beta(a,b)) - a_v C_A(a) - b_v C_B(b) \\ &= \kappa(a_v - \rho b_v) h(\alpha(a,b)) - a_v C_A(a) - b_v C_B(b) \\ &= \kappa(\rho b_v - a_v) h(\beta(a,b)) - a_v C_A(a) - b_v C_B(b)\end{aligned} \tag{131}$$

As opposed to the partisan case (see Appendix C.3) it is sufficient to assume single-peakedness in the following sense: a univariate function is called single-peaked if it is strictly increasing until reaching its unique maximum, where the latter could be equal to $\bar{b}$. Recall also Proposition 6 in the main text: if $\rho b_v > a_v$, any equilibrium is such that $a = a_0$, and vice-versa, if $a_v > \rho b_v$, any equilibrium is such that $b = b_0$.

Proposition 11. *Let $\kappa > 0$. If $\rho b_v > a_v$ and $b \mapsto \Phi(a_0, b)$ is single-peaked, there exists a unique equilibrium. Similarly, if $a_v > \rho b_v$ and $a \mapsto \Phi(a, b_0)$ is single-peaked, there exists a unique equilibrium.*

Proof. We restrict ourselves to the case $\rho b_v > a_v$, as the proof in the opposite case goes analogously. By Proposition 6 in the main text, $a = a_0$, so that any equilibrium is entirely described by b, and $b > b_0$.

Let b be an equilibrium. It satisfies

$$\left.\frac{\partial}{\partial b^i} EU^\kappa\left(a_0, b, a_0, b^i\right)\right|_{b^i=b} = \frac{\partial}{\partial b}\Phi(a_0, b) = 0 \tag{132}$$

if $b \in (b_0, \bar{b})$ and

$$\left.\frac{\partial}{\partial b^i} EU^\kappa\left(a_0, b, a_0, b^i\right)\right|_{b^i=b} = \frac{\partial}{\partial b}\Phi(a_0, b) \geq 0 \tag{133}$$

if $b = \bar{b}$. In both cases, the single-peakedness assumption implies that b is a maximum of $b \mapsto \Phi(a_0, b)$.

Now, assume for a contradiction that there exist two equilibria $b_1 \neq b_2$. Then, both are maxima of the auxiliary function, contradicting the single-peakedness assumption. □

Note that this proposition generalizes Proposition 8 of the main text: indeed, if, say, $a_v > \rho b_v$ and $a_0 \geq b_0$, then $\alpha(a, b_0) \geq 0$ for all $a \in [a_0, \bar{a}]$ and therefore $h''(a, b_0) \geq 0$ by assumption. Recalling that the cost term is strictly convex, $a \mapsto \Phi(a, b_0)$ is strictly concave.

Similar to the partisan case (see Appendix C.3), we can pin down sufficient conditions for single-peakedness for the specific function we use in our illustrating examples.

Lemma 4. *Let $\kappa > 0$ and let $h(x) = \frac{\arctan(mx)}{\arctan(m)}$.*

1. *Let $\lambda = a_v - \rho b_v > 0$. If*
   - *$\frac{f_A(c)c}{F_A(c)}$ is decreasing,*
   - *for some $r < \frac{2a_v m}{\lambda(m^2-2m+1)}$,*

$$\lim_{c\to 0}\frac{F_A(c)}{c^r} > 0, \tag{134}$$

- *and* $\bar{s}_A \geq \frac{\lambda b_0 (m^2+1)}{2a_v^2 m \arctan(m)}$,

single-peakedness of $a \mapsto \Phi(a, b_0)$ *holds.*

2. *Let* $\lambda = \rho b_v - a_v > 0$. *If*

   - $\frac{f_B(c)c}{F_B(c)}$ *is decreasing,*
   - *for some* $r < \frac{2b_v m}{\lambda(m^2-2m+1)}$,

$$\lim_{c \to 0} \frac{F_B(c)}{c^r} > 0, \tag{135}$$

   - *and* $\bar{s}_B \geq \frac{\lambda a_0 (m^2+1)}{2b_v^2 m \arctan(m)}$,

   single-peakedness of $b \mapsto \Phi(a_0, b)$ *holds.*

Proof. We restrict ourselves to the case $\lambda = a_v - \rho b_v > 0$, as the proof for the other case goes analogously. By an abuse of notation, we write $\Phi(s_A, b_0)$ instead of $\Phi(a_0 + a_v F_A(s_A), b_0)$. Then, writing $M(s_A) = a_0 + a_v F_A(s_A) - b_0$ and $T(s_A) = a_0 + a_v F_A(s_A) + b_0$,

$$\begin{aligned}\frac{\partial}{\partial s_A}\Phi(a_0, s_B) &= \lambda\kappa h'(\alpha(s_A, b_0)) \frac{2\kappa b_0 a_v f_A(s_A)}{T(s_A)^2} - a_v s_A f_A(s_A) \\ &= \underbrace{a_v \frac{f_A(s_A)}{T(s_A)^2 + m^2 M(s_A)^2}}_{>0} \lambda\kappa h'(\alpha(s_A, b_0)) \left(\frac{2m\kappa\lambda b_0}{\arctan(m)} - \left(T(s_A)^2 + m^2 M(s_A)^2\right) s_A\right).\end{aligned} \tag{136}$$

Note that $a \mapsto \Phi(a, b_0)$ is single-peaked if and only if $s_A \mapsto \Phi(s_A, b_0)$ is single-peaked. Therefore, it is sufficient to show that $\phi(s_A) = \left(T(s_A)^2 + m^2 M(s_A)^2\right)$ is strictly increasing, and eventually greater than $\frac{2m\kappa\lambda b_0}{\arctan(m)}$. This follows from the conditions in the same way as in the proof of Proposition 10. □

E SymPy calculations

[1]:
```
from sympy import *
```

[2]:
```
m,rho, sbi,bars,bv, kap = symbols(r'm \rho s_B \bar{s}_B b_v␣
↪\tilde{\kappa}',positive=True)
b = Symbol(r"\tilde{b}")
a = Symbol("a")
abar = Symbol(r"\bar{a}",positive=True)
b0 = Symbol("b_0")
FB = Function("F_B")(sbi)
fB = Function("f_B")(sbi)
FBsymb = Symbol("F_B")
fBsymb = Symbol("f_B")
```

Analyzing ϕ_B

[9]:
```
Tkap = kap*bv*FB + (1-kap)*b + b0 + a
Mkap = kap*bv*FB + (1-kap) * b + b0 - a

phi = Tkap**2 + m**2 * Mkap**2 + 2*sbi*kap*rho *bv*fB *(Tkap + m**2*Mkap)
display(collect(expand(phi),[a**2,a]))
```

$$-2\rho\tilde{\kappa}^2\tilde{b}b_v m^2 s_B^i f_B(s_B^i) - 2\rho\tilde{\kappa}^2\tilde{b}b_v s_B^i f_B(s_B^i) + 2\rho\tilde{\kappa}^2 b_v^2 m^2 s_B^i F_B(s_B^i) f_B(s_B^i) + 2\rho\tilde{\kappa}^2 b_v^2 s_B^i F_B(s_B^i) f_B(s_B^i) + 2\rho\tilde{\kappa}\tilde{b}b_v m^2 s_B^i f_B(s_B^i) + 2\rho\tilde{\kappa}\tilde{b}b_v s_B^i f_B(s_B^i) + 2\rho\tilde{\kappa}b_0 b_v m^2 s_B^i f_B(s_B^i) + 2\rho\tilde{\kappa}b_0 b_v s_B^i f_B(s_B^i) + \tilde{\kappa}^2\tilde{b}^2 m^2 + \tilde{\kappa}^2\tilde{b}^2 - 2\tilde{\kappa}^2\tilde{b}b_v m^2 F_B(s_B^i) - 2\tilde{\kappa}^2\tilde{b}b_v F_B(s_B^i) + \tilde{\kappa}^2 b_v^2 m^2 F_B^2(s_B^i) + \tilde{\kappa}^2 b_v^2 F_B^2(s_B^i) - 2\tilde{\kappa}\tilde{b}^2 m^2 - 2\tilde{\kappa}\tilde{b}^2 - 2\tilde{\kappa}\tilde{b}b_0 m^2 - 2\tilde{\kappa}\tilde{b}b_0 + 2\tilde{\kappa}\tilde{b}b_v m^2 F_B(s_B^i) + 2\tilde{\kappa}\tilde{b}b_v F_B(s_B^i) + 2\tilde{\kappa}b_0 b_v m^2 F_B(s_B^i) + 2\tilde{\kappa}b_0 b_v F_B(s_B^i) + \tilde{b}^2 m^2 + \tilde{b}^2 + 2\tilde{b}b_0 m^2 + 2\tilde{b}b_0 + a^2(m^2+1) + a\Big(-2\rho\tilde{\kappa}b_v m^2 s_B^i f_B(s_B^i) + 2\rho\tilde{\kappa}b_v s_B^i f_B(s_B^i) + 2\tilde{\kappa}\tilde{b}m^2 - 2\tilde{\kappa}\tilde{b} - 2\tilde{\kappa}b_v m^2 F_B(s_B^i) + 2\tilde{\kappa}b_v F_B(s_B^i) - -2\tilde{b}m^2 + 2\tilde{b} - 2b_0 m^2 + 2b_0\Big) + b_0^2 m^2 + b_0^2$$

Indeed, the leading coefficient is positive.

[10]:
```
asol = simplify(solve(diff(phi,a),a)[0])
phi = simplify(phi.subs(a,asol))*(m**2 +1)
display(collect(expand(phi),[b**2,b]))
```

$$-\rho^2\tilde{\kappa}^2 b_v^2 m^4 (s_B^i)^2 f_B^2(s_B^i) + 2\rho^2\tilde{\kappa}^2 b_v^2 m^2 (s_B^i)^2 f_B^2(s_B^i) - \rho^2\tilde{\kappa}^2 b_v^2 (s_B^i)^2 f_B^2(s_B^i) + 8\rho\tilde{\kappa}^2 b_v^2 m^2 s_B^i F_B(s_B^i) f_B(s_B^i) + 8\rho\tilde{\kappa}b_0 b_v m^2 s_B^i f_B(s_B^i) + 4\tilde{\kappa}^2 b_v^2 m^2 F_B^2(s_B^i) + 8\tilde{\kappa}b_0 b_v m^2 F_B(s_B^i) + \tilde{b}^2\cdot(4\tilde{\kappa}^2 m^2 - 8\tilde{\kappa}m^2 + 4m^2) + \tilde{b}\left(-8\rho\tilde{\kappa}^2 b_v m^2 s_B^i f_B(s_B^i) + 8\rho\tilde{\kappa}b_v m^2 s_B^i f_B(s_B^i) - 8\tilde{\kappa}^2 b_v m^2 F_B(s_B^i) - -8\tilde{\kappa}b_0 m^2 + 8\tilde{\kappa}b_v m^2 F_B(s_B^i) + 8b_0 m^2\right) + 4b_0^2 m^2$$

Notice that $\tilde{\kappa}^2 - 2\tilde{\kappa} + 1 = (\tilde{\kappa} - 1)^2 > 0$ for $\tilde{\kappa} \neq 1$, showing that the leading coefficient is positive. Moreover, observe that for $\tilde{\kappa} = 1$, the polynomial is actually constant in $\tilde{b}$, i.e. of degree zero.

[11]:
```
bsol = simplify(solve(Eq(diff(phi,b),0),b)[0])
display(bsol)
```

$$\frac{\rho\tilde{\kappa}b_v s_B^i f_B(s_B^i) + \tilde{\kappa}b_v F_B(s_B^i) + b_0}{\tilde{\kappa} - 1}$$

If $\tilde{\kappa} \neq 1$, after minimizing over $\tilde{b}$, we conclude that the expression is minimal for a negative $\tilde{b}$. For $\tilde{\kappa} = 1$, the expression does not change with $\tilde{b}$. In either case, we can plug in $\tilde{b} = 0$ because it minimizes the expression over $\tilde{b}$ in our admissible range $[0, b_v]$.

[12]:
```
phi = collect(expand(simplify(phi.subs(b,0))),b0)
display(phi)
```

$$-\rho^2\tilde{\kappa}^2 b_v^2 m^4 (s_B^i)^2 f_B^2(s_B^i) + 2\rho^2\tilde{\kappa}^2 b_v^2 m^2 (s_B^i)^2 f_B^2(s_B^i) - \rho^2\tilde{\kappa}^2 b_v^2 (s_B^i)^2 f_B^2(s_B^i) + 8\rho\tilde{\kappa}^2 b_v^2 m^2 s_B^i F_B(s_B^i) f_B(s_B^i) + 4\tilde{\kappa}^2 b_v^2 m^2 F_B^2(s_B^i) + 4b_0^2 m^2 + b_0 \cdot (8\rho\tilde{\kappa}b_v m^2 s_B^i f_B(s_B^i) + 8\tilde{\kappa}b_v m^2 F_B(s_B^i))$$

Observe that the coefficient in front of b_0 is strictly positive, allowing us to minimize easily over b_0.

[13]:
```
b0sol = simplify(solve(Eq(diff(phi,b0),0),b0)[0])
display(b0sol)
```

$-\tilde{\kappa}b_v\left(\rho s_B^i f_B\left(s_B^i\right)+F_B\left(s_B^i\right)\right)$

This expression is negative. Since we allow only positive values for b_0, our expression is minimal for $b_0=0$. The next step is to plug in $b_0=0$ and to obtain a polynomial in s_B, regarding F_B and f_B as constants.

[14]:
```
phi = phi.subs(b0,0).subs(fB,fBsymb).subs(FB,FBsymb)
display(collect(expand(phi),sbi))
```

$$4F_B^2\tilde{\kappa}^2b_v^2m^2+8F_B\rho\tilde{\kappa}^2b_v^2f_Bm^2s_B^i+\left(s_B^i\right)^2\left(-\rho^2\tilde{\kappa}^2b_v^2f_B^2m^4+2\rho^2\tilde{\kappa}^2b_v^2f_B^2m^2-\rho^2\tilde{\kappa}^2b_v^2f_B^2\right)$$

$$\frac{2F_Bm}{\rho f_B\left(m^2-2m+1\right)}$$

$$-\frac{2F_Bm}{\rho f_B\left(m^2+2m+1\right)}$$

We see here that that the leading coefficient is negative as long as $m^4-2m^2+1=(m^2-1)^2$ is positive, i.e. for $m\neq 1,-1$. Therefore, the polynomial expression is positive as long as s_B is between the roots of the polynomial expression. For $m=1$, the expression is always positive for $s_B\geq 0$. We calculate the roots below.

[16]:
```
sols = solve(phi,sbi)
for sol in sols:
    display(simplify(sol))
```

$$\frac{2F_Bm}{\rho f_B\left(m^2-2m+1\right)}$$

$$-\frac{2F_Bm}{\rho f_B\left(m^2+2m+1\right)}$$

The second root being negative whereas $s_B\geq 0$, this allows us to find a positivity criterion by comparing s_B to the first root.

Analyzing ψ_B

[17]:
```
psi = Tkap**2 + m**2 * Mkap **2
display(collect(expand(psi),[a**2,a]))
```

$$\tilde{\kappa}^2\tilde{b}^2m^2+\tilde{\kappa}^2\tilde{b}^2-2\tilde{\kappa}^2\tilde{b}b_vm^2F_B\left(s_B^i\right)-2\tilde{\kappa}^2\tilde{b}b_vF_B\left(s_B^i\right)+\tilde{\kappa}^2b_v^2m^2F_B^2\left(s_B^i\right)+\tilde{\kappa}^2b_v^2F_B^2\left(s_B^i\right)-2\tilde{\kappa}\tilde{b}^2m^2-2\tilde{\kappa}\tilde{b}^2-2\tilde{\kappa}\tilde{b}b_0m^2-2\tilde{\kappa}\tilde{b}b_0+2\tilde{\kappa}\tilde{b}b_vm^2F_B\left(s_B^i\right)+2\tilde{\kappa}\tilde{b}b_vF_B\left(s_B^i\right)+2\tilde{\kappa}b_0b_vm^2F_B\left(s_B^i\right)+2\tilde{\kappa}b_0b_vF_B\left(s_B^i\right)+\tilde{b}^2m^2+\tilde{b}^2+2\tilde{b}b_0m^2+2\tilde{b}b_0+a^2\left(m^2+1\right)+a\left(2\tilde{\kappa}\tilde{b}m^2-2\tilde{\kappa}\tilde{b}-2\tilde{\kappa}b_vm^2F_B\left(s_B^i\right)+2\tilde{\kappa}b_vF_B\left(s_B^i\right)-2\tilde{b}m^2+2\tilde{b}-2b_0m^2+2b_0\right)+b_0^2m^2+b_0^2$$

The leading coefficient in a being positive, we can minimize easily in a. We plug the result into ψ_B.

[18]:
```
asol = solve(psi.diff(a),a)[0]
psi = simplify(psi.subs(a,asol))
```

```
display(collect(expand(psi),[b**2,b]))
```

$$\frac{4\tilde{\kappa}^2 b_v^2 m^2 F_B^2(s_B^i)}{m^2+1} + \frac{8\tilde{\kappa} b_0 b_v m^2 F_B(s_B^i)}{m^2+1} + \tilde{b}^2 \cdot \left(\frac{4\tilde{\kappa}^2 m^2}{m^2+1} - \frac{8\tilde{\kappa} m^2}{m^2+1} + \frac{4m^2}{m^2+1}\right) + \tilde{b}\left(-\frac{8\tilde{\kappa}^2 b_v m^2 F_B(s_B^i)}{m^2+1} - \frac{8\tilde{\kappa} b_0 m^2}{m^2+1} + \frac{8\tilde{\kappa} b_v m^2 F_B(s_B^i)}{m^2+1} + \frac{8 b_0 m^2}{m^2+1}\right) + \frac{4 b_0^2 m^2}{m^2+1}$$

This expression as a polynomial in $\tilde{b}$ has a positive leading coefficient for $\tilde{\kappa} \neq 1$, and is otherwise constant in $\tilde{b}$. For $\tilde{\kappa} \neq 1$, we minimize over $\tilde{b}$.

[19]:
```
bsol = solve(psi.diff(b),b)[0]
display(bsol)
```

$$\frac{\tilde{\kappa} b_v F_B(s_B^i) + b_0}{\tilde{\kappa} - 1}$$

This solution being negative for $\tilde{\kappa} \neq 1$ and the aforementioned expression being constant in $\tilde{b}$ for $\tilde{\kappa} = 1$, we may plug in $\tilde{b} = 0$ as our lowest possible $\tilde{b}$.

[20]:
```
psi = simplify(psi.subs(b,0))
display(collect(expand(psi),[b0**2,b0]))
```

$$\frac{4\tilde{\kappa}^2 b_v^2 m^2 F_B^2(s_B^i)}{m^2+1} + \frac{8\tilde{\kappa} b_0 b_v m^2 F_B(s_B^i)}{m^2+1} + \frac{4 b_0^2 m^2}{m^2+1}$$

We minimize this expression over b_0, which is straightforward due to the positive leading coefficient.

[21]:
```
b0sol= solve(psi.diff(b0),b0)[0]
display(b0sol)
```

$$-\tilde{\kappa} b_v F_B\left(s_B^i\right)$$

This expression being negative, we may substitute $b_0 = 0$. Moreover, recall that we assumed $s_B = \bar{s}_B$, so that $F_B(s_B) = 1$, yielding the minimal ψ_B for $s_B = \bar{s}_B$.

[23]:
```
psi = psi.subs(b0,0)
psi = psi.subs(FB,1)
display(psi)
```

$$\frac{4\tilde{\kappa}^2 b_v^2 m^2}{m^2+1}$$