Voter Turnout with Peer Punishment☆

David K. Levine[1], Andrea Mattozzi[2]

Abstract

We introduce a model of turnout where social norms, strategically chosen by competing political parties, determine voters' participation. When the cost of enforcement of social norms is low, the larger party is always advantaged. Otherwise, in the spirit of Olson (1965), the smaller party may be advantaged. Party advantage depends on the convexity of the function describing the cost to the party of choosing an incentive compatible social norm. Our model shares features of the "ethical" voter model in large elections and of the pivotal voter model in small elections, and it delivers novel and empirically relevant comparative statics results.

JEL Classification Numbers: later

Keywords: Voting, Turnout, Party, Peer Punishment

☆First Version: June 2, 2013. Previously titled "Voter Participation with Collusive Parties." We are especially grateful to Juan Block, Michele Boldrin, Rohan Dutta, Tim Feddersen, Andrea Galeotti, Piero Gottardi, Johannes Hörner, Philipp Kircher, Cesar Martinelli, Salvatore Modica, Massimo Morelli, Salvatore Nunnari, Pietro Ortoleva, Thomas Palfrey, Kirill Pogorelskiy, John Roemer, Guido Tabellini, Luigi Tenco, Francesco Trebbi, participants at the 2015 Priorat Workshop, 2015 Barcelona JOCs, 2016 Cornell Political Economy Conference, 2016 ASSET, the UCL Theory Workshop, the UECE Meetings in Lisbon, the Dondena Political Economy Seminar in Bocconi, the Yale theory seminar, the MEDS theory seminar and four referees. Giovanni Andreottola provided excellent reseaearch assistance.

[1]Department of Economics, EUI and WUSTL, david@dklevine.com

[2]Department of Economics, EUI and MOVE, andrea.mattozzi@eui.eu

1. Introduction

Woman who ran over husband for not voting pleads guilty.
USA Today April 21, 2015

Social norms play a key role in voter participation. To take a few of many pieces of evidence, Della Vigna et al (2014) and Funk (2010) demonstrate that an important incentive for citizens to vote is to show others that they have voted. Gerber, Green and Larimer (2008) show that social pressure significantly increase turnout. Palfrey and Pogorelskiy (2016) provide experimental evidence that communication among voters and in particular communication within parties increases turnout. Larreguy, Marshal and Querubin (2016) show that greater monitoring capacity of parties' local mobilizers increases turnout buying. The relation between voter participation and peer pressure is also widely discussed in the sociology literature, see for example Coleman (1988).

Typically social norms are maintained by various forms of social disapproval and ostracism. This is well documented in Elinor Ostrom's work, especially Ostrom (1990). While the news article mentioned in the incipit is clearly an extreme case of punishment, a less traumatic example is represented by Ted Cruz's campaign strategy in the 2016 Iowa Presidential primaries. Voters who were most likely to support Cruz received mailings with information about their own past voting behavior and that of their neighbors, stating that they were in "voting violation" if they had not participated in all previous past elections.

In this paper we introduce a theory of turnout based on the premise that social norms affect voter participation and these norms are chosen strategically by competing political parties. In our model social norms enforced by peer pressure are endogenous, and maximize a group objective. The endogeneity is hardly questionable: we know that turnout in U.S. national elections is considerably higher in presidential election years than off-years, and in general participation rates and the social norms that lead to them adjust strategically to reflect the stakes in the elections. As for maximizing a group objective, Coleman (1988) and Ostrom (1990) as well as Olson (1965) argue that - within the limits of available monitoring and punishment - peer pressure mechanisms do a good job of solving public goods problems. Hence, a simple starting point for modelling endogenous social norms that reflect party interests is to assume that they are optimal with respect to a group objective function. But how does this work in practice?

In real elections there are many party leaders actively trying to encourage turnout. These range from the candidates themselves to party officials, donors, interest groups, lower ranking elected officials, party activists, and so on.[3] In the U.S., for example, evidence for the central role of political parties as turnout monitors and enforcers can be found in the systematic use of targeted turnout canvassing.[4] We typically think of these actors as wishing to encourage turnout to the maximum and not being so concerned with the costs that voters might incur. Yet voters do care, and it is them who largely enforce the social norms through peer networks. Hence, while leaders may communicate the importance of the election, strategic concerns and so forth, in the end they must convince voters that it is worth turning out. Furthermore, leaders often face resource constraints in doing so: for example in local races, higher level party officials direct funding to particular races.

We consider a large population of voters divided into two parties of different size. Parties choose a social norm in the form of a participation rate for their members. They do so in an effort to win an election. Individual party members, given the social norm, optimally choose whether or not to vote. Voters monitor each others voting behavior within each party and punish deviators. The total cost to parties of choosing an incentive compatible social norm is the sum of two components: the turnout cost, which is the participation cost of voting born by voters, and the monitoring cost, which is the expected cost of punishing party members who did not vote. Because both parties incur the cost of turning out voters but only the one with larger turnout wins, the resulting game is a

[3]As Larreguy, Marshal and Querubin (2016) point out: "Common mobilization strategies - including door-to-door or telephone campaigns, transporting voters to the polling station, and turnout buying - rely on the superior knowledge of local mobilizers about the political preferences of individual voters, which enables them to identify supporters that would not have otherwise turned out."

[4]Data companies such as NGP VAN, Catalist and Aristotle produced the bulk of Democrats and Republicans' detailed vote registration information aimed at vote mobilization.

type of all-pay auction. We show that this game has a unique mixed strategy equilibrium in which the party with the lowest cost of mobilizing a given fraction of voters - *the advantaged party* - gets all the surplus. We give a simple and economically intuitive formula for determining which party is advantaged, which depends on the maximal number of voters that a party is able *and* willing to mobilize. [5]

Our main finding is that when monitoring costs are small, and the benefit of winning the election is the same for both parties, the larger party is always advantaged, turns out a higher expected number of voters, and has a better chance of winning the election. These results are consistent with the predictions of ethical voter (Feddersen and Sandroni (2006) and Coate and Conlin (2004)) and follow-the-leader (Shachar and Nalebuff (1999)) type of theories. By contrast, and contrary to existing theories of political participation, when monitoring costs are large, our model delivers a result reminiscent of Olson (1965) in which the smaller party ise advantaged, turns out a higher expected number of voters and has a better chance of winning the election. As we will discuss, this is consistent with existing evidence - for example on School Board elections.

The intuition for our finding follows from uncovering that party advantage depends crucially on the convexity of the total cost to the party of choosing an incentive compatible social norm. While the turnout cost is typically increasing and convex, this is not the case for monitoring costs. As a result, the total cost of mobilization can be either convex or concave. If it is convex, as it is commonly assumed in existing theories of political participation that abstract from monitoring costs, a large party has the lowest cost of mobilization and it

[5]Notice that, in the absence of aggregate shocks, mixed stragies are essential in contests with opposing interests. That is - voting between parties competing for turnout has the flavor of matching pennies and all pay auctions, as originally shown by Hillman and Riley (2006), do not have pure strategy equilibria. Interestingly, there is a sense in which mixed equilibria - in particular when mixing is done by the parties rather than by individuals as in our model - are reflected in the reality of elections. A good example is the case of "GOTV" (Get Out The Vote) efforts by real political parties. These mobilization efforts, ranging from phone calls to driving people to the polls, are typically more successful the less predictable they are. We further elaborate on this point in Section 3.2.

is therefore advantaged. If the cost of mobilization fails to be convex, which is the case when the enforcement of social norms is costly, it is possible for a small party to have the lowest cost of mobilizing voters. Hence, introducing peer pressure and monitoring costs in a theory of turnout is not only empirically relevant, but it increases the range of predictions that the theory can deliver.

In terms of comparative statics, we show that if the electoral stakes are high enough, then lowering them encourages the small party and makes elections more competitive.[6] Perhaps more interestingly, we show that an increase in monitoring efficiency can increase turnout, a relation that has been shown to hold empirically in Larreguy, Marshal and Querubin (2016). The reverse direction, that is decreasing monitoring inefficiency leads to lower turnout and more competitive elections, is also relevant. Consider for example the discussion in political science - described, among others, in the introduction to Katz and Mair (1995) - of the breakdown of the mass party system established after the Second World War. It is argued that mass parties - political organizations able to successfully mobilize large proportions of the electorate - were supported by strong social ties. This has a natural consequence in our setting of keeping monitoring costs low. The breakdown of these ties raised monitoring costs and led to lower turnout and more competitive elections: the prologue to the breakdown of the mass party system. Finally, we use our model to investigate the strategic use of vote suppression - showing that disadvantaged parties will never find it profitable to suppress the vote of their rival, but that advantaged parties often will. This result suggests that when the Republican party is advantaged, it will be tempted to introduce voters suppression laws in the form of strict voter identification requirements in order to hold down its own turnout cost. Our finding is in contrast with the rationale often discussed in the popular press, which argues that in a closely contested state a short term Republican victory may be translated into long term advantage by introducing a voter suppression law. Following this we extend the basic model to explore under what conditions higher turnout might be welfare improving.

[6]This is reminiscent of Sayre's law: "Academic politics is the most vicious and bitter form of politics, because the stakes are so low."

In the last part of the paper, we introduce aggregate shocks and examine general contest resolution functions and incentive constraints that account for pivotality. We show that when aggregate shocks are sufficiently large or pivotality is sufficiently important, pure strategy equilibria exist. Conversely, with large electorates and small aggregate shocks only mixed strategy equilibria exist. In this precise sense, our model shares features of the "ethical" voter model of Feddersen and Sandroni (2006) and Coate and Conlin (2004) in large elections, and of the pivotal voter model of Palfrey and Rosenthal (1985) in small elections. We also examine the robustness of the comparative statics of the all-pay auction to aggregate shocks and pivotality and give conditions under which participation declines with the size of the electorate.

1.1. Literature Review

Our work builds on existing rational theories of voter participation. These theories remain controversial. The standard pivotal voter model of Palfrey and Rosenthal (1985) finds empirical support in the laboratory (see for example Levine and Palfrey (2007)) but it has difficulty explaining turnout in large scale plurality rule elections. Indeed, Coate, Conlin and Moro (2008) show that in a sample of Texas liquor referenda, elections are much less close than what would be predicted by the pivotal voter model, and Coate and Conlin (2004) show that a model of "ethical" voters better fits the data than the model of pivotal voters. Not surprisingly, the probability of being pivotal in large elections is very low as documented by Mulligan and Hunger (2003) and Shachar and Nalebuff (1999).[7] Furthermore, this probability - being proportional to the standard error - should decline roughly as the square root of the number of voters. The same applies to participation rates if voting costs are non-negative and uniformly distributed.[8] However, if we focus on post-war national elections

[7]Based on Shachar and Nalebuff (1999)'s calculations on the probability of casting a pivotal vote in presidential elections, in all but a few states a rational voter for whom it cost $1 to vote would have to win a prize larger than the wealth of the wealthiest person in the world to turnout.

[8]In a two-candidates election with an even number of voters n each casting her vote randomly, the probability of a tie approaches $\sqrt{2/n\pi}$ as n grows large. See Penrose (1945) and Chamberlain and Rothschild (1981).

in consolidated democracies with per capita income above the world average and voluntary voting, the relation between voter turnout and the size of the country is hardly consistent with the predictions of the pivotal voter model.

The most recent rational voter theories studied by economists have been the follow-the-leader model introduced by Shachar and Nalebuff (1999), and the social preference model of ethical voters introduced by Feddersen and Sandroni (2006) and Coate and Conlin (2004).[9] Roughly speaking these models assume that some or all voters choose to participate based upon whether or not the benefits of their vote to their party justifies the cost of their participation. While Shachar and Nalebuff (1999) focus on the costs to the leaders, we follow the ethical voter literature in focusing on the costs to the followers. Furthermore, we take the view that social preferences arise as a social norm. That is, rather than assuming that voters weigh the benefits to the party of their vote against the cost, we assume that each party weighs the benefits of voter turnout against the costs and chooses a social norm that is optimal for the party. In turn, this social norm is enforced through costly peer punishment.[10]

Our model of peer punishment originates in Kandori (1992)'s work on social norms in repeated games and is a variant of the mechanism design approach to collective action of Levine and Modica (2014) and Dutta, Levine and Modica (2014). In these models agents monitor each others behavior and punish deviators through ostracism and social disapproval. In our model, we hypothesize that the reason why voters want to show others that they have voted is that either they have internalized a social norm, or they expect to be rewarded for following the social norm or punished for failing to do so.[11] The idea of shifting focus from voters to parties is nothing new - a large range of literature in political economy studying parties such as elites, masses and other groups, often

[9]Ali and Lin (2013) extend the Feddersen and Sandroni (2006) model by introducing "pragmatic" voters alongside ethical voters. A pragmatic voter votes only because she wishes others to think of her as being ethical.

[10]Interestingly, Shachar and Nalebuff (1999) show that parties' effort (measured by the number of calls and visits to individuals to encourage their turnout) is positively correlated with group membership and parents' involvement in politics.

[11]See the experimental evidence of Della Vigna et al (2014).

treats these groups as single players who act in the group interest. Our political parties behave in a similar way although they must do so subject to incentive constraints - that is, parties can only make choices that are incentive compatible for its members. If punishment is adequate to induce voter turnout, then that turnout can be chosen by the party - otherwise not.

2. The Model

Consider a large population of N voters divided into two parties of size $\eta_k N$, where $k = \{S, L\}$ denotes *small* and *large*. Assume that $\eta_S < \eta_L$ and $\eta_S + \eta_L = 1$. These parties compete in an election.

Parties - either by consensus or directed by leaders - move first, and simultaneously and non-cooperatively choose a social norm in the form of a participation rate for their members. The individual party members move second and, given the social norm, optimally choose whether or not to vote in an election. For simplicity we assume symmetry in voting costs between the two parties: the general case is addressed in the supplemental appendix. Next, we describe the cost to each voter of voting, and then we introduce electoral competition between the two parties, their strategies and the equilibrium notion. We subsequently discuss the key assumptions.

2.1. The Cost of Participation

Each identical party member privately draws a type y from a uniform distribution on $[0, 1]$. This type determines a *net participation cost* of voting $c(y)$ and, based on this, the member decides whether or not to vote. This cost of voting consists of the direct cost and inconvenience (costs of time and transportion) minus the direct personal benefits such as fulfilling civic duty, the camaraderie of the polling place or expressive voting. The participation cost of voting c is continuously differentiable, strictly increasing and satisfies $c(\underline{y}) = 0$ for some $\underline{y} \in [0, 1]$. Party members for whom $y < \underline{y}$, those with a negative net cost of voting, are called *committed voters* and will always vote.

The social norm of the party is a threshold φ_k together with a rule prescribing voting if $y \leq \varphi_k$. Hence φ_k is the probability that a representative party member votes, and the expected partipation rate of the party. Since N is large we assume that φ_k is also the actual participation rate of party k. This rule is enforced through peer auditing and the possibility of imposing punishments on party members. Specifically, each member of the party is audited by another party member. For concreteness we may think of party members as forming a circular network with each party member auditing the member to their left. The auditor observes whether or not the auditee voted. If the auditee did not vote and the auditee violated the policy the auditor learns this for certain. If the auditee did not violate the policy there is a probability $1 - \theta \in [0, 1]$ that the auditor will learn this. The value of θ represents the *monitoring inefficiency*: if $\theta = 1$ then the auditor learns nothing about y; if $\theta = 0$ the auditor perfectly observes whether y is above or below the threshold φ_k. Whatever the quality of the signal, if the auditee voted or is discovered not to have violated the policy, the auditee is not punished. If the auditee did not vote and the auditor cannot determine whether or not the auditee violated the policy, the auditee is punished with a loss of utility $P_k \geq 0$.[12] A social norm φ_k is *incentive compatible* if and only if $P_k = c(\varphi_k)$. Any member with $y \leq \varphi_k$ would be willing to pay the participation cost $c(y)$ of voting rather than face the certain punishment P_k, while any member with $y > \varphi_k$ prefers to pay the expected cost of punishment θP_k over the participation cost of voting $c(y)$. The punishment itself, as it is paid by a member, is a cost to the party and we assume that the overall cost of a punishment P_k to the party is exactly P_k.

We are now ready to determine how costly it is for the party to induce additional members other than the committed voters to vote. As with participation cost we will measure all costs per capita. The *total cost* of choosing an incentive compatible social norm $\varphi_k > \underline{y}$ is denoted by $C(\varphi_k)$ with the convention that $C(\varphi_k) = 0$ for $\varphi_k \leq \underline{y}$. We can decompose the total cost $C(\varphi_k)$ into two additive components. The first component is the *turnout cost* $T(\varphi_k) = \int_{\underline{y}}^{\varphi_k} c(y)dy$, which

[12]In addition, the audits and the punishments may themselves be costly. If so we assume that these costs are proportional to the size of the punishment. The case where there is a binding upper bound on the size of punishment is covered in the Supplemental Appendix.

is the participation cost of voting to the member types who vote.[13] The second component is the *monitoring cost* $M(\varphi_k) = \int_{\varphi_k}^{1} \theta P_k dy$, which is the (expected) cost of punishing party members who did not vote.

2.2. Election and Strategies

Parties compete in an election and we assume that the side that produces the greatest expected number of votes wins a prize worth $v_L > 0$ and $v_S > 0$ to each member respectively.[14] We assume that both parties face per capita costs of turning out voters characterized by the same $\underline{y}$, total cost of participation $C(\varphi_k)$ and same punishment P.

The turnout rate φ_k is the fraction of the party that turns out, and it determines the per capita cost of voting. The outcome of the election is determined by the fraction of the electorate $b_k = \eta_k \varphi_k$ that turns out, and we refer to this as the *bid* of party k. The party that "submits the highest bid" wins. This is similar to an all-pay auction: the highest bid wins, but each party pays the cost for their bid. There is also a *tie-breaking rule* $0 \leq B_S(b_S) \leq 1$ which is a measurable function specifying the probability of the small party winning the election when $b_L = b_S$ with the probability of the large party winning $B_L = 1 - B_S$.

A strategy for party k is a probability measure represented by a cumulative distribution function F_k over bids, that is, on $[\eta_k \underline{y}, \eta_k]$. It will be convenient to denote by $F_k^0(b_k) = F_k(b_k) - \sup_{\beta < b_k} F_k(\beta)$ the size of the discrete jump in F_k at b_k (if any). We take the objective function of each party to be the total utility of members

$$U_k(b_k, F_{-k}) = \Pi_k(b_k, F_{-k}) \eta_k v_k - \eta_k C(b_k/\eta_k)$$

where $\Pi_k(b_k, F_{-k})$ is the probability that a bid b_k wins - the probability of a strictly lower bid by the other party plus the probability of winning a tie:

$$\Pi_k(b_k, F_{-k}) = \left(F_{-k}(b_k) - F_{-k}^0(b_k)\right) + F_{-k}^0(b_k) B_k(b_k).$$

[13]Coate and Conlin (2004) assume that $c(y)$ is linear so that the turnout cost of voting for $y \geq \underline{y}$ is quadratic.

[14]In Section 7 we consider the case in which the election is decided by the actual number of votes rather than the expected number of votes.

An *equilibrium* consists of strategies for both parties together with a tie-breaking rule such that the strategy of a party is optimal given the strategy of the other party and the tie-breaking rule.[15]

3. Discussion of the Assumptions

There are two key elements of the model: peer punishment to enforce the social norm and the endogenous choice of a social norm that maximizes the utility of the representative party member. We first discuss these assumptions, then examine why adding monitoring costs to a group-turnout model is important.

3.1. Peer Punishment

We assume that peer pressure and social norms take place within a party rather than globally. That is, if the social norm is that voting is a civic duty then this is not a party specific social norm. Our view is that both global social norms and party social norms matter. Here we incorporate the general social norm of civic duty into the committed voters, treating it as exogenous - as it is standard in previous literature - in order to focus on the party social norm. There is a number of reasons why the party social norm should be relevant. First, many voters are more inclined to enforce a social norm within their own party - that is with voters who are likely to vote as they do. Certainly in a high stakes election many people will put a lot of pressure on like-minded family and friend to vote, and will be less likely to bother with those who take the "wrong position." This has been observed both in the laboratory (Grosser and Schram (2006)) and in the field (Bond et al (2012)). More to the point: there is a high correlation of political beliefs within the social networks important for enforcing social norms. We see this in the positive correlation of political beliefs within families (Jennings, Stoker and Bowers (2009)), in the geographic concentration of political preferences (Chen and Rodden (2013)), and in the strength of party

[15]Notice that in this definition the tie-breaking rule is endogenous. As explained by Simon and Zame (1990), and as we shall discuss subsequently, this is necessary to guarantee existence of an equilibrium.

identity (Dunham, Arechar and Rand (2016)).[16]

3.2. Endogenous Social Norms

Our model of peer punishment is meant to capture in a very stylized way a far more elaborate informal process. In reality, the political hierarchy (candidates, party officials, donors, and activists) chooses strategically to devote resources to mobilization, and voters use this to determine an appropriate social norm through pub and dinner table conversations (not to speak of Facebook). Nevertheless we think that the model captures the important feature that the social norm adapts to the circumstances of the election.

One consequence of endogenizing social norms in a competitive environment and in the absence of aggregate shocks is that mixed stragies become essential. Notice, however, that in our model the mixing is done by the parties rather than by individuals and this is reflected in the reality of elections as in the case of "GOTV" (Get Out The Vote) efforts. Our view is that these efforts are an important part of establishing the social norm for the particular election, and indeed, GOTV efforts are variable and strategic. Furthermore, political parties have strong incentives not to advertize their GOTV effort, and in fact to keep it secret.[17] Clearly, there is little reason to do that unless indeed GOTV effort is random. Hence, the mere fact that it is secret provides evidence that - consciously or not - political parties engage in randomization when choosing social norms for particular elections.

Finally, we should observe that the necessity of party leaders establishing a social norm that is optimal from the voter point of view is not baked into our model. While we have interpreted v_k as the value of the prize to the voter, in

[16] One theory of the strength of these social networks is the skill selection model of Penn (2015).

[17] Accounts in the popular press document both the surprise over the strength of the GOTV and the secrecy surrounding it. For example "The power of [Obama's GOTV] stunned Mr. Romney's aides on election night, as they saw voters they never even knew existed turn out..." Nagourney et al (2012) or "[Romney's] campaign came up with a super-secret, super-duper vote monitoring system [...] to plan voter turnout tactics on Election Day " York (2012). Note that the secrecy at issue is not over whether or not people voted as for example voting pins: we assume that the act of voting is observable. Rather the secrecy is over the social norm that is enforced on election day.

the end it simply represents a value used to establish the social norm. If party leaders have a strong voice in this, then v_k may represent in whole or in part the value of the prize to the leaders which may well be larger than the value to individual voters. For example, the value of the prize might have two parts $v_k = w_k + \nu_k$, where w_k is the value to a party member and ν_k is some valence dimension of the leader or candidate.

3.3. Why Monitoring Costs Matter

What is gained by adding the monitoring cost to a group-turnout model? A crucial question for understanding voting outcomes is whether or not the function $C(\varphi_k)$ is convex because this determines which party has the lowest cost of mobilizing a given fraction of voters. If $C(\varphi_k)$ is convex the large party has the lowest cost of bidding. If it fails to be convex it is possible for the small party to have a lower cost of bidding. It turns out that introducing monitoring affects the convexity of $C(\varphi_k)$.

Notice that $T^{'}(\varphi) = c(\varphi_k)$ and so $T(\varphi_k)$ is both increasing and convex. Hence without monitoring cost the large party always has the lowest cost of bidding. The crucial feature of the monitoring cost is that it has the opposite implication. Since incentive compatibility requires $P = c(\varphi_k) = T^{'}(\varphi_k)$, the monitoring cost can be written as $M(\varphi_k) = \theta(1 - \varphi_k)T^{'}(\varphi_k)$. This function cannot be convex or everywhere increasing if $\theta > 0$. It is non-negative, takes on strictly positive values, yet at the endpoints is equal to zero since $M(\underline{y}) = T^{'}(\underline{y}) = c(\underline{y}) = 0$ and also $M(1) = 0$. In fact, when only committed voters vote, no monitoring is needed, while on the other hand if everyone votes there is nobody to punish. Notice that this property of M is robust to the details of the particular monitoring process. Because the monitoring cost cannot be convex it might be the case that the total cost $C(\varphi_k)$ also fails to be convex. A simple example illustrates the point.

Example. Suppose that, as for example in Coate and Conlin (2004), for $\varphi_k \geq \underline{y}$ the participation cost is linear $c(\varphi_k) = 2(\varphi_k - \underline{y})$ or equivalently that turnout cost $T(\varphi_k) = (\varphi_k - \underline{y})^2$ is quadratic. Simple algebra delivers that $C(\varphi_k) = (1 - 2\theta)(\varphi_k - \underline{y})^2 + 2\theta(1 - \underline{y})(\varphi_k - \underline{y})$. In other words, quadratic turnout cost implies quadratic monitoring cost and therefore quadratic total cost. Moreover $C'(\varphi_k) = 2\left((1-2\theta)(\varphi_k - \underline{y}) + \theta(1-\underline{y})\right) = 2\left(\theta(1-\varphi_k) + (1-\theta)(\varphi_k - \underline{y})\right) > 0$, so the total cost $C(\varphi_k)$ is strictly increasing. Furthermore, $C''(\varphi_k) = 2(1-2\theta)$ so $C(\varphi_k)$ is concave if monitoring is sufficiently inefficient, that is $\theta > 1/2$, and convex otherwise.

Clearly, the specific model of monitoring introduced above should be viewed as illustrative: it is the non-convexity of M that we use in the sequel. In the rest of the paper we shall assume, as in the example, that $C'(\varphi_k) > 0$.

Insofar as monitoring costs are significant to determining voter turnout it is important to recognize that they are something that can be measured. For example, the strength of ties in a social network seems particularly relevant not because of how many connections individuals have, but because of how well known each person is. In a closer knit community in which everyone is well known by someone, monitoring costs should be low - meaning that total costs will tend to be convex. In the bigger more anonymous communities - big apartment blocks of poor people none of whom are known by their neighbors - are likely to have high monitoring costs and hence more likely to have concave costs.

4. Equilibrium

We now wish to analyze equilibria of the model. The key concept to unravel the equilibrium is the willingness of a party to bid, that is: How many voters is the party willing to turn out if it is guaranteed the prize?

4.1. The willingness to bid

If a party turns out a fraction φ_k of its (uncommitted) voters and wins the prize it gets $v_k - C(\varphi_k)$ per capita, while if it turns out only its committed voters and loses it gets 0. If $C(1) < v_k$ then the party is willing to turn out all its voters to get the prize and its willingness to bid is $\hat{b}_k = \eta_k$. If $C(1) \geq v_k$ then the party willingness to bid is determined by the indifference condition $C(\hat{b}_k/\eta_k) = v_k$. Notice that the willingness to bid is endogenous in the sense that it depends on the parameters of the model: the cost function and the size of the party. But it is not an equilibrium quantity in the sense that the willingness to bid of a party is independent of any choices made by the other party, or indeed their characteristics. Stated differently, the willingness to bid is an endogenous cap on bids.

Let us rule out the degenerate case where both parties are equally willing to bid. We will call a party *advantaged* if it has the higher willingness to bid and *disadvantaged* otherwise. The disadvantaged party will be denoted by d so by definition $\hat{b}_d < \hat{b}_{-d}$. As it will be clear shortly, it is not always the case that the large party is advantaged. Finally, we restrict attention to the interesting case in which the small party is willing to bid more than the committed voters of the large party, that is we assume that $\hat{b}_S > \eta_L \underline{y}$ or equivalently $v_S < C(\eta_L \underline{y}/\eta_S)$.[18]

4.2. Equilibrium payoffs

Our first result characterizes equilibria.

Theorem 1. *There is a unique equilibrium. In this equilibrium the utility of the disadvantaged party is* 0 *and the utility of the advantaged party is* $U_{-d} = \eta_{-d} v_{-d} - \eta_{-d} C(\hat{b}_d/\eta_{-d})$.

A simple way to understand this result is to see that it is exactly the same result as if a second price auction were held and each party submitted their willingness to bid. The disadvantaged party submits $\hat{b}_d$, a relatively low bid, so it loses the auction and does not pay. The advantaged party wins but must match the bid of the disadvantaged party and the cost of doing so is $\eta_{-d} C(\hat{b}_d/\eta_{-d})$.

While the formal proof is in the Supplemental Appendix, the reasoning can be understood relatively easily. If the disadvantaged party (no matter whether it is the small or the large party) does not get a utility of zero then its minimum bid must have a positive probability of winning. That means that the advantaged party must make bids that are sure to lose, which would imply that the advantaged party gets zero: this is impossible since the large party can

[18]When $\hat{b}_S < \eta_L \underline{y}$ there is a unique equilibrium in which each party turns out only committed voters. This is the only case in which there is an equilibrium in pure strategies. Notice that our restriction rules out the especially uninteresting case $\eta_L \underline{y} > \eta_S$ in which it is infeasible for the small party to match the committed voters of the large party.

guarantee a positive utility by bidding slightly more than $\hat{b}_d$. Since the disadvantaged party gets zero it must be willing to bid up to $\hat{b}_d$. Otherwise, if it never bids more than, say, $b_d < \hat{b}_d$ then the advantaged party will also bid no more than b_d.[19] But if the advantaged party bids only b_d then the disadvantaged party should bid just a bit more thereby getting positive utility. Turning to the advantaged party and given the argument above, it should be clear that the advantaged party must at least some of the time bid $\hat{b}_d$ and the most it can get is $\eta_{-d}v_{-d} - \eta_{-d}C(\hat{b}_d/\eta_{-d})$ as asserted in the theorem. But it cannot get less than that, because it can get arbitrarily close to that by bidding a small enough ϵ above $\hat{b}_d$. Since the advantaged party must be indifferent between all bids it makes, this must be the equilibrium utility of the advantaged party.[20]

Theorem 1 gives a simple and general result about the utility of the two groups. It shows that it is crucial to understand which party is advantaged. Notice that in general if the prize is worth a great deal to one party and very little to the other, the party that places a high value on the prize is advantaged. This is a basic sanity check and is unsurprising and not terribly interesting. As a result we now turn out attention to the important case in which both groups value the prize equally.[21]

5. Party Advantage with a Common Prize

Let us focus on the case in which both parties value the prize equally. This is the most commonly studied case, and it would be the situation, for example, if the outcome of the election resulted in a certain number of government jobs going to the winner, or a fixed government budget being used to subsidize the

[19] Notice why the tie-breaking rule should be endogenous: if the advantaged party were to bid b_d the tie-breaking rule should assign it a win. Otherwise it should win by bidding $b_d + \epsilon$ where ϵ is the smallest strictly positive number: such a number does not exist.

[20] This part of the argument is similar to that of Baye, Kovenock and De Vries (1996) to show uniqueness of equilibrium payoff in the all-pay auction with complete information. However, this is not enough to prove Theorem 1 since in our case bids are endogenously capped by the willingness to bid.

[21] The Supplemental Appendix gives a complete characterization of equilibrium payoffs and strategies with general cost functions and arbitrary prizes.

winner.[22] More generally it represents a situation where the winner can impose an (efficient) tax on the loser. Specifically, we assume that $\eta_S v_S = \eta_L v_L = V$. Notice that the common prize assumption implies that per capita the value of the prize is always greater for the small party. How does this translate into party advantage? And how does party advantage matter for equilibrium turnout and probability of winning the election?

5.1. Party Advantage

Which party is advantaged will crucially depend on the properties of the total cost function. To see why this is the case, notice that if the small party is willing to bid $\hat{b}_S$, then it incurs an *aggregate* total cost (that is total cost per size of the party) equal to $\eta_S C(\hat{b}_S/\eta_S) = \hat{b}_S \left[C(\hat{b}_S/\eta_S)/(\hat{b}_S/\eta_S)\right]$. On the other hand, the aggregate total cost to the large party of matching a bid of $\hat{b}_S$ is $\eta_L C(\hat{b}_S/\eta_L) = \hat{b}_S \left[C(\hat{b}_S/\eta_L)/(\hat{b}_S/\eta_L)\right]$. This cost will be smaller for the large party if $C(\hat{b}_S/\eta_S)/(\hat{b}_S/\eta_S) > C(\hat{b}_S/\eta_L)/(\hat{b}_S/\eta_L)$. Hence, when the average per capita cost is increasing the large party is advantaged. Conversely, when the average per capita cost is decreasing the small party is advantaged. Since average cost increasing and decreasing corresponds to convex and concave cost respectively, we see that which party is advantaged depends crucially on the convexity of the cost function.

The case where $C(\varphi_k)$ is convex on $[\underline{y}, 1]$ is clear-cut. In this case $C(\varphi_k)$ is globally convex on $[0, 1]$ and consequently the large party is advantaged. In particular, this will be the case if monitoring costs were zero. It is only the presence of monitoring costs that raises the possibility of the small party being advantaged.

If the total cost function $C(\varphi_k)$ is globally concave on $[0, 1]$, then the large party would have a higher cost of matching the small party's willingness to bid and it would be disadvantaged. Notice, however, that the total cost function cannot be globally concave. Because of the presence of committed voters, even if $C(\varphi_k)$ is concave on $[\underline{y}, 1]$, it is neither concave nor convex on $[0, 1]$. For the remainder of the paper when we say that "C is concave" we are referring

[22]In Feddersen and Sandroni (2006) both parties value the prize equally.

only to the part of C that lies above $\underline{y}$. To identify sufficient conditions for the small party to be advantaged, define $\xi \equiv 1/\left(\max_{\varphi\in[\underline{y},\eta_S/\eta_L]}|C''(\varphi)|/C'(\underline{y})\right)$ as a measure of the curvature of $C(\varphi_k)$, where a lower ξ represents a higher curvature of C. Then we have the following theorem.

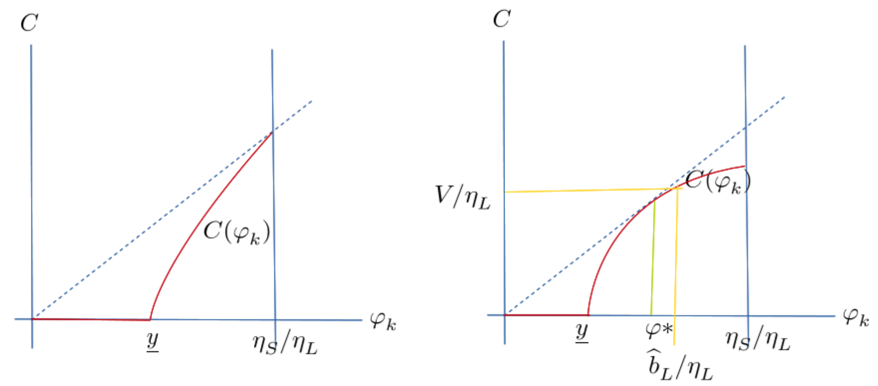
Theorem 2. *If C is convex then the large party is advantaged. If C is concave and*

a. $\sqrt{(2\xi+\underline{y})\underline{y}} < \eta_S/\eta_L$ *(enough concavity relative to the number of committed voters)*

b. $\eta_L C(\sqrt{(2\xi+\underline{y})\underline{y}}) < V < \eta_L C(\eta_S/\eta_L)$ *(intermediate value of prize)*

then the small party is advantaged. If C is quadratic then condition a. is also necessary for the small party to be advantaged.

Theorem 2 points out that for the small party to be advantaged the number of committed voters must be small enough, $C(\varphi_k)$ must be "concave enough" and V must be neither too large nor too small. Furthermore, the fewer the committed voters are, the less concave C must be for the small party to be advantaged. To understand the role of such conditions, consider the two figures below. In Figure 1, $\underline{y}$ is relatively large and while $C(\varphi_k)$ is concave, it has relatively little curvature. Average cost may be read from the graph as the slope of the line connecting the origin to the total cost curve. As shown, the average cost is always smaller than the marginal cost and hence increasing throughout the range $[0, \eta_S/\eta_L]$. For any bid of the large party, the cost of the small party to matching the bid is greater than the cost of the large party, so - despite the concavity of $C(\varphi_k)$on $[\underline{y}, 1]$ - the large party is the advantaged one.



By contrast, in Figure 2 $\underline{y}$ is relatively small and $C(\varphi_k)$ has substantial curvature. Here there is a unique point φ^* at which the average cost is equal to marginal cost. Below φ^* the average cost is increasing, while above it is decreasing. The willingness to bid of the large party, that is where $V/\eta_L = C(\hat{b}_L/\eta_L)$, it is shown in Figure 2 as well. When V is small, $\hat{b}_L/\eta_L < \varphi^*$ and conversely, as shown in the figure, when V is large. In the latter case, when $\hat{b}_L/\eta_L \in (\varphi^*, \eta_S/\eta_L)$ and the small party matches the willingness to bid of the large party, the small party must bid above $\hat{b}_L/\eta_L$ in a region where average cost is decreasing. Since in this case the small party must have lower average cost than the large party, the small party is advantaged. To understand the second condition of Theorem 2, note that if V is sufficiently small then $\hat{b}_S/\eta_S$ must necessarily lie to the left of φ^* so that the small party will be disadvantaged, while if V is quite large, $\hat{b}_L/\eta_L$ will lie to the right of η_S/η_L and the small party will be unable to match the willingness to bid of the large party and so will be disadvantaged. Hence, for the small party to be advantaged, the value of the prize has to be intermediate.

5.2. Turnout and Probability of Winning the Election

We are now ready to study how party advantage translates into equilibrium turnout and probability of winning the election. To do so, we show that the properties of the cost function, which determine party advantage, affect the equilibrium strategy of the parties.

Recall that given two cumulative distribution functions F and G, the distribution F *first-order stochastically dominates* the distribution G (we write $F \geq_{FSD} G$), if $F(x) \leq G(x)$ for all x, with strict inequality at some x. Hence, if we can provide conditions under which party k equilibrium bidding strategy FSD that of party $-k$, then we can conclude that party k will turnout more members in expectation and will also have a higher probability of winning the election. The next theorem provides such conditions.

Theorem 3. *If the cost function is either convex or it is concave and the small party is advantaged, then the advantaged party equilibrium bidding function FSD that of the disadvantaged party.*

Notice that there are cases in which the bidding function of a large advantaged party does not *FSD* that of the small party.[23] This happens precisely in the case in which the cost function is concave but the large party is nevertheless advantaged. In this case the *FSD* result can fail in a strong sense. Specifically when costs are quadratic and $\underline{y}$ is small enough we show in the Supplementary Appendix that there is an intermediate range of $V's$ and for each such V an open set of θ's for which the large party is advantaged, yet the small party turns out more members in expectation and has a higher probability of winning than the large party.[24]

A unique feature of our theory is that, when the enforcement of social norms is costless, it delivers predictions consistent with the ethical voter and follow-the-leader type of theories. In particular, in equilibrium the large party is advantaged, turns out a higher expected number of voters and has a better chance of winning the election. However, things change drastically when the enforcement of social norms is costly. In this case, contrary to existing theories of political participation and much in the spirit of Olson (1965), our model predicts that when there are few committed voters and the prize is of intermediate value, the small group turns out a higher expected number of voters, it has a

[23]In the Supplementary Appendix we prove a slightly stonger version of Theorem 3 that holds even if the prize is not common provided the value to the advantaged party is at least that of the disadvantaged party. With a common prize, concave cost, large party advantaged and small party unconstrained neither bidding function FSD that of the other party.

[24]See Proposition 1 in the Supplementary Appendix.

better chance of winning the election and, as Theorem 1 shows, it has a higher equilibrium payoff.

Clearly, for a small group to be advantaged, it must be the case that turnout is relatively low. Examples of a smaller group prevailing over a larger one, are not uncommon, but, since our theory predicts a positive probability of the disadvantaged party winning, we cannot drawn conclusions about advantage by examining the results of a single election.[25] One exception, where there is data on many similar elections, is the case of teacher unions capturing school boards. We briefly elaborate on this example in the next subsection. Further details can be found in the School Board Appendix.

School Board Elections

Consider the parties to the election as being a small party of teachers and a large party of students' parents. The school board controls resources and money that can be allocated either to teachers or to students. If the teachers win the election the money goes to them, if the parents win the election the money goes to the students. Hence we take it as a reasonable approximation that there is a common prize. School board elections are often held at a different time than other elections so that the school board is the only issue on the ballot. In these elections, teacher unions are extremely effective at getting their candidates on the board (86%). While turnout among teachers is extremely high (90%), overall turnout is very low (10-20%), which is probably due to many voters, who are not parents of students, being uninterested in the election. Since school boards generally control budgetary resources and can allocate those between teachers and pupils but they cannot set taxes or determine the overall size of the budget, we can take that the prize is not too large.

Consider ethical voter or follow-the-leader type of theories in which costs of

[25] As Kobach (1993) puts it "trooping to the polls in great numbers while most of the electorate stays home, an active minority can defeat a position held by the majority of citizes. The result is what I describe as a false majority (in which the final verdict on the proposal, had all citizens voted, would have been different)" page 350. For some anecdotal evidence on false majorities in Swiss referenda see Kobach (1993). Also, Rawnsley (2005) suggests that the outcome of the 1973 sovereignty referendum in Northern Ireland could be a case of false majority.

mobilization are assumed to be convex. The basic prediction of these theories is that the large party - parents - should prevail. That is, it is difficult to explain why teacher unions are so effective. Can we explain this by monitoring costs? First, notice that there are few committed voters, since civic duty does not extend for most people to elections that are viewed as unimportant - and the very low turnout seems consistent with this idea. Second, with respect to monitoring costs, notice that since the turnout among teachers is very high, for the teachers monitoring costs do not much matter. So the key issue is whether it is plausible that the monitoring costs of the parents are sufficiently large as to make overall cost sufficiently concave as to disadvantage them. In our theory, we have introduced only a simple model of the social network underlying monitoring costs - but it is possible to make some intuitive statements about what affects monitoring costs. How easy is it for network members to observe each other? In practice the people who are likely to be able to observe you are your friends and neighbors. In large general elections with relatively high stakes, those people are likely to be involved in the election. In small special purpose elections involving only a small fraction of the electorate, the other people involved - the other parents - are less likely to be "close" to you, as only a fraction of your friends, co-workers and neighbors are fellow parents. This suggests that indeed it makes sense to think that monitoring costs are relatively high in small special purpose elections like those for school board.[26]

5.3. Equilibrium Strategies

We now want to examine the party strategies in the unique equilibrium and discuss some non-standard features of our all-pay auction.

We observe first that there cannot be an equilibrium in pure strategies. By assuming that the small party is always willing to bid, that is $\hat{b}_S > \eta_L \underline{y}$, it

[26] All other things equal, if we look at electoral success of teachers unions against the resources controlled by the board (the size of the prize) we might expect the curve to be u-shaped: a very weak school board with little power not be worth controlling, while a very powerful school board with a lot of resources would advantage the larger party. This implies that the teacher union has an incentive to keep school boards from getting too strong - while it would be better to control a more powerful board, it is less likely to be able to do so.

cannot be an equilibrium for both parties to bid only their committed voters. Hence, if one party outbids the other in a pure strategy equilibrium, it must bid more than the committed voters of the large party. This is impossible since then the higher bidding party should cut its bid slightly, and still win for sure. The only case that is left to check is whether there could be a pure strategy equilibrium with a tie. However, if the advantaged party does not win the tie with probability one, it should bid slightly more for a sure win at small extra cost. On the other hand, if it does win the tie with probability one, then the disadvantaged party should only bid its committed voters. Finally, if there is a tie in which the large disadvantaged party bids only its committed voters and loses for sure then, by rising its bid slightly, the large party would win for sure. As a result, the unique equilibrium must be in mixed strategies.

The equilibrium strategies have both a continous part and a discrete part. The support of the continuous part for both parties runs from bidding the committed voters in the large party $\eta_L \underline{y}$ up to the willingness to pay of the disadvantaged party $\hat{b}_d$, and there are no atoms in the interior of this interval. The equilibrium continuous density function is found with standard techniques by observing that each point in the interior must be indifferent and yield $U_d = 0$ for the disadvantaged party and $U_{-d} = V - \eta_{-d} C(\hat{b}_d/\eta_{-d})$ for the advantaged party, respectively. Since the probability of winning with a bid b_k is the probability that the other party has a lower bid $F_{-k}(b_k)$ this indifference condition is $F_{-k}(b_k)V - \eta_k C(b_k/\eta_k) = U_k$ which we may differentiate to find the density:

$$f_{-k}(b_k) = C'(b_k/\eta_k)/V.$$

Note that in the special case of quadratic cost and $\theta = 1/2$, this density will be the same for both parties. This is the standard result for all-pay auctions since it is generally assumed that the cost of a bid is the bid itself.

In addition to the continuous part, the equilibrium strategies may have atoms, that is, bids with strictly positive probability. While proofs and the exact probabilities are in the Supplementary Appendix, here we want to mention that there are three bids that can have positive probability.

1. There must be a positive probability that the disadvantaged party *con-*

cedes the election: it bids only its committed voters. This result is standard for the all-pay auction.

2. A large advantaged party must bid only its committed voters with positive probability. This is because the small party when disadvantaged must be willing to bid very close to $\eta_L \underline{y}$ and since this requires turning out more than just its committed voters, it must be compensated for doing so by a positive probability of winning. This result is not standard for all-pay auctions because of the presence of committed voters.[27]
3. When the large party is advantaged and the small party is constrained ($\hat{b}_S = \eta_S$), the large party must *take the election*: it bids η_S with positive probability and wins the election for sure. This is because the small party must get zero for bidding near η_S and since it is constrained, this is possible only if it has a positive probability of losing. This result is not standard for all-pay auctions because of the willingness to pay, which representes an endogenous upper bound on bids.[28]

6. Comparative Statics

We will now investigate the comparative statics properties of our model with respect to three important variables: the value of election, the relative size of parties and the efficiency of the monitoring technology. We then turn to vote suppression and welfare. In investigating comparative statics we are particularly interested in turnout and what it implies for the competitiveness of elections. In models with aggregate shocks that support pure strategy equilibria, turnout is deterministic so it is clear what it is, and greater competitiveness can be measured by a smaller difference between parties' turnout.[29] In our model turnout is stochastic, so the meaning of greater turnout and higher competitiveness should be qualified. For turnout, in addition to First-Order Stochastic Dominance (FSD) in the equilibrium bid distributions and expected turnout, we

[27]In the all-pay auction literature, a similar result holds if one of the bidders enjoys an "head start". See, for example, Siegel (2014) and references therein.

[28]See Szech (2015) for the role tie-breaks and bid-caps in all-pay auctions.

[29]See Feddersen and Sandroni (2006) and Herrera, Morelli and Nunnari (2015).

consider expected turnout cost and peak turnout. *Peak turnout* is the highest equilibrium turnout of a party: we know that this is the same for both parties and equal to the disadvantaged party's willingness to bid $\hat{b}_d$. For competiveness in addition to the expected vote differential we examine the *bid differential*, that is $\hat{b}_{-d} - \hat{b}_d$.

6.1. High and Small Value Elections

We assumed that $V > \eta_S C(\eta_L \underline{y}/\eta_S)$, which implies that the small party is willing to outnumber the committed voters of the large party. If in addition $V > \eta_L C(\eta_S/\eta_L)$ the large party is willing to outnumber the small party. If these two conditions are satisfied we say the election is a *high value election.* This is a natural model of elections where the stakes are high such as elections for national leader or important referenda such as Brexit. Notice, however, that while in a high value election both parties are willing to mobilize the number of voters in the small party, in equilibrium neither party does so. Indeed we know that since $V > \eta_S C(\eta_L \underline{y}/\eta_S)$ the equilibrium is in mixed strategies - and in particular the small party has a chance of winning. Notice the implication that regardless of the polls, events such as "Brexit wins" or "Trump wins" always have a positive probability of occurring in equilibrium.

To further analyze turnout in high value elections, let us use the terminology introduced in the previous section, and say that the small party *concedes the election* if it mobilize only its committed voters. Furthermore, we say that the large party *takes the election* if it bids η_S, that is the most voters feasible for the small party. Then while both parties are willing to mobilize all their voters the next result shows that what they do mobilize in equilibrium is far less.

Theorem 4. *In a high value election the probabilities that the small party concedes and the large party takes the election increase in V, and approach* 1 *in the limit. The bid distribution of the small party declines and the bid distribution of the large party increases in V in the sense of FSD. The expected vote differential increases in V while expected turnout cost remains constant.*

We refer to the fact that mobilization of the small party is decreasing and its probability of concession increasing in V as the *discouragement effect.* Since the large party is willing and able to outnumber the small party, the small party becomes discouraged and, as the stakes increase, turns out fewer and fewer voters. Nevertheless, because the stakes are very high and despite the fact that the small party is turning out very few voters, with very high probability the large party turns out enough voters to guarantee victory against the small party. Notice that the expected turnout cost of the small party declines and the expected turnout cost of the large party increases, but the two effects exactly offset each other. A simple version of this result is standard for all-pay auctions. In particular, in a standard complete information all-pay auction, as the ratio between players valuations increases, the player with low valuation bids zero with a probability increasing in this ratio.

It is interesting to contrast a low value election in which $V > \eta_S C(\eta_L \underline{y}/\eta_S)$ with a high value election.[30] Turnout of both parties in a high value election is substantially higher than in a low value election and the turnout of both parties stochastically dominates their turnout in a low value election. This is consistent with suggestive evidence of higher participation in national than in local elections, and with empirical evidence showing that electoral participation will be higher in elections where stakes are high.[31]

In the high value election we can see clearly that there is a discontinuity in the surplus when the parties are of near equal size. As $\eta_L \to 1/2$ the surplus of the large party approaches $V - (1/2)C(1)$, that is, in the limit it does not approach zero. Hence a small change in party size shifting a small party into a large party causes the surplus of that party to jump from zero to a strictly positive value and conversely. Moreover neither the probability of concession

[30]In a low value election the small party is not willing to outnumber the committed voters of the large party, and hence the large party is trivially advantaged. In this case, both parties turn out only committed voters, and the large party wins for sure. In a high value election the large party is also advantaged and wins with very high probability, while the small party with very high probability turns out only its committed voters. Notice, however, that in the low value election the small party loses for sure, while in the high value election the small party has a positive chance of winning. Furthermore, in the low value election the large party turns out only committed voters, while in a high value election the large party with very high probability turns out enough voters to guarantee a win.

[31]See, Andersen, Fiva and Natvik (2014).

nor the probability of the large party taking the election approache zero and for large V both are close to one. In other words, a small change in the party size causes a party that was conceding with positive probability to stop conceding and instead take the election with positive probability. The discontinuity is important if we step back from the model and consider a broader setting in which parties choose platforms in an effort to compete for members prior to the election: we see that a small shift in the relative sizes of the parties can have disproportionate consequences, suggesting that the competition over platforms may be a fierce one.[32]

6.2. Consequences of Increased Competitiveness

In a standard all-pay auction competitivess - the bid differential - is measured by the difference by how much the two parties value the prize. As this approaches zero the surplus of the advantaged party also vanishes. In our setting when cost is concave and the small party is advantaged this need not be the case. Concavity and small party advantaged imply that $\eta_L C(\eta_S/\eta_L) > \eta_S C(1)$, which means that the cost to the large party of matching the maximum turnout of the small party is bigger than the cost to the small party of turning out all voters. If the value of winning the elections is in between $\eta_L C(\eta_S/\eta_L) > V > \eta_S C(1)$, the small party is advantaged but constrained: it would like to turnout more voters but cannot do so. As V increases the willingess to bid of the small party remains at η_S, while the willingness to bid of the large party increases, reducing the bid differential - increasing competitiveness. Nevertheless the surplus accruing to the small party does not approach zero. When instead $V > \eta_L C(\eta_S/\eta_L)$, the advantage switches to the large party and the surplus of the small party drops abruptly to zero. This is interesting from the point of view of agenda setting, for example a referendum proposed by the small party - they want on the one

[32]A final observation is that these sharp results about high value elections are sensitive to the assumption that the party bidding the highest turnout wins the elections for sure. As we discuss in Section 7 in Theorem 10 when there is noise in the outcome and parties have nearly equal size, high enough value elections behave closer to what we would expect in the case of an exact tie. That is, with a small amount of noise both parties have nearly an equal chance of winning, and each gets a surplus equal to about half the difference between the value of the prize and the cost of turning out all the voters.

hand to ask for a large prize, but if they make it just a bit too big they can lose everything.[33]

6.3. Disagreement

Next, following Feddersen and Sandroni (2006), we can consider the effect of the level of "disagreement" within the electorate on equilibrium outcomes, by varying the relative size of parties. Intuitively we expect that greater disagreement should mean more fiercely contested elections with higher turnout. Specifically since $\eta_S = 1 - \eta_L$, as η_S grows we approach a situation where party supporters are evenly divided, that is, the level of disagreement in society increases. How does this affect the equilibrium outcome? In the ethical voters model of Feddersen and Sandroni (2006) or in the group-turnout model with aggregate shocks studied in Herrera, Morelli and Nunnari (2015) greater disagreement does indeed lead to higher turnout and more competitiveness. With our proposed measures of turnout and competitiveness we can show:

Theorem 5. *If C is convex or C is concave and the small party is advantaged, disagreement increases the peak turnout, the expected turnout cost and decreases the bid differential.*

6.4. Monitoring Inefficiency

Finally, we turn to monitoring inefficiency. Recall this is denoted by θ: if $\theta = 1$ then the auditor learns nothing about the auditee's type; if $\theta = 0$ the auditor perfectly observes whether the auditee has violated the norm or not. There are a variety of possible effects of monitoring inefficiency on turnout and competitiveness, some of which are explored in the Supplementary Appendix. Here we focus on the most interesting case of a high value election. Recall that $C(\varphi_k) = T(\varphi_k) + \theta(1 - \varphi_k)T'(\varphi_k)$ so that increasing monitoring inefficiency acts as a multiplier on the monitoring cost. To gain a bit of intuition, suppose that we increase the multipler of the turnout cost T. This would act as a

[33]One example of this may have been the heavy defeat of Proposition 16 in California in 2010: This was a ballot initiative sponsored by Pacific Gas and Electric Company that was designed to reduce competition from local governments.

multiplier on total cost C, and it would be equivalent to reducing the size of the prize V. In this case, we know from Theorem 4 that it increases the small party turnout, it reduces the large party turnout, and consequently makes the election more competitive. Notice, however, the counter-intuitive fact that increased cost actually raises small party turnout - this is why competitiveness increases. Indeed, we might have expected that a less valuable prize is less worth fighting over, so that the election would be less competitive. If η_S is large enough this is the case. In fact, unlike T and C, as we have observed before, $(1-\varphi_k)T'(\varphi_k)$ must eventually decrease since it is zero at $\varphi_k = 1$. When η_S is large enough, the support of bids overlaps far enough into this decreasing range and hence we get the following result.

Theorem 6. *In a high value election, an increase in monitoring inefficiency decreases the turnout of the advantaged party in terms of FSD. Furthermore, there exists* $0 < \underline{\eta} < \overline{\eta} \leq 1/2$ *such that for* $\underline{\eta} < \eta_S < \overline{\eta}$ *the expected turnout of the disadvantaged party decreases in monitoring inefficiency while the expected vote differential also decreases.*

As in the small η_S case, competitiveness increases. However, when η_S is large, turnout decreases and competitiveness increases because turnout of the small party decreases less than the decrease in turnout of the large party.

There is some direct data on the effect of monitoring inefficiency on turnout cost: Larreguy, Marshal and Querubin (2016) found that increased monitoring inefficiency of local mobilizers decreases turnout buying for two parties of similar size as Theorem 6 suggests. Another application concerns the idea that in Western Europe, over the period since World War II, the social ties underlying the party system have broken down. One possible interpretation of this is that monitoring has become more inefficient. For example, in the old days labor union members in the UK socialized in pubs and old money socialized in clubs, with the resulting strong social ties keeping monitoring costs low for the Labor and Conservative party, respectively. This is consistent with the concept of "mass parties" in the political science literature - see, for example, the discussion of the literature in Katz and Mair (1995). For a considerable period after World War II, Western Europe was dominated by large mildly left-wing parties of various flavors of labor or Christian Democrats. Theorem 6 supports the idea that there is a connection between the breakdown in social ties - meaning less efficient monitoring - and the decline in these parties as measured by declining turnout, more competitive elections.[34]

6.5. Vote Suppression

The 1993 Ed Rollins scandal suggests that sometimes party effort can be directed to supress the votes of the opposition.[35] More recently, it has been argued that voter identification laws increase voting costs for relatively poor Democratic voters (in particular Hispanics, Blacks, and mixed-race Americans) with relatively little effect on whites and on the political right.[36] Our model can be used to investigate the strategic use of vote suppression. Suppose that parties are unconstrained (that is, $C(1)\eta_S > V$ and parties are willing to turn out all their voters) and that each party can slightly increase the participation cost of the opposing party from $c(y)$ to $\tilde{c}(y) = c(y) + h(y)$ where $h(y) > 0$, by incurring a cost of S.

Theorem 7. *[Cesar Martinelli] If* $\sup_y h(y)$ *and* S *are sufficiently small then only the advantaged party will suppress votes.* [37]

Our theory suggests that when the Republican party is advantaged, it will be tempted to introduce voters suppression laws in the form of strict voter identification requirements. It will do so in order to hold down its own turnout cost. The rationale discussed in the popular press revolves instead about taking strategic advantage: in a closely contested state a short term Republican victory may be translated into long term advantage by introducing a voter suppression

[34] Gray and Caul (2000) relates post-war turnout decrease with the decline of mobilizing actors such as labor parties and trade unions, and Knack (1992) connects the decline of American voter turnout with a weakened enforcement of social norms.

[35] Ed Rollins was the campaign consultant of Christine Todd Whitman during the 1993 New Jersey gubernatorial election. Rolins claimed to TIME magazine that he secretly paid african american ministers and Democratic campaign workers in order to suppress voter turnout.

[36] See Hajnal, Lajevardi and Nielson (2016) and Ingraham (2016).

[37] This theorem was suggested to us by Cesar Martinelli during the 2015 Priorat Workshop.

law.[38] To explore this we take a strong measure of what has been argued to be voters suppression - strict photo id laws. The table below reports the states with these laws, the year the law was introduced and the 2012 vote differential between Romney and Obama as a measure of which party is advantaged.[39]

State	Year	Republican Advantage
Alabama	2014	+23
Kansas	2011	+22
Tennessee	2011	+20
Texas	1990	+16
Mississippi	2011	+12
Indiana	2005	+10
Georgia	1977	+8
Virginia	1996	+3
North Carolina	2013	+3
Wisconsin	2011	-7

Interestingly, the GOP holds an overwhelming electoral advantage in most of the states with strict photo id laws: the median Republican advantage in these states is more than 10 points. This is consistent with our theory and not with the strategic advantage theory. Only in the case of Virginia, North Carolina and Wisconsin - all of which have had both Republican and Democratic governors in recent years, and two of which only recently have introduced voter suppression laws - does the strategic theory seem to have merit.

6.6. Welfare

In our model, effort spent voting and monitoring voters is pure waste. From this pespective, vote suppression, for example, would always raise welfare. Furthermore, if the prize is of equal value to the two parties and there is no uncertainty about party sizes then a mechanism of flipping a coin (and allowing

[38] For example, from Levy (2016) writing in *Mother Jones* "In order to mitigate their waning political popularity, Republicans have [...] passed an unconstitutional voter suppression law to weaken the voting power of African Americans and other Democratic-leaning voters."

[39] Data are taken from https://en.wikipedia.org/wiki/Voter_ID_laws_in_the_United_States.

the committed voters to vote, but ignoring the results of that vote) is a strict welfare improvement over voting. Clearly, these results are heavily influenced by the fact that increased participation increases costs and thereby tends to reduce welfare. However, there are a number of reasons to believe that higher participation might be welfare improving. In this section we explore some of these arguments.[40]

A possible reason for preferring higher participation is that, in a model of incomplete information, higher participation would lead to greater information aggregation. However, once the absolute number of voters is reasonably large the additional improvement in information aggregation from doubling or tripling the number of voters seems quite modest. A more common reason given why high participation is important is because *it increases the sense of participation in government*. What concretely this might mean beyond the relatively meaningless statement that each vote creates a utility benefit for someone is hard to say. One possibility is that it signals the willingness or likelihood that citizens will abide by the democratic rules or that they will fight off attempts at coup d'etats and the like. In our view, one of the most reasonable arguments for preferring higher participation is that when turnout is too low there is a concrete possibility that a minority non-democratic party may come to power. This latter possibility is relatively straightforward to model and not inconsistent with the idea that there might be broader benefits such as signaling.

Suppose that depending on turnout b_k there is a probability that a third party who will refer to as *blackshirts* might win the election. Denote this probability by $\beta(b_S, b_L)$ and suppose that each party suffers a per capita loss of ζ in the event in which the blackshirts win the election. Extending the model to the the possibility of blackshirts winning the election leads to a new cost function $\tilde{C}(b_k, b_{-k}) = C(b_k/\eta_k) + \zeta\beta(b_S, b_L)$. This differs from the original model for two reasons: First, the participation rate of the opposing party $-k$ now enters the cost function of party k. Second, the sizes of the parties have an impact on per capita costs. To see how these differences will affect the welfare implications of

[40]We thank Guido Tabellini for pushing us in this direction and helping us to explore some of these ideas.

the model, consider a simple example in which the number of blackshirts who will cast votes is uncertain and uniformly distributed with density $1/\lambda$, and institutions are such that in order to achieve power, blackshirts must collect an absolute majority. Then

$$\beta(b_S, b_L) = \beta_0 - \lambda(b_S + b_L - \underline{y}),$$

where $\beta_0 \leq 1$ denotes the probability that blackshirts are enough to outnumber the commited voters. Furthermore, we assume that $\beta(\eta_S, \eta_L) \geq 0$ so that there is a positive chance that blackshirts are enough to outnumber all the voters in both parties.[41] This additively separable functional form has the advantage that the marginal benefit of reducing the chance of blackshirts winning is independent of the participation rate of the other parties, so that we may apply our existing analysis of equilibrium. In particular, we may analyze equilibrium using the overall per-capita cost function for party k

$$\tilde{C}(b_k, b_{-k}) = C(b_k/\eta_k) + \zeta\left[\beta_0 - \lambda(b_S + b_L - \underline{y})\right].$$

If $\lambda\eta_k\zeta < C'(b_k/\eta_k)$, then $\tilde{C}$ is increasing in b_k, despite the fact that β is decreasing. In this case our all-pay auction applies unchanged. The welfare analysis can change substantially, however, if $\tilde{C}$ is decreasing. Indeed, if $\lambda\eta_k\zeta > C'(b_k/\eta_k)$, the overall per-capita cost of voting $\tilde{C}(b_k, b_{-k})$ is *decreasing* in b_k. In this case the earlier welfare conclusions will effectively be reversed.

Three further observations may be of interest. First, in addition to parties we may wish to consider an additional social network of "citizens" to which everyone belongs and which also can create incentives for voting. Such a network may have as its objective keeping the blackshirts out of power. Because this social network strictly prefers greater voter participation, we may wish to interpret the committed voters as those for whom the citizen incentives are binding, which is

[41]It should be noted that turnout in 1933 German federal elections was 89% and the Nazist party fell barely short of an absolute majority with 44% of the votes. Clearly the election was not free from intimidation, but of course this is part of the reason for turning out a great many voters to defeat the blackshirts.

to say that they vote because it is their "civic duty." This would lead to a theory in which the number of committed voters is endogenous and would depend on the strength of the blackshirts and whether the other existing parties are willing to turn out voters. Second, the idea that the strength of blackshirts might vary is an important one. Since the Second World War voter participation has fallen. There are many explanations of this, including the weakening of parties ability to mobilize voters.[42] But note also that during the same period the possibility of a blackshirts surge diminished substantially as democratic institutions became increasingly stronger (ranging from peacefully accepting the election outcome, to the strength of the courts and independence of the press). In the early years of European democracy, institutions were weak, the chance of blackshirts high, and a high voter turnout was perfectly reasonable. As time has gone on and institutions got stronger, better established, and confidence in them has been growing, there has been less reason for high turnout to keep blackshirts out of power.[43] Finally, the role of charimatic leaders - high leader valence that effectively increases the size of the pie - should not be overlooked. This increases turnout: whether this is good or bad from a welfare point of view depends on whether blackshirts are a concern.

7. Aggregate Uncertainty and Pivotality

The all pay auction is the limiting case of a contest that is decided by a conflict resolution function in which the probability of winning the election is a continuous function of the expected number of voters each party turns out. However, the outcome of the election is decided by the actual number of votes rather than the expected number of votes, and in this case the conflict resolution function is derived from the binomial distribution.[44] Furthermore, there are other reasons such that the outcome of an election may be random

[42] See Gray and Caul (2000) and Knack (1992) among others.

[43] It should be noted that while comparisons are sometimes made between modern populist movements such as UKIP in the UK, 5 Star in Italy and Trump in the US, with the exception of the Golden Dawn in Greece and unlike the Nazi and Fascist movements, these parties do not have anti-democratic paramilitary forces.

[44] See, for example, Palfrey and Rosenthal (1985) and Levine and Palfrey (2007).

such as correlation in the individual voting costs, random errors in the counting of votes, the way in which ballots are validated or invalidated, or intervention by courts. In this section we want to incorporate such randomness in our model, as well as account for the consequent probability of being pivotal.[45]

We start by giving a formulation of a contest model that enables us to compute the probability of being pivotal. We define two partial conflict resolution functions: $P_k^0(b_k, b_{-k})$, the probability of winning conditional on all voters except one following the social norm b_k/η_k and *the remaining voter not voting*, and $P_k^1(b_k, b_{-k})$, the probability of winning conditional on all voters except one following the social norm b_k/η_k and *the remaining voter voting*. These should be differentiable and non-decreasing in b_k and satisfy $P_S^i + P_L^i = 1$, where $i \in \{0, 1\}$. These two functions enable us to compute an overall conflict resolution function and the probability of being pivotal. The overall conflict resolution function is $p_k(b_k, b_{-k}) = (b_k/\eta_k)P_k^1(b_k, b_{-k}) + (1 - (b_k/\eta_k))P_k^0(b_k, b_{-k})$ and the probability of being *pivotal* is $Q_k(b_k, b_{-k}) = P_k^1(b_k, b_{-k}) - P_k^0(b_k, b_{-k})$.

To analyze incentives with pivotality we start by identifying what an individual voter would like to do in the absence of punishment. This depends on what voters from both parties are doing. Notice that at the time a party chooses its punishment level and social norm it does not know the social norm of the other party. Hence, party members consider the expected probability of being pivotal. For any given bid b_k and mixed strategy of the other party F_{-k}, we may define the unique pivotal cutoff $\gamma_k(\varphi_k, F_{-k})$, which represents the type of voter who is indifferent between bearing the cost of voting in order to improve the party's chance of victory and abstaining. The incentive constraint when there is punishment for not voting establishes that the net cost of voting, which is the direct cost c minus the benefit due to pivotality, must be less than or equal to the punishment for not voting. Finally, from the incentive constraint we can derive the monitoring cost, as the cost of punishing a non-voter who was not supposed to vote after having received a wrong signal.

There remains the issue of what happens if the social norm calls for less

[45] We omit most of technical details that the interested reader can find in the Supplementary Appendix.

participation than would be individually optimal in the presence of the pivotality incentive. For example, for voters with $\varphi_k < y < \gamma_k(\varphi_k, F_{-k})$ the social norm calls on y to not to vote, but in fact y would like to. It seems natural in this case to assume that there is no cost of getting a voter not to vote so that for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ we have that monitoring costs are equal to zero. As before, the goal of each party is to maximize per capita utility. The next theorem takes care of existence of an equilibrium of this extended model.

Theorem 8. *An equilibrium of the conflict resolution model exists.*

We are now ready to compare the benchmark deterministic model with the extended model with aggregate uncertainty.

7.1. Conflict Resolution versus All-Pay Auction

There are a number of cases of interest. First, as Feddersen and Sandroni (2006), Coate and Conlin (2004) and Herrera, Morelli and Nunnari (2015) show, in the absence of pivotality, if there is sufficient aggregate noise, and if the cost function is convex then pure strategy equilibria exist. In this case we are relatively far from the all-pay auction model. We discuss this case, and the existence of pure strategy equilibria in the model with aggregate uncertainty, in the Supplementary Appendix. Second, it may be that the conflict resolution functions are very close to the all-pay auction - that is the higher bid has a very high probability of winning. This is the case, for example, if the partial conflict resolution functions are derived as in Palfrey and Rosenthal (1985) from independent binomial voting choices in a large population. To make this precise, consider an infinite sequence of conflict resolution models $P_k^{iN}(b_k, b_{-k})$. We say that this sequence *converges to the all-pay auction* if for all $\epsilon > 0$ and $b_k > b_{-k}+\epsilon$ we have $P_k^{0N}(b_k, b_{-k}) \to 1$ uniformly, and and $Q_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) \to 0$ uniformly.

Theorem 9. *Suppose that $P_k^{iN}(b_{k,-k})$ converges to the all-pay auction, that F_k^N are equilibria of the conflict resolution models and that F_k is the unique equilibrium of the all-pay auction. Then $F_k^N \to F_k$.*

In particular if there is little aggregate uncertainty then the parties must use mixed strategies very close to those in the all-pay auction case. If the only source of noise is independent cost draws by party members and the population is large, then our all-pay auction equilibrium is a good approximation to the "true" equilibrium - and we can draw obvious conclusions, such as while the disadvantaged party may get some surplus, this is small relative to that of the advantaged party. Notice also the implication that since pure strategy equilibria cannot converge to a mixed equilibrium, if aggregate noise is too small then the equilibrium must be mixed. Stated differently, regardless of the conflict resolution function, the outcome of the election must be uncertain (unless the small party concedes). If the conflict resolution function is noisy there are pure strategy equilibria but the conflict resolution function itself introduces substantial uncertainty into the outcome. If it is not noisy, the players must randomize and this introduces instead endogenous noise in the outcome. *The outcome of elections are intrinsically unpredictable.*

Theorem 9 emphasizes the continuity between the conflict resolution and model and the all-pay auction model. There is also an important discontinuity. Recall that in the all-pay auction model regardless of the size of the prize, the large party never turns out more voters than the small party is able to. This is not the case in the conflict resolution model, quite the opposite:

Theorem 10. *Suppose $V \to \infty$. Then for all $\epsilon \in (0,1)$, $F_k(1-\epsilon) \to 0$.*

With a very large prize even a small degree of uncertainty about the outcome pushes both parties to make very high bids. If the parties are of similar size this implies that they have nearly equal chance of winning. This fact, in conjunction with Theorem 9, enables us to draw an important conclusion about voter turnout and population size. Suppose the only source of noise is due to the independent draws of costs. We first fix N and then make the size of the prize large enough that both parties will turn out most of their voters. If we now fix the size of the prize and increase the absolute number of voters keeping their proportions fixed, Theorem 9 implies that equilibrium must approach the equilibrium of the all-pay auction. Hence, from Theorem 4, the turnout of the large party must decline to η_S/η_L and that of the small party to $\underline{y}$, that is participation declines with the absolute size of the electorate. Notice that this is also true in pivotal voter models, because the probability of pivotality declines reducing the individual attractiveness of participation. However, in our model, the declining participation rate is for a distinctively different reason. A high participation rate provides "insurance" against losing the election by "accident." As the size of the electorate increases - and noise in the outcome declines - there is less reason to insure against such accidental loss.

7.2. Pivotality and Monitoring Costs

We would expect that pivotality would play a decisive role if monitoring costs are large. In our model so far monitoring costs are only the cost to the voter of being punished and incentive compatibility implies this is bounded by the cost of voting. However, it is natural to think that punishment may entail other costs to the group. Punishing a member of a party may be costly not only to that party member but to other party members as well. For example if the punishment is ostracism not only does the member being punished not get to enjoy drinking beer with his fellows, but they do not get to enjoy drinking beer with him. For this reason it is natural to consider an objective function $p_k(b_k, b_{-k})v_k - T(b_k/\eta_k) - \psi M_k(b_k/\eta_k, F_{-k})$ where $\psi \geq 1$ indicate that the overall social cost to the party of imposing a punishment P_k is ψP_k.

How does the model behave when $\psi \to \infty$, that is when monitoring costs grow unbounded? Clearly, the answer to this question depends on whether or not it would pay to turn out the entire electorate. Indeed, since there is no cost of monitoring at $\varphi_k = 1$, the very high costs of monitoring can potentially be avoided by choosing a very high participation rate. Suppose in particular that $C(1) = T(1) < v_k$ so that it would pay to turn out the entire electorate without monitoring cost. Then we cannot rule out the possibility that even for very large ψ the equilibrium might involve participation rates close to 1 and very much higher than the pivotal cutoff $\gamma_k(\varphi_k, F_{-k})$. By contrast, suppose that $T(1) > v_k$ so that it would never pay to turn out the entire electorate. Then

Theorem 11. *If* $C(1) = T(1) > v_k$ *then as* $\psi \to \infty$ *we have* $F_k^{\psi}(|\varphi_k - \gamma_k(\varphi_k, F_{-k}^{\psi})| \leq \epsilon) \to 1$.

Notice that this does not necessarily imply that the limit of our model when monitoring costs grow unbounded is an equilibrium in the sense of Palfrey and Rosenthal (1985) since we allow correlation devices within parties, but rather a correlated equilibrium with pivotality of the type studied by Pogorelskiy (2015).

8. Conclusion

We have examined a model that captures the importance of social norms and peer pressure in voter turnout. The resulting theory does not discard the major existing theories - the ethical voter and pivotal voter model - rather it nests both with ethical considerations being important in large elections and pivotal considerations in small elections. The theory also makes a rich new set of predictions - relating, for example, empirical results on monitoring cost to turnout. One key prediction concerns the case in which monitoring cost is large and committed voters few: in this case, unlike in ethical voting and follow-the-leader theories, the small group may be advantaged. This may explain why there are many referenda where special interests do well: for example, the type of commercial gambling permitted on Indian reservations, school budgets, the working environment for prison guards and so forth. In the case of school board elections, where many similar elections have been held, we show that the data are generally consistent with our theory of small group advantage.

There is a number of concluding observations to make.

When monitoring costs are taken into account, it is likely that the cost function is relatively flat. When this is the case the amount of exogenous uncertainty needed for the existence of pure strategy equilibrium is large. For example: if the cost function is linear and the outcome is determined by a Tullock contest success function as in Herrera, Morelli and Nunnari (2015) in order for pure strategy equilibria to exist it must be that when one party outbids the other by a factor of two it must nonetheless have at least a 25% chance of losing.

If there is too little exogenous uncertainty our results show how endogenous uncertainty must arise. Hence it should not be a surprise that although the night before the election polls showed that Brexit would lose and the pound went up substantially, Brexit nevertheless won. We were struck by a debate among pollsters in the days before Trump won the U.S. Presidential election: was the probability of Clinton winning 73% or 99%? What was striking about the debate was that neither camp seemed to realize that the outcome had to be uncertain because strategic turnout must be uncertain. Even if all sources of noise are removed, as in our all-pay auction setting, endogenous uncertainty about turnout must emerge to compensate for the lack of exogenous uncertainty.

Between the pure strategy equilibrium with substantial exogenous uncertainty and the all-pay auction with no exogenous uncertainty lies a no-womans land where the landscape is still largely unexplored. Reflecting on the likely nature of exogeous uncertainty in real elections, it seems that the issue is less one of "mistakes" such as vote counting errors than what might be described as "voter enthusiasm." Here we think the auction theoretic approach has further potential: it is natural to think of moving from an all-pay auction with known values to one with independent private values and to interpret the private realization of the value of the prize to a party as "voter enthusiasm" for the issue or candidate. In such a model equilibrium takes the form of bidding schedules. When there is little uncertainty about the private valuation, the model is much like a Harsanyi purification of our mixed strategy all-pay auction equilibrium. As uncertainty increases we would expect the equilibrium to behave more like the pure strategy equilibria of the exogenous uncertainty model.

Our model applies more generally to a situation where two groups compete by turning out members - for example in street demonstrations or strikes. The model potentially also has applications to models of lobbying by bribery as in Hillman and Riley (2006), Acemoglu (2001), or Levine and Modica (2015) - although the welfare analysis is quite different as the "votes" which are wasted in a model of voting (or demonstrations) are income to a politician in a model of lobbying by bribery.

References

Acemoglu, S. and J. A. Robinson (2001): "Inefficient Redistribution," *American Political Science Review* 95: 649-661.

Akerlof, G. A. (1990): "A Theory of Social Custom, of Which Unemployement May Be One Consequence," *Quarterly Journal of Economics* 94: 749–775.

Ali, S. N. and C. Lin (2013): "Why People Vote: Ethical Motives and Social Incentives," *American Economic Journal: Microeconomics* 5: 73–98.

Andersen, J. J., J. H. Fiva and G. J. Natvik (2014): "Voting When the Stakes are High," *Journal of Public Economics* 110: 157-166.

Baye, M. R., D. Kovenock and C. De Vries (1996): "Rigging the Lobbying Process: An Application of the All-Pay Auction," *American Economic Review* 83: 289-294.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler (2012): "A 61-million-person experiment in social influence and political mobilization," *Nature* 489: 295-298.

Chamberlain, G., M. Rotschild (1981): "A Note on the Probability of Casting a Decisive Vote," *Journal of Economic Theory* 25: 152–162.

Coate, S., M. Conlin (2004): "A Group Rule–Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review* 94: 1476–1504.

Coate, S., M. Conlin and A. Moro (2008): "The Performance of Pivotal-Voter Model in Small-Scale Elections: Evidence from Texas Liquor Referenda," *Journal of Public Economics* 92: 582–596.

Coleman, James S. (1988): "Social Capital in the Creation of Human Capital," *American Journal of Sociology* 94: S95-S120.

Della Vigna, S., J. A. List, U. Malmendier and G. Rao (2016): "Voting to Tell Others," *Review of Economic Studies* forthcoming.

Dunham, Y., A. A. Arechar and D. G. Rand (2016): "Unity for Democrats But Not Republicans: The Temporal Dynamics of Intra-Party Bias in US Electoral Politics," SSRN 2846915.

Dutta, R., D. K. Levine and S. Modica (2014): "Collusion, Randomization and Leadership in Groups," EUI.

Dutta, R., D. K. Levine and S. Modica (2016): "Collusion Constrained Equilibrium," EUI.

Dutta, R., D. K. Levine and S. Modica (2016): "Damned if you do and damned if you don't: Competing Networks," EUI.

Feddersen, T., A. Sandroni (2006): "A Theory of Participation in Elections," *American Economic Review* 96: 1271–1282.

Funk, P. (2010): “Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System,” *Journal of the European Economic Association* 8: 1077–1103.

Gerber A. S., D. P. Green, C. W. Larimer (2018): “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment,” *American Political Science Review* 102: 33-48.

Gray, M. and M. Caul (2000): “Declining Voter Turnout in Advanced Industrial Democracies, 1950 to 1997,” *Comparative Political Studies* 33: 1091-1122.

Grosser, J. and A. Schram (2006): “Neighborhood information exchange and voter participation: An experimental study,” *American Political Science Review* 100: 235-248.

Hajnal, Z., N Lajevardi and L. Nielson (2016): “Voter Identification Laws and the Suppression of Minority Votes,” UCSD.

Herrera, H., M. Morelli, and S. Nunnari (2016): “Turnout Across Democracies,” *American Journal of Political Science* 60: 607-624.

Hillman, A. L. and J. Riley (2006): “Politically Contestable Rents and Transfers,” *Economics and Politics* 1: 17-39.

Ingraham, C. (2016): “New evidence that voter ID laws ‘skew democracy’ in favor of white Republicans,” W*ashington Post*, February 4.

Jennings, M. K., L. Stoker, and J. Bowers (2009): “Politics across Generations: Family Transmission Reexamined,” *The Journal of Politics* 71:782-799.

Katz, Richard S., and Peter Mair (1995): “Changing models of party organization and party democracy the emergence of the cartel party,” *Party Politics* 1: 5-28.

Kandori, M. (1992): “Social Norms and Community Enforcement,” *The Review of Economic Studies* 59: 63-80.

Knack, S. (1992): “Civic Norms, Social Sanctions, and Vote Turnout,” *Rationality and Society* 4: 133-156.

Kobach, K. (1993): *The Referendum: Direct Democracy in Switzerland,* Aldershot.

Larreguy, H., J. Marshall and P. Querubin (2016): “Parties, Brokers and Voter Mobilization: How Turnout Buying Depends Upon the Party’s Capacity to Monitor Brokers,” *American Political Science Review* 110: 160-179.

Levine, D. K. and S. Modica (2013): “Peer Discipline and the Strength of Organizations,” WUSTL.

Levine, D. K. and S. Modica (2014): ”Peer Discipline and Incentives Within Groups,” EUI.

Levine, D. K. and S. Modica (2015): “Size, Fungibility, and the Strength of Organizations,” EUI.

Levine, D., T. Palfrey (2007): “The Paradox of Voter Participation? A Laboratory Study,” *American Political Science Review* 101: 143–158.

Levy, Pema (2016): “North Carolina Republicans Make Brazen Bid for Permanent Power After Losing Governor’s Race,” *Mother Jones* December 15.

Mulligan, C. B. and C. G. Hunter (2003): “The Empirical Frequency of a Pivotal Vote,” *Public Choice* 116: 31-54.

Nagourney, A., A. Parker, J. Rutenberg and J. Zelenynov (2012): “How a Race in the Balance Went to Obama,” *New York Times*, November 7.

Palfrey, T. R. and K. Pogorelskiy (2016): “Voting with Communciation: An Experimental Study,” Caltech.

Palfrey, T. R. and H. Rosenthal (1985): “Voter Participation and Strategic Uncertainty,” *American Political Science Review* 79: 62-78.

Penn, Maggie (2015): “Inequality, Social Context, and Value Divergence,” *Journal of Politics,* forthcoming.

Penrose, L. S. (1946): “The Elementary Statistics of Majority Voting,” *Journal of the Royal Statistical Society* 109: 53-57.

Pogorelskiy, K. (2015): “Correlated Equilibria in Voter Turnout Games,” Caltech.

Rawnsley, G. (2005): *Political Communication and Democracy,* Palgrave Macmillan.

Olson, M. (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups,* Harvard University Press.

Ostrom, E. (1990): *Governing the Commons: The Evolution of Institutions for Collective Action,* Cambridge University Press.

Chen, J. and J. Rodden (2013): “Unintentional gerrymandering: Political geography and electoral bias in legislatures,” *Quarterly Journal of Political Science* 8: 239-269.

Shachar, R. and B. Nalebuff (1999): "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review* 89: 525-547.

Siegel, R. (2014): "Asymmetric Contests with Head Starts and Nonmonotonic Costs," *American Economic Journal: Microeconomics* 6: 59-105.

Simon, L. K. and W. R. Zame (1990): "Discontinuous Games and Endogenous Sharing Rules," *Econometrica* 58: 861-872.

Szech, N. (2015): "Tie-Breaks and Bid-Caps in All-Pay Auctions," *Games and Economic Behavior* 92: 138-149.

York, B. (2012): "In Boston, stunned Romney supporters struggle to explain defeat," W*ashington Examiner*, November 7.

FOR ONLINE PUBLICATION

School Board Appendix

The basic source of information about school board success, turnouts and timing are the empirical studies by Moe (2003) and Moe (2006). We supplement this with some facts about the LAUSD, a large urban school district. From achieve.lausd.net/about we learn that the district has 84,000 employees (about half teachers) and 640,000 students, while the population of the district, which is approximately the same as the city of Los Angeles, is 3.9 million. If we assume three students per family, we have about 210,000 families of students. This gives a rough estimate of the large party (the families of students) being 2.5 or more times the size of the small party (the employees). Turnout is reported in http://www.scpr.org/news/2015/05/12/51612/how-low-voter-turnout-impacts-lausd-schools/ as about 10-20%, while in LA county, for which data is available from http://www.laalmanac.com/election/el18.htm, there are about 4.7 million registered voters out of a population of 10 million, so about half the population are registered voters. This means that roughly 300,000 votes are cast in a school board election. Assuming 90% turnout among employees and spouses means about 150,000 votes, about half of the total. It also implies that turnout among the roughly 420,000 parents is about 36%. Note that this assumes that all parents are registered voters - but of course this includes many people who are ineligible to vote. However, if only half the parents are registered voters (about the same proportion as the county as a whole) then turnout among parents would be higher, about 72%.

Is there a common prize?

As noted in the text, the turnover rate may be different between parents and teachers - and it is natural to think that the prize might be worth more to teachers since they get the benefits over their career while parents only get benefits while their children are in school. Some roughh calculations show that this is probably not a big difference. Assume an election is held every three years. Because this is a relatively short time we ignore discounting. We assume that everyone lives in the district and that nobody leaves the district. We assume that teachers arrive in the district at the time they take the job while parents arrive two years before their first student enter school. From https://nces.ed.gov/pubs2007/ruraled/tables/tablea3_8.asp?referrer=report the average experience of a public school teacher is about 14 years. We assume that every teacher remains on the job the same length of time then retires so that the length of career is 28 years. We assume each parent has three children and four years between children. This means that they will have children in school for 21 years, plus we add two years for arriving in the district early, so they remain in the voting population for 23 years. Hence the average length of a teacher career is about 22% longer than the average length of a parent being in the interested population. However, what is relevant is how this determines the turnover rate during the three years between elections. After three years with a constant departure rate 3/28 of the teachers will have left and 3/23 of the parents. They leave in a continuous stream not all at the end, so the fraction of the prize lost due to departure will be half this amount: 3/56 for teachers and 3/46 for parents. Hence among those party members who are able to vote in the current election teachers will claim 53/56 of the prize and parents 43/46. That means that the ratio of prize value of teachers to parents is $53/43 * 46/56 = 1.012$ meaning that the value of the prize to the teachers is about 1.2% greater than to the parents.

To put that number in perspective, observe that with no committed voters - the case most favorable to the small party - and quadratic turnout cost we have for the large party $V = \hat{b}_L^2/\eta_L$ and for the small party for whom the prize is worth tV that $tV = \hat{b}_S^2/\eta_S$. Hence $\hat{b}_L = \sqrt{V\eta_L}$ and $\hat{b}_S = \sqrt{tV\eta_S}$, and $\hat{b}_S/\hat{b}_L = \sqrt{t\eta_S/\eta_L}$. In other words, for the small party to be advantaged in an ethical voter model without monitoring costs and quadratic turnout cost we need $t > \eta_L/\eta_S$. In the case of teacher and parents, we know that η_L/η_S is about 2.5 or more so it would require a very big prize advantage for the teachers to be advantaged in the absence of monitoring costs.

Supplemental Appendix Part 1: The General All Pay Auction

The Model

Each identical member of party k privately draws a type y from a uniform distribution on $[0, 1]$. This type determines a cost of voting $c_k(y)$, possibly negative, and based on this the member decides whether or not to vote. The cost of voting c_k is continuously differentiable, has $c_k'(y) > 0$ and satisfies $c_k(\underline{y}) = 0$ for some $0 \leq \underline{y}_k \leq 1$. Voters for whom $y < \underline{y}_k$ are called *committed voters.*

The party can impose punishments $0 \leq P_k \leq \overline{P}_k$ on members. The social norm of the party is a threshold φ_k together with a rule prescribing voting if $y \leq \varphi_k$. This rule is enforced through peer auditing and punishment. Each member of the party is audited by another party member. The auditor observes whether or not the auditee voted. If the auditee did not vote and the party member did not violate the policy (that is, $y > \varphi_k$) there is a probability $1 - \theta_k$ that the auditor will learn this. Whatever the quality of the signal, if the auditee voted or is discovered not to have violated the policy, the auditee is not punished. If the auditee did not vote and the auditor cannot determine whether or not the auditee violated the policy, the auditee is punished with a loss of utility P_k. Pivotality is assumed not to matter so the *social norm is incentive compatible* if and only if $P_k = c_k(\varphi_k)$, in which case any member with $y \leq \varphi_k$ would be willing to pay the cost $c_k(y)$ of voting rather than face the certain punishment P_k, while any member with $y > \varphi_k$ prefers to pay the expected cost of punishment $\theta_k P_k$ over the cost of voting $c_k(y)$.

The overall cost of a punishment P_k to the party is $\psi_k P_k$ where $\psi_k \geq 1$. Naturally the punishment itself as it is paid by a member is a cost to the party. However, there may be other costs: for example, if the punishment is ostracism this may not only be costly to the member punished, but also to other party members who might otherwise have enjoyed the company of the ostracized member.

The Cost of Turning Out Voters

The total cost of turning out voters $C_k(\varphi_k)$ has two parts. The *participation cost* $T_k(\varphi_k) = \int_{\underline{y}_k}^{\varphi_k} c_k(y)dy$ is the total cost of voting to the members who vote.

Notice that $T_k'(\varphi_k) = c_k(\varphi_k)$ and so $C_k(\varphi_k)$ is increasing and convex. The *monitoring cost* $M_k(\varphi_k) = \int_{\varphi_k}^1 \psi_k \theta_k P_k dy$ is the (expected) cost of punishing party members who did not vote. As incentive compatibility requires $P_k = c_k(\varphi_k) = T_k'(\varphi_k)$, this can be written as $M_k(\varphi_k) = \psi_k \theta_k T_k'(\varphi)$.

Since $c_k(y)$ is strictly increasing we may define the unique $\overline{y}_k$ to be such that $c_k(\overline{y}_k) = \overline{P}_k$ where recall that $\overline{P}_k$ is the maximum feasible punishment, or $\overline{y}_k = 1$ if $c_k(1) \leq \overline{P}_k$. Observe that those for whom $y > \overline{y}_k$ will not vote regardless of the social norm. The feasible turnout rates φ_k are therefore those in the range $\underline{y}_k \leq \varphi_k \leq \overline{y}_k$, so our interest is on the behavior of $C_k(\varphi_k)$, $T_k(\varphi_k)$, $M_k(\varphi_k)$ in this range.

All Pay Auction

We now suppose that a population of N voters is divided into two parties $k = S, L$ of size $\eta_k N$ where $\eta_S + \eta_L = 1$. These parties compete in an election. The side that produces the greatest expected number of votes wins a prize worth $v_L > 0$ and $v_S > 0$ to each member respectively. We make the generic assumption that $\eta_S \underline{y}_S \neq \eta_L \underline{y}_L$ and $\eta_S \overline{y}_S \neq \eta_L \overline{y}_L$. We define the *large* party L to be the one with the largest possibility for turning out voters $\eta_L \overline{y}_L > \eta_S \overline{y}_S$, with S the *small* party. We define the *most committed* party to be the one with the largest number of committed voters, that is, with the largest value of $\eta_k \underline{y}_k$ and the *least committed* party to be the one with the smallest value. We will assume that $C_k'(\varphi_k) > 0$ since it is the standard assumption in the literature and the non-increasing case is harder to characterize and seems less interesting. For notational convenience we assume that for $\varphi_k < \underline{y}_k$ the cost is $C_k(\varphi_k) = 0$.

A strategy for party k is a probability measure represented as a cumulative distribution function F_k on $[\eta_k \underline{y}_k, \eta_k \overline{y}_k]$ where we refer to $b_k = \eta_k \varphi_k$ as the *bid.* A tie-breaking rule is a measurable function B_S from $[\max_k \eta_k \underline{y}_k, \eta_S \overline{y}_S]^2 \to [0, 1]$ with $B_S(\eta_S \varphi_S, \eta_L \varphi_L) = 0$ for $\eta_S \varphi_S < \eta_L \varphi_L$ and $B_S(\eta_S \varphi_S, \eta_L \varphi_L) = 1$ for $\eta_S \varphi_S > \eta_L \varphi_L$ with $B_L = 1 - B_S$. We say that F_S, F_L are an *equilibrium* if there is a tie-breaking rule B_S such that

$$\int v_k B_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) F_k(d\eta_k\varphi_k) F_{-k}(d\eta_{-k}\varphi_{-k}) - \int C_k(\varphi_k) F_k(d\eta_k\varphi_k) \geq$$
$$\int v_k B_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) \tilde{F}_k(d\eta_k\varphi_k) F_{-k}(d\eta_{-k}\varphi_{-k}) - \int C_k(\varphi_k) \tilde{F}_k(d\eta_k\varphi_k)$$

for all cdfs $\tilde{F}_k$ on $[\eta_k \underline{y}_k, \eta_k \overline{y}_k]$. We note that by the Lesbesgue decomposition theorem the cdf F_k may be decomposed into a density for a continuous random variable f_k and a discrete density ϕ_k along with a singular measure (such as a Cantor measure) that fortunately can be ruled out in equilibrium.

Let $\hat{\varphi}_k$ satisfy $C_k(\hat{\varphi}_k) = v_k$ or $\hat{\varphi}_k = \overline{y}_k$ if there is no solution. Hence, $\hat{\varphi}_k$ represents the most fraction of voters the party is willing and able to turn out. We make the generic assumption that the party sizes are such that $\eta_L\hat{\varphi}_L \neq \eta_S\hat{\varphi}_S$. We define the *disadvantaged* party d to be the party for which $\eta_d\hat{\varphi}_d < \eta_{-d}\hat{\varphi}_{-d}$, where $-d$ is the *advantaged* party.

Notice that we have three measures of the "strength" of a party: the overall possibility of turning out voters (large or small), the number of committed voters (most or least committed) and the willingness to turn out voters (advantaged or disadvantaged). Our theorem shows that each of these plays a role in determining the outcome of elections.

We let ℓ denote the party with the most committed voters.

Theorem. [1 in text] *There is a unique mixed equilibrium. The disadvantaged party earns zero and the advantaged party earns $v_{-d} - C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d) > 0$. If $\hat{\varphi}_k \leq (\eta_{-k}/\eta_k)\underline{y}_{-k}$ then the election is* uncontested*: the least committed party k is disadvantaged,* concedes *the election by bidding $\eta_k\underline{y}_k$ and the most committed party $-k$* takes *the election by bidding $\eta_{-k}\underline{y}_{-k}$.*

If $\hat{\varphi}_k > (\eta_{-k}/\eta_k)\underline{y}_{-k}$ for $k \in \{S, L\}$ then the election is contested: *in $(\max_k \eta_k\underline{y}_k, \eta_d\hat{\varphi}_d)$ the mixed strategies of the parties have no atoms, and are given by continuous densities*

$$f_k(\eta_k\varphi_k) = C'_{-k}((\eta_k/\eta_{-k})\varphi_k)/(\eta_{-k}v_{-k}).$$

In these contested elections there are three points that may have atoms: each *party may turn out only its committed voters and the advantaged party may take the election by turning out $\eta_d\hat{\varphi}_d$ with positive probability. The possible cases are as follows:*

1) The only party that concedes the election with positive probability is the disadvantaged party which does so by bidding $\eta_d\underline{y}_d$ with probability $F_d^0(\eta_d\underline{y}_d) = 1 - C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)/v_{-d} + C_{-d}((\eta_L/\eta_{-d})\underline{y}_L)/v_{-d}$

2) The only time an advantaged party turns out only its committed voters with positive probability is if it is also the most committed party in which case this probability is $F_{-d}^0(\eta_{-d}\overline{y}_{-d}) = C_d((\eta_{-d}/\eta_d)\underline{y}_{-d})/v_d$

3) The advantaged party takes the election by turning out $\eta_d\hat{\varphi}_d$ with positive probability only if $\hat{\varphi}_S = \overline{y}_S$ in which case this probability is $F_{-d}^0(\eta_S\overline{y}_S) = 1 - C_{-d}((\eta_S/\eta_L)\overline{y}_S)/v_S$. This is the only case in which the tie-breaking rules matters: when both parties bid $\eta_S\overline{y}_S$ the large party must win with probability 1.

Proof. A party will never submit a bid $\eta_k\varphi_k$ for which $\eta_k\underline{y}_k < \eta_k\varphi_k < \eta_{-k}\underline{y}_{-k}$ since such a bid will be costly but losing, and neither party will submit a bid for which $\eta_k\varphi_k > \eta_k\hat{\varphi}_k$ since to do so would cost more than the value of the prize. Since $C_k(\underline{y}_k) = 0 < v_k$, then it follows that bids must either be $\max_k \eta_k\underline{y}_k$ or in the range $(\max_k \eta_k\underline{y}_k, \eta_d\hat{\varphi}_d]$. If $v_k \leq C\left(\underline{y}_{-k}\eta_{-k}/\eta_k\right)$, it follows that $\eta_k\hat{\varphi}_k \leq \eta_{-k}\underline{y}_{-k}$. In this case party k will only mobilize committed voters, that is will bid $\eta_k\underline{y}_k$, and the other party can win with probability 1 by bidding $\eta_{-k}\underline{y}_{-k}$. Consider now the case of $v_k > C\left(\underline{y}_{-k}\eta_{-k}/\eta_k\right)$ for both parties. In the range $(\max_k \eta_k\underline{y}_k, \eta_d\hat{\varphi}_d)$ there can be no atoms by the usual argument for all-pay auctions: if there was an atom at $\eta_k\varphi_k$ then party $-k$ would prefer to bid a bit more than $\eta_k\varphi_k$ rather than a bit less, and since consequently there are no bids immediately below $\eta_k\varphi_k$ party k would prefer to choose the atom at a lower bid. This also implies that party k with the least committed voters cannot have an atom at $\eta_{-k}\underline{y}_{-k}$: if $-k$ has an atom there, then k should increase its atom slightly to break the tie. If the $-k$ does not have an atom there, then k should shift its atom to $\eta_k\underline{y}_k$ since it does not win either way.

Next we observe that in $(\max_k \eta_k\underline{y}_k, \eta_d\hat{\varphi}_d)$ there can be no open interval with zero probability. If party k has such an interval, then party $-k$ will not submit bids in that interval since the cost of the bid is strictly increasing it would do strictly better to bid at the bottom of the interval. Hence there would have to be an interval in which neither party submits bids. But then, for the same reason, it would be strictly better to lower the bid for bids slightly above the interval.

Let U_k be the equilibrium expected utility of party k. In equilibrium the disadvantaged party must earn zero since it must make bids with positive probability arbitrarily close to $\eta_d\hat{\varphi}_d$, while the advantaged party gets at least $U_{-d} \geq \eta_{-d}v_{-d} - \eta_{-d}C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d) > 0$ since by bidding slightly more than $\eta_d\hat{\varphi}_d$ it can win for sure, but gets no more than that since it must make bids with positive probability arbitrarily close to $\eta_d\hat{\varphi}_d$. We conclude that the equilibrium payoff of the advantaged party must be exactly $U_{-d} = \eta_{-d}v_{-d} - \eta_{-d}C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)$.

From the absence of zero probability open intervals in $(\max_k \eta_k \underline{y}_k, \eta_d\hat{\varphi}_d)$ it follows that the indifference condition for the advantaged party

$$v_{-d}F_d(\eta_d\varphi_d) - C_{-d}((\eta_d/\eta_{-d})\varphi_d) = v_{-d} - C_d((\eta_d/\eta_{-d})\hat{\varphi}_d)$$

must hold for at least a dense subset. For the disadvantaged party we have

$$v_dF^0_{-d}(\eta_{-d}\varphi_{-d}) - C_d((\eta_{-d}/\eta_d)\varphi_{-d}) = 0$$

for at least a dense subset. This uniquely defines the cdf for each party in that range:

$$F_d(\eta_d\varphi_d) = 1 - \frac{C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d) - C_{-d}((\eta_d/\eta_{-d})\varphi_d)}{v_{-d}}$$

for $\eta_d\varphi_d \in (\max_k \eta_k \underline{y}_k, \eta_d\hat{\varphi}_d)$, and

$$F_{-d}(\eta_{-d}\varphi_{-d}) = \frac{C_d((\eta_{-d}/\eta_d)\varphi_{-d})}{v_d}$$

for $\eta_{-d}\varphi_{-d} \in (\max_k \eta_k \underline{y}_k, \eta_d\hat{\varphi}_d)$. As these are differentiable they can be represented by continous density functions which are found by taking the derivative. Evaluating $F_d(\eta_d\varphi_d)$ at $\max_k \eta_k \underline{y}_k$ gives $F^0_d(\eta_d\underline{y}_d) = 1 - C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)/v_{-d} + C_{-d}(\max_k \eta_k \underline{y}_k/\eta_{-d})/v_{-d}$. Note that $F_d(\max_k \eta_k \underline{y}_k)$ is always strictly positive and may or may not be smaller than 1. To see this, notice that $\eta_{-d}\hat{\varphi}_{-d} > \eta_d\hat{\varphi}_d$

implies $C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)/v_{-d} < C_{-d}(\hat{\varphi}_{-d})/v_{-d} \leq 1$. Since the disadvantaged party has an atom at $\max_k \eta_k \underline{y}_k$ if an only if $\max_k \eta_k \underline{y}_k = \eta_d \hat{\varphi}_d$ we see that the disadvantaged party has an atom at $\eta_d \underline{y}_d$ with probability $F_d^0(\eta_d \underline{y}_d) = 1 - C((\eta_d/\eta_{-d})\hat{\varphi}_d)/v_{-d} + C(\max_k \eta_k \underline{y}_k/\eta_{-d})/v_{-d}$ and no other atom.

As for the advantaged party, if -$d = S$ then $\eta_L \overline{y}_L > \eta_S \overline{y}_S \geq \eta_S \hat{\varphi}_S > \eta_L \hat{\varphi}_L$ implies that $F_S(\eta_L \hat{\varphi}_L) = C_L(\hat{\varphi}_L)/v_L = 1$. If instead $-d = L$ then $F_L(\eta_S \hat{\varphi}_S) = C_S(\hat{\varphi}_S)/v_S$. If $\hat{\varphi}_S \leq \overline{y}_S$then this is 1 and there is no atom, otherwise there must be an atom of size $1 - C_S(\overline{y}_S)/v_S$. Turning to $\max_k \eta_k \underline{y}_k$ we see that the atom there is given by

$C_d((\eta_{-d}/\eta_d)(\eta_\ell/\eta_{-d})\underline{y}_\ell)/v_d = C_d((\eta_\ell/\eta_d)\underline{y}_\ell)/v_d$. If $\ell = d$ this is $C_d(\underline{y}_d)/v_d = 0$ if $\ell = -d$ this is

$$C_d((\eta_{-d}/\eta_d)\underline{y}_{-d})/v_d > C_d(\underline{y}_d) = 0.$$

□

Comparative Statics

Next we examine the comparative statics of the model using a notion of decreased turnout in terms of first-order stochastoc dominance.

Corollary 1. *We have the following*

1. If the value of the prize to the least committed party is small enough then that party is disadvantaged and concedes the election with probability one. If the value of the prize to the large party is large enough then it is advantaged and it takes the election with very high probability, while the small party concedes the election.

In contested elections:

2. Increasing the value of the prize of the advantaged party increases the surplus of the advantaged party (and hence welfare), increases the probability of the advantaged party winning, decreases the turnout of the disadvantaged party and has no effect on the turnout of the advantaged party. The reverse is true for decreasing the valuation of the advantaged party provided it remains advantaged.

3. Decreasing the valuation of the disadvantaged party increases the surplus of the advantaged party (and hence welfare), decreases the turnout of the advantaged party and if $\hat{\varphi}_d < \overline{y}_d$ decreases the turnout of the disadvantaged party. The *reverse is true for increasing the valuation of the disadvantaged party provided it remains disadvantaged.*

Proof. 1. If $v_{-\ell} \leq C_\ell((\eta_\ell/\eta_{-\ell})\underline{y}_\ell)$, it follows that $\eta_{-\ell}\hat{\varphi}_{-\ell} \leq \eta_\ell \underline{y}_\ell$ and the party with the least committed voters always concedes the election. In other words if the value of the prize to the party with the least committed voters is small enough then that party is disadvantaged and concedes the election. On the other hand as $v_L \to \infty$ then $\hat{\varphi}_k = \overline{y}_L$ so that the large party is advantaged. The probability that the small party concedes is then $P_S(\eta_S \underline{y}_S) = 1 - C_L((\eta_S/\eta_L)\overline{y}_S)/v_L$. Hence, the probability that the small party concedes goes to one at a rate that is bounded independently of the value of the prize to the small party. In other words, in a very high value election, the small party turns out only its committed voters and the large party acts preemptively turning as many voters as the small party is capable of turning out.

In a contested election:

2. If v_{-d} increases then $\hat{\varphi}_d$ does not change. The equilibrium payoff of the advantaged party is $\eta_{-d}v_{-d} - \eta_{-d}C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)$ so increases. However, the equilibrium bidding strategy of the advantaged party and its expected payment do not depend on v_{-d}. Hence, it must be the case that the expected probability of winning of the advantaged party increases with v_{-d}. Moreover, since the bidding strategy of the advantaged party does not change, neither does its turnout. Finally, the density of the disadvantaged party $f_d(\eta_d\varphi_d) = C'_{-d}((\eta_d/\eta_{-d})\varphi_d)/(\eta_{-d}v_{-d})$ falls with the extra weight accumulating at the atom where it concedes the election, clearly lowering the turnout (and providing an alternative argument as to why the probability of the advantaged party winning must increase).

3. Suppose that v_d decreases. Then $\hat{\varphi}_d$ weakly decreases since $C_{-d}(\hat{\varphi}_d) = v_d$ and C_{-d} is assumed to be increasing and strictly decreases if $\hat{\varphi}_d < \overline{y}_d$. The density of the advantaged party $f_{-d}(\eta_{-d}\varphi_{-d}) = C'_d((\eta_{-d}/\eta_d)\varphi_{-d})/(\eta_d v_d)$ increases by a fixed ratio and the probability that it turns out only its committed voters, which equals $C_d((\eta_{-d}/\eta_d)\underline{y}_{-d})/v_d$, increases by exactly the same ratio. This means that the cdf has shifted to the left reducing turnout.

If $\hat{\varphi}_d < \overline{y}_d$, then an increase in $\hat{\varphi}_d$ strictly decreases the surplus $\eta_{-d}v_{-d} -$

$\eta_{-d} C_{-d}((\eta_d/\eta_{-d})\hat{\varphi}_d)$ of the advantaged party. Moreover, $f_d(\eta_d \varphi_d)$ is unchanged, while the range is strictly smaller, with the extra weight accumulating where the disadvantaged party concedes the election so the turnout of the disadvantaged party declines. $\square$

Common Prize

We now assume that both parties have identical costs $C_k(\varphi_k) = C(\varphi_k)$ with common fraction of committed voters $\underline{y}_k = \underline{y}$ and that both are able to turn out all voters $\overline{y}_k = 1$. Moreover, the prize is a common one: $v_k = V/\eta_k$.

Define $\xi \equiv 1/\left(\max_{\varphi \in [\underline{y}, \eta_S/\eta_L]} |C''(\varphi)|/C'(\underline{y})\right)$ as a measure of the curvature of $C(\varphi_k)$, where a lower ξ represents a higher curvature of C.

Theorem. [2 in text] *If C is convex then the large party is advantaged. If C is concave and*

a. $\sqrt{(2\xi + \underline{y})\underline{y}} < \eta_S/\eta_L$ *(enough concavity relative to the number of committed voters)*

b. $\eta_L C(\sqrt{(2\xi + \underline{y})\underline{y}}) < V < \eta_L C(\eta_S/\eta_L)$ *(intermediate value of prize)*

then the small party is advantaged.

If C is quadratic then condition a. is also necessary for the small party to be advantaged.

Proof. The case of C is convex is covered in the text. Now consider the case of a concave C. For the small party to be advantaged we need decreasing average per-capita cost, and this can only be the case if the curvature of C is high enough. Indeed, let φ^* the point at which marginal and average cost are equal, that is the unique solution to $C(\varphi) = C'(\varphi)\varphi$. Above φ^* the average cost is decreasing, and for an interior solution φ^* to exists we need that $C(\eta_S/\eta_L) > C'(\eta_S/\eta_L)\eta_S/\eta_L$. We claim that a sufficient condition for the small party to be advantaged is $\varphi^* < \hat{b}_L/\eta_L < \eta_S/\eta_L$. To see that the latter condition implies $\hat{b}_S > \hat{b}_L$, notice that since $\hat{b}_L < \eta_S$, it is feasible for the small party to match the willingness to pay of the large party. Since above φ^* the average cost is decreasing, we have that for $b_S/\eta_S > \hat{b}_L/\eta_L > \varphi^*$ it follows that $C(b_S/\eta_S)/\left(b_S/\eta_S\right) < C(\hat{b}_L/\eta_L)/\left(\hat{b}_L/\eta_L\right)$.

In particular, if the small party bids exactly $b_S = \hat{b}_L$, we have that

$$C(\hat{b}_L/\eta_S)/\left(\hat{b}_L/\eta_S\right) < C(\hat{b}_L/\eta_L)/\left(\hat{b}_L/\eta_L\right) = V/\hat{b}_L,$$

where the last equality follows form the definition of $\hat{b}_L$. Rearranging we get

$$C(\hat{b}_L/\eta_S)\eta_S < V$$

or $\hat{b}_S > \hat{b}_L$.

Since φ^* is not a primitive of the model, in order to compute an upper bound $\bar{\varphi}$ on φ^*. First, notice that

$$\left(C(\varphi) - C'(\varphi)\varphi\right)/C'(\underline{y}) = \left(\int_{\underline{y}}^{\varphi} C'(x)dx - C'(\varphi)\varphi\right)/C'(\underline{y}) =$$

$$\left(\int_{\underline{y}}^{\varphi} [C'(\underline{y}) + \int_{\underline{y}}^{x} C''(z)dz]dx - \varphi \int_{\underline{y}}^{\varphi} C''(z)dz - \varphi C'(\underline{y})\right)/C'(\underline{y}) =$$

$$-\underline{y} + \left(\int_{\underline{y}}^{\varphi}\int_{\underline{y}}^{x} C''(z)dzdx - \varphi\int_{\underline{y}}^{\varphi} C''(z)dz\right)/C'(\underline{y}) =$$

$$-\underline{y} + \left(\int_{\underline{y}}^{\varphi}\int_{\underline{y}}^{x} C''(x)dxdz - \varphi\int_{\underline{y}}^{\varphi} C''(z)dz\right)/C'(\underline{y}) =$$

$$-\underline{y} + \left(\int_{\underline{y}}^{\varphi}(x-\underline{y})C''(x)dx - \int_{\underline{y}}^{\varphi} \varphi C''(z)dz\right)/C'(\underline{y}) =$$

$$-\underline{y} + \int_{\underline{y}}^{\varphi}(x-\underline{y}-\varphi)\left(C''(x)/C'(\underline{y})\right)dx \geq -\underline{y} - \int_{\underline{y}}^{\varphi}(x-\underline{y}-\varphi)/\xi dx,$$

where $\xi \equiv 1/\left(\max_{\underline{y}\leq\varphi\leq\eta_S/\eta_L}|C''(\varphi)|/C'(\underline{y})\right)$ is a measure of the curvature of C (a lower ξ represents a higher curvature of C), and the inequality follows from the fact that if the RHS is non-negative the LHS must be non-negative. The solution $\bar{\varphi}$ to

$$-\underline{y} - \int_{\underline{y}}^{\varphi} (x - \underline{y} - \varphi)/\xi dx = 0$$

is therefore an upper bound on φ^*. Simplifying the LHS we get

$$-\underline{y} + (\overline{\varphi}^2 - \underline{y}^2)/2\xi = 0,$$

which yields

$$\overline{\varphi} = \sqrt{(2\xi + \underline{y})\underline{y}}.$$

Since $C(\bar{\varphi}) < V/\eta_L$ implies $\hat{b}_L/\eta_L > \overline{\varphi} \geq \varphi^*$, and $C(\eta_S/\eta_L) > V/\eta_L$ implies $\hat{b}_L < \eta_S$, we get the sufficient conditions stated in the theorem. Notice that condition a) establish an upper bound on the LHS, and since $\underline{y} < \eta_S/\eta_L$, for any $\underline{y}$ there is some ξ sufficiently large such that $\sqrt{(2\xi + \underline{y})\underline{y}} < \eta_S/\eta_L$. If C is quadratic, $C''(x)/C'(\underline{y})$ is constant and hence $\bar{\varphi}$ is equal to φ^*. In this case condition a) is $\varphi^* < \hat{b}_L/\eta_L < \eta_S/\eta_L$, which is also necessary for the small party to be advantaged. □

Recall that given two cumulative distribution functions F and G, the distribution F *first-order stochastically dominates* the distribution G (we write $F \geq_{FSD} G$), if $F(x) \leq G(x)$ for all x, with strict inequality at some x. Hence, if we can provide conditions under which party k equilibrium bidding strategy FSD that of party $-k$, then we can conclude that party k will turnout more members in expectation and will also have a higher probability of winning the election. The next theorem provides such conditions.

Theorem. [3 in text] *If the cost function is either convex or it is concave and the small party is advantaged, then the advantaged party equilibrium bidding function FSD that of the disadvantaged party: this holds even if the prize is not common provided the value to the advantaged party is at least that of the disadvantaged party: $\eta_{-d}v_{-d} \geq \eta_d v_d$. With a common prize, concave cost, large party advantaged and small party unconstrained $V = \eta_d C(\hat{b}_d/\eta_d)$ neither bidding function FSD that of the other party.*

Proof. At $\hat{b}_d$ we have $F_d(\hat{b}_d) = \hat{F}_{-d}(\hat{b}_d) = 1$ so this is irrelevant for FSD. For

$\eta_S \underline{y} \leq b < \eta_L \underline{y}$ we have $F_L(b) = 0$ while $F_S(b) > 0$ if and only if the small party is disadvantaged. Hence when the small party is disadvantaged its bidding schedule cannot FSD that of the large party, while if it is advantaged this range is irrelevant for FSD.

It remains to examine the range $\eta_L \underline{y} \leq b < \hat{b}_d$. In this range the equilibrium bid distributions are given by

$$F_d(b) = 1 - \frac{\eta_{-d} C(\hat{b}_d/\eta_{-d})}{\eta_{-d} v_{-d}} + \frac{\eta_{-d} C(b/\eta_{-d})}{\eta_{-d} v_{-d}}$$

$$F_{-d}(b) = \frac{\eta_d C(b/\eta_d)}{\eta_d v_d}.$$

Hence for FSD of the advantaged party, we must have

$$1 - \frac{\eta_{-d} C(\hat{b}_d/\eta_{-d})}{\eta_{-d} v_{-d}} + \frac{\eta_{-d} C(b/\eta_{-d})}{\eta_{-d} v_{-d}} - \frac{\eta_d C(b/\eta_d)}{\eta_d v_d} > 0.$$

Since $\eta_{-d} v_{-d} \geq \eta_d v_d$ this is true if

$$1 - \frac{\eta_{-d} C(\hat{b}_d/\eta_{-d})}{\eta_d v_d} + \frac{\eta_{-d} C(b/\eta_{-d})}{\eta_d v_d} - \frac{\eta_d C(b/\eta_d)}{\eta_d v_d} > 0$$

and in the common value case this is if and only if. Moreover since $\eta_d v_d \geq \eta_d C(\hat{b}_d/\eta_d)$ this is true if

$$1 - \frac{\eta_{-d} C(\hat{b}_d/\eta_{-d})}{\eta_d C(\hat{b}_d/\eta_d)} + \frac{\eta_{-d} C(b/\eta_{-d})}{\eta_d C(\hat{b}_d/\eta_d)} - \frac{\eta_d C(b/\eta_d)}{\eta_d C(\hat{b}_d/\eta_d)} > 0$$

and if and only if the disadvantaged party is not constrained. This is equivalent to

$$\left(\eta_{-d} C(b/\eta_{-d}) - \eta_d C(b/\eta_d)\right) - \left(\eta_{-d} C(\hat{b}_d/\eta_{-d}) - \eta_d C(\hat{b}_d/\eta_d)\right) > 0.$$

Let $g(\eta, b) \equiv \eta C(b/\eta)$. The derivative with respect to η is $g_\eta(\eta, b) = C(b/\eta) - (b/\eta)C'(b/\eta)$ so the cross partial is $g_{\eta b}(\eta, b) = -(b/\eta^2)C\text{''}(b/\eta)$. Observe that the sufficient condition may be written as

$$0 < (g(\eta_{-d}, b) - g(\eta_d, b)) - \left(g(\eta_{-d}, \hat{b}_d) - g(\eta_d, \hat{b}_d)\right)$$

$$= \int_{\eta_d}^{\eta_{-d}} \left(g_\eta(\eta, b) - g_\eta(\eta, \hat{b}_d)\right) d\eta$$

$$= -\int_{\eta_d}^{\eta_{-d}} \int_b^{\hat{b}_d} g_{\eta b}(\eta, b') d\eta db' = \int_{\eta_d}^{\eta_{-d}} \int_b^{\hat{b}_d} (b'/\eta^2) C\text{''}(b'/\eta) d\eta db'$$

This is positive if $\eta_{-d} > \eta_d$ and C is convex or if $\eta_d > \eta_{-d}$ and C is concave, which gives the primary result. On the other hand, in the case of a common prize, the large party advantaged, and the small party unconstrained, it is negative and gives the exact sign of $F_d(b) - F_{-d}(b)$ (it is necessary and sufficient). Hence, since the difference between F_d and F_{-d} is positive for $\eta_S \underline{y} \leq b < \eta_L \underline{y}$ and negative for $\eta_L \underline{y} \leq b < \hat{b}_d$ neither bidding schedule FSD the other. □

Proposition 1. *Suppose that*

$$\frac{2\sqrt{(2-\underline{y})\underline{y}}}{(1-\underline{y})^2} < \eta_S/\eta_L$$

and that cost is quadratic. Then there exists an open set of V's and for any such V there are bounds $1/2 < \underline{\theta} < \theta^* < \overline{\theta} \leq 1$ such that

a. for $\theta > \underline{\theta}$ cost is concave on $[\underline{y}, 1]$

b. for $\overline{\theta} > \theta > \theta^*$ the small party is advantaged

c. for $\theta^* > \theta > \underline{\theta}$ the large party is advantaged yet the small party turns out more expected voters and has a higher probability of winning the election.

Proof. Recall the quadratic case $C(\varphi_k|\theta) = (1-2\theta)(\varphi_k - \underline{y})^2 + 2\theta(1-\underline{y})(\varphi_k - \underline{y})$. Hence $C'(\underline{y}) = 2\theta(1-\underline{y})$ and $C''(\varphi_k) = 2(1-2\theta)$ giving for $\theta > 1/2$ that $\xi \equiv \theta(1-\underline{y})/(2\theta-1)$. The assumption on $\underline{y}$ implies $\sqrt{(2-\underline{y})\underline{y}} < \eta_S/\eta_L$ which suffices for the sufficient concavity condition for small party advantage $\sqrt{(2\xi+\underline{y})\underline{y}} < \eta_S/\eta_L$ to be satisfied at $\theta = 1$. Notice that the derivative of C with respect to θ is $2(\varphi_k - \underline{y})(1-\varphi_k) > 0$. Hence if $V < \eta_S(1-\underline{y})^2$ then $\hat{b}_S < \eta_S$ for $1/2 \leq \theta \leq 1$.

Moreover, at $\theta = 1$ the aggregate cost of $\sqrt{(2-\underline{y})\underline{y}}$ to large party is at worst $2\eta_L\sqrt{(2-\underline{y})\underline{y}}$, which by assumption is smaller than $\eta_S(1-\underline{y})^2$. Hence there exists a V satisfying $\eta_L\sqrt{(2-\underline{y})\underline{y}} < V < \min\{\eta_S(1-\underline{y})^2, \eta_L C(\eta_S/\eta_L|\theta = 1)\}$. For such a V the small party is advantaged at $\theta = 1$ and $\hat{b}_S < \eta_S$ for $1/2 \leq \theta \leq 1$. Fix such a V. The willingness to bid is the solution of

$$(1-2\theta)(b_k/\eta_k - \underline{y})^2 + 2\theta(1-\underline{y})(b_k/\eta_k - \underline{y}) = V/\eta_k.$$

Since this is quadratic in b_k it can be solved by the quadratic formula from which it is apparent that $\hat{b}_k(\theta)$ is a continuous function. This implies as well that the strategies are continuous in θ, since the support of the continuous part of the density is continuous as is the upper bound. We can also conclude that $\hat{b}_S = \hat{b}_L = b$ if and only if

$$(1-2\theta)(b - \eta_S\underline{y})^2 + 2\theta(1-\underline{y})\eta_S(b - \eta_S\underline{y}) = \eta_S V$$

and

$$(1-2\theta)(b - \eta_S\underline{y})^2 + 2\theta(1-\underline{y})\eta_S(b - \eta_S\underline{y}) - \eta_S V =$$
$$(1-2\theta)(b - \eta_L\underline{y})^2 + 2\theta(1-\underline{y})\eta_L(b - \eta_L\underline{y}) - \eta_L V.$$

The latter equation is linear in b since the b^2 terms are the same on both sides. Hence the equation has at most two solutions (b, θ), and it has a solution $b(\theta)$ which is a rational function of θ. Substituting that into the first equation we find that those values of θ for which $\hat{b}_S = \hat{b}_L$ are zeroes of a rational function. Hence, either there must be a finite number of zeroes or the function must be identically equal to zero. But it cannot be identically zero since $\hat{b}_S - \hat{b}_L$ is negative at $\theta = 1/2$ and positive at $\theta = 1$. We conclude that there is some point θ^* at which $\hat{b}_S = \hat{b}_L$ and the small group is advantaged for $\theta^* < \theta < \overline{\theta}$ for some $\overline{\theta}$, while the large group is advantaged for $\theta_0 < \theta < \theta^*$ for some θ_0.

Since C is concave in $\theta* < \theta < \overline{\theta}$ the small party follows a strategy that FSD that of the large party. It follows that in the limit at $\theta*$ the strategy of the small party either FSD that of the large party or is the same as that of the large party. However, for $\theta > \theta*$ the small party plays $\eta_L \underline{y}$ with probability zero while the large party plays it with probability

$$1 - \frac{C((b_L/\eta_S))}{\eta_S V} + \frac{C((\eta_L/\eta_S)\underline{y})}{\eta_S V} \to \frac{C((\eta_L/\eta_S)\underline{y})}{\eta_S V} > 0$$

so in the limit the two strategies are not identical. Since at $\theta*$ the strategy of the small party FSD that of the large party it has a strictly higher probability of winning and strictly higher expected turnout. Since the probability of winning and expected turnout are continuous functions of the strategies which are continuous in θ is follows that this remains true in an open neighborhood of θ^*. □

Theorem. [4 in text] *In a high value election the probabilities that the small party concedes and the large party takes the election increase with V, and approach* 1 *in the limit. The bid distribution of the small party declines and the bid distribution of the large party increases in* V *in the sense of FSD. The expected vote differential increases in* V *while expected turnout cost remains constant.*

Proof. In a high value election the small party is constrained and the large party is advantaged. In this case probability of concession by the small party is $F_S^0(\eta_S \underline{y})V/\eta_L - C(\eta_S \underline{y}/\eta_L) = U_L/\eta_L = V/\eta_L - C(\eta_S/\eta_L)$ giving $F_S^0(\eta_S \underline{y}) = 1 - (\eta_L/V)\left(C(\eta_S/\eta_L) - C(\eta_S \underline{y}/\eta_L)\right)$ which is indeed increasing in V. The probability the large party takes the election is $(1 - F_L^0(\eta_S))\, V/\eta_S - C(\eta_S/\eta_L) = 0$ from which $F_L^0(\eta_S) = 1 - (\eta_S/V)C(\eta_S/\eta_L)$, also increasing in V. Since the support and shape of the cost function in the mixing range do not change, raising V simply lowers the densities by a common factor, meaning that these shifts reflect stochastic dominance as well.

Total surplus is $V - \eta_{-d} C(\hat{b}_d/\eta_{-d})$. Since one party gets the prize for certain, this implies the expected turnout cost is $\eta_{-d} C(\hat{b}_d/\eta_{-d})$ and in a high value election $\hat{b}_d$ remains constant at η_S. □

Theorem. [5 in text] *If* C *is convex or if* C *is concave and the small party is advantaged, disagreement increases the peak turnout, the expected turnout cost and decreases the bid differential.*

Proof. The case in which $C(1)\eta_L \leq V$ is immediate since $\hat{b}_S = \eta_S < \eta_L = \hat{b}_L$ and the result follows. If instead, $C(1)\eta_S > V$, neither party is constrained. Given the definition of willingness to bid $\eta_k C(\hat{b}_k/\eta_k) - V = 0$, we can apply the implicit function theorem and find that

$$\frac{d\hat{b}_k}{d\eta_k} = -\frac{C(\hat{b}_k/\eta_k) - \hat{b}_k C'(\hat{b}_k/\eta_k)/\eta_k}{C'(\hat{b}_k/\eta_k)} =$$

$$(\hat{b}_k/\eta_k)\frac{C'(\hat{b}_k/\eta_k) - C(\hat{b}_k/\eta_k)/(\hat{b}_k/\eta_k)}{C'(\hat{b}_k/\eta_k)}.$$

If C is convex, $\hat{b}_L > \hat{b}_S$, and marginal cost is larger than average cost for both parties. As a result, we have that

$$\frac{d\hat{b}_k}{d\eta_k} > 0.$$

Hence increasing η_S and decreasing η_L implies that the bid differential decreases and peak turnout increases. If instead C is concave and the small group is advantaged (that is, $\hat{b}_S > \hat{b}_L$), marginal cost is smaller than average cost for both parties. As a result

$$\frac{d\hat{b}_k}{d\eta_k} < 0.$$

Hence increasing η_S and decreasing η_L implies that the bid differential increases and peak turnout increases. Total surplus is $V - \eta_{-d}C(\hat{b}_d/\eta_{-d})$. Since one party gets the prize for certain, expected turnout cost is $\eta_{-d}C(\hat{b}_d/\eta_{-d})$. Differentiate this with respect to η_{-d} to find

$$C(\hat{b}_d/\eta_{-d}) - (\hat{b}_d/\eta_{-d})C'(\hat{b}_d/\eta_{-d}) + C'(\hat{b}_d/\eta_{-d})d\hat{b}_d/d\eta_{-d}$$

$$(\hat{b}_d/\eta_{-d})\left(\frac{C(\hat{b}_d/\eta_{-d})}{\hat{b}_d/\eta_{-d}} - C'(\hat{b}_d/\eta_{-d})\right) + C'(\hat{b}_d/\eta_{-d})d\hat{b}_d/d\eta_{-d}$$

If C is convex the bracketed expression is negative and $d\hat{b}_d/d\eta_{-d} \leq 0$, hence the entire expression is negative when disagreement decreases. If C is concave and the small party is advantaged the bracketed expression is positive and $d\hat{b}_d/d\eta_{-d} \geq 0$, hence the entire expression is positive when disagreement increases. $\square$

Theorem. [6 in text] *An increase in monitoring inefficiency decreases peak turnout. In a high value election it decreases in FSD the turnout of the advantaged party and there exists* $0 < \eta_0 < \underline{\eta} < \overline{\eta} \leq 1/2$ *such that for* $\eta_S < \eta_0$ *the turnout of the disadvantaged party increases in FSD and the expected vote differential decreases, while for* $\underline{\eta} < \eta_S < \overline{\eta}$ *the expected turnout of the disadvantaged party decreases while the expected vote differential also decreases.*

Proof. If the election is not high value the disadvantaged party is unconstrained. Hence, given the definition of willingness to bid $\eta_k C(\hat{b}_k/\eta_k) - V = 0$, we can apply the implicit function theorem and find that

$$\frac{d\hat{b}_d}{d\theta} = -\frac{\eta_d dC(\hat{b}_d/\eta_d)/d\theta}{C'(\hat{b}_d/\eta_d)} = -\frac{\eta_d \theta(1 - \hat{b}_d/\eta_d)c(\hat{b}_d/\eta_d)}{C'(\hat{b}_d/\eta_d)} < 0$$

Hence as θ decreases, that is as monitoring efficiency increases, so it does peak turnout. In a high value election the peak turnout $\hat{b}_d$ is fixed at η_S and the small party is disadvantaged. Examining the equilbrium bid distributions we have

$$F_S(b) = 1 - \frac{\eta_L C(\eta_S/\eta_L)}{V} + \frac{\eta_L C(b/\eta_L)}{V}$$

$$F_L(b) = \frac{\eta_S C(b/\eta_S)}{V}$$

while $C(\varphi_k) = T(\varphi_k) + \theta(1 - \varphi_k)T^{'}(\varphi_k)$. Examining $F_L(b)$ first, we see that $dF_L/d\theta > 0$ which is the condition for a decrease in FSD. For $F_S(b)$ we have

$$\frac{dF_S}{d\theta} = -\frac{\eta_L}{V}\left((1 - \eta_S/\eta_L)T^{'}(\eta_S/\eta_L) - (1 - b/\eta_L)T^{'}(b/\eta_L)\right)$$

Notice that for φ_k sufficiently close to $\underline{y}$ we must have $(1-\varphi_k)T^{'}(\varphi_k)$ increasing, say for $\underline{y} < \varphi_k < \varphi_0$. Hence for $\eta_S/\eta_L < \varphi_0$ we have $dF_S/d\theta < 0$ for $b \leq \eta_S$. This is the condition for an increase in FSD. Since F_L stochastically dominates F_S and F_L decreases while F_S increases it follows that the expected vote differential must decrease.

Consider next that as $\eta_S \to 1/2$, it follows that $(1-\eta_S/\eta_L)T^{'}(\eta_S/\eta_L) \to 0$. Hence for any fixed b it is eventually true that $dF_S(b)/d\theta > 0$. It follows that, for sufficiently large η_S, the expected turnout of the small party must decline with θ. Since the derivative of expected turnout is a continous function of θ, it follows that there is a value of $\underline{\eta}$ such that expected turnout of the small party is constant with θ while for larger η_S it declines. At $\underline{\eta}$ the expected vote differential must decline with θ since the small party expected turnout is constant and the large party expected turnout declines. Since the derivative of the expected vote differential is also continuous in η_S it follows that for η_S larger than but close enough to $\underline{\eta}$, the small party expected turnout declines and the expected vote differential does as well. □

Theorem. [7 in text] *If* $\sup_y h(y)$ *is sufficiently small then only the advantaged party will suppress votes. If* S *is sufficiently small it will choose to do so.*

Proof. Recall that we are focusing on the case in which parties are unconstrained (that is, $C(1)\eta_S > V$). We first show that voter suppression raises total cost. Let $\tilde{C}(\varphi_k)$ be the total cost after vote suppression. Since $c(\underline{y}) = 0$, and $\tilde{c}(\underline{y}) = h(\underline{y}) > 0$, then $\underline{\tilde{y}} < \underline{y}$. For $\underline{\tilde{y}} < \varphi_k \leq \underline{y}$ we have $\tilde{C}(\varphi_k) > 0 = C(\varphi_k)$. For $\varphi_k > \underline{y}$, we have that

$$\tilde{C}(\varphi_k) = \tilde{C}(\underline{y}) + \int_{\underline{y}}^{\varphi_k} [c(y) + h(y)]dy + \theta(1-\varphi_k)[c(\varphi_k) + h(\varphi_{k,})] >$$

$$> \int_{\underline{y}}^{\varphi_k} c(y)dy + \theta(1-\varphi_k)c(\varphi_k) = C(\varphi_k)$$

Next, we show that raising total cost leads to the result. If h is sufficiently small then the disadvantaged party cannot suppress enough votes to become advantaged, so vote suppression never changes which party is advantaged. The disadvantaged party therefore gets zero payoff regardless of whether it suppresses votes or not, hence it will not pay a positive cost to do so. On the other hand, if h is sufficiently small and parties were unconstrained, that is, $\hat{b}_k < \eta_k$, increasing the cost of the disadvantaged party must decrease its willingness to bid. Since the cost function of the advantaged party remains unchanged when the disadvantaged party does not suppress votes, its surplus therefore goes up. Hence if S is small enough it is worth paying. $\square$

Supplementary Appendix Part 2: Contests

We suppose that a population of N voters is divided into two parties $k = S, L$ of size $\eta_k N$. These parties compete in a contest. Each party produces an expected number of votes $\eta_k \varphi_k$ and may win a prize worth $v_k > 0$ to each voter respectively. Now, however, the probability of the small party winning the prize is given by a conflict resolution function $p_S(\eta_S \varphi_S, \eta_L \varphi_L) \in [0, 1]$ and we define $p_L(\eta_L \varphi_L, \eta_S \varphi_S) = 1 - p_S(\eta_S \varphi_S, \eta_L \varphi_L)$.

A strategy for party k is a probability measure represented as a cumulative distribution function F_k on $[0, 1]$. Each party faces a per capita costs of turning out voters characterized by a cost function $C_k(\varphi_k, F_{-k})$ which represents the cost of turning out a fraction φ_k of voters when pivotality is determined by φ_k, F_{-k}. For a topology to the space of cumulative distribution functions on $[0, 1]$ we use the weak topology: the corresponding notion of convergence is that $F_k^N \to F_k$ if the expectation of every continous function on $[0, 1]$ converges.[46]

We assume that $p_S(\eta_S \varphi_S, \eta_L \varphi_L)$ and $C_k(\varphi_k, F_{-k})$ are continuous on $[0, 1]$. Note that in general both depend on the absolute size of the population N. The latter assumption amounts to assuming that $\overline{y}_k = 1$, that is, that the punishment is sufficiently large that it is possible (although possibly very expensive) to coerce all voters into voting.[47] Note that we assume nothing regarding the monotonicity of C_k, this is important for the result on convergence to pivotal equilibrium since when we allow pivotality, and monitoring costs are very high, C_k will be continuous but certainly not monotone - it can be expensive to get voters not to vote when they are motivated to vote because of the chance they are pivotal.

[46] This is called the weak topology by probability theorists and the weak* topology in functional analysis **?**.

[47] Otherwise we would have $\overline{y}_k$ depending on the strategy of the other party F_{-k} creating the problem discussed in Dutta, Levine and Modica (2014) when there are constraints on party behavior that depend on the choice of the other party.

We say that F_S, F_L are an *equilibrium* of the conflict resolution model if

$$\int v_k p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) F_k(d\eta_k\varphi_k) F_{-k}(d\eta_{-k}\varphi_{-k}) - \int C_k(\varphi_k, F_{-k}) F_k(d\eta_k\varphi_k) \quad \geq$$
$$\int v_k p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) \tilde{F}_k(d\eta_k\varphi_k) F_{-k}(d\eta_{-k}\varphi_{-k}) - \int C_k(\varphi_k, F_{-k}) \tilde{F}_k(d\eta_k\varphi_k)$$

Theorem. [8 in text] *An equilibrium of the conflict resolution model exists.*

Proof. This is essentially the theorem of Glicksberg (1952), except that we do not require $C_k(\varphi_k, F_{-k})$ to be linear in F_{-k}. However inspection of Glicksberg's proof shows that only continuity in F_{-k} is needed - Glicksberg uses only the fact that the objective function is weakly concave in F_k so that the best-response correpondence is convex-valued and the fact that it is jointly continuous in F_k, F_{-k} so that it is upper hemi-continous. Weak concavity in F_k follows here as it does in Glicksberg because the objective function is linear in F_k - the linearity of the objective in F_{-k} is used by Glicksberg only to establish continuity in F_k, F_{-k} which we have by assumption. $\square$

Continuity

Consider an infinite sequence of conflict resolution models $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L)$, $C_S^N(\varphi_S, F_L)$, $C_L^N(\varphi_L, F_S)$ where $C_k^N(\varphi_k, F_{-k})$ differentiable in φ_k on $(\underline{y}_k, 1]$ with $(1/\epsilon) > C_k^{N\prime}(\varphi) > \epsilon$ for some $\epsilon > 0$ an all-pay auction with costs $C_k(\varphi_k)$ also differentiable on $(\underline{y}_k, 1]$ with $(1/\epsilon) > C_k'(\varphi) > \epsilon$. In addition we assume that in the limiting case $\eta_S\hat{\varphi}_S \neq \eta_L\hat{\varphi}_L$. We say that the sequence of conflict resolution models *converges to the all-pay auction* if for all $\epsilon > 0$ and $\eta_S\varphi_S > \eta_L\varphi_L + \epsilon$ we have $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L) \to 1$ uniformly, and $\eta_S\varphi_S < \eta_L\varphi_L - \epsilon$ implies $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L) \to 0$ uniformly, and $C_k^N(\varphi_k, F_{-k}) \to C_k(\varphi_k)$ uniformly.

Theorem. [9 in text] *Suppose that* $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L), C_S^N(\varphi_S, F_L), C_L^N(\varphi_L, F_S)$ *converge to the all-pay auction* $C_S(\varphi_S), C_L(\varphi_L)$*, that* F_k^N *are equilibria of the conflict resolution models and that* F_k *is the unique equilibrium of the all-pay auction. Then* $F_k^N \to F_k$.

To prove this theorem we use the standard approach to prove that equilibrium correspondences are upper hemi-continuous: we show that for any strategy by party k, the utility received for large N is uniformly close to the utility received in the limit. Hence, any strictly profitable deviation in the limit would also have to be strictly profitable for large N. If the limit of p_k^N was a continuous function this would be completely straightforward: since the convergence of the objective functions is assumed to be uniform and the integrals defining expected utility would converge for any fixed p_k by a standard argument this would give uniform convergence of the objective functions. The complication is that the limit is not a continuous function: it is discontinuous when there is a tie. Suppose, however, that we can show that the equilibria F_k^N had the property that for any ϵ and large enough N there is a uniform bound such that the probability of an ϵ neighborhood is at most $\overline{\mathscr{F}}\epsilon$ - basically that F_k^N is converging to a limit with a bounded continuous density. Then for any choice of φ_{-k} it would be the case that for large enough N the utility of F_k^N would be at most ϵ-different for φ_k outside of an ϵ neighborhood of $(\eta_k/\eta_{-k})\varphi_{-k}$ and that the probability of being in that neighborhood is also of order no more than ϵ. This implies that the utility of F_k^N is of order ϵ different than the utility of F_k^∞, which is what is needed for the standard argument to work. This argument is not completely adequate because there can be two points where there is an atom in the limit, but these cases can be covered by the appropriate choice of tie-breaking rule.

Note that while this is stated as an upper hemi-continuity result, since the equilibrium of the all-pay auction is unique and we have existence it is in fact a continuity result. We can summarize it by saying that if the conflict resolution model is close enough to winner-take-all and pivotality is not important then the equilibria of the conflict resolution model are all close to the unique equilibrium of the all-pay auction.

The theorem allows N to be any abstract index of a sequence. A particularly interesting application is the one in which N represents population size and we consider that the conflict resolution function is binomial arising from independent draws of type by the different voters. In this case an application of Chebychev's inequality gives the needed uniform convergence of $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L)$. The convergence of costs when pivotality is accounted for is shown below.

Proof. Since the space of distributions is compact in the given topology there is a convergent subsequence. Hence it is sufficient to assume $F_k^N \to F_k^\infty$ and show that $F_k^\infty = F_k$. We do this by showing that F_k^∞ is an equilibrium relative to the tie-breaking rule such that the advantaged party wins if there is a tie where the disadvantaged party turns out all voters and the party that can turn out the most committed voters wins when it turns out exactly its committed voters. Since the equilibrium of the all-pay auction is unique, it must then be that F_k^∞ is in fact F_k. Note that it then follows that F_k^∞ is also an equilibrium with respect to only the first half of the tie-breaking rule, since that is the case for F_k.

By assumption for any ϵ^2 the convergence of $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L)$ is uniform on the set of (φ_S, φ_L) with $|\eta_S\varphi_S - \eta_L\varphi_L| \geq \epsilon^2$. Hence we may assume that for any ϵ and for large enough N if $\eta_S\varphi_S > \eta_L\varphi_L + \epsilon^2$ then $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L) > 1 - \epsilon^2$ and $\eta_S\varphi_S < \eta_L\varphi_L - \epsilon^2$ then $p_S^N(\eta_S\varphi_S, \eta_L\varphi_L) < \epsilon^2$.

Let d be the disadvantaged party in the all-pay auction. We observe the obvious fact that F_k^∞ places no weight above $(\eta_d/\eta_k)\hat{\varphi}_d$ nor below $\underline{y}_k$ so we certainly have convergence outside these intervals. It is similarly obvious that for any $\gamma > 0$ there is an N large enough that F_k^N places no weight above $(\eta_d/\eta_k)\hat{\varphi}_d + \gamma$ nor below $\underline{y}_k - \gamma$, so in examining F_k^N we may restrict attention to those intervals. The assumption on the slopes of $C_S^N(\varphi_S, F_L), C_L^N(\varphi_L, F_S)$ implies that there are constants $\infty > \overline{D}, \underline{D} > 0$ such that for any ϵ and κ and arge enough N for $\varphi_k + \kappa\epsilon > \varphi_k' > \varphi_k$

$$C_k^N(\varphi_k', F_{-k}) - C_k^N(\varphi_k, F_{-k}) < \overline{D}\epsilon$$

and for $\varphi_k \geq \underline{y}_k$ and $\varphi_k' > \varphi_k + \epsilon/\kappa$

$$C_k^N(\varphi_k', F_{-k}) - C_k^N(\varphi_k, F_{-k}) > \underline{D}\epsilon$$

Let ℓ denote the party with the largest value of $\eta_k\underline{y}_k$. Consider first the intervals $((\eta_\ell/\eta_k)\underline{y}_\ell, (\eta_d/\eta_k)\hat{\varphi}_d)$. Fix a point $\tilde{\varphi}_S = (\eta_L/\eta_S)\tilde{\varphi}_L$ in this interval where there is a tie and consider an ϵ open square around of this point, $\Phi_S \times \Phi_L$ (we may assume that these open intervals are entirely contained in the set in question by choosing ϵ small enough). Choose $\epsilon \max_k v_k < \overline{D}$ at least. Consider that one of the parties k has no greater than a $1/2$ chance of winning the contest in this interval. If $\mathscr{F}_{-k}$ is the probability F^N_{-k} assigns to Φ_{-k} then, if k shifts any weight in Φ_k to the top of the interval, he gains at least $(1/2)\mathscr{F}_{-k}v_k - \epsilon^2 v_k - \overline{D}\epsilon$, so that if $\mathscr{F}_k > 0$ then $\mathscr{F}_{-k} \leq (2/v_k)(\overline{D} - \epsilon v_k)\epsilon < \overline{\mathscr{F}}\epsilon$ where $\overline{\mathscr{F}} = (2/v_k)\overline{D}$. If $\mathscr{F}_k = 0$ certainly $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$. Hence, we see that there is a constant $\overline{\mathscr{F}}$ such that in each square of the type described we must have $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ for at least one of the two parties k.

Now consider $\mathscr{F}_k$ for the other party for which this bound is not necessarily satisfied, and consider φ_{-k} lying below Φ_{-k}. If such a bound is not satisfied we may assume that $\mathscr{F}_k > 2\epsilon^2$. Shifting to the top of the interval yields a gain by the previous argument of at least $\mathscr{F}_k v_{-k} - \epsilon^2 v_{-k} - \overline{D}(\tilde{\varphi}_{-k} - \varphi_{-k} - \epsilon)$. From this we see that there is another constant $\underline{\epsilon} > 0$ such that party $-k$ places no weight on the interval of φ_{-k} such that $\tilde{\varphi}_{-k} - \epsilon \geq \varphi_{-k} \geq \tilde{\varphi}_{-k} - \epsilon - \underline{\epsilon}\mathscr{F}_k$.. If $\underline{\epsilon}\mathscr{F}_k > \epsilon/\kappa$ then party k shifting all the weight in Φ_k to $\tilde{\varphi}_{-k} - \epsilon - \underline{\epsilon}\mathscr{F}_k$ causes k to gain $\underline{D\epsilon}\mathscr{F}_k - (\overline{\mathscr{F}}\epsilon + \epsilon^2)v_k \leq 0$. Hence, there is a constant $\overline{\mathscr{F}}$ such that $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ for both parties k. If $\hat{\varphi}_d < 1(= \overline{y}_d)$ then exactly the same argument works when we extend the upper limit slightly $((\eta_L/\eta_k)\underline{y}_L, (\eta_d/\eta_k)\hat{\varphi}_d) + \gamma)$, and we already know it is true for $((\eta_d/\eta_k)\hat{\varphi}_d) + \gamma, \eta_d/\eta_k]$ so the bound holds for $((\eta_\ell/\eta_k)\underline{y}_\ell, \eta_d/\eta_k]$.

Suppose instead that one party is turning out all voters at the upper bound: in this case it must be disadvantaged and the other party must be larger. We observe that since for any γ we already know that F^N_{-d} places no weight above $(\eta_d/\eta_{-d}) + \gamma$, so F^∞_{-d} places no weight above η_d/η_{-d}. In the intervals $((\eta_d/\eta_k) - \epsilon, \eta_d/\eta_k]$ if W_{-d} is the probability that $-d$ wins conditional on that interval then the gain to $-d$ by shifting to $(\eta_d/\eta_k) + \epsilon$ is $(1 - W_{-d})v_{-d} - 2\epsilon\overline{D}$. There are two possibilities: either $\mathscr{F}_{-d} = 0$ in which case we have $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ for both parties in $((\eta_\ell/\eta_k)\underline{y}_\ell, 1]$, or the gain is non-negative implying that $W_{-d} \geq 1 - (2\overline{D}/v_{-d})\epsilon = 1 - \underline{W}\epsilon$. Now consider the lower bound $(\eta_\ell/\eta_k)\underline{y}_\ell$. The argument above that one party has to satisfy $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ remains valid since it relies on deviating to the top of the interval and the upper bound on the cost derivative $\overline{D}$ which is globally valid. The party with the most committed voters does not bid below $(\eta_\ell/\eta_k)\underline{y}_\ell - \gamma$ so if it is the party that satisfies $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ then the other party has a chance of winning by bidding below $(\eta_\ell/\eta_k)\underline{y}_\ell$ of at most $2\overline{\mathscr{F}}\gamma$, while if it were to bid $\underline{y}_S$ it would save nearly $\underline{D}((\eta_\ell/\eta_k)\underline{y}_\ell - \underline{y}_S)$ so for small enough γ it would not choose to bid in this interval. Hence we conclude that the party with the least committed voters must satisfy the bound $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ in $(\underline{y}_\ell, (\eta_d/\eta_k)\hat{\varphi}_d)$.

Now we are in a position to consider the sequence of equilibrium expected utilities U_k^N which, if necessary by passing to a subsequence, may be assumed to converge to some U_k^∞. For any ϵ we observe that in the compact region in which the difference between bids is at least ϵ the objective function in the limit is continuous, so in this region the integral defining expected utility converges to the identical value as computed from the limit distributions. In case $\hat{\varphi}_d < 1 (= \overline{y}_d)$ we also see that the region where the difference between bids is smaller than ϵ as that area is made up of at most $1/\epsilon$ squares each with probability no greater than $\overline{\mathscr{F}}^2\epsilon^2$ the probability of that region is at most $\overline{\mathscr{F}}^2\epsilon$ for all large enough N so that this does not matter for computing utility in the limit. Hence in this case $U_k^\infty = U_k$ the utility computed from the limit distributions (and in this case the tie-breaking rule does not matter). In case $\hat{\varphi}_d = 1 (= \overline{y}_d)$ if we exclude a neighborhood of the tie at η_d/η_k again utility converges to the right limit, moreover, we have shown that in the region near the tie in equilibrium (for $N < \infty$) $W_{-d} \geq 1 - \underline{W}\epsilon$ which gives the same result in the limit as $N \to \infty$ as the tie-breaking rule that $-d$ always wins the tie.

Now consider deviations against the limit distributions. For deviations to a point where the bound $\mathscr{F}_k \leq \overline{\mathscr{F}}\epsilon$ is satisfied by the opponent a strict gain with respect to the limit distribution of the opponent translates immediately in the usual way to a strict gain for large N so this is impossible. The same reasoning applies in case $\hat{\varphi}_d = 1 (= \overline{y}_d)$ to deviations by the advantaged party to η_S/η_L since it must win before the limit is reached with probability at least $W_{-d} \geq 1 - \underline{W}\epsilon$.

Finally, in the case $\hat{\varphi}_d = 1 (= \overline{y}_d)$ if is profitable for the disadvantaged party to deviate to 1 since by the tie-breaking rule it loses for sure it could equally well make a strict profit by bidding slightly less than 1. Nor can it be advantageous for the small party to deviate to $(\eta_L/\eta_S)\underline{y}_L$ since by the tie-breaking rule it loses for sure.

Hence we conclude that F_k^∞ is in fact an equilibrium with respect to the proposed tie-breaking rule. $\square$

High Value Elections and the Impact of Electorate Size on Turnout

Now we consider an alternative conceptual experiment: we hold fixed the size of the electorate, but allow the size of the prize to be very large.

Theorem. [10 in text] *Suppose $v_k \to \infty$. Then $F_k(1-\epsilon) \to 0$.*

Proof. Recall that we have assumed that $\overline{y} = 1$. Suppose that with positive probability party k chooses φ smaller than $1-\epsilon$. Because the other party turns out at least $\underline{y}$ and the conflict resolution is fixed and continuous (N is fixed) this implies that the difference in the probability of losing between φ and 1 is bounded away from 0 by $r > 0$ independent of v_k. Hence party k gains at least $rv_k - \max_{\{\varphi_k, F_{-k}\}} C_k^N(\varphi_k, F_{-k})$, which is positive for v_k large enough. $\square$

Pure Strategy Equilibrium

When the conflict resolution function is "nearly" winner-take-all and pivotality is not too important, we are close to the all-pay auction and hence the equilibrium is in mixed strategies. By contrast if the objective functions $p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})v_k - C_k(\varphi_k, F_{-k})$ are single-peaked in $\eta_k\varphi_k$ (for example: p_k is concave and C_k convex, at least one strictly) then a standard argument shows that there is a pure strategy equilibrium. Indeed, when the objective functions are single-peaked there is a unique optimum for each party and so the party cannot mix - all equilibria must be pure strategy equilibria. This is basically the case considered in Coate and Conlin (2004).[48]

It is worth considering what is needed for $p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})$ to be concave in the symmetric case in which $p_L(\eta_L\varphi_L, \eta_S\varphi_S) = p_S(\eta_L\varphi_L, \eta_S\varphi_S)$. Symmetry

[48] However, the model in Coate and Conlin (2004) differs slightly from the one here: in their model the size of the parties is random. Conceptually this poses a problem for a model of peer punishment within parties - it is not clear how a collusive agreement can be reached among a party whose members are not known. Coate and Conlin (2004) also use a different objective function than we do: they assume that the "party" maximizes total utility so that when a random event causes the party to be large the party is "happier" than when it is small. This is not necessary, they could consider (as implicitly we do) a party that maximizes per capita rather than total utility.

implies $p_k(1/2, 1/2) = 1/2$. Concavity implies $p_k(1/4, 1/2) \geq (1/2)p_k(0, 1/2) + (1/2)p_k(1/2, 1/2) \geq 1/4$. In other words, when one party sets a target of a 2-1 majority, it must nonetheless have at least a 25% chance of losing: in this sense concavity means "a great deal" of variance in the outcome. Hence we have the broad picture that when pivotality is not important small aggregate uncertainty leads to non-trivial mixing in equilibrium, while large aggregate uncertainty leads to pure strategies in equilibrium.

Example 1. Suppose that the types y_k have both a common and an idiosyncratic component where the common component may be correlated between the two parties. We have indexed types by a uniform distribution on $[0, 1]$. It is convenient in developing an example with a common component to index types by z_k drawn from continous strictly increasing cdfs $G_k(z)$ on $[0, \infty)$. The original index y_k can then be recovered from the formula $y_k = G_k(z_k)$. In our example the objective function will be concave in z_k - this implies that it is single-peaked in y_k. With the index z_k the party chooses a type threshold ζ_k. We assume that N is large - the proof of a formal convergence result here is straightforward since the limiting conflict resolution function is continuous.

The specific example is defined by a parameter $0 < \beta$. We assume that costs are sufficiently high relative to the prize so that $\hat{\varphi}_k < \beta/(1 + \beta)$. Each voter i in party k takes an iid draw u_i from a uniform distribution on $[0, 1]$. A single independent common draw ν is taken also from a uniform on $[0, 1]$. We set $\nu_S = \nu^{1/\beta}$ and $\nu_L = (1 - \nu)^{1/\beta}$ and a voter's type is $z_{ik} = \beta u_i/(1 + \beta)\nu_k$. We let ζ_k denote the threshold for voting in terms of z_k. Because we are assuming that N is large we ignore pivotality so that it follows that the cost of turning out voters is $C_k(G_k(\zeta_k)) \geq 0$.

Conditional on the common shock ν_k the expected fraction of voters that turns out is $\Pr(z_k \leq \zeta_k|\nu_k) = \Pr(u_i \leq ((1 + \beta)/\beta)\zeta_k\nu_k|\nu_k)$. For $\zeta_k \leq \beta/(1 + \beta)$ this is $\Pr(z_k \leq \zeta_k|\nu_k) = ((1 + \beta)/\beta)\zeta_k\nu_k$ (since the RHS is no greater than 1). Observe first that $\Pr(z_k \leq \zeta_k) = \int((1 + \beta)/\beta)\zeta_k\nu^{1/\beta}d\nu = \zeta_k$ from which we can conclude that for $\zeta_k \leq \beta/(1 + \beta)$ we have $y_k = z_k$. Since it cannot be optimal to choose $\varphi_k > \hat{\varphi}_k$ and $\hat{\varphi}_k \leq \beta/(1 + \beta)$ we see that for $\varphi_k \leq \hat{\varphi}_k$ the expected fraction of voters who turn out conditional on the common shock ν_k

is $((1+\beta)/\beta)\varphi_k\nu_k$.

Because we are assuming that N is large we suppose that the actual fraction of voters who turn out is exactly $((1+\beta)/\beta)\varphi_k\nu_k$. Hence party k wins the election if $\eta_k\varphi_k\nu_k > \eta_{-k}\varphi_{-k}\nu_{-k}$. Taking logs, this reads $\log(\eta_k\varphi_k/(\eta_{-k}\varphi_{-k})) + (1/\beta)(\log(\nu)-\log(1-\nu)) > 0$. Since for a uniform ν on $[0,1]$ the random variable $\log(\nu) - \log(1-\nu)$ follows a logistic distribution the probability of winning is the Tullock contest success function

$$\frac{1}{1+(\eta_{-k}\varphi_{-k}/(\eta_k\varphi_k))^\beta} = \frac{(\eta_k\varphi_k)^\beta}{(\eta_k\varphi_k)^\beta + (\eta_{-k}\varphi_{-k})^\beta}.$$

A sufficient condition for this to be concave is that $\beta \leq 1$ so that if $C_k(\varphi_k)$ is strictly increasing when it is strictly positive, continous and (at least for $\varphi_k \leq \hat{\varphi}_k$) convex then there are only pure strategy equilibria. By contrast as $\beta \to \infty$ the distribution of $\nu^{1/\beta}$ approaches a point mass at 1 and we approach the case of the all-pay auction and for large β there can be no pure strategy equilibrium. However, under the assumption that the cost function satisfies a diminishing hazard rate condition, Herrera, Morelli and Nunnari (2015) show that pure strategy equilibria exist for relatively large β and give a detailed breakdown of the comparative statics of the equilibrium.

Pivotality

Up to this point we have assumed that voters vote only because either they prefer to do so, the commited voters, or because they face punishment if they do not: these motivations are reflected in the incentive constraint $c_k(y) \leq P_k$ that voters should vote provided the cost of doing so is less than or equal to the cost of being punished for not doing so. This formulation ignores the traditional focus in Palfrey and Rosenthal (1985) on the individual incentive to vote based on the chance of being pivotal. We turn now to combining the more traditional incentives of pivotality with the possibility of punishment.

We start by giving a formulation of the contest model that enables us to compute the probability of being pivotal. Rather than a single conflict resolution function we define two partial conflict resolution functions: $P_k^0(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})$, the probability of winning conditional on all voters except one following the social norm φ_k and *the remaining voter not voting*, and $P_k^1(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})$, the probability of winning conditional on all voters except one following the social norm φ_k and *the remaining voter voting*. These should be differentiable and non-decreasing in φ_k. This two functions enable us to compute both the overall conflict resolution function and the probability of being pivotal: the overall conflict resolution function is $p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) = \varphi_k P_k^1(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) + (1 - \varphi_k)P_k^0(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})$ and the probability of being *pivotal* is $Q_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) = P_k^1(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) - P_k^0(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})$. It is convenient in what follows to view the strategies F_k as measures rather than functions.

To analyze incentives with pivotality we start by identifying what and individual voter would like to do in the absence of punishment. This depends on what voters from both parties are doing. For any given social norm φ_k and mixed strategy of the other party F_{-k} we may define the pivotal cutoff $\gamma_k(\varphi_k, F_{-k})$ by the solution to $c_k(\gamma_k) = Q_k(\eta_k\varphi_k, \eta_{-k}F_{-k})v_k$. This represents the type of voter who is indifferent between bearing the cost of voting in order to improve the party's chance of victory and abstaining. Since $c_k(y)$ is differentiable and has a strictly positive derivative the solution is unique and continuous. We can now determine the incentive constraint when there is punishment for not voting. For voters who would not otherwise vote, that is, $y \geq \gamma_k(\varphi_k, F_{-k})$, the incentive constraint is $c_k(y) - Q_k(\eta_k\varphi_k, \eta_{-k}F_{-k})v_k \leq P_k$. This says that the net cost of voting, which is the direct cost c_k minus the benefit because of pivotality $Q_k v_k$, must be less than or equal to the punishment for not voting. Notice that the mixed strategy of the other party F_{-k} appears in the incentive constraint since P_k must be chosen before the realization of φ_{-k} is known.

From the incentive constraint we can derive the monitoring cost for $\varphi_k \geq \gamma_k(\varphi_k, F_{-k})$ as the cost of punishing a non-voter who was not supposed to vote after having received a wrong signal

$$M_k(\varphi_k, F_{-k}) = \psi\theta(1 - \varphi_k)\left(c_k(\varphi_k) - Q_k(\eta_k\varphi_k, \eta_{-k}F_{-k})v_k\right).$$

Notice that if φ_k is pivotal in the sense that $\varphi_k = \gamma_k(\varphi_k, F_{-k})$ then $M_k(\varphi_k, F_{-k}) = 0$ and the function M_k is continuous.

There remains the issue of what happens if the social norm calls for less participation than would be individually optimal in the presence of the pivotality incentive $\varphi_k < \gamma_k(\varphi_k, F_{-k})$. For voters with $\varphi_k < y < \gamma_k(\varphi_k, F_{-k})$ the social norm calls on y to not to vote, but in fact y would like to. This case is not covered by the basic model and there is more than one modeling possibility. One is to assume that there is no cost of getting a voter not to vote, in which case $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ and $M_k(\varphi_k, F_{-k}) = 0$. In this case we may write $M_k(\varphi_k, F_{-k}) = \psi\theta(1-\varphi_k)\max\{0, (c_k(\varphi_k) - Q_k(\eta_k\varphi_k, \eta_{-k}F_{-k})v_k)\}$ which is obviously continuous, although scarcely linear in F_{-k}. However, all that is required for the results that follow is that $M_k(\varphi_k, F_{-k})$ is non-negative for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$.

To summarize: the goal of the party is to maximize per capita utility $p_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})v_k - T_k(\varphi_k) - M_k(\varphi_k, F_{-k})$. We already showed that equilibrium distributions F_S, F_L exist. A useful technical result is that the model chosen of monitoring costs for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ does not matter because in equilibrium neither party ever chooses such a low participation rate.

Lemma 1. *Basic Lemma: If F_S, F_L are equilibrium distributions then $F_k(\varphi_k < \gamma_k(\varphi_k, F_{-k})) = 0$.*

Proof. Start by assuming that $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ and $M_k(\varphi_k, F_{-k}) = 0$, we will show later that this assumption does not matter. Then the objective function is

$$U_k = \left(\varphi_k P_k^1(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) + (1-\varphi_k)P_k^0(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})\right)v_k - T_k(\varphi_k)$$

and differentiating with respect to φ_k we get

$$\frac{dU_k}{d\varphi_k} = -c(\varphi_k) +$$

$$+\left(Q_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k}) + \varphi_k\eta_k\frac{dP_k^1(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})}{d\eta_k\varphi_k} + (1-\varphi_k)\eta_k\frac{dP_k^0(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})}{d\eta_k\varphi_k}\right)v_k.$$

Since $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ we have $Q_k(\eta_k\varphi_k, \eta_{-k}\varphi_{-k})v_k = c_k(\gamma_k(\varphi_k, F_{-k})) > c_k(\varphi_k)$ so that $dU_k/d\varphi_k > 0$, so that certainly F_k puts no weight on a neighborhood of φ_k. Now we drop the assumption that for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ we have $M_k(\varphi_k, F_{-k}) = 0$. Notice that we may increase φ_k until the first time that $\tilde{\varphi}_k = \gamma_k(\tilde{\varphi}_k, F_{-k})$ is satisfied, and it follows that $U_k(\tilde{\varphi}_k) - M_k(\tilde{\varphi}_k, F_{-k}) > U_k(\varphi_k) - M_k(\varphi_k, F_{-k})$ since the derivative is strictly positive. But $M_k(\tilde{\varphi}_k, F_{-k}) = 0$ so φ_k is strictly worse than $\tilde{\varphi}_k$. $\square$

One implication of the final step of the proof of the lemma is that the set of equilibria for any definition of $M_k(\varphi_k, F_{-k}) \geq 0$ for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ contains the equilibria for the corresponding model with $M_k(\varphi_k, F_{-k}) = 0$ for $\varphi_k < \gamma_k(\varphi_k, F_{-k})$ - and in particular since equilibria in the former model exist, they also exist in the latter.

Our main interest is in what equilibrium with pivotality looks like. There are two limits of interest. The first is the large population case which has been our focus. In the case of iid draws of y as $N \to \infty$ the fact that $\underline{y}_k > 0$ forces the probability of being pivotal to converge to zero uniformly. It follows that $M_k(\varphi_k, F_{-k})$ converges uniformly to $M_k(\varphi_k) = \psi\theta(1-\varphi_k)c_k(\varphi_k)$ and so Theorem 9 tell us that we approach the unique all-pay auction equilibrium: as we expect, pivotality is not important when the electorate is large.

The second limit of interest is the case of large monitoring costs - in this case we expect pivotality to play the decisive role. Specifically we would like to show that as $\psi \to \infty$ we approach a correlated equilibrium of the purely pivotal model. However, this need not be the case - there is no cost of monitoring at $\varphi_k = 1$ and we have assumed $\overline{y} = 1$, so the very high costs of monitoring can potentially be avoided by choosing a very high participation rate. Suppose in particular that $C_k(1) < v_k$ so that it would pay to turn out the entire electorate without monitoring cost. Then we cannot rule out the possibility that even for very large ψ equilibrium might involve participation rates close to 1 and very much higher than the pivotal cutoff $\gamma_k(\varphi_k, F^{\psi}_{-k})$.

By contrast, suppose that $T_k(1) > v_k$. It follows directly from Lemma 1 that

Theorem. [11 in text] *If $T_k(1) > v_k$ then as $\psi \to \infty$ we have $F^{\psi}_k(|\varphi_k - \gamma_k(\varphi_k, F^{\psi}_{-k})| \leq \epsilon) \to 1$.*

Notice that this does not necessarily imply that the limit is an equilibrium in the sense of Palfrey and Rosenthal (1985) since we allow correlation devices within parties, but rather a correlated equilibrium with pivotality of the type studied by Pogorelskiy (2015).

References

Glicksberg, I. (1952): "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points," *Proceedings of the American Mathematical Society* 3: 170-174.

Moe, T. M. (2003): "Teacher Unions and School Board Elections," Hoover Institute.

Moe, T. M. (2006): "Political Control and the Power of the Agent," *Journal of Law, Economics, and Organization* 22: 1-29.