Voter Turnout with Peer Punishment☆

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

Abstract

We introduce a model where social norms of voting participation are strategically chosen by competing political parties and determine voters' turnout. Social norms must be enforced through costly peer monitoring and punishment. 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. Our model shares features of the ethical voter model and it delivers novel and empirically relevant comparative statics results.

JEL Classification Numbers: later

Keywords: Voting, Turnout, Party, Peer Punishment

☆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, the editor and four referees. Giovanni Andreottola provided excellent research 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

In this paper we study voter turnout in two-party elections and investigate the theoretical relation between party size and electoral advantage. We do this in the context of a novel theoretical model that formalizes a well-documented empirical fact: peer pressure plays a key role in enforcing social norms of voting participation.[3] Social norms are typically maintained by various forms of social disapproval and ostracism (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.[4]

We model voter participation by viewing a political party as facing a public goods problem: individual voters bear the cost of participation, but the benefit of an electoral victory is shared with all party members. Our core assumption is that each party is able to design a mechanism to overcome this public goods problem by providing incentives to individual voters in the form of social pressure. The optimal solution to this mechanism design problem is to excuse high cost voters from voting. How costly it is to do so depends on how difficult it is to monitor individual voter costs. This difficulty plays a key role in determining the outcome of elections: transparent costs favor a large party. By contrast, we show that difficulty in monitoring voter costs may favor a small party.

In existing turnout models a large party has a natural advantage. Since the marginal cost of turning out an additional voter is increasing in the fraction of party members who turned out, a large party can turn out the same number of

[3] We discuss this literature in Section 6.2.

[4] The mailing contained the following statement: "You are receiving this election notice because of low expected voter turnout in your area. Your individual voting history as well as your neighbors' are public record. Their scores are published below, and many of them will see your score as well. Caucus on Monday to improve your score and please encourage your neighbors to caucus as well."

voters as a small party at lower cost. While in these models a small party makes a stronger effort than a large party, this is not enough to overcome the natural advantage of the large party. In our model this is likewise the case when voter costs are transparent. When this is not the case, however, some non-voters must be punished, and for any given number of voters turned out, a small party has fewer non-voters to punish. This can lead to an overall advantage for the small party when its value of winning the election is intermediate and committed voters are few.

Our mechanism design approach in which social norms are chosen to maximize a group objective can also be found in Townsend (1994) and in the ethical voter model of Feddersen and Sandroni (2006) and Coate and Conlin (2004).[5] From an empirical perspective, Coleman (1988) and Ostrom (1990) as well as Olson (1965) all provide evidence that - within the limits of available monitoring and punishment - peer pressure mechanisms do a good job of solving public goods problems. Palfrey and Pogorelskiy (2016) provide experimental evidence that communication among voters and in particular communication within parties increases turnout - that is, enables them to attain more advantageous social norms.

Our model delivers a rich set of intuitive comparative statics on equilibrium turnout with respect to the value of the election, its closeness, and to changes in the relative size of the groups. We clarify which results carry over from existing models and which do not. Furthermore, with our model we are able to address questions that could not be asked with existing turnout models. In particular, we examine the consequences of changes in difficulty of monitoring voter costs. We observe, for example, that rule changes that lower turnout costs may also make monitoring more difficult and so have the perverse effect of lowering turnout. This prediction is in line with empirical evidence on the effect of postal voting on turnout in Swiss elections (Funk (2010)) and on the effect of the monitoring capacity of parties' local mobilizers on turnout buying in Mexico (Larreguy, Marshal and Querubin (2016)).[6] We observe, as well,

[5]For the relation between our model and the ethical voter model see sections 4 and 6.

[6]Our predictions are also consistent with the failure of the 2000 UK policy experiment

that in a high stakes election a small increase in party size that gives a party a slight majority will result in a discontinuous upward jump in utility. This suggests that competition over platforms may be fierce indeed.

2. The Model

A continuum of voters is divided into two parties $k = \{S, L\}$ denoting *Small* and *Large*, respectively. The fraction of the voting population belonging to party k is η_k where $0 < \eta_S < \eta_L$ and $\eta_S + \eta_L = 1$. The two parties compete in an election for a common prize worth V to the party that produces the greatest number of votes and V/η_k to each member of party k. We assume, in other words, that the prize is fungible - such as taxes, subsidies, or government jobs - so that the collective value of the prize to a party does not depend on the size of the party.[7]

Voting is costly and individuals face the same distribution of voting costs independent of party. Each voter privately and independently draws a type y from a uniform distribution on $[0, 1]$, which determines a *net participation cost* of voting $c(y)$. This cost of voting is net of social pressure: it consists of the direct cost and inconvenience (costs of time and transportation) minus the direct personal benefits such as fulfilling civic duty or expressive voting. The participation cost of voting c is continuously differentiable, strictly increasing and satisfies $c(\underline{y}) = 0$ for some $\underline{y} \in [0, 1]$. Voters for whom $y < \underline{y}$ have a negative net cost of voting, they are called *committed voters* and will always vote.

Our main innovation and core assumption is that each party is able to design a mechanism providing incentives to individual voters in the form of social pressure. We model this as an ability to impose penalties that are costly to individual party members. The mechanism has two parts. The first part is a threshold φ_k together with a rule for party members prescribing voting if

of setting in-store poll booths to reduce voting costs and hence increase turnout. See https://www.theguardian.com/uk/2000/aug/17/juliahartleybrewer.

[7]A party is generally made up of individuals with different interests: we do not attempt to model the internal decision making of the party, but simply note that in practice V represents a composite of those interests. We discuss the assumption that V is independent of the size of the party in Section 6.

$y \leq \varphi_k$. That is all, voters with sufficiently low cost of voting are expected to vote. We refer to this as the *social norm* for the party. Notice that φ_k is the probability that a representative party member votes and, since there are a continuum of voters, φ_k is also the turnout rate of the party. This first part of the mechanism is already present in Feddersen and Sandroni (2006). The second part of the mechanism is novel and is a punishment $P_k \geq 0$ representing social disapproval for failing to comply with the social norm.

The ability to apply punishments is limited by imperfect information. While there is no difficulty in determining whether or not a party member voted, individual costs of voting are not transparent and the party can only observe a noisy binary signal about whether or not a non-voter followed the social norm. In particular, among voters that failed to vote there are two types: high cost voters who were "excused" according to the social norm and low cost voters who were not. If the non-voting member violated the social norm, that is, $y < \varphi_k$, a negative signal is received with probability $\pi_0 \in [0, 1]$ and a positive signal with corresponding probability $1 - \pi_0$. If the non-voting member did not violate the social norm, that is, $y > \varphi_k$, then a negative signal is received with a lesser probability $\pi_1 \in [0, \pi_0]$. A positive signal should be thought of as "producing a good excuse for not voting" so that a non-voter who is genuinely excused is more likely to be able to produce a good excuse than one who is not.[8]

The outcome of the election is determined by the fraction of the electorate $b_k = \eta_k \varphi_k$ that each party turns out, and sometimes we will refer to this as the *bid* of party k. The party that turns out more of its members wins. In case of a tie we assume that the large party wins.[9] This is similar to an all-pay auction:

[8] At the extremes, costs are either public information, so that $\pi_1 = 0$ and $\pi_0 = 1$, or costs are private information and the signal is uninformative, so that $\pi_0 = \pi_1$. Notice that in this latter case the party can provide incentive for voting only by punishing all non-voters.

[9] If we assume that in case of a tie each party has an equal chance of winning an equilibrium may fail to exist for an uninteresting reason. As we shall see there can be equilibria in which the large party with positive probability bids preemptively - and this is the only case in which the tie-breaking rule matters. The large party bids preemptively when it turns out more voters than there are in the small party. However, it would always benefit from mobilizing slightly fewer voters. Hence we must allow the large party to bid preemptively by mobilizing the exact number of voters in the small party - meaning if there is a tie it must win. An alternative approach is to follow Simon and Zame (1990) and allow the tie-breaking rule to

the party that "submits the highest bid" wins, but each party pays the cost for their bid. If a party bids only its committed voters it bids $\underline{b}_k \equiv \eta_k \underline{y}$ at zero cost and if it bids all its voters it bids η_k.

We assume that the mechanism is designed - either by benevolent party leaders or by consensus of the party - to maximize the common *ex ante* utility of party members. Let $\Pi_k(\varphi_k)$ be the probability of winning the election as a function of party turnout. If all group members adhere to the social norm party utility is given by the expected benefit of winning $\Pi_k(\varphi_k)V$ minus the direct cost of participation $\eta_k \int_0^{\varphi_k} c(y)dy$ *and* the cost of punishing "excused" party members $\eta_k \int_{\varphi_k}^1 \pi_1 P_k dy$. This latter cost arises precisely because the signal is noisy. The overall utility of the party is therefore $U_k \equiv \Pi_k(\varphi_k)V - \eta_k \left[\int_0^{\varphi_k} c(y)dy + \int_{\varphi_k}^1 \pi_1 P_k dy\right]$. The two parties simultaneously choose their mechanisms in an effort to maximize party utility. This is a game between mechanisms in the sense of Myerson (1982) or Dutta, Levine and Modica (2016a). An equilibrium consists of probability distributions over mechanisms for both parties such that, given the mechanism of the other party, each party finds its own mechanism optimal.

3. The Optimal Punishment

In choosing a mechanism $\{\varphi_k, P_k\}$ the party must ensure that party members are willing to comply with the social norm. That is, the maximization of U_k is subject to incentive constraints. In particular the members who are "expected" to vote, that is, $y \leq \varphi_k$ must be willing to do so. As voters are infinitesimal, they receive no direct benefit from the public good so if they vote they simply bear the net cost $c(y)$. If they choose not to vote they avoid the cost, but instead face an expected cost of punishment equal to the probability of being "caught without a good excuse" times the punishment $\pi_0 P_k$. Hence the incentive constraints are that for $y \leq \varphi_k$ we have $c(y) \leq \pi_0 P_k$. In addition, the punishment cannot be set so high that voters who are not expected to vote are tempted to do so. That is, there are additional incentive constraints for these voters that the cost of

be endogenous: this leads to the same equilibrium described here.

voting should be greater than the expected punishment, that is, for $y > \varphi_k$ we have $c(y) \geq \pi_1 P_k$.

A mechanism that maximizes utility must minimize the monitoring cost $\int_{\varphi_k}^1 \pi_1 P_k dy$ for any social norm φ_k. This means that P_k should be chosen as small as possible provided that incentive constraints are not violated. Clearly, it cannot be optimal for the party to choose $\varphi_k < \underline{y}$. If instead $\varphi_k = \underline{y}$ and just committed voters are asked to vote, incentive compatibility is not an issue and P_k should be taken equal to zero. It is only when $\varphi_k > \underline{y}$ that incentives must be provided to members who find voting costly. As $c(y)$ is increasing we see that if the incentive constraint is satisfied by the marginal voter $y = \varphi_k$ then it is satisfied for all $y \leq \varphi_k$. Hence the relevant constraint is $c(\varphi_k) \leq \pi_0 P_k$ and the punishment P_k is minimized when $P_k = c(\varphi_k)/\pi_0$. Notice that when this punishment is chosen the incentive constraints for $y > \varphi_k$ do not bind. If we define the monitoring difficulty of voting as $\theta = \pi_1/\pi_0 \in [0, 1]$ then an optimal mechanism has least monitoring cost equal to $\int_{\varphi_k}^1 \pi_1 P_k dy = \theta(1 - \varphi_k)c(\varphi_k) \equiv M(\varphi_k)$ which we refer to for simplicity as the *monitoring cost.*

Recall that in addition to the monitoring cost there is the direct cost of participation. This has two parts: $\int_0^{\underline{y}} c(y)dy$ that is a negative constant representing a benefit that we may ignore and $\int_{\underline{y}}^{\varphi_k} c(y)dy \equiv T(\varphi_k)$, which we refer to as the *turnout cost* and represents the added cost of turning out voters beyond those who are committed. Hence, for $\varphi_k \geq \underline{y}$ we define the total expected cost of turning out voters to be the sum of the turnout and monitoring cost $C(\varphi_k) = T(\varphi_k) + M(\varphi_k)$ which is increasing in φ_k. For $\varphi_k < \underline{y}$ it is convenient to define $M(\varphi_k) = 0$ since in any case it is never optimal to choose $\varphi_k < \underline{y}$. With these conventions party utility can be rewritten as

$$U_k \equiv \Pi_k(\varphi_k)V - \eta_k C(\varphi_k) - \eta_k \int_0^{\underline{y}} c(y)dy =$$

$$\Pi_k(\varphi_k)V - \eta_k \left[\int_{\underline{y}}^{\varphi_k} c(y)dy + \theta(1 - \varphi_k)c(\varphi_k) \right] - \eta_k \int_0^{\underline{y}} c(y)dy.$$

4. Main Results

What is gained by adding monitoring to an otherwise standard group-turnout model? A large group has a natural advantage since, for a given social norm, it can turn more voters out. In particular, to turn out a fraction b of the electorate, the large party can choose a smaller social norm $\varphi_L = b/\eta_L$ than the small party $\varphi_S = b/\eta_S$. Hence, focusing only on turnout costs, the large party has a cost advantage in turning out voters since the marginal cost of turnout $T'(\varphi) = c(\varphi)$ is increasing. By contrast, the large party faces a disadvantage in monitoring cost since it will need to monitor and punish a greater proportion of non-voters. To determine the combined effect of turnout and monitoring costs consider the total expected cost of mobilizing a fraction b of the electorate, which equals

$$\eta_k C(b/\eta_k) = b\frac{C(b/\eta_k)}{(b/\eta_k)} = bAC(b/\eta_k)$$

where $AC(\varphi_k)$ denotes the average cost. We see immediately that the party with the lower average cost will have a cost advantage. If $AC(\varphi_k)$ is declining in φ this will be the small party while if $AC(\varphi_k)$ is increasing it will be the large party. Alternatively, we can think in terms of the concavity of the expected cost function $C(\varphi_k)$. If $C(\varphi_k)$ is convex, then $AC(\varphi_k)$ is increasing and the large party has an advantage. If it is concave, then $AC(\varphi_k)$ is declining and the small party has an advantage.

When monitoring is perfect, i.e. $\theta = 0$, we are in the existing world of the ethical voter model. In this case, $AC(\varphi_k)$ is increasing and the large party always has an advantage. By contrast, if $\theta > 0$, the monitoring cost $M(\varphi_k) = \theta(1 - \varphi_k)c(\varphi_k)$ is non-negative, takes on strictly positive values, yet at the endpoints is equal to zero. In particular, when only committed voters vote, no monitoring is needed, while on the other hand if everyone votes there is nobody to punish.[10] Because the monitoring cost cannot be convex, it might be the case that average costs are decreasing giving the small party a cost advantage.

[10] Notice that this property of monitoring cost is robust to the details of the particular monitoring process.

4.1. Equilibrium

In order to characterize the equilibrium, we should first determine the highest fraction of the electorate $\bar{b}_k$ a party is willing to turn out. At this upper bound, either the party reached full turnout or the utility from winning the election is equal to 0. That is, if $\eta_k C(1) < V$ then the party is willing to turn out all its voters and $\bar{b}_k = \eta_k$. If instead $\eta_k C(1) \geq V$, then $\bar{b}_k$ is the unique solution to $\eta_k C(\bar{b}_k/\eta_k) = V$. We refer to $\bar{b}_k$ as the *willingness to bid*.

We say that the party with the smaller willingness to bid is *disadvantaged*, denoted by the subscript d, and the party with the higher value is *advantaged*, denoted by $-d$. Except where explicitly stated we also assume that the small party is willing to turn out at least the number of committed voters of the large party, that is, $\bar{b}_S > \eta_L \underline{y}$.[11] In the next theorem we characterize payoffs in the unique equilibrium.

Theorem 1. *There is a unique equilibrium. In this equilibrium neither party uses a pure strategy, the utility of the disadvantaged party is* 0 *and the utility of the advantaged party is* $V - \eta_{-d} C(\bar{b}_d/\eta_{-d})$.

While a complete proof of all of our theorems, including a characterization of the equilibrium strategies, can be found in the Online Appendix, the intuition for the result is fairly straightforward and follows from basic properties of auction theory. In a second price auction the disadvantaged party loses the auction and gets 0 while the advantaged party gets the difference between the value of the prize and the cost of matching the bid of the disadvantaged party. We know that this result continues to hold for the all-pay auction - although the equilibrium strategies are mixed rather than pure. We can now move to one of the main results of our paper: what determines party advantage.

4.2. Which party is advantaged?

Intuitively, the large party having a large number of committed voters is naturally advantaged. On the other hand, we also know that increasing average costs of turning out supporters favors the small party. The expected cost

[11] When $\bar{b}_S < \eta_L \underline{y}$ 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.

function $C(\varphi)$, however, is not a fundamental of our model: it depends on the distribution of costs in the population and on the monitoring difficulty θ. We aim here to establish how these economic fundamentals interact to determine party advantage. Specifically, we will establish that, regardless of the distribution of costs, for the small party to be advantaged three conditions must be satisfied: *i)* monitoring costs must be high; *ii)* the small party must not be "too small"; *iii)* the value of winning the election must be of intermediate size.

Theorem 2. *For* any *individual cost function $c(y)$ with corresponding committed voters $\underline{y}$ there exist four constants $\underline{\theta}^S < 1, \underline{\eta}_S < 1/2$ and $\overline{V} > \underline{V}^S > 0$ such that if all the conditions $\theta > \underline{\theta}^S, \eta_S \geq \underline{\eta}_S$ and $\underline{V}^S < V < \overline{V}$ hold the small party is advantaged. Conversely if $\underline{y} > 0$ for any values of the other parameters there exist three constants $\underline{\theta}^L > 0, \overline{\eta}_S > 0$ and $\underline{V}^L > 0$ such that if any of the conditions $\theta < \underline{\theta}^L$, $\eta_S < \overline{\eta}_S$, $V < \underline{V}^L$ or $V > \overline{V}$ are satisfied then the large group is advantaged.*

Notice that the reason for the intermediate value of the prize is very intuitive: if the prize is small parties turnout will be low and committed voters will play a disproportionate role favoring the large party. If the value of the prize is high parties will be willing to turn out many voters and the large party has more voters. It is only in the intermediate case that the small party may be advantaged.[12] Notice that the theorem has the following implication: if monitoring costs are high, the two parties are relatively close in size and willing to turn out most but not all of their voters then it is the small and not the large party that is advantaged. Furthermore, if any one of these conditions fails sufficiently badly the large party is advantaged. In particular if $\theta = 0$, that is monitoring is perfect and we are in the existing world of the ethical voter model with a common prize for each group, the smaller party can never be advantaged.[13]

[12]The theorem provides sufficient conditions for party advantage and does not cover all parameter configurations. Indeed, there are intermediate cases in which the identity of the advantaged party depends on the shape of the cost function $c(y)$.

[13]In the case of a common per capita prize for each group, as in Feddersen and Sandroni (2006), the large group is always advantaged regardless of monitoring costs. We discuss this point and different assumptions about a common prize in Section 6.

4.3. Who wins?

Advantage is defined in terms of willingness to turn out supporters. From Theorem 1 we know that this is the same as a utility advantage: the advantaged party receives a positive utility and the disadvantaged party receives no utility. Does it translate also into an advantage in terms of winning the election? To what extent does the advantaged party turn out more voters and have a better chance of winning? As turnout is stochastic for both parties, a natural measure is first order stochastic dominance (FOSD). If the equilibrium bidding function of one party FOSD that of the other then it has a higher chance of winning the election and in a strong sense it turns out more voters.

Party advantage, as we shall see, is not enough to guarantee FOSD. Hence we introduce the notion of *strong advantage.* Convexity of $C(\varphi)$ is a simple sufficient condition for large party advantage and it is natural to view this as a strong advantage. On the other hand, in the presence of committed voters $C(\varphi)$ cannot be concave, so for the small party we introduce the weaker notion of incremental concavity - that $C(\varphi)$ be concave for $\varphi \geq \underline{y}$. We define strong advantage for the small party as the combination of small party advantage (that is a larger willingness to turn out supporters) and incremental concavity.[14] Equipped with this definition we have the following result:

Theorem 3. *The equilibrium bidding function of a strongly advantaged party FOSD that of the disadvantaged party.*

To relate strong party advantage with the distribution of costs in the population and the monitoring difficulty θ, we denote by $G(c)$ the cdf of costs for an individual so that the net participation cost of voting can be expressed as $c(\varphi) = G^{-1}(\varphi)$, $\varphi = G(c)$ and the support is $[c(0), c(1)]$. We denote the density of $G(c)$ by $g(c)$, and we assume it is continuously differentiable, strictly positive, and has a single "top" in the sense that it is either single peaked or a it is a limiting case such as the uniform where the density is flat at the top. The key

[14]We discuss the case of incrementally concave costs and an advantaged large party in the Online Appendix. Indeed, with incrementally concave costs and an advantaged large party, it might be the case that the small disadvantaged party turns out more members in expectation and has a higher probability of winning than the large advantaged party.

determinant of strong advantage is how many relatively low cost and relatively high cost voters there are. In order to show this formally we introduce two measures based on the density of relatively low cost voters $\underline{\gamma}$ and of relatively high cost voters $\overline{\gamma}$. First define

$$\mu(c) = \frac{(g(c))^2}{2\,(g(c))^2 + (1 - G(c))\,g'(c)},$$

then $\underline{\gamma} = \min_{c \geq 0} \mu(c)$ and $\overline{\gamma} = \max\{0, 1 - \max_{c \geq 0} \mu(c)|\}$. In the Online Appendix we prove the following result.

Proposition 1. *The measures* $\underline{\gamma}$ and $\overline{\gamma}$ *satisfy* $0 \leq \underline{\gamma}, \overline{\gamma} \leq 1/2$. *If* $g(c)$ *is weakly decreasing then* $\underline{\gamma} = 1/2$. *If* $g(c)$ *is weakly increasing then* $\overline{\gamma} = 1/2$. *If the density* $g(c)$ *shifts to the right then* $\overline{\gamma}$ *is constant and* $\underline{\gamma}$ *decreases; if the density shifts to the right holding fixed* $c(1)$ *then* $\overline{\gamma}$ *increases. Furthermore, increasing dispersion by a change of scale around the mode increases both* $\underline{\gamma}$ *and* $\overline{\gamma}$.

The proposition asserts that if there are many relatively low cost voters then $\underline{\gamma}$ is large and if there are many relatively high cost voters then $\overline{\gamma}$ is large. If there are many of both type then both may be large: in particular for the uniform $\underline{\gamma} = \overline{\gamma} = 1/2$. Intuitively we expect that having many relatively low cost voters, that is high $\underline{\gamma}$, is similar to having many committed voters and so it should favor the large party. The next theorem makes this precise and also shows that, conversely, having many relatively high cost voters, that is high $\overline{\gamma}$ favors the small party.

Theorem 4. *The large party is strongly advantaged if and only if* $\theta < \underline{\gamma}$. *Cost is incrementally concave (a necessary condition for small party strong advantage) if and only if* $\theta > 1 - \overline{\gamma}$.

In particular a necessary condition for the large party to be strongly advantaged is $\theta < 1/2$ and similarly $\theta > 1/2$ is necessary for the small party to be strongly advantaged. These conditions are also sufficient in the uniform case. More broadly for a downward sloping density $\theta < 1/2$ is sufficient for the large party to be strongly advantaged and for a upward sloping density $\theta > 1/2$ is necessary for the small party to be strongly advantaged.[15] When there is a single peak in the interior, increasing dispersion by raising both $\underline{\gamma}$ and $\overline{\gamma}$ favors whichever party has the monitoring cost advantage. That is, if $\theta < 1/2$ it favors the large party and if $\theta > 1/2$ it favors the small party.

Discussion. While our results are based on the neutral assumption that costs are the same for both parties this is not essential. With differential costs the all-pay auction is still equivalent to the second price auction and anything that lowers a party's costs are to their advantage.[16] In particular: small party advantage rests not on high monitoring costs, but on high monitoring costs for the large party - if the small party has lower monitoring costs this is also to its advantage. So, for example, a rural minority may have an advantage because urban voters have high monitoring costs although the rural voters have low monitoring costs.

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. One case where we have data on many similar elections is that of teacher unions capturing school boards. These have been studied by Moe (2003, 2006) who indicates that these elections are often the only ballot issue and that the unions - the small party - are consistently successful at defeating the parents - the large party. Since the stakes are control over budgetary resources the common prize model is not unreasonable and turnout is low indicating that civic duty is probably not an important reason for voting. Although other explanations are possible it seems likely that the fact that the interested voters in the large party (the parents) are a scattered fraction of the overall population makes monitoring difficult. That is, these elections seem likely to satisfy our conditions for small party advantage.

There is also a strategic lesson here for small parties. Consider a fixed cost per voter of turning out: for example, the cost of busing voters to the polls or a voter ID law. An increase in the fixed cost shifts the distribution of voting costs

[15] Recall that additional conditions are required for small party advantage: the small party should not be too small and the prize must be of intermediate value.

[16] Our earlier working paper, which is available upon request, analyzes this case.

to the right, raising $\underline{\gamma}$ and leaving $\overline{\gamma}$ unchanged. It also decreases the number of committed voters $\underline{y}$. The former decreases the chances of large party strong advantage, and the latter increases the chances of small party advantage: that is, higher fixed costs for both parties favor the small party.

It is natural to try to raise costs for the other party. What this analysis shows is that it is enough for the small party to raise the fixed costs of voting - make it more difficult and unpleasant - for everyone. For this to work, however, two other things must be true: the stakes must be sufficiently low and monitoring costs sufficiently high.[17] On the other hand, if the stakes become large enough this policy can fail catastrophically: a small change from $V < \overline{V}$ to $V > \overline{V}$ will abruptly shift party advantage from the small to the large.

5. Comparative Statics

We will now investigate the effects on turnout and closeness of elections of three important variables: the value of election, the relative size of parties and the efficiency of the monitoring technology.

5.1. The value of elections

If $V > \eta_L C(\eta_S/\eta_L)$ the large party is willing to outnumber the entire small party: in this case we say the election is a *high value election.* This is a natural model when the stakes are high such as elections for national leader or important referenda such as Brexit. We will say that the small party *concedes the election* if it turns out only its committed voters. Furthermore, we say that the large party *preempts the election* if it turns out the most voters feasible for the small party, that is η_S. The next result shows that while both parties are willing to turn out all their voters, this will not occur in equilibrium.

Theorem 5. *In a high value election the probabilities that the small party concedes and the large party preempts the election increase in V, and approach* 1

[17]We argue below in section 5.2 that over time monitoring costs have probably increased. If so it would pay small parties to try to increase participation costs, and this may explain why the small party in the U.S., the Republican party, has increasingly engaged in efforts to raise costs through voter ID laws and the like.

in the limit. As V increases the bid distribution of the small party declines in FOSD and the bid distribution of the large party increases in FOSD. The expected vote differential increases in V while the 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*, which is standard in all pay auctions. 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. 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.

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 nor the probability of the large party taking the election approach 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 preempt 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.

It is interesting to contrast a *low value election* in which $V < \eta_S C(\eta_L \underline{y}/\eta_S)$ and only committed voters turn out with a high value election. As Theorem 5 shows in a high value election election both parties turn out more than their committed voters with positive probability and the large party does so with probability one and indeed has positive probability of turning out as many voters as there are in the small party. Hence turnout is substantially higher in a high value than 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.[18]

5.2. Monitoring difficulty in high value elections

Finally, we turn to monitoring difficulty θ. Here we focus on the important case of a high value election. Our intuition is that increasing monitoring difficulty should decrease turnout. The following theorem shows that if the small party is neither too large nor too small this is true and that in addition elections are closer.

Theorem 6. *In a high value election, an increase in monitoring difficulty θ decreases the turnout of the advantaged (large) party in terms of FOSD. 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 (small) party decreases in monitoring difficulty in terms of FOSD while the expected vote differential also decreases.*

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. That increased monitoring difficulty decreases turnout may also help to explain why measures designed to increase turnout by lowering participation costs may actually have the perverse effect of decreasing turnout because they also raise monitoring costs. Voting at a polling place is a relatively visible and easy to monitor activity. Voting by post, internet, or indeed in the supermarket is not so much so. Hence lowering the inconvenience of voting by allowing it to take place away from the polling place is an example of a reform that may have the perverse effect of reducing turnout. There is evidence that this is indeed the case: Funk (2010) shows this was the case when postal voting was introduced in Switzerland, and the 2000 UK policy experiment of setting in-store poll booths also failed to increase turnout.

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

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.[19] 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.[20]

6. Related Literature and Discussion of the Model

In our model there is no room for pivotality of the individual voter to play a role. Although they do well in the laboratory for small elections (see for example Levine and Palfrey (2007)) pivotal voters models such Palfrey and Rosenthal (1985) have difficulty in explaining turnout in mass elections.[21] Consequently attention has turned to follow-the-leader models such as Shachar and Nalebuff (1999) or models of ethical voters such as Feddersen and Sandroni (2006) and Coate and Conlin (2004).[22] While Shachar and Nalebuff (1999) focus on the

[19]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).

[20]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.

[21]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 that 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).

[22]Herrera, Morelli and Nunnari (2015) examine all three models. 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. Morris and Shadmer (2017) study a model where "voters" with heterogeneous beliefs are encouraged to participate by a "party" that provides incentives. In their model incentives

costs to the leaders, we follow the ethical voter literature in focusing on the costs to the followers. Our model can be interpreted as an ethical voter model in which in addition to voting out of a sense of civic duty, voters also engage in peer pressure out of a sense of civic duty. As a result, when in the mechanism design problem monitoring cost is zero, the solution is the same as that in the ethical voter model.

In the remainder of this section we examine the four key elements of our model: symmetry of the fundamentals, the use of peer punishment to enforce an endogenous social norm, the absence of exogenous uncertainty, and the fact that costs are bounded.

6.1. Symmetry of the fundamentals

As a benchmark we have assumed symmetry in voting costs between the two parties and in that they compete for a common prize worth the same to each party. This is a useful benchmark model: voters in both parties face identical *ex ante* participation costs and a common prize is efficiency neutral. Naturally if one side has a cost advantage or values the prize more highly than the other party it will be more able and willing to turn out voters and this will give it an electoral advantage.

A common prize makes sense when the outcome of the election are taxes, subsidies and other transfer payments. By contrast, if civil rights and law changes are at stake it may make sense to assume that the benefit of winning is the same for all members of a party. It is less certain in this case that the benefit should be the same for both parties.

In the literature, assumptions about a common prize vary. Shachar and Nalebuff (1999) and Herrera, Morelli and Nunnari (2015) who are interested in the same turnout issues we are assume a common prize. Coate and Conlin (2004) assume that the per capita value of the prize is independent of the size of the party, but may differ between the two parties: this is appropriate in their setting of liquor referenda. Feddersen and Sandroni (2006) assume the

have an opposite effect on optimistic and pessimistic voters, while our incentives encourage all voters.

per capita value of the prize is the same for everyone. Finally, Palfrey and Rosenthal (1985) and Levine and Palfrey (2007) also assume the per capita value of the prize is the same for everyone - but study a pivotal voter model in which the aggregate size of the prize to the party is of less importance.

The case of a common per capita prize make a useful contrast to that of a common prize: this provides an additional advantage to a large party which has more members to enjoy the per capita value. Indeed it provides enough advantage to the large party that it is advantaged in our model regardless of monitoring costs.[23] In all likelihood reality lies in between a common prize and a per capita prize: typically elections involve a mix of issues, some involving taxes and transfers, other involving rights. Esteban and Ray (2011) consider a model with both types of prizes to describe the surge of ethnic conflicts and Esteban, Mayoral and Ray (2012) bring their theoretical predictions to the data showing that both types of prizes are important.[24]

6.2. Endogenous social norms and peer punishment

In our setting a choice of mechanism can be described as the choice of a social norm φ_k with the implied punishment $P_k = c(\varphi_k)/\pi_0$ or equivalently the choice of a punishment P_k with the implied choice of a social norm $\varphi_k = c^{-1}(\pi_0 P_k)$. Both the strength and the weakness of mechanism design theory is abstraction of the mechanism designer and the distinction between the design and the implementation. In our context we can ask: how does it all work in practice? Who designs the mechanism and who carries out the punishments? These are distinct questions which we now address.

---

[23]If $V_k = \eta_k v$ the objective function is equivalent to $\Pi_k(b_k, F_{-k})v - C(b_k/\eta_k)$. Since for a given bid b the smaller party must turn out more voters $b/\eta_S > b/\eta_L$ and expected costs are increasing, it follows that the cost of a bid is always lower for the large party, hence it is advantaged.

[24]In the Online Appendix we show that our results about small group advantage are robust in the sense that they hold for any non-trivial mix of a common and per capita prize. It is only in the extreme case of a pure per capita prize that the large group is advantaged regardless of monitoring cost. The theory of group size has been explored also in settings where the benefit of an electoral victory itself provides incentives due to private attributes. See, for example, Esteban and Ray (2001) and Nitzan and Ueda (2011).

A possible story one can tell about the mechanism designer is that of Feddersen and Sandroni (2006) in their discussion of ethical voters. Each voter individually solves the mechanism design problem in their head, and, all being ex ante identical, all reach the same conclusion about the optimal social norm. Here the design problem is completely decentralized. While this narrative is sound from a conceptual point of view, we think it is not so likely in practice.

It is useful to think of the design parameter P_k as an indication of how important it is to vote. This has to be agreed upon by the party and communicated to the party members. Parties are typically made up of several layers with party officials at the top, grassroots organizers and turnout brokers in the middle, and voters at the bottom. In principle officials instruct organizers who then canvass voters. In practice, however, the voters have to be convinced. Consequently there is communication up and down with party officials arguing that the election is of great importance with pushback from the voters if that is not in fact the case. Through such a process we imagine a consensus is formed over the importance of voting in this particular election.

Once a consensus is reached, that is, a P_k is agreed upon, who does the actual punishment? Here we think the existing empirical literature provides a compelling answer: this is done by the voters themselves - friends, family and so forth - through various forms of peer pressure. To take a few of many pieces of evidence, Della Vigna et al (2014) 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 from the household or neighbors significantly increases turnout. Amat et al (2018) using elections data of Spain's Second Republic show that turnout was driven by political parties and trade unions' social pressure. The relation between voter participation and peer pressure is also widely discussed in the sociology literature, see Coleman (1988). The example of attacking a husband with an automobile for not voting is merely the most extreme example of which we are aware.

If voters punish each other, we may ask why? Here our story is no different than that in the ethical voters model: people adhere to the mechanism, that is they punish as prescribed, because they want to "do their bit" out of social conscience and because it is their self-interest to do so. As we have modelled punishment as being without cost to those carrying out the punishment, this is unproblematic. What happens if, as may be the case, punishment is costly to the punishers? First, a strict incentive to punish can be given by punishing those who fail to punish. This is a part of real social norms - see, for example, Skarbek (2014). A simple model of multiple punishment rounds in Levine and Modica (2014) shows that this amounts to add a multiplicative factor to monitoring costs - equivalent to increasing the parameter θ in our model. In other words, our model may be interpreted as a reduced form of a model with repeated punishment. The existence of repeated punishment rounds in practice is reinforced by the empirical literature that, as indicated above, suggests peer pressure affecting voter participation primarily originates from close social ties - individuals that interact repeatedly. That is, friends, family and neighbors may find it worthwhile punishing non-voters because they expect they will be treated similarly in a similar situation.[25] There is a second reason that people may carry out costly punishments. As in the ethical voter model, we have allowed a benefit from civic duty as an offset to the cost of participation: this is the source of committed voters in our model. That same sense of civic duty would equally be a reason for voters to carry out low cost punishments against those who are seen to have failed in their own civic duty.

A related issue is that what is punished is the appearance of guilt despite the fact that everyone is innocent and everybody knows this. This, however, is partly a matter of interpretation. That is, the "bad signal" may occur not from the lack of a good excuse, but from deviant preferences - a second draw from the cost distribution as it were - that leads an individual not to vote even when their "objective" excuse is a bad one. For example, perhaps the only accepted excuse is to be in the hospital, but on election day the particular voter cannot resist going to the beach rather than voting, hence getting a bad signal. In this way "guilt" and "innocence" become relative terms, and there is a sense in which those who get bad signals may indeed be "guilty."

[25] See Ali and Miller (2016) for a dynamic model showing that third party punishments in the form of ostracism may improve cooperation if ostracism is tempered by forgiveness.

Our model of punishment advances the ethical voter model. In the ethical voter model no party would engage in additional, and costly, GOTV strategies. Here we provide a microfoundation for why parties may be dissatisfied with turnout, but still find it too costly to increase it. At the same time we endogenously characterize which cost range will be targeted by the GOTV strategy: this defines the group that will, in equilibrium, feel compelled to follow the norm. Notice that unlike Feddersen and Sandroni (2006) we assume that civic duty is to party only. That is, voters only encourage like minded voters to vote. 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)), in the field (Bond et al (2012)), and 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. 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 identity (Dunham, Arechar and Rand (2016)).[26]

6.3. Endogenous versus exogenous uncertainty

There are different assumptions used in the literature about the way in which voting determines the outcome of an election. Palfrey and Rosenthal (1985) assume as do we that the parties are of fixed size and the party with the most votes wins. Other models introduce aggregate shocks and assume that these are sufficiently large to guarantee the existence of a pure strategy equilibrium. In our model voter turnout is also random but this is endogenous due to the use of mixed strategies by the parties. 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

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

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 advertise their GOTV effort, and in fact to keep it secret.[27] 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.

In general and in Shachar and Nalebuff (1999) and Coate and Conlin (2004) exogenous random turnout leads to a model in which the probability of winning depends not only on bids but also on the size of the two parties. Herrera, Morelli and Nunnari (2015) use a more standard contest resolution function in which the probability of winning depends only on the bids and this is also the case in the specific application of Feddersen and Sandroni (2006).

Despite this wide variety of assumptions on conflict resolution the existing literature assumes that monitoring costs are absent and conclude that the large party is advantaged. They all study pure strategy equilibria. As each paper makes special assumptions we give a general result in the Online Appendix for a common prize, convex common costs, and a standard contest resolution function where the probability of winning depends only on the bids: pure strategy equilibrium advantages the large party which turns out more voters and a gets greater utility than the small party. This is the same result we find for our mixed equilibrium.[28]

[27] 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.

[28] There is an analogous result for concave costs, but it is of lesser interest since pure strategy equilibria are not so likely to exist in that case.

6.4. Bounded costs

We have assumed that costs are bounded by $c(1)$. A consequence of this is that if the prize V is sufficiently large a party that could insure victory by doing so would turn out all of its voters. Moreover while we show in the Online Appendix that the equilibrium probability that either party turns out all of its voters is zero in a high stakes election, there is a small probability that the turnout of the small party is close to 100% . Empirically this has little meaning since measured turnout is as a fraction of the voting age population not, as in the model, as a fraction of people actively contemplating voting. Nevertheless it might be judged unreasonable that a party would turn out nearly all of its voters: in reality we would expect that some voters would have such high costs that it would not be worth turning them out regardless of how high the stakes might be. It turns out that this does not matter very much. In the Online Appendix we show that while the equilibrium is more difficult to compute with unbounded costs the equilibrium strategies and utilities are close to those with bounded costs even when V is very large.

7. 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 model corresponds to the special case in which monitoring costs are zero and, when the electorate is small, pivotality can be incorporated into the incentive constraints for individual voters. The theory also makes a rich new set of predictions - relating, for example, 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.[29] This may explain why there are many referenda where special interests do well: for example, the type

[29] We should mention that minority advantage is also present in Casella and Turban (2014) albeit for an entirely different reason - they study a model in which votes can be bought and sold in market.

of commercial gambling permitted on Indian reservations, school budgets, the working environment for prison guards and so forth.

Our model applies more generally to a situation where two groups compete by turning out members. It has the potential to organize in a common framework a variety of puzzles in the literature on voting, lobbying and conflict. Why are small groups much more effective at lobbying or bribery as in as in Hillman and Riley (2006), Acemoglu (2001), or Levine and Modica (2017)? What is the nature of small and large party advantage with a broader range of contest success functions? What does the theory tell us about which contest success functions most successfully trade off getting the prize to the right party while minimizing the cost of electoral competition? Opening the black box of the party to understand the mechanism design that overcomes the free-rider problem as we have done can help identifying key elements common to all of these issues.

References

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

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

Ali, S. N. and D. A. Miller (2016): "Ostracism and Forgiveness," *American Economic Review* 106(8): 2329–2348.

Amat, F., Boix C., Munoz J. and T. Rodon (2018): "From Political Mobilization to Electoral Participation: Turnout in Barcelona in the 1930s," Mimeo.

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

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.

Casella, Alessandra, and Sébastien Turban (2014): "Democracy undone. Systematic minority advantage in competitive vote markets," *Games and Economic Behavior* 88: 47-70.

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

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 (2016): “Collusion Constrained Equilibrium,” EUI.

Esteban, J., L. Mayoral and D. Ray (2012): “Ethnicity and Conflict: An Empirical Study,” *American Economic Review* 102(4): 1310–1342.

Esteban, J. and D. Ray (2001): "Collective action and the group size paradox," *American Political Science Review* 95: 663-672.

Esteban, J. and D. Ray (2011): “Linking Conflict to Inequality and Polarization,” *American Economic Review* 101 (4): 1345–74.

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.

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.

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 (2016): ”Peer Discipline and Incentives Within Groups,” *Journal of Economic Behavior and Organization*, 23: 19-30]

Levine, D. K. and S. Modica (2017): “Size, Fungibility, and the Strength of Lobbying Organizations,”*European Journal of Political Economy*, 49: 71-83

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

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.

Morris, S., M. Shadmehr (2017): “Reward and Punishment in a Regime Change Game,” mimeo.

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

Myerson, R. B. (1982): “Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics* 10: 67-81.

Nitzan, S. and K. Ueda (2011): "Prize sharing in collective contests," *European Economic Review* 55: 678-687.

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

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.

Palfrey, T. R. and K. Pogorelskiy (2016): “Voting with Communication: 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.

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

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.

Skarbek, D. (2014): *The social order of the underworld: How prison gangs govern the American penal system*, Oxford University Press.

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.

Townsend, Robert M. (1994): "Risk and insurance in village India," *Econometrica*: 539-539.

Online Appendix (Not for publication)

Equilibrium

Party strategies are probability distributions over mechanisms. Given that the objective functions can be written in terms of cost functions that depend only on bids, we may as well take them to be probability distributions over bids. Formally, we take a strategy for party k to be a probability measure represented by a cumulative distribution function F_k over bids, that is, on $[\underline{b}_k, \eta_k]$. The objective function of each party is $\Pi_k(b_k, F_{-k})V - \eta_k C(b_k/\eta_k)$ where $\Pi_k(b_k, F_{-k})$ is the probability that a bid b_k wins. With this formulation we have an all-pay auction model. We do not assume common prize, we allow arbitrary $V_k \geq 0$ and common cost $C(\varphi)$ which is continuous and strictly increasing for $\varphi \geq \underline{y}$. As in the text we assume $\overline{b}_L \neq \overline{b}_S$.

Proposition 2. [1 in text] *There is a unique equilibrium. In this equilibrium neither party uses a pure strategy, the utility of the disadvantaged party is* 0 *and the utility of the advantaged party is* $V_{-k} - \eta_{-d}C(\overline{b}_d/\eta_{-d})$.

We record the additional facts not reported in the text but are used subsequently in the Appendix which are the equilibrium strategies. Let $F^0_k(b)$ *record the size of the atom at* b *(if any). In* $(\eta_L \underline{y}, \overline{b}_d)$*:*

$$F_d(b_d) = 1 - \frac{\eta_{-d}C(\overline{b}_d/\eta_{-d}) - \eta_{-d}C(b_d/\eta_{-d})}{V_{-k}}$$

$$F_{-d}(\eta_{-d}\varphi_{-d}) = \frac{\eta_d C(b_{-d}/\eta_d)}{V_k}$$

The disadvantaged party has a single atom at $F^0_d(\eta_d \underline{y}) = 1 - \eta_{-d}C(\overline{b}_d/\eta_{-d})/V_{-k} + \eta_{-d}C(\eta_L \underline{y}/\eta_{-d})/V_{-d}$. *The advantaged party if it is large has an atom at* $F^0_L(\overline{b}_S) = 1 - \eta_S C(\overline{b}_S/\eta_S)/V_L$, *and whichever party is advantaged has an atom at* $F^0_{-d}(\underline{y}\eta_L/\eta_d) = \eta_d C(\underline{y}\eta_L/\eta_d)/V_d$.

Proof. S will never submit a bid b_k for which $\eta_S \underline{y} < b_k < \eta_L \underline{y}$ since such a bid will be costly but losing, and neither party will submit a bid for which $b_k > \overline{b}_k$ since to do so would cost more than the value of the prize. It follows that k

must either bid $\eta_k \underline{y}$ or in the range $[\eta_L \underline{y}, \overline{b}_d]$. If $V_S \leq \eta_S C\left(\underline{y}\eta_L/\eta_S\right)$, it follows that $\overline{b}_S \leq \eta_L \underline{y}$. In this case S will only turn out committed voters, that is will bid $\eta_S \underline{y}$, and L wins with probability 1 by bidding $\eta_L \underline{y}$. This case is ruled out in the text.

Consider now the case $V_S > \eta_S C\left(\underline{y}\eta_L/\eta_S\right)$. In the range $(\eta_L \underline{y}, \overline{b}_d)$ there can be no atoms by the usual argument for all-pay auctions: if there was an atom at b_k then party $-k$ would prefer to bid a bit more than b_k rather than a bit less, and since consequently there are no bids immediately below b_k party k would prefer to choose the atom at a lower bid. This also implies that S cannot have an atom at $\eta_L \underline{y}$: if L has an atom there, then S should increase its atom slightly to break the tie. If L does not have an atom there, then S should shift its atom to $\eta_S \underline{y}$ since it does not win either way.

Next we observe that in $(\eta_L \underline{y}, \overline{b}_d)$ 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 $\overline{b}_d$, while the advantaged party gets at least $U_{-d} \geq V_{-d} - \eta_{-d} C(\overline{b}_d/\eta_{-d}) > 0$ since by bidding slightly more than $\overline{b}_d$ it can win for sure, but gets no more than that since it must make bids with positive probability arbitrarily close to $\overline{b}_d$. We conclude that the equilibrium payoff of the advantaged party must be exactly $U_{-d} \geq V_d - \eta_{-d} C(\overline{b}_d/\eta_{-d})$.

From the absence of zero probability open intervals in $(\eta_L \underline{y}, \overline{b}_d)$ it follows that the indifference condition for the advantaged party

$$F_d(b_d)V_d - \eta_{-d} C(b_d/\eta_{-d}) = V_d - \eta_{-d} C(\overline{b}_d/\eta_{-d})$$

must hold for at least a dense subset. Similarly for the disadvantaged party

$$F_{-d}(b_{-d})V_{-d} - \eta_d C(b_{-d}/\eta_d) = 0$$

for at least a dense subset. This uniquely defines the cdf for each party in $(\eta_L \underline{y}, \overline{b}_d)$:

$$F_d(b_d) = 1 - \frac{\eta_{-d} C(\overline{b}_d/\eta_{-d}) - \eta_{-d} C(b_d/\eta_{-d})}{V_d}$$

$$F_{-d}(b_{-d}) = \frac{\eta_d C(b_{-d}/\eta_d)}{V_{-d}}.$$

As these are differentiable they can be represented by continuous density functions which are found by taking the derivative.

Evaluating $F_d(b_d)$ at $\eta_d \underline{y}$ gives $F_d^0(\eta_d \underline{y}) = 1 - \eta_{-d} C(\overline{b}_d/\eta_{-d}/V_{-d}) + \eta_{-d} C(\eta_L \underline{y}/\eta_{-d})/V_{-d}$. Since $F_d(\overline{b}_d) = 1$ and we already proved that S has no atom at $\eta_L \underline{y}$ this is in fact the only atom for the disadvantaged party.

As for the advantaged party, if -$d = S$ then $\eta_L > \eta_S \geq \overline{b}_S > \overline{b}_L$ implies that $F_S(\overline{b}_L) = \eta_L C(\overline{b}_L/\eta_L)/V_L = 1$. If instead $-d = L$ then $F_L(\overline{b}_S) = \eta_S C(\overline{b}_S/\eta_S)/V_L$. If $\overline{b}_S < \eta_S$ then this is 1 and there is no atom, otherwise there must be an atom of size $F_L^0(\overline{b}_S) = 1 - \eta_S C(\overline{b}_S/\eta_S)/V_L$. Turning to $\eta_L \underline{y}$ we see that the atom there is given by

$$F_{-d}^0 = \frac{\eta_d C(\underline{y} \eta_L/\eta_d)}{V}$$

since the advantaged group never bids less. □

Theorem. [2 in text] *For* any *individual cost function $c(y)$ with corresponding committed voters $\underline{y}$ there exist $\underline{\theta}^S < 1, \underline{\eta}_S < 1/2$ and $\overline{V} > \underline{V}^S > 0$ such that if all the conditions $\theta > \underline{\theta}^S, \eta_S \geq \underline{\eta}_S$ and $\underline{V}^S < V < \overline{V}$ hold the small party is advantaged. Conversely if $\underline{y} > 0$ for any values of the other parameters there exist $\underline{\theta}^L > 0, \overline{\eta}_S > 0$ and $\underline{V}^L > 0$ such that if any of the conditions $\theta < \underline{\theta}^L$, $\eta_S < \overline{\eta}_S$, $V < \underline{V}^L$ or $V > \overline{V}$ are satisfied then the large group is advantaged.*

Proof. For completeness we allow in this theorem the possibility that $\overline{b}_S < \eta_L \underline{y}$, that is the small party may or may not be willing to turn out at least the number of committed voters of the large party. To prove the first half of the theorem, observe that marginal cost is $C'(\varphi) = (1-\theta)c(\varphi) + \theta(1-\varphi)c'(\varphi)$ so if $\theta = 1$ then $C'(1) = 0$. Since average cost at $\varphi = 1$ is $C(1) > 0$, average cost is strictly larger than marginal cost at $\theta = 1, \varphi = 1$. Therefore from continuity it must be so for θ, φ both sufficiently close to 1. That is for $1 \geq \theta > \underline{\theta}^S$ and $1 > \overline{\varphi} > \varphi > \underline{\varphi}$ average cost is declining. Having fixed $\overline{\varphi}$ we may choose $\underline{\eta}_S < 1$ large enough that for $\eta_S \geq \underline{\eta}_S$ the small party is large enough to outbid $\overline{\varphi}\eta_L$, that is $1/2 > \underline{\eta}_S \geq \overline{\varphi}(1 - \underline{\eta}_s)$. Hence if we choose the prize V so that the large party's maximal willingness to turn out lies in this range, that is, $\overline{\varphi} > \overline{b}_L/\eta_L > \underline{\varphi}$ then the small party must be advantaged as is able to outbid the large party and has a lower average cost of matching the large party bid. For the second half of the Theorem, the large party is advantaged for $\theta = 0$ hence by continuity for small θ. For $\eta_S < \eta_L\underline{y}$ the small party is unable to overcome the committed voters of the large party. If $V < C(\eta_L\underline{y}/\eta_S)$ then the small party is unwilling to bid. If $V > \eta_L C(\eta_S/\eta_L) = \overline{V}$ then $\overline{b}_L > \eta_S$ so the large party is surely advantaged. □

Who Wins?

Theorem. [3 in text] *The equilibrium bidding function of a strongly advantaged party FOSD that of the disadvantaged party.*

Proof. At $\overline{b}_d$ we have $F_d(\overline{b}_d) = \hat{F}_{-d}(\overline{b}_d) = 1$ so this is irrelevant for FOSD. 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 S is disadvantaged. Hence when S is disadvantaged its bidding schedule cannot FOSD that of L, while if it is advantaged this range is irrelevant for FOSD.

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

$$F_d(b) = 1 - \frac{\eta_{-d}C(\overline{b}/\eta_{-d})}{V} + \frac{\eta_{-d}C(b/\eta_{-d})}{V}$$

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

Hence for FOSD of the advantaged party, we must have

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

Moreover since $\eta_d v_d \geq \eta_d C(\bar{b}_d/\eta_d)$ this is true if

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

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

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

Let $t(\eta, b) \equiv \eta C(b/\eta)$. The derivative with respect to η is $t_\eta(\eta, b) = C(b/\eta) - (b/\eta)C'(b/\eta)$ so the cross partial is $t_{\eta b}(\eta, b) = -(b/\eta^2)C''(b/\eta)$. Observe that the sufficient condition may be written as

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

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

$$= -\int_{\eta_d}^{\eta_{-d}} \int_{b}^{\bar{b}_d} t_{\eta b}(\eta, b') d\eta db' = \int_{\eta_d}^{\eta_{-d}} \int_{b}^{\bar{b}_d} (b'/\eta^2) C''(b'/\eta) d\eta db'$$

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, L advantaged, and S 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 FOSD the other. □

The next proposition studies the case where costs are incrementally concave and yet the large party is advantaged. It shows how the *FOSD* result can fail in the strong sense that the disadvantaged small party turns out more members in expectation and has a higher probability of winning than the large advantaged party.

Proposition 3. *Suppose that cost is quadratic so that for $\theta > 1/2$ it is incre-*

mentally concave. For any η_S there exists a $\underline{\varphi} > 0$ such that for any $\underline{y} < \underline{\varphi}$ there is 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 $\overline{\theta} > \theta > \theta^*$ the small party is advantaged

b. 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 $C(\varphi_k) = (1-2\theta)(\varphi_k - \underline{y})^2 + 2\theta(1-\underline{y})(\varphi_k - \underline{y})$. Hence $C'(\underline{y}) = 2\theta(1-\underline{y})$ and $C''(\varphi_k) = 2(1-2\theta)$. We fix $\theta > 1/2$ so that cost is incrementally concave.

We first establish that for sufficiently small $\underline{y}$ there is a range of $V's$ such $\eta_L \underline{y} < \overline{b}_S < \eta_S$ for $1/2 \leq \theta \leq 1$ and such that S is advantaged at $\theta = 1$.

Since the derivative of C with respect to θ is $2(\varphi_k - \underline{y})(1-\varphi_k) > 0$ the greatest willingness to bid is at $\theta = 1/2$ and the least is at $\theta = 1$. At $\theta = 1/2$ the utility of S is $V - (1-\underline{y})(\varphi_k - \underline{y})$ and so $\overline{b}_k < 1$ for $V < (1-\underline{y})^2 = \overline{V}_S$. At $\theta = 1$ the utility of S is $V + (\varphi_k - \underline{y})^2 - 2(1-\underline{y})(\varphi_k - \underline{y})$ so $\overline{b}_k > \eta_L \underline{y}$ for $V > 2(1-\underline{y})(\eta_L \underline{y}/\eta_S - \underline{y}) - (\eta_L \underline{y}/\eta_S - \underline{y})^2 = \underline{V}_S$.

Set $\theta = 1$ and let φ^* be defined by $A(\varphi^*) = A(\eta_L \varphi^*/\eta_S)$. Some algebra yields $\varphi^* = \sqrt{(\eta_S/\eta_L)\underline{y}(2-\underline{y})}$. This will be less than η_S/η_L provided $\underline{y}(2-\underline{y}) < \eta_S/\eta_L$. At $\theta = 1$ the utility for L is $V - \eta_L \left(-(\varphi_L - \underline{y})^2 + 2(1-\underline{y})(\varphi_L - \underline{y})\right)$. Hence L would like to bid greater than $\eta_L \varphi^*$when

$$V > \eta_L \left(-(\sqrt{(\eta_S/\eta_L)\underline{y}(2-\underline{y})} - \underline{y})^2 + 2(1-\underline{y})(\sqrt{(\eta_S/\eta_L)\underline{y}(2-\underline{y})} - \underline{y}) \right) = \underline{V}_L.$$

It is smaller than η_S when $V < \eta_L \left(-(\eta_S/\eta_L - \underline{y})^2 + 2(1-\underline{y})(\eta_S/\eta_L - \underline{y})\right) = \overline{V}_L$. Hence for V in this range and $\theta = 1$ S is advantaged.

We observe that when $\underline{y} = 0$ we have $\overline{V}_S = 1, \underline{V}_S = 0, \overline{V}_L = \eta_S(2 - \eta_S/\eta_L) > \eta_S$ and $\underline{V}_L = 0$. This establishes that for sufficiently small $\underline{y}$ there is a range of $V's$ such that $\eta_L \underline{y} < \overline{b}_S < \eta_S$ for $1/2 \leq \theta \leq 1$ and such that tS is advantaged at $\theta = 1$. Fix such a V.

Define the desire to bid as the solution of

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

and for S at least this is also the willingness to bid, and it will be the willingness to bid of L provided the constraint $b_L < \eta_L$ is satisfied. Since the last equation is quadratic in b_k it can be solved by the quadratic formula from which it is apparent that $\overline{b}_k(\theta)$ 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 $\overline{b}_S = \overline{b}_L = b$ 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 a unique 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 $\overline{b}_S = \overline{b}_L$ 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 $\overline{b}_S - \overline{b}_L$ is negative at $\theta = 1/2$ and positive at $\theta = 1$. We conclude that there is some point θ^* at which $\overline{b}_S = \overline{b}_L$ and S is advantaged for $\theta^* < \theta < \overline{\theta}$ for some $\overline{\theta}$, while L is advantaged for $\theta_0 < \theta < \theta^*$ for some θ_0.

Since C is incrementally concave in $\theta^* < \theta < \overline{\theta}$ and S is advantaged there, it follows that S follows a strategy that FOSD that of Ly. Hence in the limit at θ^* the strategy of the small party either FOSD that of the large party or is the same as that of the large party. However, for $\theta > \theta^*$, S plays $\eta_L\underline{y}$ with probability zero while L 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 θ^* the strategy of S FOSD that of L, 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 it follows that this remains true in an open neighborhood of θ^*. $\square$

When is Advantage Strong?

Recall from the text that $G(c)$ is the cdf of costs for an individual so that $c(\varphi) = G^{-1}(\varphi)$, $\varphi = G(c)$ and the support is $[c(0), c(1)]$. We denote the density of $G(c)$ by $g(c)$, and we assume it is continuously differentiable, strictly positive, and has a single "top" in the sense that it is either single peaked or a it is a limiting case such as the uniform where the density is flat at the top. To define $\underline{\gamma}, \overline{\gamma}$ we first defined

$$\mu(c) = \frac{(g(c))^2}{2(g(c))^2 + (1 - G(c))\, g'(c)},$$

then $\underline{\gamma} = \min_{c \geq 0} \mu(c)$ and $\overline{\gamma} = \max\{0, 1 - \max_{c \geq 0} \mu(c)|\}$. For the purpose of deriving the properties of $\underline{\gamma}, \overline{\gamma}$ it will be convenient instead to define

$$\lambda(c) = -\frac{(1 - G(c))\, g'(c)}{(g(c))^2}$$

and observe that $\mu(c) = 1/(2 - \lambda(c))$ so that $\mu(c)$ and $\lambda(c)$ share the same monotonicity properties. We take $\underline{\lambda} = \min_{c \geq 0} \lambda(c) \leq 0$ the smallest possible value of $\lambda(c)$ and $\overline{\lambda} = \max_{c \geq 0} \lambda(c) \geq 0$ the largest. In the uniform case $g'(c) = 0$ so $\lambda(c) = 0$. If the density is increasing then $\lambda(c) \leq 0$ so $\overline{\lambda} = 0$ and if it is decreasing $\underline{\lambda} = 0$. Equivalent to the definition in the text

$$\underline{\gamma} = \frac{1}{2 - \underline{\lambda}}$$

and define $\overline{\gamma} = 0$ if $\overline{\lambda} > 1$ and

$$\overline{\gamma} = 1 - \frac{1}{2 - \overline{\lambda}}$$

otherwise. Hence $\underline{\gamma}$ is an increasing function of $\underline{\lambda}$ and $\overline{\gamma}$ is a decreasing function of $\overline{\lambda}$. Since $\underline{\lambda} \leq 0$ and $\overline{\lambda} \geq 0$ we have $0 \leq \underline{\gamma}, \overline{\gamma} \leq 1/2$. The properties of $\underline{\gamma}, \overline{\gamma}$

for the uniform, increasing and decreasing cases can be read directly from the results for $\underline{\lambda}, \overline{\lambda}$: both 1/2 for the uniform case, $\overline{\gamma} = 1/2$ in the increasing case, and $\underline{\gamma} = 1/2$ in the decreasing case. For the single-peaked case we now prove

Proposition 4. *a. If the density shifts to the right then $\overline{\lambda}$ is constant and $\underline{\lambda}$ decreases ($\underline{\gamma}$ decreases); if the density shifts to the right holding fixed $c(1)$ then $\overline{\lambda}$ decreases ($\overline{\gamma}$ increases);*

b. Increasing dispersion by a change of scale around the mode increases $\underline{\lambda}$ ($\underline{\gamma}$ increases) and decreases $\overline{\lambda}$ ($\overline{\gamma}$ increases).

Proof. (a) We consider first the case of shifting the density to the right holding fixed $c(1)$. The only interesting case is when the peak c_g is interior, that is, satisfies $c(1) > c_g > 0$. Consider a $h(c)$ also with upper support $c(1)$ with mode c_h. Suppose that for some positive constants Δ, ζ we have $c_h > c_g + \Delta$ and for $c > c_h$ we have $h(c) = \zeta g(c - \Delta)$ (density shifts right). Notice the scaling factor ζ is needed since holding fixed the upper bound $c(1)$ mass is lost as we shift g to the right. We prove that $\overline{\lambda}_h < \overline{\lambda}_g$.

Notice that since by assumption of a single peak $g'(c_g) = h'(c_h) = 0$, we can define $\overline{\lambda}$ without loss of generality only for values of c to the right of the mode. Hence we have that

$$\overline{\lambda}_h = \max_{c(1) \geq c \geq c_h} -\frac{\int_c^{c(1)} h(\xi) d\xi h'(c)}{(h(c))^2} = \max_{c(1) \geq c \geq c_g + \Delta} -\frac{\int_c^{c(1)} \zeta g(\xi - \Delta) d\xi \zeta g'(c - \Delta)}{(\zeta g(c - \Delta))^2}$$

and after a change of variable $\tilde{c} = c - \Delta$ we have

$$= \max_{c(1) - \Delta \geq \tilde{c} \geq c_g} -\frac{\int_c^{c(1) - \Delta} g(\tilde{\xi}) d\tilde{\xi} g'(\tilde{c})}{(g(\tilde{c}))^2} < \max_{c(1) \geq c \geq c_g} -\frac{\int_c^{c(1)} g(\tilde{\xi}) d\tilde{\xi} g'(\tilde{c})}{(g(\tilde{c}))^2} = \overline{\lambda}_g.$$

This gives the result for fixed $c(1)$. Focus on the key result

$$\overline{\lambda}_h = \max_{c(1) - \Delta \geq \tilde{c} \geq c_g} -\frac{\int_c^{c(1) - \Delta} g(\tilde{\xi}) d\tilde{\xi} g'(\tilde{c})}{(g(\tilde{c}))^2}; \overline{\lambda}_g = \max_{c(1) \geq c \geq c_g} -\frac{\int_c^{c(1)} g(\tilde{\xi}) d\tilde{\xi} g'(\tilde{c})}{(g(\tilde{c}))^2}$$

For $\overline{\lambda}$ there are two effects of a right shift: the range over which the integral of

$g(\tilde{\xi})d\tilde{\xi}g'(\tilde{c})$ in the numerator is taken is shorter for h and the maximum is taken over a narrower range. There is no analogous result for $\underline{\lambda}$. For $\underline{\lambda}$ the range of the integral remains the same, but rather than a maximum over $c(1) - \Delta \geq \tilde{c} \geq c_g$ we take minimum over $c_g \geq \tilde{c} \geq c(0) - \Delta$. Hence the minimum is taken over a larger range, offsetting the effect of the shorter range of the integral and the combination of the two is ambiguous.

For an ordinary right shift (that is, not holding fixed $c(1)$) the range of the integral does not change. For $\overline{\lambda}$ the range over which the maximum is is taken does not change, so the right shift is neutral. For $\underline{\lambda}$ the range over which the minimum is taken increases so the minimum becomes more negative.

(b) We first prove the result for $\overline{\lambda}$. Consider a $h(c)$ also with upper support $c(1)$ with mode $c_h = c_g$. Suppose that for some positive constants $\sigma > 1, \zeta$ for $c > c_h$ we have $h(c) = \zeta g(c_g + (c - c_g)/\sigma)$ (greater dispersion to the right of the mode).

We have

$$\overline{\lambda}_h = \max_{c(1) \geq c \geq c_g} -\frac{\int_c^{c(1)} h(\xi)d\xi h'(c)}{(h(c))^2}$$

$$= \max_{c(1) \geq c \geq c_g} -\frac{\int_c^{c(1)} \zeta g(c_g + (\xi - c_g)/\sigma)d\xi(1/\sigma)\zeta g'(c_g + (c - c_g)/\sigma)}{(\zeta g(c_g + (c - c_g)/\sigma))^2}$$

and after a change of variable $\tilde{c} = c_g + (c - c_g)/\sigma$ we have

$$= \max_{c_g + (c(1) - c_g)/\sigma \geq \tilde{c} \geq c_g} -\frac{\int_c^{c_g + (c(1) - c_g)/\sigma} g(\tilde{\xi})d\tilde{\xi}g'(\tilde{c})}{(g(\tilde{c}))^2}.$$

Furthermore, since $\sigma > 1$ and $c_g < c(1)$ we have $c_g + (c(1) - c_g)/\sigma = [(\sigma - 1)/\sigma]c_g + [1/\sigma]c(1) < c(1)$ so

$$\overline{\lambda}_h < \max_{c(1) \geq c \geq c_g} -\frac{\int_c^{c(1)} g(\tilde{\xi})d\xi g'(\tilde{c})}{(g(\tilde{c}))^2} = \overline{\lambda}_g.$$

Here again there are two effects, a shorter range of integral and a shorter range over which the maximum is taken, both lowering $\overline{\lambda}$. In the case of $\underline{\lambda}$ it is also the case that both the range of the integral and range over which the minimum is taken shrink: hence the minimum must increase. □

Theorem 4 is equivalent to

Proposition 5. *a. cost is incrementally convex if and only if* $\theta < 1/(2-\underline{\lambda})$

b. cost is incrementally concave if and only if $\overline{\lambda} < 1$ *and* $\theta > 1/(2-\overline{\lambda})$.

Proof. We report expected cost $C(\varphi) = \int_{\underline{y}}^{\varphi} c(y)dy + \theta(1-\varphi)c(\varphi)$ and its first two derivatives $C'(\varphi) = (1-\theta)c(\varphi) + \theta(1-\varphi)c'(\varphi)$, and $C''(\varphi) = (1-2\theta)c'(\varphi) + \theta(1-\varphi)c''(\varphi)$. Observe that $c(\varphi) = G^{-1}(\varphi)$ so

$$c'(\varphi) = \frac{1}{g(G^{-1}(\varphi))}$$

$$c''(\varphi) = -\frac{g'(G^{-1}(\varphi))}{\left(g(G^{-1}(\varphi))\right)^3}$$

and hence we can rewrite $C''(\varphi)$ as

$$C''(\varphi) = \frac{1-\theta(2-\lambda(c))}{g(c)}.$$

Hence $C''(\varphi) > 0$ if and only if $\theta < 1/(2-\lambda(c))$ from which the result follows. □

Proposition 6. [Example in text] *Suppose participations costs c are normalized to lie in* $[0,1]$ *and have a density function* $g(c) = \alpha c^{\alpha-1}$ *where* $\alpha > 0$. *If* $\alpha < 1$ *then* $\underline{\gamma} = 1/2$ *and* $\overline{\gamma} = 0$. *If* $\alpha > 1$, *then* $\underline{\gamma} = 0$ *and* $\overline{\gamma} = 1/2$.

Proof. For $\alpha < 1$ the density is decreasing so $\underline{\lambda} = 0$ and for $\alpha > 1$ it is increasing so $\overline{\lambda} = 0$. We have

$$\lambda(c) = -(\alpha-1)\frac{(1-c^{\alpha})\,\alpha c^{\alpha-2}}{(\alpha c^{\alpha-1})^2} = -\frac{\alpha-1}{\alpha}(1-c^{\alpha})\,c^{-\alpha}$$

which goes to infinity in absolute value as $c \to 0$. Hence $\overline{\lambda} = \infty$ and $\underline{\lambda} = -\infty$ giving the required result. □

High and Low Value Elections

Theorem. [Theorem 5 in text] *In a high value election the probabilities that the small party concedes and the large party preempts the election increase in*

V, and approach 1 in the limit. As V increases the bid distribution of the small party declines in FOSD and the bid distribution of the large party increases in FOSD. The expected vote differential increases in V while the expected turnout cost remains constant.

Proof. In a high value election S is constrained and L is advantaged. The probability L preempts is $F_L^0(\eta_S) = 1 - (\eta_S/V)C(1)$, increasing in V. The probability of concession by S is $F_S^0(\eta_S \underline{y}) = 1 - \eta_L C(\eta_S/\eta_L)/V$ increasing in V.

Since changing V with $\overline{b}_d = \eta_S$ the support and shape of the cost function in the mixing range do not change, so raising V simply lowers the densities by a common factor, meaning that these shifts reflect stochastic dominance as well. The FOSD result implies the increased vote differential.

Total surplus is $V - \eta_{-d} C(\overline{b}_d/\eta_{-d})$. Since some party certainly gets the prize this implies the expected turnout cost is $\eta_{-d} C(\overline{b}_d/\eta_{-d})$ and in a high value election $\overline{b}_d$ remains constant at η_S, so expected turnout cost is $\eta_L C(\eta_S/\eta_L)$ independent of V . □

Monitoring Difficulty in High Value Elections

Theorem. [Theorem 6 in text] *In a high value election, an increase in monitoring difficulty θ decreases the turnout of the advantaged (large) party in terms of FOSD. 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 (small) party decreases in monitoring difficulty in terms of FOSD 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 $\eta_k C(\overline{b}_k/\eta_k) - V = 0$, we can apply the implicit function theorem and find that

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

Hence as θ decreases, that is as monitoring efficiency increases, so it does peak turnout. In a high value election the peak turnout $\overline{b}_d$ is fixed at η_S and S is disadvantaged. Examining the equilibrium 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 FOSD. For $F_S(b)$ we have

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

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 FOSD. 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$S$ must decline with θ. Since the derivative of expected turnout is a continuous function of θ, it follows that there is a value of $\underline{\eta}$ such that expected turnout of S is constant with θ while for larger η_S it declines. At $\underline{\eta}$ the expected vote differential must decline with θ since S expected turnout is constant and L 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 $\underline{\eta}$, S expected turnout declines and the expected vote differential does as well. □

Symmetry of the Fundamentals

Let $\rho \in [0,1]$ be a measure of the mix of issues between transfers and laws where $\rho = 0$ mans the election is purely about transfers and $\rho = 1$ means it is purely about laws. Examples of transfers include control over natural resources, the division of government jobs, the division of a fixed budget, taxes and subsidies and limitations on competition such as trade restrictions or occupational licensing. Examples of laws include civil rights, laws concerning abortion, criminal law, defense expenditures, non-trade foreign policy and policies concerning monuments. We suppose that $V_k = v(\eta_k, \rho)$ where $v(1/2, \rho) = V$. We take pure transfers to mean a common prize so that $v(\eta_k, 0) = V$ and pure laws to mean a common per capita prize so that $v(\eta_k, 1) = 2V\eta_k$. We assume that $v(\eta_k, \rho) \geq 0$ twice continuously differentiable with $v_\eta(\eta_k, \rho) \geq 0$. Define the prize elasticity with respect to party size $\gamma(\eta_k, \rho) = d(\log v(\eta_k, \rho)/d\log \eta_k) = v_\eta(\eta_k, \rho)\eta_k/v(\eta_k, \rho)$. Then for pure transfers we have $\gamma(\eta_k, 0) = 0$ for for pure laws we have $\gamma(\eta_k, 1) = 1$. It is natural to assume then that $\gamma_\rho(\eta_k, \rho) > 0$: that as the importance of laws as an issue increases the prize elasticity with respect to party size goes up. This implies in addition that $v_\rho(\eta_k, \rho) > 0$ for $\eta_k > 1/2$ and $v_\rho(\eta_k, \rho) < 0$ for $\eta_k < 1/2$. That is, as the importance of laws as an issue increases the value of prize to the large party goes up and to the small party goes down. It follows directly that increasing the importance of laws improves the advantage (positive or negative) of the large party by raising its willingness to bid and lowering that of the small party.

Example. Suppose that the election has a mix of transfer and legal issues so that $v(\eta_k, \rho) = (1-\rho)+2\rho\eta_k$ where $0 \leq \rho \leq 1$ is the relative importance of legal issues. Then $\gamma(\eta_k, \rho) = 2\rho\eta_k/\left((1-\rho)+2\rho\eta_k\right)$ and $\gamma(\eta_k, 0) = 0$, $\gamma(\eta_k, 1) = 1$ and the derivative is

$$\gamma_\rho(\eta_k, \rho) = \frac{2\eta_k\left((1-\rho)+2\rho\eta_k\right)+2\rho\eta_k(1-2\eta_k)}{\left((1-\rho)+2\rho\eta_k\right)^2} > 0.$$

Proposition 7. *If $\rho > 1$ then there are cost functions, prize values, party sizes, and monitoring difficulty for which the small party is advantaged.*

Proof. Willingness to bid is $v(\eta_k, \rho) - \eta_k C(b_k/\eta_k) = 0$ or $1-(\eta_k/v(\eta_{k,}\rho))C(b_k/\eta_k) = 0$. Using the implicit function theorem we find

$$db_k/d\eta_k = -\frac{(v(\eta_k, \rho) - v'(\eta_k, \rho)\eta)C(b_k/\eta_k)/v(\eta_k, \rho)^2 - (1/v(\eta_k, \rho))C'(b_k/\eta_k)(b_k/\eta_k)}{C'(b/\eta)/v(\eta_k, \rho)}$$

so that the sign determined by $C'(\varphi_k)\varphi_k - (1-\gamma(\eta_k, \rho))C(\varphi_k)$. If the parties are of near equal size and this is positive or $V > (1/2)C(1)$ then L is advantaged, if the parties are of near equal size, $V < (1/2)C$ and this is negative, S is advantaged. If $\rho = 0$ so the election is purely about transfers then this is $C'(\varphi_k)\varphi_k - C(\varphi_k)$

so which party is advantaged depends on whether average cost is increasing or decreasing as we know. If $\rho = 1$ so the election is purely about laws this is $C'(\varphi_k)\varphi_k$ which is always positive so L is always advantaged. In the intermediate cases there are always parameter values for which S is advantaged. Take the quadratic case with no committed voters where $C(\varphi_k) = (1-2\theta)\varphi_k^2 + 2\theta\varphi_k$ At $\theta = 1$ this is $C(\varphi_k) = -\varphi_k^2 + 2\varphi_k$ and $C'(\varphi_k) = -2\varphi_k + 2$. Hence

$$C'(\varphi_k)\varphi_k - (1-\gamma(\eta_k,\rho))C(\varphi_k) = (-2\varphi_k + 2)\,\varphi_k - (1-\gamma(\eta_k,\rho))\left(-\varphi_k^2 + 2\varphi_k\right).$$

$$= -\varphi_k^2 + \gamma(\eta_k,\rho))\left(-\varphi_k^2 + 2\varphi_k\right)$$

$$= -(1+\gamma(\eta_k,\rho))\varphi_k^2 + \gamma(\eta_k,\rho))2\varphi_k.$$

Notice that for positive $\gamma(\eta_k,\rho)$ and small φ_k this is necessarily positive. However, as $\varphi_k \to 1$ this approaches $-(1-\gamma(\eta_k,\rho))$ which is strictly negative for $\rho < 1$, so also for $\varphi_k < 1$ but close to 1. □

The proof shows that with quadratic cost given $\rho < 1$ if there are sufficiently few committed voters, if $V < (1/2)C(1)$ but close enough (intermediate size prize), parties of similar enough size (small party not too small) and θ near enough 1 (high monitoring costs) the small party is advantaged. This is the same qualitatively as in the $\rho = 0$ case: however, quantitatively the criteria are much more stringent.

Endogenous versus Exogenous Uncertainty

Suppose that the probability of winning the election for party k is given by $P(b_k, b_{-k})$ non-decreasing in b_k. This must satisfy the identity $P(b_k, b_{-k}) = 1 - P(b_{-k}, b_k)$. Suppose there is a common prize the value of which we may normalize to 1 and common cost $C(\varphi)$. The objective function of party k is therefore $P(b_k, b_{-k}) - \eta_k C(b_k/\eta_k)$.

Proposition 8. *In any pure strategy equilibrium b_k, b_{-k} (if one exists) if $C''(\varphi) > 0$ then $b_L > b_S$ and the large party receives strictly greater utility than the small party; if $b_L \leq \eta_S$ and $C''(\varphi < 0)$ then $b_S > b_L$ and the small party receives strictly greater utility than the large party.*

Proof. In the convex case if $b_L > \eta_S$ then certainly L turns out more than Sy, so in both cases we may assume $b_L \leq \eta_S$. Consider that the utility to party k from playing b_{-k} rather than b_k must not yield an improvement in utility. That is

$$P(b_k, b_{-k}) - \eta_k C(b_k/\eta_k) \geq (1/2) - \eta_k C(b_{-k}/\eta_k)$$

or

$$P(b_k, b_{-k}) - (1/2) \geq \eta_k C(b_k/\eta_k) - \eta_k C(b_{-k}/\eta_k).$$

For party $-k$ this reads

$$P(b_{-k}, b_k) - (1/2) \geq \eta_{-k} C(b_{-k}/\eta_{-k}) - \eta_{-k} C(b_k/\eta_{-k})$$

and using $P(b_{-k}, b_k) = 1 - P(b_k, b_{-k})$

$$(1/2) - P(b_k, b_{-k}) \geq \eta_{-k} C(b_{-k}/\eta_{-k}) - \eta_{-k} C(b_k/\eta_{-k})$$

or

$$P(b_k, b_{-k}) - 1/2 \leq \eta_{-k} C(b_k/\eta_{-k}) - \eta_{-k} C(b_{-k}/\eta_{-k})$$

so the inequalities for the two parties are

$$\eta_k C(b_k/\eta_k) - \eta_k C(b_{-k}/\eta_k) \leq \eta_{-k} C(b_k/\eta_{-k}) - \eta_{-k} C(b_{-k}/\eta_{-k}).$$

Suppose without loss of generality that $b_k \geq b_{-k}$ so both sides are non-negative.

We work through the convex case. If $k = S$ we see that we must have

$$\eta_S C(b_k/\eta_S) - \eta_S C(b_{-k}/\eta_S) \leq \eta_L C(b_k/\eta_L) - \eta_L C(b_{-k}/\eta_L).$$

Consider the function $\eta_k C(b_k/\eta_k) - \eta_k C(b_{-k}/\eta_k)$ and differentiate it with respect to η_k to find

$$C(b_k/\eta_k) - C(b_{-k}/\eta_k) - ((b_k/\eta_k)C'(b_k/\eta_k) - (b_{-k}/\eta_k)C'(b_{-k}/\eta_k))$$

which may also be written as

$$C(b_k/\eta_k) - (b_k/\eta_k)C'(b_k/\eta_k) - (C(b_{-k}/\eta_k) - (b_{-k}/\eta_k)C'(b_{-k}/\eta_k)) \, .$$

Consider the function $C(\varphi) - \varphi C'(\varphi)$ and differentiate with respect to φ to find

$$-\varphi C''(\varphi) < 0.$$

This implies

$$C(b_k/\eta_k) - (b_k/\eta_k)C'(b_k/\eta_k) - (C(b_{-k}/\eta_k) - (b_{-k}/\eta_k)C'(b_{-k}/\eta_k)) < 0$$

which in turn implies

$$\eta_L C(b_k/\eta_L) - \eta_L C(b_{-k}/\eta_L) < \eta_S C(b_k/\eta_S) - \eta_S C(b_{-k}/\eta_S)$$

a contradiction, so we conclude that $k = L$, that is, $b_L > b_S$.

If $b_L > b_S$ suppose that L were to lower its bid to b_S. It would then have a 1/2 chance of winning - at least the equilibrium utility of S - and a cost lower than the equilibrium cost of S, so bidding b_S yields L more than the equilibrium utility of S. Hence the equilibrium utility of L must be larger than that of S.

Finally if $C(\varphi)$ is concave then the role of the two parties in determining the equilibrium bids is reversed, so we conclude that $b_S > b_L$. □

Bounded Costs

We compare two participation cost functions: $c(y)$, $\xi(y)$ where for some $\eta_S/\eta_L < \overline{y} < 1$ and $y \leq \overline{y}$ we have $c(y) = \xi(y)$ while for $\overline{y} < y \leq 1$ we have $c(y) < \xi(y)$. The cost function $c(y)$ is bounded, but we allow $\xi(1) = \infty$. It follows that the corresponding expected cost functions $C(y), \Xi(y)$ share the same property that $y \leq \overline{y}$ we have $C(y) = \Xi(y)$ while for $\overline{y} < y \leq 1$ we have $C(y) < \Xi(y)$ and $C(y)$ is bounded while $\Xi(y)$ need not be

Proposition 9. *If c has high stakes so $V > \max\{\eta_L C(\eta_S/\eta_L), \eta_S C(1)\}$ and ξ has high costs $\Xi(1) > V/\eta_S$ then the large party is advantaged. The equilibrium strategies and payoffs of the small party are the same for c, ξ. For the large* *party for low bids $b \le \eta_S \overline{y}$ the strategies are the same for c, ξ. The probability of a high bid under ξ is approximately the same as the atom at η_S under c*

$$\left[1 - F_L^{\xi}(\eta_S \overline{y})\right] - F_L^{0c}(1) = \eta_S \left[C(1) - C(\overline{y})\right]/V$$

as are the equilibrium payoffs

$$\eta_L \left(C(1) - C(\overline{y})\right) > U_L^{\xi} - U_L^{c} > 0.$$

Proof. As L never bids more than η_S and $\overline{y} > \eta_S/\eta_L$ only c is relevant for computing the payoffs of L; this implies in particular that the strategy of S is the same for c or ξ. Moreover, L is advantaged for both c, ξ. This follows from $V > \eta_L C(\eta_S/\eta_L)$ meaning L is willing to bid more than η_S which is the most S can bid. Since L is advantaged for c, ξ, S gets 0 in either case. For L bids below $\eta_S \overline{y}$ we have $F_L^c(b) = \eta_S C(b/\eta_S)/V = \eta_S \Xi(b/\eta_S)/V = F_L^{\xi}(b)$.

Under c, S is willing to bid η_S (by high stakes) while under χ, S is willing to bid $\eta_S \overline{y} < \overline{b}_S < \eta_S$. The first part $\eta_S \overline{y} < \overline{b}_S$ follows from $V - \eta_S \Xi(\overline{y}) = V - \eta_S C(\overline{y}) > V - \eta_S C(1) > 0$ and the second part $\overline{b}_S < \eta_S$ follows from the high cost assumption $V - \eta_S \Xi(1) < 0$.

We now compute the probability L makes a high bid $1 - F_L^{\xi}(\eta_S \overline{y})$. Since $F_L^{\xi}(\eta_S \overline{y})V - \eta_S D(\overline{y}) = 0$ we have $1 - F_L^{\xi}(\eta_S \overline{y}) = 1 - \eta_S C(\overline{y})/V$. By contrast $F_L^{0c}(1)$ satisfies $(1 - F_L^{0c}(1))V - \eta_S C(1) = 0$ so $F_L^{0c}(1) = 1 - \eta_S C(1)/V$. These two give the desired result

$$\left[1 - F_L^{\xi}(\eta_S \overline{y})\right] - F_L^{0c}(1) = \eta_S \left[C(1) - C(\overline{y})\right]/V$$

Finally we compute $U_L^{\xi} - U_L^{c} = \eta_L \left(C(1) - C(\overline{b}_S/\eta_S)\right)$. Hence indeed $\eta_L \left(C(1) - C(\overline{y})\right) > U_L^{\xi} - U_L^{c} > 0$ □