The Calculus of Ethical Voting*

Timothy Feddersen and Alvaro Sandroni
MEDS Department, Kellogg School of Management
2001 Sheridan Road, Evanston IL 60208

May 23, 2005

Abstract

Explaining participation in large elections is a particularly difficult problem for standard game-theoretic models. If voting is costly then, since the likelihood of a vote being pivotal is very small, standard game-theoretic models predict low levels of turnout. We analyze a model of participation in elections in which voting is costly and no vote is ever pivotal. In our model *ethical* agents are motivated to participate when they determine that agents of their type are morally obligated to do so. Unlike previous duty-based models of participation, an ethical agent's moral obligation to vote is determined endogenously as a function of the behavior of other agents. In order to predict outcomes we develop a solution concept we call *consistency* linking agents' preferences with actual behavior in a manner analogous to Nash equilibrium. The resulting model delivers high turnout and comparative statics that are consistent with strategic behaviour.

*We thank many audiences for their comments. Sandroni gratefully acknowledges the financial support from the National Science Foundation Grant SES 0109650. All errors are ours.

1. Introduction

Behavior in laboratory experiments and in the field does not always correspond to the predictions of decision and game theory. Economists, game theorists, psychologists and political scientists have considered a variety of alternative approaches to explain these anomalies (see, for example, Simon (1955), Kahnemann and Tversky (1979), Becker (1976), Rabin (1993)). In this paper, we relax one of the central assumptions in game theory: that agents' payoffs are exogenously determined by the outcomes of the game. We develop a model in which agents' payoffs are also determined by endogenously generated preferences over actions. Agents' preferences over actions depend upon how they believe they *should* behave given the behavior of other agents. Because agents are not playing a standard game, existing solution concepts do not readily apply. We develop a solution concept we call *consistency* that links agents' preferences with actual behavior in a manner analogous to Nash equilibrium.

Rather than developing a general model our focus here is on demonstrating some of the basic features of our approach in the context of a large election. We have chosen to analyze participation in large elections because this is a particularly difficult problem for standard game-theoretic models. If voting is costly then, since the likelihood of a vote being pivotal is very small, standard game-theoretic models predict low levels of turnout (e.g., Palfrey and Rosenthal (1985)). Typically, we observe high levels of turnout. The discrepancy between theory and prediction in the case of turnout is taken by some as evidence against the applicability of rational choice theory to politics (see for example, Green and Shapiro (1996)). But a claim that voters are not instrumentally rational runs afoul of considerable evidence of strategic behavior by voters. For example, turnout levels are correlated with the expected closeness of the election (Blais (2000)); small increases in costs to vote can have significant impact on turnout levels (Riker and Ordeshook (1968)); and behavior in multicandidate elections shows compelling evidence of strategic voting (Franklin, Niemi and Whitten (1994)). The evidence for strategic behavior by voters is as puzzling as the fact of large scale turnout and for the same reason. Not only is the incentive to vote minimal but so is the incentive to vote strategically. The challenge is to provide a model that not only explains participation but also explains the comparative statics that seem to be consistent with strategic behavior.

Our approach is related to earlier work by Harsanyi. Harsanyi (1992) assumes that a fraction of the population are "rule-utilitarians." A rule-utilitarian is an agent who receives a payoff for acting according to a strategy profile (a rule) that, if taken by all agents, maximizes social welfare. In an earlier paper, Harsanyi (1977) provides an example in which rule-utilitarians are motivated to vote. In his example there are two candidates, one of whom is assumed to maximize social welfare if elected. A critical assumption is that a fixed fraction of the population are voting for the socially inferior candidate. Harsanyi argues that rule-utilitarians will vote at a level to ensure the victory of the superior candidate. However, he provides no explanation for why some are voting for the inferior candidate. The rule-utilitarian model is problematic because if everyone agreed on the social welfare function then the inferior candidate would receive no votes. In that case the superior candidate may be elected with just one vote thereby minimizing the social cost of voting.

We model a two-candidate election in which a large population of agents may vote for candidate 1, vote for candidate 2 or abstain. Voting costs vary within the population and a single vote is never pivotal. The winner of the election is the candidate that receives the majority of votes cast. Like Harsanyi, we assume that agents have preferences over the candidates, the cost of the election and they also care about how they should behave. Unlike Harsanyi, we assume that agents' preferences are not necessarily related to social welfare (hence, agents are not necessarily rule utilitarians) and we allow for heterogenous preferences. There are two types of agents in our model: those who prefer candidate 1 (type 1) and those who prefer candidate 2 (type 2). The proportion of agents of each type is uncertain. Fixing the probability of winning for each candidate, all agents prefer to minimize the cost of the election.

In our model each agent has an action he *should* take and receives an additional payoff from taking this action. Riker and Ordershook (1968) analyze a model of participation in which agents receive a "duty" payoff $D > 0$ when they vote for their preferred candidate. In their model the action an agent should take is determined exogenously. In contrast, the action an agent should take is determined endogenously in our model. In particular, it may be (and often is) the case that some agents receive an additional payoff for voting while others receive a payoff for *not voting.*

Given a preference type, let a *rule* define how each agent should behave (i.e. a rule is an arbitrary mapping from voting costs to the action space). We say that agents who act as they should are "doing their part." Some agents receive a payoff $D > 0$ for doing their part, but not all of them do their part. Only those for whom it is not too costly. Let a *social outcome* be a probability that candidate

1 is elected and an expected cost of the election. Given the behavior of agents with preference type $j \neq i$, agents with preference type i evaluate each rule by the social outcome they produce. That is, agents rank rules according to their preferences. A behavior profile is *consistent* if the behavior of each agent follows from the agent's preferred rule (i.e. the rule with maximum ranking).

In a companion paper (Feddersen and Sandroni (2005)), we consider a simple example and demonstrate that, in this example, the participation rate of the minority is greater than the majority, and yet the majority's preferred candidate wins the election with probability greater than a half. We also show that expected turnout is increasing (and expected margin of victory is decreasing) in the level of disagreement.

In this paper, we generalize the example in Feddersen and Sandroni (2005) and develop a methodology that incorporates ethical reasoning in voting models. We show that if agents' behavior is consistent then there exists a cut-off point for each type such that agents with voting costs below this threshold should (and will) vote for their favored candidate. Agents with voting costs above the threshold abstain. As noted above, it is possible that agents with voting costs below D will abstain because they reason they should. We also show that there exists a one-to-one correspondence between the cut-off points that define a consistent profile and a pure strategy equilibrium of a suitably defined two-player game that we call 2PG. So, 2PG can be used to find consistent behavior profiles. However, a mixed strategy equilibrium in 2PG cannot be mapped into a consistent behavioral profile. This creates a difficulty because, as is often the case in games, an equilibrium in pure strategies may not exist in 2PG. We show that a pure strategy Nash equilibrium exists in 2PG for an interesting class of elections and is unique.

Finally, we demonstrate that, in general, the participation rate of the minority is indeed greater than the majority and that the majority does wins the election with probability greater than a half. However, the inverse relationship between expected turnout and margin of victory (as the level of disagreement changes) does not necessarily hold. We show an example where these two variables are not inversely related.

This paper is organized as follows: In section 1.1 we discuss the related literature within political science and economics. In section 2, we present our model. In section 2.1, we explain some of our modelling choices. In section 3, we show the relationship between a consistent behavior profile and a pure strategy equilibrium in $2PG$. We present an illustration of the properties of the model using a simple example in section 3.1. In section 4, we show the existence and uniqueness of pure strategy equilibrium in a class of elections in 2PG. In section 5, we show the general properties of our model. We conclude the paper in section 6. Proofs are in appendixes A and B. In appendix C, we provide some formal justification for some of our assumptions.

1.1. Literature Review

Extensive literature reviews of the literature on participation in elections can be found in Aldrich (1993), Palfrey and Rosenthal (1983, 1985), Feddersen and Pesendorfer (1999), Blais (2000), Mueller (2003) and Feddersen (2004). For our purposes it is useful to highlight Blais (2000) who reports that a sense of duty plays an important role in the decision to vote. He writes:

> I conclude that for many people voting is not only a right, it is also a duty. And the belief that in a democracy every citizen should feel obliged to vote induces many people to vote in almost all elections. That sense of duty is not shared by everyone. It may vary from one country to another. It can also vary over time. Blais (2000, page 113).

There are a variety of formal models of turnout and the surveys above discuss the formal literature. The standard rational actor models (e.g. Palfrey and Rosenthal (1983, 1985)) rely on events of insignificant likelihood such as the chances of being pivotal in a large election. These models provide useful insights, but they cannot explain large scale turnout when voting is costly.

Models by Morton (1991), Nalebuff and Shachar (1999), and Ulhaner (1989) assume that elites produce participation by providing incentives for voting. The elite-based models do not make the underlying mechanism generating turnout explicit. Our model can deliver the microfoundations for persuasive elites because agents care about how they should behave and, therefore, elites might provide credible information about the importance of the election and thereby increase turnout.[1] However, large scale turnout may still occur in the absence of elites. In this paper we focus on a basic model without elites and several other important institutional details.

The papers most closely related to our effort here are the decision theoretic models of Riker and Ordershook (1968) and Harsanyi's (1977, 1980 and 1992)

[1] Given that the interests of elites are not necessarily aligned with the interests of the electorate, the results in models where elites have to persuade the electorate may be quite different from models in which elites provide incentives for turnout.

models of rule-utilitarianism. We can subsume the Riker and Ordeshook model by considering preferences such that agents will understand they should always vote, no matter the cost. We can also subsume Harsanyi's (1992) model by assuming that all agents are rule-utilitarians and that the level of disagreement is zero. However, under these extreme assumptions the comparative statics results are not consistent with the empirical regularities observed in large elections. As demonstrated in section 5, the model delivers more powerful comparative static results when agents endogenously decide which behavioral rule they must follow and there is sufficient heterogeneity of preferences in the electorate.

In addition to models of turnout there is also a large related literature on the provision of public goods.[2] Standard models predict contributions levels below observed. Altruism has been proposed as an explanation for these anomalies (see, among others, Hirschleifer (1985), Monroe (1994) and Sugden (1984)). However, in models in which the marginal effect of a contribution decreases as total contributions increases, a "crowding out" effect is predicted. Thus, predicted contributions are still low. The crowding out effect is particularly severe in elections. If all voters prefer the same candidate then a single vote will crowd out all others. Hence, simple altruism cannot explain observed contribution patterns. Contributions will be larger if, as assumed by Becker (1977) and Andreoni (1989), agents receive a "warm glow" payoff for contributing for the public good which is independent of the contribution of others. As in Riker and Ordeshook (1968), the warm glow effect is exogenous. There is a warm glow aspect to our model. While we do not assume that agents receive a warm glow payoff for voting, agents do receive an additional payoff for doing their part.

In a political setting, Roemer (2003) develops a model in which agents decide on how much to contribute to a political party. The problem of contributing to a candidate is similar to the problem of costly voting because an individual contribution, like an individual vote, is seldom pivotal. Roemer develops the idea of a Kantian equilibrium. Like in our model, contributors make contributions because they understand they have a duty to do so, and the required contributions are determined endogenously. Unlike our model, each person contributes at a level such that he or she is indifferent between raising and lowering everyone's contribution. This mechanism explains why agents contribute even when their individual contribution has no impact on the ultimate outcome.

[2] The crucial difference is that the optimal level of contributions for a public good may or may not be large, but the optimal level of turnout is necessarily small. This follows because any election outcome can be reproduced, at lower social cost, by a representative sample of voters.

Samuelson (1993), Bernheim and Stark (1988), and Bergstrom (1995), among others, address the problem of how ethical preferences may arise as the consequence of evolutionary pressures.

2. The Model

We model an election with two candidates, 1 and 2. There is a continuum of agents who must choose one of three actions: vote for candidate 1, vote for 2 or abstain. Let A be the set of these three actions. The election is decided by majority rule.

Each agent has a cost of voting that is a realization of a random variable $\tilde{c}$ distributed over $[c^l, c^u]$, $0 \leqslant c^l < c^u$ with a continuous strictly positive density function $f_{\tilde{c}}$ on $[c^l, c^u]$. The cost of voting can be taught of as, for example, the time spent in the process. We assume that the disutility from voting is equal to the cost and that each agent's cost of voting is independent of any other random variable in this model. Each agent knows her own realized voting costs, but not the realization of other agents' voting costs.

There are two types of agents. Type 1 agents prefer candidate 1 and type 2 prefer candidate 2. All agents prefer the social cost of voting to be minimized.[3] Formally, type 1 agents have a utility function given by

$$wp - \vartheta(\phi), \tag{2.1}$$

where $\vartheta : [0,1] \to \Re_+$ is a strictly increasing, convex, twice differentiable function, p is the probability candidate 1 wins the election, ϕ is the expected social cost of voting and $w \in \Re_+$ is a parameter of the model that we call the *importance of the election*.[4] Type 2 agents have an utility function given by

$$w(1-p) - \vartheta(\phi). \tag{2.2}$$

[3] The social cost of voting is given by

$$\int_{c^l}^{c^u} c\beta(c) f_{\tilde{c}}(c)\partial c$$

where $\beta(c)$ is the fraction of the population with costs c that voted.

[4] Holding everything else constant, the higher the value of w the higher the expected social cost that agents would be willing to trade for an increase in the chances that their favored candidate wins. Thus, w parameterizes the importance of the difference between the two candidates relative to the social cost of voting.

No single agent decides who will be elected or the social cost of voting. Each agent's preferences reflect the choices he would make if he was a social planner. More formally, let a lottery be a random variable that determines the elected candidate and the social cost of voting. We assume that *if* he had the opportunity, an agent type 1 (2) would choose among these lotteries according to 2.1 (2.2).[5]

We assume that the fraction of agents of each type is unknown. The fraction of type 1 agents within the population is given by a random variable $\tilde{k}$ with expected value k. The parameter k may be thought of as indicating the *level of disagreement* within the population. When k is small almost everyone in the population agrees that candidate 2 is preferred to candidate 1. When k is close to .5 the society is nearly divided on the question of which candidate is preferable.

The model as defined so far is fairly standard. It is known that in voting games, as above, with a continuum of agents and costly voting there are generically no equilibria in which a positive fraction of the population participates. We now alter this standard game. We assume that every agent has an action that he understands to be the action he *should* take (and some agents derive positive utility from acting as they should). As will become clear, agents endogenously determine the actions they should take as well as the actions they do take. Hence, our assumption differs in a fundamental way from an assumption that agents receive positive utility for taking some exogenously determined action, e.g., voting for one of the candidates. We now define these ideas formally.

We call a function $e_i : (c^l, c^h) \longrightarrow A$, $i \in \{1, 2\}$, a *rule*. *If* a rule e_i is chosen by agents with preferences type $i \in \{1, 2\}$, then $e_i(c)$ denotes the action an agent type i with cost c *should* take. As mentioned above, we say that agents who take the action they should are "doing their part." Some agents (called *ethical* agents) receive a payoff $D > c^l$ for doing their part. Other agents (called *abstainers*) receive zero payoff for doing their part. The fraction of ethical agents in group $i \in \{1, 2\}$ is given by the random variable $\tilde{q}_i$ with expected value q_i.

The fact that an agent has an action that he should take does not guarantee that he *will* take it. If a rule e_i is chosen by agents with preferences type $i \in \{1, 2\}$ then $s_i^{e_i} : (c^l, c^h) \times \{0, D\} \longrightarrow A$ such that

$$s_i^{e_i}(c, d) = \begin{cases} \text{abstain} & if \quad d \leq c; \\ e_i(c) & \quad c > d. \end{cases}$$

denotes the action each type i agent will take.[6] That is, ethical agents will do

[5]The assumption of two types is for simplicity only.

[6]The domain of a rule could also be $(c^l, c^h) \times \{0, D\}$. This would not change any results.

their part if and only if the cost of the required action is sufficiently small. All others abstain.

We assume that agents' rank rules according to their preferences. Formally, let $s_i : (c^l, c^h) \times \{0, D\} \longrightarrow A$ be a behavioral profile for type $i \in \{1, 2\}$ agents. Given a pair (s_1, s_2) of behavioral profiles, let $p(s_1, s_2)$ be the probability that candidate 1 wins the election and $\phi(s_1, s_2)$ be the expected social voting cost. Type 1 agents rank rules as follows. Given a strategy profile s_2 for type 2 agents, the higher the value of $R_1(e \mid s_2)$,

$$R_1(e \mid s_2) \equiv wp(s_1^e, s_2) - \vartheta(\phi(s_1^e, s_2)),$$

the higher agents with preferences 1 rank rule e. Analogously, given the strategy profile s_1, the higher the value of $R_2(e \mid s_1)$,

$$R_2(e \mid s_1) \equiv w(1 - p(s_1, s_2^e)) - \vartheta(\phi(s_1, s_2^e)),$$

the higher agents with preferences 2 rank rule e. That is, rules are ranked according to the outcomes the rule produces, taking as given the actions of agents with different preferences.

We say that a behavior profile $(s_1^*,\ s_2^*\ )$ is *consistent* when each agent's action is determined by a rule that is maximally ranked for her type, given the behavior of other types.

Definition 1 ((Consistency Requirement)). *The pair* $(s_1^*,\ s_2^*\ )$ *is a consistent behavior profile if there exists a pair of rules* $(e_1^*,\ e_2^*)$ *such that*

1.
$$R_1(e_1^* \mid s_2^*) \geqslant R_1(e \mid s_2^*) \text{ for all rules } e : (c^l, c^h) \longrightarrow A;$$
$$R_2(e_2^* \mid s_1^*) \geqslant R_2(e \mid s_1^*) \text{ for all rules } e : (c^l, c^h) \longrightarrow A.$$

2.
$$s_i^* = s_i^{e_i^*},\ i \in \{1, 2\}.$$

Part 1 of the definition ensures that all agents choose a rule that is maximally ranked given the behavior of the other types. Part 2 ensures that the behavior of all agents of the same type is determined by the same maximally ranked rule.

2.1. Comments on the Model

In our model agents have preferences about which candidate wins and the social cost of turnout. These preferences can be determined by asking agents to decide *separately* on the probability that candidate 1 is elected and the social cost of the election. Informally, we have in mind that agents' preferences reflect not only their self-interest but also their religious, ethical or philosophical perspectives. For example, these preferences might be associated with a notion of social welfare (utilitarianism), a concern for distributive justice or support for human rights.[7]

We assume that, fixing the probability candidate 1 wins the election, all agents prefer lower turnout since it produces lower social costs. Readers may wonder if this assumption is at odds with the fact that so many people worry about low turnout. Expressions of concern about low turnout need not represent a concern for turnout per se. Instead, people may be concerned by what low turnout signals about society e.g., lack of civic mindedness. *If* the winner of the election is fixed (independently of the votes) along with, for example, the level of civic-mindedness then a desire to maximize turnout seems hard to justify. Even so, including agents who prefer high turnout (or are indifferent about turnout) is not problematic. We only require that *some* agents prefer to minimize social cost. We do not include agents that prefer high turnout in this paper because these agents vote if and only if their cost to vote is smaller than their payoff for doing their part. In effect, we would be assuming what we are trying to explain.

2.1.1. Ranking of Rules

In the model section, we show that the ranking of rules is ultimately implied by the agents' preferences. Taking as given the behavior of types different than their own, type i agents rank rules as if they had used the following thought experiment. Assume that a rule e defines the actions that all type i agents should take. The outcome associated with e is the probability candidate i wins and an expected social cost of voting. The ranking of rules corresponds to the ranking of associated outcomes as given by the preferences type i.

There are several implications of these rankings that require explanation. First, rules are ranked according to the outcome produced when they define the actions that should be taken. Consider the following alternative: agents rank rules according to the outcome that would be produced if all agents in their group were

[7] See Roemer (1996) for an exposition of these ideas.

to act as they should under the rule. This alternative ranking would be written

$$R_1(e \mid s_2) \equiv wp(e, s_2) - \vartheta(\phi(e, s_2)).$$

This ranking might seem natural but it would generally make it impossible to find behavior profiles that satisfy our consistency requirement. A simple example illustrates the problem. Suppose that type 2 agents are all voting for candidate 2 and that the only way type 1 agents can win is if all such agents, including abstainers, vote for candidate 1. Then, if rules were ranked under the assumption that every type 1 agent acts as they should, the best rule would require all such agents to vote for 1 because by so doing 1 will be elected.[8] However, it is common knowledge among both types that 2 will be elected because abstainers (or ethical agents with voting costs higher than D) will not act as they should and vote for candidate 1. Instead they will abstain. Under this alternative ranking agents would be evaluating rules under premises they know to be false. This violates part 2 of our consistency definition. We avoid such contradictions by ensuring that when agents rank rules, they take into account agents' incentives for doing their part. Under our definition the best rule for agents type 1 in this example is for all of them to abstain.

We also assume that when agents rank rules they take as given the behavior of agents with preferences different from their own. In appendix C, we provide a formal foundation for this assumption. It will be helpful now to discuss informally what we might have done instead. At one extreme, we might have assumed an agent ranks rules taking as given the behavior of all other agents. In our model with a continuum of agents the act of an individual agent does not have any impact on either the probability a candidate wins or on the social cost of voting. So, the behavior of this agent can be exogenously determined. In suitably redefined model with countably many agents this agent would abstain. Since we already have a set of agents who abstain it would be redundant to include such types. At the other extreme, we might have assumed that some agents rank rules as if they define the actions everyone should take. A simple example will illustrate the problem with such an assumption. Under majority rule type 1 agents would rank as best a rule (applied to all agents) that would direct one agent (with cost $c < D$) to vote for candidate 1 and everyone else to abstain. Type 1 agents would evaluate this rule under the assumption that everyone (including type 2 agents) accepts the rule and behaves accordingly. The result would be that candidate 1 wins the

[8] Here we assume that voting costs are sufficiently low.

election at a social cost of 0. Analogously, type 2 agents would rank as best a rule that directs one agent to vote for candidate 2 and everyone else to abstain. This clearly violates part 2 of our definition of consistency. That is, agents would evaluate rules assuming that everyone will accept the same rule when they know this will not be the case. In Appendix C, we consider a model in which each agent could, a priori, take as given the behavior of *any* group of agents. We show that in our two candidate model, consistency implies that we can assume, without loss of generality, that agents take as given the behavior of agents with preferences different from their own.

3. 2PG

In this section, we show that there exists a one-to-one correspondence between a consistent behavior profile and a Nash-equilibrium, in pure strategies, of a suitably chosen two player game that we call 2PG. The *only* role of 2PG is to simplify the computation of a consistent behavior profile.

Proposition 1. *Let* $(s_1^*,\ s_2^*\ )$ *be a consistent behavior profile. There is a pair of cut-off points* $c_i^* \in (c^l,\ D]$, $i \in \{1,2\}$, *such that for almost all costs* $c \in [c^l,\ c^u]$*;*

$$\begin{aligned} s_i^*(c,D) = e_i^*(c) = \textit{vote for } i \quad & \textit{if } c \in [c^l,\ c_i^*]; \\ s_i^*(c,D) = e_i^*(c) = \textit{abstain} \quad & \textit{if } c \in (c_i^*,\ c^u]. \end{aligned}$$

Proof: *See Appendix A.*

Proposition 1 shows that if $(s_1^*,\ s_2^*)$ is a consistent behavior profile then there exists a pair of cut-off points $(c_1^*, c_2^*) \in (c_l, D)$ such that type i ethical agents should and will vote for i when their voting cost is smaller than c_i^*. They also should and will abstain when their voting cost is greater than c_i^*. Therefore, there is no loss of generality in assuming that a behavior profile can be defined by a pair of cut-off points.

The intuition behind this result is that the least costly way to achieve a given level of turnout is to have low-cost agents vote and high-cost agents abstain. The fact that the cut-off point c_i^*, $i \in \{1,2\}$, is strictly less than D follows from the fact that agents with costs above D necessarily abstain.

We now determine which pairs of cut-off points define consistent behavior. Assume that type i ethical agents vote if and only if their cost is below some cut-off point $c_i \leqslant D$. Then, the fraction of type i ethical agents who vote is

$$\sigma_i(c_i) \equiv \int_{c^l}^{c_i} f_{\tilde{c}}(x)\partial x.$$

The maximum fraction of type i ethical agents who vote is

$$\sigma^u \equiv \int_{c^l}^{\min\{D,c^u\}} f_{\tilde{c}}(x)\partial x.$$

If a fraction σ_i of type i ethical agents vote, the expected social cost of voting associated with this behavior is

$$C_i(\sigma_i) \equiv k_i q_i \int_{c^l}^{c_i(\sigma_i)} f_{\tilde{c}}(x)x\partial x.$$

where $c_i(\sigma_i)$ is the inverse function of $\sigma_i(c_i)$. The expected social cost of voting is

$$C_1(\sigma_1) + C_2(\sigma_2).$$

Candidate 1 wins the election if he gets the most votes. Hence, under the assumption that, for $i = 1, 2$, a fraction $\sigma_i > 0$ of type i ethical agents vote (for i), candidate 1 wins if

$$\tilde{k}\tilde{q}_1\sigma_1 \geqslant (1-\tilde{k})\tilde{q}_2\sigma_2 \;\Leftrightarrow\; \frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1} \leqslant \frac{\sigma_1}{\sigma_2}.$$

So, candidate 1 wins the election with probability

$$F_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1}{\sigma_2}\right),$$

where $F_{\tilde{x}}(x) \equiv P(\tilde{x} \leqslant x)$ is the cumulative distribution function of a random variable $\tilde{x}$.

Analogously, candidate 2 wins the election with probability

$$F_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\sigma_2}{\sigma_1}\right).$$

We now define 2PG. To avoid confusion, we call the decision makers in 2PG *players* while the voters in our underlying model are *agents or types*. Players do not represent anyone in the model. As mentioned above, 2PG is merely a device used to compute consistent behavior profiles.

A pure strategy for player $i \in \{1,2\}$ is a real number $\sigma_i \in (0, \sigma^u]$. The payoff for player 1 is given by

$$\Pi_1(\sigma_1 \mid \sigma_2) \equiv wF_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1}{\sigma_2}\right) - \vartheta\left(C_1(\sigma_1) + C_2(\sigma_2)\right).$$

The payoff for player 2 is given by

$$\Pi_2(\sigma_2 \mid \sigma_1) \equiv wF_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\sigma_2}{\sigma_1}\right) - \vartheta\left(C_1(\sigma_1) + C_2(\sigma_2)\right).$$

The pair (σ_1^*, σ_2^*) is an *equilibrium* if

$$\sigma_1^* = \arg\max_{\sigma_1 \in (0,\sigma_1^u]} \Pi_1(\sigma_1 \mid \sigma_2^*) \text{ and } \sigma_2^* = \arg\max_{\sigma_2 \in (0,\sigma_2^u]} \Pi_2(\sigma_2 \mid \sigma_1^*).$$

Proposition 2. *A pair of cut-points (c_1^*, c_2^*) defines consistent behavior if and only if the pair (σ_1^*, σ_2^*), $\sigma_i^* = \sigma_i(c_i^*)$, $i \in \{1,2\}$, is an equilibrium in 2PG.*

Proposition 2 shows that 2PG can be used to obtain a consistent behavior profile. Given an equilibrium (σ_1^*, σ_2^*) of 2PG, the pair of cut-off points $(c_1^*, c_2^*) = (c_1(\sigma_1^*), c_2(\sigma_2^*))$ defines consistent behavior. Conversely, any pair of cut-off points (c_1^*, c_2^*) that defines consistent behavior can be obtained by an equilibrium (σ_1^*, σ_2^*) of 2PG. Proposition 2 follows directly from the definitions above. Therefore, we omit the proof.

3.1. Example

In this subsection we consider the example in Feddersen and Sandroni (2005). The focus here, however, is to show the precise way in which the consistent rule profile $(\sigma_1^*,\ \sigma_2^*)$ can be derived, in closed-form solution, as a function of the parameters $(k, w, \bar{c}, D)$. In this example, the fraction of ethical agents in each type group, $\widetilde{q}_1$ and $\widetilde{q}_2$, are independent and uniformly distributed over $[0,1]$. Voting costs $\tilde{c}$ are also uniformly distributed over $[0,1]$. In addition, $\vartheta(x) = x$ and the fraction of type 1 agents $\widetilde{k}$ is deterministic and equal to $k < 0.5$, so candidate 1 is supported by a minority.

The cumulative distribution and density function of $\frac{\widetilde{q}_1}{\widetilde{q}_2}$ (F and f, respectively) are given by

$$\begin{array}{llll} \bar{F}(z) = \frac{z}{2} & \text{if } z \leqslant 1; & f(z) = \frac{1}{2} & \text{if } z \leqslant 1; \\ \bar{F}(z) = 1 - \frac{1}{2z} & \text{if } z \geqslant 1. & f(z) = \frac{1}{2z^2} & \text{if } z \geqslant 1. \end{array}$$

Let $\bar{k} \equiv \frac{k}{1-k}$. We first consider the case $D > \frac{\bar{c}}{2}$. If σ_1 is part of a consistent profile then the first order conditions of the maximization problem (for type 1) in the consistency requirement

$$wf\left(\bar{k}\frac{\sigma_1}{\sigma_2}\right)\bar{k}\frac{1}{\sigma_2} - k\left(\frac{\bar{c}}{2}\right)\sigma_1 \begin{cases} = 0 & \text{if} \quad \sigma_1 \in (0,1) \\ \geq 0 & \text{if} \quad \sigma_1 = 1. \end{cases}$$

must be satisfied.

Note that $f(z) = f(\frac{1}{z})\frac{1}{z^2}$. So, the first order conditions of the maximization problem (for type 2) can be written as

$$wf(\bar{k}\frac{\sigma_1}{\sigma_2})\frac{\bar{k}\sigma_1}{(\sigma_2)^2} - (1-k)\left(\frac{\bar{c}}{2}\right)\sigma_2 \begin{cases} = 0 & \text{if} \quad \sigma_2 \in (0,1) \\ \geq 0 & \text{if} \quad \sigma_2 = 1. \end{cases}$$

Remark 1. *There is no consistent profile such that $\sigma_2 = 1$ and $\sigma_1 < 1$*

To see this observe that the first order conditions imply the following contradiction:

$$k\left(\frac{\bar{c}}{2}\right)\sigma_1 = wf\left(\bar{k}\sigma_1\right)\bar{k} > wf(\bar{k}\sigma_1)\bar{k}\sigma_1 \geqslant (1-k)\left(\frac{\bar{c}}{2}\right) > k\left(\frac{\bar{c}}{2}\right)\sigma_1.$$

Thus, there are three types of consistent rule profiles to consider: interior ($\sigma_1 < 1$ and $\sigma_2 < 1$); a corner ($\sigma_1 = 1$ and $\sigma_2 = 1$); and a semi-corner ($\sigma_1 = 1$ and $\sigma_2 < 1$).

We first look for an interior profile. Dividing the first order conditions for σ_1 by condition for σ_2 yields $\frac{\sigma_1}{\sigma_2} = \sqrt{\frac{1}{\bar{k}}}$. This implies that the expected turnout of the majority is higher than the minority (i.e, $\bar{k}\frac{\sigma_1}{\sigma_2} < 1$) and, hence, $f(\bar{k}\frac{\sigma_1}{\sigma_2}) = 0.5$. This simplifies the first order conditions and gives an unique solution

$$\sigma_1^* = \sqrt{\frac{w}{\bar{c}}}\frac{1}{\sqrt[4]{k(1-k)}} \text{ and } \sigma_2^* = \sqrt{\frac{w}{\bar{c}}}\sqrt[4]{\frac{k}{(1-k)^3}}.$$

The requirement that $\sigma_1^* < 1$ and $\sigma_2^* < 1$ implies

$$\frac{\bar{c}}{w} > \frac{1}{\sqrt{k(1-k)}}.$$

So, an interior consistent profile only exists if the expected cost to vote (relative to the importance of the election) is sufficiently high. If this restriction is not satisfied we either have a semi-corner or a corner solution.

A semi-corner solution exists and is unique when the expected cost to vote (relative to the importance of the election) is at an intermediary level, satisfying

$$\frac{k}{(1-k)^2} < \frac{\bar{c}}{w} \leqslant \frac{1}{\sqrt{k(1-k)}}$$

To see this note that, in the semi-corner case, dividing the first order conditions for 1 by the condition for 2 yields $\sigma_2 \geqslant \sqrt{\bar{k}}$. So, $f(\bar{k}\frac{1}{\sigma_2}) = 0.5$. Hence, the unique solution is

$$\sigma_1^* = 1 \text{ and } \sigma_2^* = \sqrt[3]{\frac{wk}{\bar{c}(1-k)^2}}.$$

A corner solution is consistent if and only if $wf\left(\bar{k}\right)\bar{k} \geqslant k\left(\frac{\bar{c}}{2}\right)$ and $wf(\bar{k})\bar{k} \geqslant (1-k)\left(\frac{\bar{c}}{2}\right)$. However, $f\left(\bar{k}\right) = 0.5$. Therefore, $\sigma_1^* = \sigma_2^* = 1$ is consistent if and only if

$$\frac{\bar{c}}{w} \leq \frac{k}{(1-k)^2}.$$

We now consider the case $0 < D \leqslant \frac{\bar{c}}{2}$. Then a pair $(\sigma_1^*, \sigma_2^*) \in (0, \frac{2D}{\bar{c}}]\text{x}(0, \frac{2D}{\bar{c}}]$ is a consistent rule profile if

$$R_1(\sigma_1^*, \sigma_2^*) \geqslant R_1(\sigma_1, \sigma_2^*) \text{ for } \sigma_1 \in (0, \frac{2D}{\bar{c}}];$$

$$R_2(\sigma_1^*, \sigma_2^*) \geqslant R_2(\sigma_1^*, \sigma_2) \text{ for } \sigma_2 \in (0, \frac{2D}{\bar{c}}].$$

A consistent rule profile must can be obtained in an analogous way as the computation conducted above for the case $D > \frac{\bar{c}}{2}$. We, therefore, omit the algebra and report only the results. The result for this case are displayed in Table 1.

Table 1. Consistent rule profile for $D \leq \frac{\bar{c}}{2}$

$\sigma_1^* = \sqrt{\frac{w}{\bar{c}}}\frac{1}{\sqrt[4]{k(1-k)}}$, $\sigma_2^* = \sqrt{\frac{w}{\bar{c}}}\sqrt[4]{\frac{k}{(1-k)^3}}$	if	$\sqrt{\frac{w}{\bar{c}}}\frac{1}{\sqrt[4]{k(1-k)}} < \frac{D}{2\bar{c}}$
$\sigma_1^* = \frac{D}{2\bar{c}}$, $\sigma_2^* = \sqrt[3]{\frac{wkD}{2\bar{c}^2(1-k)^2}}$	if	$\sqrt[3]{\frac{wkD}{2\bar{c}^2(1-k)^2}} < \frac{D}{2\bar{c}} \leqslant \sqrt{\frac{w}{\bar{c}}}\frac{1}{\sqrt[4]{k(1-k)}}$
$\sigma_1^* = \sigma_2^* = \frac{D}{2\bar{c}}$	if	$\frac{D}{2\bar{c}} \leqslant \sqrt[3]{\frac{wkD}{2\bar{c}^2(1-k)^2}}$.

Direct investigations of this consistent profile reveals that the participation rates of the minority are greater than the participation rates of the majority (i.e., $\sigma_2^* \leq \sigma_1^*$), but the majority has a higher chance of winning the election (i.e., $k\sigma_1^* \leq (1-k)\sigma_2^*$). In section 5, we show that these properties hold in general.

The total expected turnout in the election is

$$T \equiv E(k\sigma_1^*\widetilde{q}_1 + (1-k)\sigma_2^*\widetilde{q}_2) = 0.5(k\sigma_1^* + (1-k)\sigma_2^*).$$

The expected margin of victory is:

$$MV \equiv E\left|\frac{(1-k)\widetilde{q}_2\sigma_2^* - k\widetilde{q}_1\sigma_1^*}{(1-k)\widetilde{q}_2\sigma_2^* + k\widetilde{q}_1\sigma_1^*}\right| =$$

$$\frac{k\sigma_1^*}{(1-k)\sigma_2^*}\left(2\ln 2 - 1 - \ln\left(1 + \frac{(1-k)\sigma_2^*}{k\sigma_1^*}\right)\right) + \frac{(1-k)\sigma_2^*}{k\sigma_1^*}\ln\left(1 + \frac{k\sigma_1^*}{(1-k)\sigma_2^*}\right).$$

The probability of victory for candidate 2 (supported by the majority) is:

$$PV \equiv F\left(\frac{\sigma_2^*(1-k)}{\sigma_1^* k}\right).$$

The expected margin of victory and the probability of victory for candidate 2 are both increasing functions of $\frac{(1-k)\sigma_2^*}{k\sigma_1^*} \geqslant 1$. Hence, the comparative statics results for them are identical.

It follows from the equations above (and the closed-form solutions for σ_1^* and σ_2^*) that T, MV and PV can be obtained as a function of the parameters $(k, w, \bar{c}, D)$. Direct investigation of these formulas show that the expected turnout (T) is increasing in the level of disagreement (k) while the expected margin of victory is decreasing in k. Hence, in this example, changes in the level of disagreement leads to an inverse relationship between expected turnout and margin of victory

4. Existence and Uniqueness of Equilibrium

In general, a pure strategy equilibrium may not exist in 2PG. For example, assume that fraction of the population in each group and the fractions of ethicals within the group is fixed, i.e., $\widetilde{k} = k$, $\widetilde{q}_1 = q_1$ and $\widetilde{q}_2 = q_2$ where $kq_1 > (1-k)q_2$. Then it is easy to see that an equilibrium in pure strategies does not exist. The problem is as follows: assume that the election is tied (i.e., $\sigma_1 k q_1 = \sigma_2(1-k)q_2$). Then, each candidate wins the election with equal probability and $\sigma_2 < \sigma_1 \leq 1$. So, there is an incentive for player 2 to slightly increase turnout and win the election. On the other hand, if the election is not tied then the winning player has an incentive to slightly decrease turnout so that it will still win, but at a smaller cost.

The absence of a pure strategy equilibrium in 2PG is especially problematic because mixed strategies in 2PG do not correspond to a consistent profile in the underlying game. A non-degenerate mixed strategy, for player i, in 2PG is a random variable distributed over $[0, \sigma^u]$ that does not put all weight on a single point. To see the problem of mapping a non-degenerate mixed strategy in 2PG into a consistent behavior profile first note that we don't consider rules in which agents randomize. Assume for the moment that we did allow rules that specify, for each agent, an independent random variable over the set of actions. In our model with a continuum of agents independent randomizations by each agent are equivalent to a rule with no randomization. Moreover, a rule in which a measurable set of agents actions are selected at random would not be maximally ranked. Consider a rule that requires a positive measure set of type i agents to randomize between voting and abstaining with probabilities strictly between zero and one. Next, consider an alternative rule which is different only for these types. Instead of randomizing, the low cost types vote and the high cost types abstain, with the cut-off point chosen so that the expected vote share is identical in both cases. It follows that, fixing the behavior of the other types, the probabilities of winning for each candidate are unchanged but the expected social cost of voting is strictly smaller under the alternative rule.

An alternative way to generate mixed strategy-like behavior would be to allow agents of the same type to choose different rules. For example, assume that for a given σ_2 there exist two possible values σ_1^A and σ_1^B that maximize Π_1. This is equivalent to saying that there are two rules (defined by the cut-off points $c_1(\sigma_1^A)$ and $c_1(\sigma_1^B)$) that maximize R_1, under the assumption that *all* type 1 agents accept this rule. If some type 1 agents were to choose one rule and the rest chooses the other then each type 1 agent would have chosen a maximally ranked rule given the behavior of type 2 agents. However, the behavior generated among all type 1 agents would be inconsistent with either rule. Type 1 agents would be evaluating rules under false assumptions. Part 2 of the consistency definition requires that all agents with the same type behave as if they had chosen the same rule to avoid this problem.

Finally, a mixed strategy in 2PG could be mapped into a (more broadly defined) rule that uses a private correlating device for some group. All types in each group would accept the rule announced by the correlating device. However, it would be necessary that only types in the group observe the realization of their correlating device and that agents in the other group cannot infer the realization of this device. The existence of such private correlating device is problematic in the context of a large election.

Given the problems associated with mixed strategies we now identify sufficient conditions for the existence and uniqueness of a pure strategy equilibrium in 2PG. Consequently, we show that a consistent behavior profile exists in a class of elections. All proofs for the results in this section are in appendix A.

A sufficient condition to guarantee the existence of an equilibrium in pure strategies is that the payoff functions for both players are strictly concave. Strict concavity implies that the best response for each player is a continuous function of the other players' strategy. This in turn guarantees existence of an equilibrium through a fixed point argument.[9] Fortunately, it turns out that the negative expected aggregate voting cost functions are strictly concave.

Lemma 1. *The function $C_i : (0, \sigma^u] \to \Re_+$ is strictly increasing, twice continuously differentiable and strictly convex.*

The intuition behind this result is that turnout increases as the cut-point for voting increases. The marginal cost increases as the cut-point increases because higher cost types are required to vote. So, marginal costs increases with turnout.

By definition, the expected probability of winning functions for each candidate are concave if the random variables

$$\tilde{r} \equiv \frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1} \text{ and } \frac{1}{\tilde{r}} = \frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}$$

are *concave* as defined below.

Definition 2. *A random variable, $\tilde{x}$, that takes values on a subset of $\Re_+$, is concave if the cumulative density function $F_{\tilde{x}}(x)$ is a concave function of x on $\Re_+$.*

Any random variable on $\Re_+$ whose density function is non-increasing is concave. For example, the uniform distribution over (0, 1) is concave.

The payoff function of each player is the probability of winning minus the expected social cost of voting. So, if $\tilde{r}$ and $\frac{1}{\tilde{r}}$ are concave then the payoff function

[9] A technical difficulty with this argument is that the best response functions are defined in a non-compact set $(0, \sigma_i^u]$. We solve this problem in appendix A.

of both player are strictly concave. The next lemma is critical in determining when $\tilde{r}$ and $\frac{1}{\tilde{r}}$ are *both* concave.

Lemma 2. *Let $\tilde{x}$ and $\tilde{y}$ be two independent random variables taking values on $\Re_+$. If $\tilde{x}$ is concave then $\tilde{x}\tilde{y}$ is concave.*

The key to lemma 1 is that no assumptions over the distribution of $\tilde{y}$ is required. For this result the assumption of independence can be relaxed. It is sufficient to assume that $\tilde{x}$ conditional on $\tilde{y} = y$ is concave for every $y \in \Re_+$.

Lemma 3. *If $\tilde{k}$ is concave then $\frac{\tilde{k}}{1-\tilde{k}}$ is concave.*

Assumption A: $\tilde{k}$, $\tilde{q}_1$ and $\tilde{q}_2$ are independent random variables. At least one of the pairs $(\tilde{q}_1, \tilde{q}_2)$, $\left(\tilde{k}, \tilde{q}_2\right)$, $\left(1 - \tilde{k}, \tilde{q}_1\right)$ is such that both random variables are concave.

Given lemmas 2 and 3, the concavity of $\tilde{r}$ and $\frac{1}{\tilde{r}}$ follows immediately from assumption A. Existence of a pure strategy equilibrium in 2PG is established.

Clearly, assumption A is satisfied if, as assumed in section 3.1, $\tilde{k}$ is deterministic while $\tilde{q}_1$ and $\tilde{q}_2$ are independent and uniformly distributed over $(0, 1)$.

Proposition 3. *Under assumption A, a pure strategy equilibrium exists in 2PG and is unique.*

The uniqueness proof is involved and we do not have a simple intuition to offer for it. This result is in sharp contrast with models of fairness (see, for example, Sugden (1984)) in which agents are motivated to do their part if and only if they believe some others agents will also do their part. These models tend to have multiple equilibria (e.g. one in which nobody does their part and another some do). Our model does not necessarily have this multiplicity problem because agents take into account the incentives of other agents to do their part when they evaluate behavioral rules.

4.1. Comments on Assumption A

Assumption A is sufficient but not necessary to guarantee existence of a pure strategy equilibrium in 2PG. For example, assume that $\tilde{k} = 0.5$, $\tilde{q}_1 = \tilde{q}_2 = 1$ and $\tilde{c} \sim U(0, 1)$ then $\sigma_1^* = \sigma_2^* = 1$ is an equilibrium. A pure strategy equilibrium can be obtained under weaker assumptions than assumption A. As mentioned above, the assumption that $\tilde{q}_1$ and $\tilde{q}_2$ are independent can be relaxed. It suffices that the conditional on $\tilde{q}_j = q_j$, $j \neq i$, $\tilde{q}_i$ is concave.

Proposition 3 is sharp in the sense that under small relaxations of the concavity assumptions may lead to elections where a consistent profile fails to exists (see example 1 in Appendix A). However, the concavity assumption can be replaced by other distributional assumptions. For example, it would suffice to assume that the density function of $\tilde{q}_i$, $i \in \{1, 2\}$, is single peaked (a proof is available upon request).[10]

4.2. Role of Uncertainty

If the only source of uncertainty in the model was over the size of the groups ($\tilde{k}$) then $\tilde{r}$ and $\frac{1}{\tilde{r}}$ would not be concave for all $k \in (0, 1)$. It is important to have at least two sources of uncertainty in order to ensure existence of equilibrium in 2PG. This is the rationale for introducing uncertainty over the fraction of ethical agents in each group. However, it is reasonable to assume that there are other independent sources of uncertainty. For example, the probability a vote is counted for a candidate may be itself a random variable. Let $\tilde{\lambda}_i$ be the fraction of votes for candidate $i \in \{1, 2\}$ that are counted.[11] The probability of winning the election for each candidate will be concave if

$$\frac{(1-\tilde{k})\tilde{q}_2\tilde{\lambda}_2}{\tilde{k}\tilde{q}_1\tilde{\lambda}_1} \text{ and } \frac{\tilde{k}\tilde{q}_1\tilde{\lambda}_1}{(1-\tilde{k})\tilde{q}_2\tilde{\lambda}_2}$$

are *concave* random variables. It follows immediately from Lemma 1 that these two random variables are concave as long as $\tilde{\lambda}_i$ are concave and independent from $\tilde{q}_i$ and $\tilde{k}_i$, $i \in \{1, 2\}$. Lemma 1 also shows that, under assumption A, these two variables are concave if $\tilde{\lambda}_i$ is independent from $\tilde{q}_i$ and $\tilde{k}_i$, no matter what is the distribution of $\tilde{\lambda}_i$.

5. Comparative Statics Results

In this section, we examine the general properties of elections in our model. To simplify the exposition, we assume that the density of the costs of voting are

[10] We would also require that $c^l = 0$.

[11] Here we assume that voters do not know whether their vote will count or not. Agents who know that their vote will not be counted will abstain. This follows not because the probability that their vote is pivotal is zero, but because these voters reason they should abstain.

identical for both groups; that the fraction of agents in group 1 $\tilde{k}$ is a constant k; and that the average fraction of ethical agents in each group is the same and equal to $q \in (0,1)$. All proofs for the results in this section are in appendix B.

Proposition 4. *The participation rate of the minority is greater than the majority, but the majority has a higher chance of winning the election than the minority. So, if $k < 0.5$ then $\sigma_1^* \geqslant \sigma_2^*$ and $k\sigma_1^* \leqslant (1-k)\sigma_2^*$.*

In the Palfrey and Rosenthal (1983, 1985) model, even if the majority is overwhelmingly large, the minority may be just as likely to win a large election as the majority (See Campbell (1999), Borgers (2004), Krasa and Polborn (2004) and Taylor and Tildram (2005)). In the purely decision theoretic model of Riker and Ordeshook (1968), the participation rate of the majority is equal to the minority (they are both $min\{\frac{2D}{\bar{c}}, 1\}$). Proposition 4 shows that our model delivers results that differ from the models of Palfrey and Rosenthal and Riker and Ordeshook.

Proposition 5. *Expected turnout goes to zero as one of the groups become an overwhelming majority. So, T goes to 0 as k goes to 0.*

When the size of the minority becomes arbitrarily small our model becomes like Harsanyi's in which there is essentially only one type of preference. In that case, turnout goes to zero. Proposition 5 shows the importance of heterogeneity of preferences as a factor in explaining turnout. It also demonstrates why Harsanyi's model will not be able to explain turnout.

Proposition 6. *Under assumption A, expected turnout T is non-decreasing in the importance of the election (w).*[12]

Proposition 6 is consistent with the evidence that turnout is higher for presidential elections than state elections (see Teixeira (1987)).

Proposition 7. *Under assumption A, the expected margin of victory is decreasing in the level of disagreement (k).*

[12] An alternative version of this result which does not require assumption A but places some restrictions on the distribution of cost (e.g. a uniform distribution) is available from the authors upon request.

In the main example in section 3.1, expected margin of victory is decreasing in the level of disagreement. This produces an inverse correlation between margin of victory and turnout that has often been the focus of study in the empirical literature (see Matsusaka (1991), Nalebuff and Shachar (1999) and Blais (2000)). However, this result does not generalize. Example 2 (see Appendix B) shows that expected turnout may decrease as the disagreement level increases from a point to another. However, expected turnout cannot monotonically decrease in the whole range of disagreement levels because (by proposition 5) it approaches zero when the level of disagreement vanishes.

6. Conclusion

We have developed a model in which the actions agents should take are determined endogenously and influence payoffs. The application of this methodology to the problem of turnout in large elections provides testable implications and predicts variations in expected turnout and margin of victory as a function of costs to vote, level of disagreement within the society, importance of the election and agents' incentives to do their part. Coate and Conlin (2002) have structurally estimated our model and compared it to alternative models of turnout, with encouraging results. We believe that the methodology developed here can be applied in a variety of settings providing new theoretical results and explaining empirical observations that appear to be inconsistent with rational actor models.

7. Appendix A: Proofs

Proof of proposition 1 Without loss generality, we restrict attention to the case $D > c^u$. The proof for the case in which $D \leqslant c^u$ is analogous. Assume, by contradiction, that there is a strictly positive set of costs C such that $e_i^*(c) = \{\text{vote for } j\}$ if $c \in C$, $j \neq i$. Consider an alternative rule in which agents who are required to vote for j are instead required to vote for i. This rule would be ranked higher by type i than e_i^* because the expected voting cost is the same and the chances that i wins the election is increased. Therefore, almost surely, $e_i^*(c) = \{\text{vote for } i \text{ or abstain}\}$.

Let c_i^* be the largest cost in $[c^l, c^u]$ such that for almost every cost $c \leqslant c_i^*$, $e_i^*(c) = \{\text{vote for } i\}$. The proof is clearly complete if $c_i^* = c^u$. So, consider the case $c_i^* < c^u$. The proof would also be complete if $e_i^*(c) = \{\text{abstain}\}$ for almost all costs $c > c_i^*$. So, consider a strictly positive measure set of costs $\bar{C} \subseteq (c_i^*, c^u]$ such that

$e_i^*(c) = \{\text{vote for } i\}$ if $c \in \bar{C}$. There exists $\bar{c}_i > c_i^*$ such that $e_i^*(c) = \{\text{vote for } i\}$ if $c \in \bar{C} \bigcap (\bar{c}_i, c^u]$. By the definition of c_i^* there is a strictly positive measure set of costs $\ddot{C}$ in $(c_i^*, \bar{c}_i)$ such that $e_i^*(c) = \{\text{abstain}\}$ if $c \in \ddot{C}$. Pick a subset of $\ddot{C}$ and $\bar{C} \bigcap (\bar{c}_i, c^u]$ with the same strictly positive measure. Consider an alternative rule in which the required actions for the costs in $\ddot{C}$ and $\bar{C} \bigcap (\bar{c}_i, c^u]$ are exchanged, but is otherwise identical to e_i^*. This rule would have a higher ranking than e_i^* because the expected voting cost is reduced and the chance that i wins the election is the same. This contradicts the definition of e_i^*. ■

Proof of lemma 1: Let $\bar{D} = \min\{D, c^u\}$. Let $c_i : [0, \sigma^u] \to [c^l, c^u]$ be the inverse of $\sigma_i : (c^l, \bar{D}] \to (0, \sigma^u]$. Let $H_i(c_i) = k_i q_i \int_{c^l}^{c^u} f_{\tilde{c}_i}(x) x \partial x$. By definition, $C_i(\sigma_i) = H_i(c_i(\sigma_i))$. So, in the interval $(0, \sigma^u)$,

$$C_i'(\sigma_i) = H_i'(c_i(\sigma_i)) \cdot c_i'(\sigma_i) = k_i q_i f_{\tilde{c}_i}(c_i(\sigma_i)) \cdot c_i(\sigma_i) \frac{1}{f_{\tilde{c}_i}(c_i(\sigma_i))} = k_i q_i c_i(\sigma_i) > 0.$$

Therefore, $C_i''(\sigma_i) = \frac{k_i q_i}{f_{\tilde{c}_i}(c_i(\sigma_i))} > 0$. ■

Proof of lemma 2:

$$F_{\tilde{x}\tilde{y}}(z) = P(\tilde{x}\tilde{y} \leqslant z) = \int_0^\infty P(\tilde{x} \leqslant \frac{z}{y} \mid \tilde{y} = y) dF_{\tilde{y}}(y) = \int_0^\infty F_{\tilde{x}}(\frac{z}{y}) dF_{\tilde{y}}(y).$$

If $\lambda \in [0, 1]$, $z \in \Re_+$ and $\bar{z} \in \Re_+$ then, $F_{\tilde{x}\tilde{y}}(\lambda z + (1-\lambda)\bar{z}) =$

$$\int_0^\infty F_{\tilde{x}}(\frac{\lambda z + (1-\lambda)\bar{z}}{y}) dF_Y(y) = \int_0^\infty F_{\tilde{x}}(\lambda \frac{z}{y} + (1-\lambda)\frac{\bar{z}}{y}) dF_{\tilde{y}}(y) \leqslant$$

$$\int_0^\infty \left(\lambda F_{\tilde{x}}(\frac{z}{y}) + (1-\lambda) F_{\tilde{x}}(\frac{\bar{z}}{y}) \right) dF_{\tilde{y}}(y) =$$

$$\lambda \int_0^\infty F_{\tilde{x}}(\frac{z}{y}) dF_{\tilde{y}}(y) + (1-\lambda) \int_0^\infty F_{\tilde{x}}(\frac{\bar{z}}{y}) dF_{\tilde{y}}(y) = \lambda F_{\tilde{x}\tilde{y}}(z) + (1-\lambda) F_{\tilde{x}\tilde{y}}(\bar{z}).$$

■

Proof of lemma 3 By assumption $F_{\tilde{k}}$ is concave. So, $F_{\tilde{k}}$ is differentiable except, at most, countably many points and the right and left derivative are well defined and decreasing. Moreover,

$$F_{\frac{\tilde{k}}{1-\tilde{k}}}(x) = P\left(\frac{\tilde{k}}{1-\tilde{k}} \leqslant x \right) = P\left(\tilde{k} \leqslant \frac{x}{1+x} \right) = F_{\tilde{k}}(\frac{x}{1+x}).$$

So, the right and left derivative of $F_{\frac{\tilde{k}}{1-\tilde{k}}}$ and $F_{\tilde{k}}$ are such that

$$F'_{\frac{\tilde{k}}{1-\tilde{k}}}(x) = F'_{\tilde{x}}(\frac{x}{1+x})(\frac{1}{1+x})^2.$$

So, $F_{\frac{\tilde{k}}{1-\tilde{k}}}$ is differentiable except, at most, countably many points and the right and left derivative are well defined and decreasing. So, $F_{\frac{\tilde{k}}{1-\tilde{k}}}$ is concave.

■

Lemma 4. *Let $\tilde{x}$ be a random variable taking values on $\Re_+$. If $\tilde{x}$ and $\frac{1}{\tilde{x}}$ are concave then $\tilde{x}$ and $\frac{1}{\tilde{x}}$ have continuous and strictly positive density functions.*

Proof : By assumption $F_{\tilde{x}}$ is concave. So, $F_{\tilde{x}}$ is differentiable except, at most, countably many points and the right and left derivatives of $F_{\tilde{x}}$ are well defined in all points. Hence, $\tilde{x}$ has a non-increasing density function $f_{\tilde{x}}$ which is well defined at any point $x \in \Re_+$ such that $F_{\tilde{x}}$ is differentiable. In these points, $f_{\tilde{x}}(x) = F'_{\tilde{x}}(x)$. By analogous argument, $\frac{1}{\tilde{x}}$ has a non-increasing density function $f_{\frac{1}{\tilde{x}}}$.

Let $x \in \Re_+$. By definition,

$$F_{\frac{1}{\tilde{x}}}(x) = P\left(\frac{1}{\tilde{x}} \leqslant x\right) = P\left(\tilde{x} \geqslant \frac{1}{x}\right) = 1 - F_{\tilde{x}}(\frac{1}{x}).$$

So, $F_{\frac{1}{\tilde{x}}}$ is differentiable at $x \in \Re_+$ if and only if $F_{\tilde{x}}$ is differentiable at $\frac{1}{x} \in \Re_+$. Moreover, whenever the density functions are defined,

$$f_{\frac{1}{\tilde{x}}}(x) = F'_{\frac{1}{\tilde{x}}}(x) = F'_{\tilde{x}}(\frac{1}{x})\frac{1}{x^2} = f_{\tilde{x}}(\frac{1}{x})\frac{1}{x^2}.$$

The density functions $f_{\tilde{x}}$ and $f_{\frac{1}{\tilde{x}}}$ are positive, non-increasing functions. They can be discontinuous at, at most, countably many points. This discontinuity must be of the form of a "jump." Assume, by contradiction, that $f_{\tilde{x}}$ is not continuous at $a \in \Re_+$. Then,

$$\lim_{x \uparrow a} f_{\tilde{x}}(x) > \lim_{x \downarrow a} f_{\tilde{x}}(x),$$

where by $\lim_{x \uparrow a} f_{\tilde{x}}(x)$ and $\lim_{x \downarrow a} f_{\tilde{x}}(x)$ it is meant the limit of $f_{\tilde{x}}(x)$ when $x > a \in \Re_+$ approaches $a \in \Re_+$, and the limit of $f_{\tilde{x}}(x)$ when $x < a \in \Re_+$ approaches $a \in \Re_+$, respectively. Therefore,

$$\lim_{x \downarrow \frac{1}{a}} f_{\frac{1}{\tilde{x}}}(x) \;=\; \lim_{x \downarrow \frac{1}{a}} f_{\tilde{x}}(\frac{1}{x})\frac{1}{x^2} \;=\; \lim_{\frac{1}{x} \uparrow a} f_{\tilde{x}}(\frac{1}{x})a^2 >$$

$$\lim_{\frac{1}{x}\downarrow a} f_{\tilde{x}}(\frac{1}{x})a^2 = \lim_{x\uparrow \frac{1}{a}} f_{\tilde{x}}(\frac{1}{x})\frac{1}{x^2} = \lim_{x\uparrow \frac{1}{a}} f_{\frac{1}{\tilde{x}}}(x) \Longrightarrow \lim_{x\downarrow \frac{1}{a}} f_{\frac{1}{\tilde{x}}}(x) > \lim_{x\uparrow \frac{1}{a}} f_{\frac{1}{\tilde{x}}}(x).$$

This contradicts the fact that $f_{\frac{1}{\tilde{x}}}$ is non-increasing. By analogous argument, $f_{\frac{1}{\tilde{x}}}$ is also a continuous function.

To demonstrate that the density functions are strictly positive, assume, by contradiction, that $f_{\tilde{x}}(a) = 0$ at some point $a \in \Re_+$. Then, $f_{\tilde{x}}(\frac{1}{x}) = 0$ for all $x \in \Re_+$ such that $\frac{1}{x} \geqslant a$. Therefore, $f_{\frac{1}{\tilde{x}}}(x) = f_{\tilde{x}}(\frac{1}{x})\frac{1}{x^2} = 0$ for all $x \leqslant \frac{1}{a}$. However, $f_{\frac{1}{\tilde{x}}}$ is non-increasing. Therefore, $f_{\frac{1}{\tilde{x}}}(x) = 0$ for all $x \in \Re_+$. A contradiction.

■

A direct corollary of lemmas $1 - 4$ is that under assumption A, both players payoff functions are strictly concave and continuously differentiable in 2PG.

Proof of proposition 3. Existence. Consider an "$\varepsilon-$election" in which there are three types: partisan, ethical, and abstainers. Partisans always vote. The proportions of partisans in group i is $\tilde{q}_i\varepsilon$, $\varepsilon > 0$. The proportions of ethical types is $\tilde{q}_i(1-\varepsilon)$. If σ_i is the fraction of ethical agents in group $i \in \{1, 2\}$ who vote then $\hat{\sigma}_i = \sigma_i(1-\varepsilon) + \varepsilon$ is the fraction of agents in group i who vote. Consider an $\varepsilon-$2PG. Player 1 chooses $\sigma_1 \in [0, \sigma^u]$. Player 2 chooses $\sigma_2 \in [0, \sigma^u]$. The payoff for player 1 is given by

$$\Pi_1^{\varepsilon}(\sigma_1 \mid \sigma_2) \equiv wF_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\hat{\sigma}_1}{\hat{\sigma}_2}\right) - \vartheta\left(C_1(\hat{\sigma}_1) + C_2(\hat{\sigma}_2)\right).$$

The payoff for player 2 is given by

$$\Pi_2^{\varepsilon}(\sigma_2 \mid \sigma_1) \equiv wF_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right) - \vartheta\left(C_1(\hat{\sigma}_1) + C_2^{\varepsilon}(\hat{\sigma}_2)\right).$$

Given that the payoff functions of both players are strictly concave and the action space is compact, there exists an equilibrium $(\sigma_1^{\varepsilon *}, \sigma_2^{\varepsilon *})$ in $\varepsilon-$2PG.

Let the derivative of $\Pi_1^{\varepsilon}(\sigma_1 \mid \sigma_2)$ be $g_1^{\varepsilon}(\sigma_1 \mid \sigma_2) \equiv (1-\varepsilon)\cdot$

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1(1-\varepsilon)+\varepsilon}{\sigma_2(1-\varepsilon)+\varepsilon}\right)\frac{1}{\sigma_2(1-\varepsilon)+\varepsilon} - \vartheta'\left(C_1(\sigma_1) + C_2(\sigma_2)\right)\cdot C_1'(\sigma_1).$$

Let the derivative of $\Pi_2^{\varepsilon}(\sigma_2 \mid \sigma_1)$ be $g_2^{\varepsilon}(\sigma_2 \mid \sigma_1) \equiv (1-\varepsilon)\cdot$

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1(1-\varepsilon)+\varepsilon}{\sigma_2(1-\varepsilon)+\varepsilon}\right)\frac{(\sigma_1(1-\varepsilon)+\varepsilon)}{(\sigma_2(1-\varepsilon)+\varepsilon)^2} - \vartheta'\left(C_1(\sigma_1) + C_2(\sigma_2)\right)\cdot C_2'(\sigma_2).$$

Consider a convergent subsequence of $(\sigma_1^{\varepsilon*},\, \sigma_2^{\varepsilon*})$ with limits $(\sigma_1^*,\, \sigma_2^*)$ as $\varepsilon \to 0$. Assume, by contradiction, that $\sigma_1^* = 0$ and $\sigma_2^* = 0$. Without loss of generality, assume that there exists an infinite subsequence such that

$$\frac{\sigma_2^{\varepsilon*}(1-\varepsilon)+\varepsilon}{\sigma_1^{\varepsilon*}(1-\varepsilon)+\varepsilon} \leqslant 1.$$

A completely analogous argument applies for an infinite subsequence in which this inequality is reversed. However, in this subsequence

$$\lim_{\varepsilon\to 0} g_1^{\varepsilon}(\sigma_1^{\varepsilon*} \mid \sigma_2^{\varepsilon*}) \geqslant \left(f_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}(1) \right) \cdot (\infty) - \vartheta'(0) \cdot C_2'(0) = \infty.$$

This is a contradiction because by the first order conditions of player 1, if $\sigma_1^{\varepsilon*} < \sigma_1^u$ then $g_2^{\varepsilon}(\sigma_2^{\varepsilon*} \mid \sigma_1^{\varepsilon*}) \leqslant 0$.

Assume, by contradiction, that $\sigma_1^* = 0$ and $\sigma_2^* > 0$. The first order conditions of player 1 imply that if ε is small enough, $g_1^{\varepsilon}(\sigma_1^{\varepsilon*} \mid \sigma_2^{\varepsilon*}) \leqslant 0$. So,

$$\lim_{\varepsilon\to 0} g_1^{\varepsilon}(\sigma_1^{\varepsilon*} \mid \sigma_2^{\varepsilon*}) = f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(0)\, \frac{1}{\sigma_2^*} - \vartheta'(0) \cdot C_1'(0) \leqslant 0 \Rightarrow f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(0) < \infty.$$

The first order conditions for group 2 imply that if ε is small enough, $g_2^{\varepsilon}(\sigma_2^{\varepsilon*} \mid \sigma_1^{\varepsilon*}) \geqslant 0$. So, as ε and $\sigma_1^{\varepsilon*}$ goes to zero and $\sigma_2^{\varepsilon*}$ goes to σ_2^*,

$$\lim_{\varepsilon\to 0} g_2^{\varepsilon}(\sigma_2^{\varepsilon*} \mid \sigma_1^{\varepsilon*}) = f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(0) \cdot 0 - \vartheta'\left(C_2(\sigma_2^*)\right) \cdot C_2'(\sigma_2^*) \geqslant 0.$$

This implies that $\vartheta'\left(C_2(\sigma_2^*)\right) \cdot C_2'(\sigma_2^*) \leqslant 0$ which is a contradiction because both $\vartheta'\left(C_2(\sigma_2^*)\right)$ and $C_2'(\sigma_2^*)$ are strictly positive. By analogous argument, the assumption that $\sigma_1^* > 0$ and $\sigma_2^* = 0$ also leads to a contradiction. So, $\sigma_1^* > 0$ and $\sigma_2^* > 0$. Therefore, by continuity, $\lim_{\varepsilon\to 0}(\sigma_1^{\varepsilon*}, \sigma_2^{\varepsilon*}) = (\sigma_1^*, \sigma_2^*)$ is an equilibrium of 2PG.

Uniqueness Let the derivative of $\Pi_1(\sigma_1 \mid \sigma_2)$ be $g_1(\sigma_1 \mid \sigma_2) \equiv$

$$w f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1}{\sigma_2}\right) \frac{1}{\sigma_2} - \vartheta'\left(C_1(\sigma_1) + C_2(\sigma_2)\right) \cdot C_1'(\sigma_1).$$

Let the derivative of $\Pi_2(\sigma_2 \mid \sigma_1)$ be $g_2(\sigma_2 \mid \sigma_1) \equiv$

$$w f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1}{\sigma_2}\right) \frac{\sigma_1}{(\sigma_2)^2} - \vartheta'\left(C_1(\sigma_1) + C_2(\sigma_2)\right) \cdot C_2'(\sigma_2) =$$

$$wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\sigma_2}{\sigma_1}\right)\frac{1}{\sigma_1} - \vartheta'\left(C_1(\sigma_1)+C_2(\sigma_2)\right)\cdot C_2^{'}(\sigma_2).$$

In equilibrium, $\sigma_1^* > 0$ and $\sigma_2^* > 0$ and the first order condition of both players optimization problem must be satisfied. That is,

$$g_1(\sigma_1^* \mid \sigma_2^*) = 0 \text{ if } \sigma_1^* < \sigma^u;\ g_1(\sigma_1^* \mid \sigma_2^*) \geqslant 0 \text{ if } \sigma_1^* = \sigma^u \tag{7.1}$$

and

$$g_2(\sigma_2^* \mid \sigma_1^*) = 0 \text{ if } \sigma_2^* < \sigma^u;\ g_2(\sigma_2^* \mid \sigma_1^*) \geqslant 0 \text{ if } \sigma_2^* = \sigma^u. \tag{7.2}$$

An interior equilibrium is one in which $\sigma_1^* < \sigma^u$ and $\sigma_2^* < \sigma^u$.

Fact 1. Assume that $(\sigma_1^*,\ \sigma_2^*)$ is an interior equilibrium. Then, there is no other interior equilibrium.

Proof : Let $(\bar{\sigma}_1,\ \bar{\sigma}_2)$ is another interior equilibrium. Let $\bar{r}$ be $\frac{\bar{\sigma}_1}{\bar{\sigma}_2}$. Let r^* be $\frac{\sigma_1^*}{\sigma_2^*}$. Let $h(\sigma_1,\sigma_2) = \vartheta'\left(C_1(\sigma_1)+C_2(\sigma_2)\right)$ and note that we will use this notation from here on. Dividing equations (9.1) and (9.2) we get

$$\frac{\sigma_2^*}{\sigma_1^*} = \frac{C_1^{'}(\sigma_1^*)}{C_2^{'}(\sigma_2^*)} \text{ and } \frac{\bar{\sigma}_2}{\bar{\sigma}_1} = \frac{C_1^{'}(\bar{\sigma}_1)}{C_2^{'}(\bar{\sigma}_2)}.$$

So, we have only two cases to consider. a) $\bar{\sigma}_2 > \sigma_2^*$ and $\bar{\sigma}_1 > \sigma_1^*$. b) $\bar{\sigma}_2 < \sigma_2^*$ and $\bar{\sigma}_1 < \sigma_1^*$. We will restrict the proof to (a). The proof for case (b) is analogous.

By equations (9.1) and (9.2),

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(r^*) = h(\sigma_1^*,\sigma_2^*)C_1^{'}(r^*\sigma_2^*)\sigma_2^* \text{ and } wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(\bar{r}) = h(\bar{\sigma}_1,\bar{\sigma}_2)C_1^{'}(\bar{r}\bar{\sigma}_2)\bar{\sigma}_2.$$

Assume, by contradiction, that $\bar{r} \geqslant r^*$. Then, $h(\bar{\sigma}_1,\bar{\sigma}_2)C_1^{'}(\bar{r}\bar{\sigma}_2)\bar{\sigma}_2 =$

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(\bar{r}) \leqslant wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(r^*) = h(\sigma_1^*,\sigma_2^*)C_1^{'}(r^*\sigma_2^*)\sigma_2^* < h(\bar{\sigma}_1,\bar{\sigma}_2)C_1^{'}(\bar{r}\bar{\sigma}_2)\bar{\sigma}_2.$$

Therefore, $\bar{r} < r^*$. From equation (9.2) we get $wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{1}{\bar{r}}\right) =$

$$h(\bar{\sigma}_1,\bar{\sigma}_2)\bar{\sigma}_1 C_2^{'}(\bar{\sigma}_2) > h(\sigma_1^*,\sigma_2^*)\sigma_1^* C_2^{'}(\sigma_2^*) = wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{1}{r^*}\right) \Rightarrow r^* < \bar{r}.$$

This is a contradiction because we have already shown that $\bar{r} < r^*$.

Fact 2 Assume that $\sigma_1^* = \sigma_2^* = \sigma^u$ is a equilibrium. Then, it is the only equilibrium.

Proof: Equations 9.1 and 9.2 imply that

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(1)\frac{1}{\sigma^u} \geqslant h(\sigma^u,\sigma^u)C_1^{'}(\sigma^u) \text{ and } wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}(1) \geqslant h(\sigma^u,\sigma^u)C_2^{'}(\sigma^u).$$

Assume, by contradiction, that $\bar{\sigma}_1 = \sigma^u$ and $\bar{\sigma}_2 < 1$ is an equilibrium. Then,

$$wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}(1) \leqslant wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\bar{\sigma}_2}{\sigma^u}\right)\frac{1}{\sigma^u} = h(\sigma^u,\bar{\sigma}_2)C_2^{'}(\bar{\sigma}_2) < h(\sigma^u,\sigma^u)C_2^{'}(\sigma^u).$$

This is a contradiction. By analogous argument, the assumption that $\bar{\sigma}_1 < 1$ and $\bar{\sigma}_2 = 1$ is an equilibrium also leads to a contradiction. Assume that $(\bar{\sigma}_1,\ \bar{\sigma}_2)$ is an interior equilibrium. Then,

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\bar{\sigma}_1}{\bar{\sigma}_2}\right)\frac{1}{\bar{\sigma}_2} = h(\bar{\sigma}_1,\bar{\sigma}_2)C_1^{'}(\bar{\sigma}_1) \ \leqslant h(\sigma^u,\sigma^u)C_1^{'}(\sigma^u) \leqslant wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(1)\frac{1}{\sigma^u}.$$

Therefore, $\bar{\sigma}_1 > \bar{\sigma}_2$. Moreover,

$$wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\bar{\sigma}_2}{\bar{\sigma}_1}\right)\frac{1}{\bar{\sigma}_1} = h(\bar{\sigma}_1,\bar{\sigma}_2)C_2^{'}(\bar{\sigma}_2) \leqslant h(\sigma^u,\sigma^u)C_2^{'}(\sigma^u) \leqslant wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}(1).$$

Hence, $\bar{\sigma}_2 > \bar{\sigma}_1$. A contradiction.

Fact 3. If $\sigma_1^* = \sigma^u$, $\sigma_2^* < \sigma^u$ is a equilibrium then $\bar{\sigma}_1 < \sigma^u$, $\bar{\sigma}_2 = \sigma^u$ is not a equilibrium.

Proof : Assume that (σ_1^*,σ_2^*) and $(\bar{\sigma}_1,\bar{\sigma}_2)$ are equilibria. By 9.1 and 9.2,

$$wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma^u}{\sigma_2^*}\right)\frac{1}{\sigma_2^*} \geqslant h(\sigma^u,\sigma_2^*)C_1^{'}(\sigma^u);\ wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma^u}{\sigma_2^*}\right)\frac{1}{\sigma_2^*} = h(\sigma^u,\sigma_2^*)C_2^{'}(\sigma_2^*)\frac{\sigma_2^*}{\sigma^u}.$$

Therefore, $C_2^{'}(\sigma_2^*) \geqslant C_1^{'}(\sigma^u)$. By the same argument, $C_1^{'}(\bar{\sigma}_1) \geqslant C_2^{'}(\sigma^u)$. Hence, $C_2^{'}(\sigma_2^*) \geqslant C_2^{'}(\sigma^u)$. A contradiction.

Fact 4. If $(\sigma_1^*,\ \sigma_2^*)$ is an interior equilibrium then $\bar{\sigma}_1 = \sigma^u$, $\bar{\sigma}_2 < \sigma^u$ is not a equilibrium (analogously, $\hat{\sigma}_1 < \sigma^u$, $\hat{\sigma}_2 = \sigma^u$ is not a equilibrium).

Proof : Assume, by contradiction, that (σ_1^*,σ_2^*) and $(\bar{\sigma}_1,\bar{\sigma}_2)$ are equilibria. Dividing equations (9.1) and (9.2) we get

$$\frac{\sigma_2^*}{\sigma_1^*} = \frac{C_1^{'}(\sigma_1^*)}{C_2^{'}(\sigma_2^*)} \text{ and } \frac{\bar{\sigma}_2}{\sigma^u} \geqslant \frac{C_1^{'}(\sigma^u)}{C_2^{'}(\bar{\sigma}_2)}.$$

Therefore,

$$\bar{\sigma}_2 C_2^{'}(\bar{\sigma}_2) \geqslant C_1^{'}(\sigma^u)\sigma^u > C_1^{'}(\sigma_1^*)\sigma_1^* = \sigma_2^* C_2^{'}(\sigma_2^*) \Rightarrow \bar{\sigma}_2 > \sigma_2^*.$$

Thus, $wf_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma^u}{\bar{\sigma}_2}\right)\frac{1}{\bar{\sigma}_2} \geqslant$

$$h\left(\sigma^u, \bar{\sigma}_2\right) \cdot C_1^{'}(\sigma^u) > h(\sigma_1^*, \sigma_2^*)C_1^{'}(\sigma_2^*) \ = f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}\left(\frac{\sigma_1^*}{\sigma_2^*}\right)\frac{1}{\sigma_2^*}. \qquad (*)$$

Analogously, $f_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\bar{\sigma}_2}{\sigma^u}\right)\frac{1}{\sigma^u} =$

$$h(\sigma^u, \bar{\sigma}_2)C_2^{'}(\bar{\sigma}_2) > h(\sigma_1^*, \sigma_2^*)C_2^{'}(\sigma_2^*) = wf_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\sigma_2^*}{\sigma_1^*}\right)\frac{1}{\sigma_1^*}.$$

Therefore, $\frac{\bar{\sigma}_2}{\sigma^u} < \frac{\sigma_2^*}{\sigma_1^*} \Rightarrow \frac{\sigma^u}{\bar{\sigma}_2} > \frac{\sigma_1^*}{\sigma_2^*}$. This is in contradiction with $(*)$ and $\bar{\sigma}_2 > \sigma_2^*$.

A direct consequence of facts $1-4$ is the uniqueness of equilibrium.

■

While existence of equilibrium can be obtained under weaker assumptions than assumption A, proposition 3 is sharp in the following sense. Consider the example below

Example 1. *Assume that* $n \geqslant 2$, $w = 1$, $\tilde{k}(n) = \frac{1}{1+n^2}$, $\tilde{q}_i(n) \sim U(\frac{1}{n}, 1)$, $i \in \{1, 2\}$, $\widetilde{q}_1(n)$ *and* $\widetilde{q}_2(n)$ *are independent,* $D > 4$, $\vartheta(x) = x$ *and* $\tilde{c} \sim 4U(0, 1)$.

In this example, $\tilde{q}_i(n)$ converges, in distribution, to the concave random variable $q \sim U(0, 1)$ as n goes to infinity. By proposition 3, a pure strategy equilibrium exists in the limit when $q \sim U(0, 1)$. However, a equilibrium in pure strategies does not exist in 2PG for any $n < \infty$. This example show that an equilibrium may fail to exist when $\tilde{q}_i(n)$ is "almost concave." A proof of this claim is available from the authors upon request.

8. Appendix B: General Comparative Statics

In this section, we demonstrate our comparative statics results.

Proof of proposition 4: The cost functions can be written as follows: $C_1(\sigma_1) = qkC(\sigma_1)$ and $C_2(\sigma_2) = q(1-k)C(\sigma_2)$. The proposition is clearly true if $\sigma_1^* = \sigma_2^* = \sigma^u$ or if $\sigma_1^* = \sigma^u$ and $\sigma_2^* < \sigma^u$. However, $\sigma_1^* < \sigma_1^u$ and $\sigma_2^* = \sigma_1^u$ cannot be an equilibrium because then

$$kC^{'}(\sigma^u) \geqslant kC^{'}(\sigma_1^*) \geqslant \frac{\sigma^u}{\sigma_1^*}(1-k)C^{'}(\sigma^u) \geqslant (1-k)C^{'}(\sigma^u) > kC^{'}(\sigma^u).$$

If (σ_1^*, σ_2^*) is an interior equilibrium then

$$\frac{\sigma_2^*}{\sigma_1^*} = \frac{k}{1-k}\frac{C'(\sigma_1^*)}{C'(\sigma_2^*)} \Rightarrow \frac{C'(\sigma_2^*)\sigma_2^*}{C'(\sigma_1^*)\sigma_1^*} = \frac{k}{1-k} \leqslant 1 \Rightarrow \sigma_2^* \leqslant \sigma_1^*.$$

This shows the first part of proposition 4. The second part is clearly true if $\sigma_1^* = \sigma_2^* = \sigma^u$. If $\sigma_1^* = \sigma^u$ and $\sigma_2^* < \sigma^u$ is an equilibrium then

$$\frac{\sigma_2^*}{\sigma^u} \geqslant \frac{k}{1-k}\frac{C'(\sigma^u)}{C'(\sigma_2^*)} \Rightarrow (1-k)\sigma_2^* \geqslant k\frac{C'(\sigma^u)\sigma^u}{C'(\sigma_2^*)} \geqslant k\sigma^u.$$

If (σ_1^*, σ_2^*) is an interior equilibrium then

$$\frac{\sigma_2^*}{\sigma_1^*} = \frac{k}{1-k}\frac{C'(\sigma_1^*)}{C'(\sigma_2^*)} \Rightarrow \frac{(1-k)\sigma_2^*}{k\sigma_1^*} = \frac{C'(\sigma_1^*)}{C'(\sigma_2^*)} \geqslant 1.$$

■

Proof of propostion 5: Assume, by contradiction, that there exists $\varepsilon > 0$ and subsequence such that $\sigma_2^* \geqslant \varepsilon$ as k goes to 0. Then, 2 wins the election with arbitrarily high probability no matter what player 1 does. If, instead player 2 votes at $0.5\sigma_2^*$ then 2 will also win the election with arbitrarily high probability. The difference in benefits is arbitrarily small, but the difference in costs $(C_2(\sigma_2^*) - C_2(0.5\sigma_2^*))$ is bounded away from zero (greater than $C_2'(\varepsilon)0.5\varepsilon$). A contradiction. ■

Proof of propositon 6 : Assume without loss of generality that $k \leqslant 0.5$ (the proof for $k \geqslant 0.5$ is analogous). So, there is no semi-corner equilibrium in which $\sigma_2 = \sigma^u$. Assume that (σ^u, σ^u) is an equilibrium for some $\bar{w}$. Then,

$$\bar{w} f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}}(1)\frac{1}{\sigma^u} \geqslant h(\sigma^u, \sigma^u) \cdot C_1'(\sigma^u) \text{ and } \bar{w} f_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}(1)\frac{1}{\sigma^u} \geqslant h(\sigma^u, \sigma^u) \cdot C_2'(\sigma^u).$$

So, (σ^u, σ^u) is an equilibrium for all $w \geqslant \bar{w}$. With some abuse of notation, let $\bar{w}$ be the smallest w such that (σ^u, σ^u) is an equilibrium. Assume that $\dot{\sigma}_1 = \sigma^u$, $\dot{\sigma}_2 < \sigma^u$ is a semi-corner equilibrium at $\dot{w} < \bar{w}$. We now show that the equilibrium at $\ddot{w}$, $\dot{w} < \ddot{w} < \bar{w}$ is also a semi-corner equilibrium. Note that the function

$$j(\sigma_2) \equiv \frac{\sigma^u \cdot h(\sigma^u, \sigma_2) \cdot C_2'(\sigma_2)}{f_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}}\left(\frac{\sigma_2}{\sigma^u}\right)}$$

is increasing and continuous in σ_2. Therefore, if $j(\sigma^u) = \bar{w}$ and $j(\dot{\sigma}_2) = \dot{w}$ then $j(\sigma_2) = \ddot{w}$ has a solution $\ddot{\sigma}_2$ such that $\sigma^u > \ddot{\sigma}_2 > \dot{\sigma}_2$. By definition,

$$\ddot{w} f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}} \left(\frac{\sigma^u}{\ddot{\sigma}_2} \right) \frac{1}{h(\sigma^u, \ddot{\sigma}_2) \cdot \ddot{\sigma}_2} = C_2'(\ddot{\sigma}_2) \frac{\ddot{\sigma}_2}{\sigma^u} > \frac{C_2'(\dot{\sigma}_2)\dot{\sigma}_2}{\sigma^u} \geqslant C_1'(\sigma^u),$$

where the last inequality comes from the fact that $(\sigma^u, \dot{\sigma}_2)$ is an equilibrium. Therefore, $(\sigma^u, \ddot{\sigma}_2)$ is an equilibrium at $\ddot{w}$. Hence, if $(\sigma^u,\ \dot{\sigma}_2)$ is a semi-corner equilibrium at $\dot{w} < \bar{w}$ then there exists a segment $[\tilde{w}, \bar{w})$ in which the equilibrium is semi-corner and expected turnout is non-decreasing in w.

Consider a segment $(0, \tilde{w})$ in which the equilibrium (σ_1^*, σ_2^*) is interior. Then,

$$\frac{\sigma_2^*}{\sigma_1^*} = \frac{C_1'(\sigma_1^*)}{C_2'(\sigma_2^*)}$$

implies that (σ_1^*, σ_2^*) are either both increasing or decreasing in $w \in (0, \tilde{w})$. Assume that both (σ_1^*, σ_2^*) are decreasing in $w \in (0, \tilde{w})$. Note that

$$\begin{aligned} w f_{\frac{\tilde{k}\tilde{q}_1}{(1-\tilde{k})\tilde{q}_2}} \left(\frac{\sigma_2^*}{\sigma_1^*} \right) &= \sigma_1^* \cdot h(\sigma_1^*, \sigma_2^*) \cdot C_2'(\sigma_2^*); \\ w f_{\frac{(1-\tilde{k})\tilde{q}_2}{\tilde{k}\tilde{q}_1}} \left(\frac{\sigma_1^*}{\sigma_2^*} \right) &= \sigma_2^* \cdot h\left(\sigma_1^*, \sigma_2^*\right) \cdot C_1'(\sigma_1^*). \end{aligned}$$

The first equation implies that $\frac{\sigma_2^*}{\sigma_1^*}$ is strictly decreasing in w while the second would imply that $\frac{\sigma_2^*}{\sigma_1^*}$ is strictly increasing in w. ■

Proof of proposition 7: Assume without loss of generality that $k \leqslant 0.5$ (the proof for $k \geqslant 0.5$ is analogous). So, there is no semi-corner equilibrium in which $\sigma_2 = \sigma^u$.

$$w f\left(\frac{\sigma_2}{\bar{k}\sigma_1}\right)\frac{1}{\bar{k}\sigma_1} - (1-k)\left(\frac{\bar{c}}{2}\right)\sigma_2 = 0;\ \text{if } \sigma_2 \in (0,1).$$

$$w f\left(\bar{k}\frac{\sigma_1}{\sigma^u}\right)\frac{\bar{k}\sigma_1}{(\sigma_2)^2} - (1-k)\left(\frac{\bar{c}}{2}\right)\sigma_2 = 0;\ \text{if } \sigma_2 \in (0,1). \tag{8.1}$$

The expected margin of victory is a positive function of $m \equiv \frac{\sigma_2(1-k)}{\sigma_1 k}$. If the equilibrium is a corner solution then m is clearly decreasing in k. If the equilibrium is a semi-corner solution then $w f(m).\frac{1}{\sigma^u} = k h(\sigma^u, \sigma_2) C'(\sigma_2)$, where f is the density of $\frac{\tilde{q}_1}{\tilde{q}_2}$. The function $\frac{wf(m)}{\sigma^u h(\sigma^u,\sigma_2)C'(\sigma_2)}$ is decreasing in σ_2. So, in a semi-corner equilibrium, σ_2 is decreasing in k and, therefore, $m = \frac{\sigma_2(1-k)}{\sigma^u k}$ is decreasing in k.

Now consider an interior equilibrium. Assume, by contradiction, that as k increases, m increases or stays constant. Then, $\frac{\sigma_2}{\sigma_1}$ increases. However, in an interior equilibrium we have $\frac{\sigma_2}{\sigma_1}\frac{k}{(1-k)} = \frac{C'(\sigma_1)}{C'(\sigma_2)}$. So, $\frac{C'(\sigma_1)}{C'(\sigma_2)}$ increases. Hence, either both σ_1 and σ_2 increase or they both decrease. If they both increase a contradiction obtains because $f(m)$ decreases and $wf(m) = k\sigma_1 h(\sigma_1,\sigma_2)C'(\sigma_2)$. If they both decrease then a contradiction obtains because $f\left(\frac{1}{m}\right)$ increases and $wf\left(\frac{1}{m}\right) = (1-k)\sigma_2 h(\sigma_1,\sigma_2)C'(\sigma_1)$. ■

In our companion paper (Feddersen and Sandroni (2005)) we show an example such that expected turnout monotonically increases in the level of disagreement. The example below that this does property deos not generalize and expected turnout may *not* increase monotonically on the level of disagreement.

Example 2. *The fraction of ethical agents in each group, $\widetilde{q}_1$ and $\widetilde{q}_2$, are independent with density*

$$f(x) = \begin{array}{l} 4 \text{ if } x \in (0,\ 0.1); \\ \frac{2}{3} \text{ if } x \in (0.1,\ 1). \end{array}$$

The level of disagreement is fixed $(\tilde{k} = k)$. Voting costs are given by $\tilde{c} \sim \frac{1}{2}\ U(0,$ $1)$. The payoff D is greater than 0.5, $\vartheta(x) = x$ *and* $w = \frac{100}{44}$.

In this example, $q = \frac{7}{20}$. The density of $\frac{\tilde{q}_1}{\tilde{q}_2}$ and $\frac{\tilde{q}_2}{\tilde{q}_1}$ is given by

$$\bar{f}(x) = \begin{array}{c} 1.4 \text{ if } x \leqslant 0.1; \\ \frac{39}{135} + \frac{1}{90x^2} \text{ if } 0.1 \leqslant x \leqslant 1; \\ \frac{39}{135x^2} + \frac{1}{90} \text{ if } 1 \leqslant x \leqslant 10; \\ \frac{1.4}{x^2} \text{ if } 10 \leqslant x. \end{array}$$

Let $k = \frac{1}{11}$. Then, $\bar{k} = \frac{k}{1-k} = \frac{1}{10}$. Note that $\sigma_1^* = \sigma_2^* = 1$ is an equilibrium because

$$\frac{14}{44} = w\bar{f}(\frac{1}{\bar{k}})\frac{1}{\bar{k}} \geqslant q(1-k) = \frac{14}{44}.$$

Let $k = \frac{1}{10}$. Then, $\bar{k} = \frac{k}{1-k} = \frac{1}{9}$. So, $\sigma_1^* = \sigma_2^* = 1$ is not an equilibrium because

$$\frac{100}{44}\left(\frac{39}{135} + \frac{81}{90}\right)\frac{1}{9} = w\bar{f}(\frac{1}{\bar{k}})\frac{1}{\bar{k}} < q(1-k) = \frac{7}{20}\left(1 - \frac{1}{10}\right).$$

It follows that expected turnout decreases when k increases from $\frac{1}{11}$ to $\frac{1}{10}$.

9. Appendix C: A General Model

In this section, we provide a more general model that clarifies the relationship between an agent's preferences and the set of agents whose behavior he takes as given when ranking rules. We say that each agent has a preference and also a *group.* An agent's group is the set of agents (including himself) that he understands to be obligated by the same rule taking as given the behavior of everyone else.

In the model presented in section 2, we assumed that the agents' group consist of all agents who share his preferences. This assumption is technically convenient but obscures an important distinction between agents who have identical preferences and agents who understand to be obligated by the same rule. We provided an informal explanation for this assumption in section 2.1.1, but in this appendix we provide a formal foundation for it. We show that our consistency solution concept implies that it is without loss of generality to consider groups in which all agents have the same preferences.

We now consider a model that generalizes the one presented in section 2. The main innovation is that we no longer assume that agents take as given the behavior of those agents whose preferences are differ from their own. The notation from section 2 is still used here.

Let $T = \{1, 2\}\text{x}(c^l, c^h)\text{x}\{0, D\}$ be the set of types determining an agent's preferences, voting costs and whether she is an abstainer or ethical.[13] Let $\Lambda(T)$ be the set of subsets of T. We say that an element of $\Lambda(T)$ is a *group*. Let

$$g : T \longrightarrow \Lambda(T)$$

be a *group function* mapping each type into a group.

Let $e : g(t) \longrightarrow A$ be a *rule.* Let $E(t)$ be the set of all rules. If type $t \in T$ chooses rule $e_t \in E(t)$ then $e_t(t')$ denotes the action $t' \in g(t)$ should take. Given a rule $e \in E(t)$, let $s^e : g(t) \longrightarrow A$ be a function such that for any type $t' = (i, c, d) \in g(t)$

$$s^e(t') \equiv \begin{cases} \text{abstain} \\ e(t') \end{cases} if \begin{array}{l} d \leq c; \\ c > d. \end{array}$$

That is, s^e is the behavior of types in group $g(t)$ under the assumption that e specifies the actions that should be chosen. Let $s_{-g(t)} : (g(t))^c \longrightarrow A$ be a function

[13] The definition of a type here differs from the previous definition. In section 2 a type determines only the agents' preferences.

determining the behavior of types outside group $g(t)$. The pair $(s^e, s_{-g(t)})$ defines a behavioral profile for all agents (because the behavior of types in group $g(t)$ and outside group $g(t)$ is defined).

We now define the ranking of rules. Let $p(s^e, s_{-g(t)})$ be the probability that candidate 1 wins the election; and let $\phi(s^e, s_{-g(t)})$ be the expected social voting cost (under the assumption that behavior is defined by the pair $(s^e, s_{-g(t)})$). Agent with preferences 1 (i.e., $t = (1, c, d)$) rank rules $e \in E(t)$ as follows: The higher the value of $R_1(e \mid s_{-g(t)})$,

$$R_1(e \mid s_{-g(t)}) \equiv wp(s^e, s_{-g(t)}) - \vartheta(\phi(s^e, s_{-g(t)})),$$

the higher the ranking of e. Analogously, agents with preferences 2 ($t = (2, c, d)$) rank rules according to

$$R_2(e \mid s_{-g(t)}) \equiv w(1 - p(s^e, s_{-g(t)})) - \vartheta(\phi(s^e, s_{-g(t)})).$$

That is, taking the behavior of types outside the agent's group as given, each type $t = (i, c, d)$ ranks rules based on the outcome produced if the rule defined the actions that should be taken by agents in her group.

Let $s : T \longrightarrow A$ be a behavior profile, where $s(t) \in A$ is the action taken by type $t \in T$. Given a group $h \in \Lambda(T)$ let $s_h : h \longrightarrow A$ be the restriction of s to $h \in \Delta(T)$. That is, $s_h(t) = s(t)$ if $t \in h$.

Definition 3 ((Consistency Requirement)). *The behavior profile s is consistent with the group function g if for every type $t \in T$, $t = (i, c, d)$, there exists a rule $e^* \in E(t)$ such that*

1. $$R_i(e^* \mid s_{-g(t)}) \geqslant R_i(e \mid s_{-g(t)}) \text{ for all } e \in E(t);$$

2. $$s_{g(t)} = s^{e^*}.$$

Assumption 1 Given a type $t = (i, c, d) \in T$ we assume that there exists an open set $C(t)$ of $[c^l, c^u]$ such that for any $c' \in C(t)$, $t' = (i, c', d) \in g(t)$.

Assumption 1 ensures that no group has measure zero. The motivation for this assumption is that a type $t \in T$, with a zero measure group $g(t)$, could choose any rule in $E(t)$ because the behavior of agents in $g(t)$ has no consequences. So, the behavior of these types can be exogenously defined. Let T_i be the set of all types with preferences $i \in \{1, 2\}$.

Proposition 8. *Let s be a behavior profile consistent with a group function g that satisfies assumption* 1. *Then, the behavior profile s is consistent with a group function $\tilde{g}$ such that for any type $t \in T_i$, $i \in \{1, 2\}$, $\tilde{g}(t) \subseteq T_i$.*

Proposition 8 shows that if a behavior profile is consistent with some group function (that satisfies assumption 1) then the same behavior profile is consistent with a group function in which each group contains only agents with the same preferences. Hence, there is no loss of generality in assuming that agents in each group share the same preferences. In the model of section 2, we focused on the group function g^* such that $g^*(t) = T_i$ for any $t \in T_i$. That is, we focused on the group function such that *all* agents who share the same preferences are in the same group.

In principle, we could also consider a model in which each agents' group consists of a strict subset of the agents who share their preferences. However, we believe that this generalization is not likely to deliver better results than the model of section 2. Consider an extreme case in which each agents' group has very small measure. We conjecture that, as in Palfrey and Rosenthal (1985), expected turnout will be small because agents will take as given the behavior agents in other groups who are voting for their favored candidate.

9.1. Proof of proposition 8.

We start the proof with the following lemmata. The first results show that agents with preferences i either vote for i or abstain.

Fact 1 Given a type $t = (i, c, d)$, $s(t') \in \{\text{abstain, vote for } i\}$ for almost all types $t' \in g(t)$.

Proof: Consider a rule $e \in E(t)$, $t = (1, c, d)$, that requires a (strictly positive measure) set of ethical types in $g(t)$, with costs below D, to vote for candidate 2. Consider an alternative rule $\tilde{e} \in E(t)$ which is identical to e except that it requires these types to abstain. The expected social cost of voting is strictly smaller under $\tilde{e}$ than under e, i.e. $\phi(s^{\tilde{e}}, s_{-g(t)}) < \phi(s^e, s_{-g(t)})$. On the other hand, the probability that 1 wins the election is not smaller under $\tilde{e}$ than under e. So, $R_1(\tilde{e} \mid s_{-g(t)}) > R_1(e \mid s_{-g(t)})$.

The same argument applies if we replace 1 for 2. So, given a type t with preferences i, for almost all ethical types $t' \in g(t)$, $c' \leqslant D$, $e^*(t') \in \{\text{abstain, vote for } i\}$.

All abstainers and ethical types whose voting costs are greater than D abstain. So, for almost all types $t' \in g(t)$, $s^{e^*}(t') \in \{\text{abstain, vote for } i\}$. The conclusion now follows from part 2 of the definition of consistency. ■

Fact 2 For almost all types $t \in T$, $t = (i, c, d)$, $s(t) \in \{\text{abstain, vote for } i\}$.

Proof: By property 1 of g and the fact that $[c_l, c^u]$ is compact, there is a finite set of types $A \subset T$, such that $T \subset \bigcup_{t \in A} g(t)$. The conclusion follows from fact 1.

■

The next result shows that given a type with preference i, almost all types in her group with preferences $j \neq i$ must abstain.

Fact 3 Given a type $t \in T$, $t = (i, c, d)$, for almost all types $t' = (j, c', d') \in g(t)$, $j \neq i$, $s(t') \in \{\text{abstain}\}$.

Proof: By fact 1, for almost all types $t' \in g(t)$, $s(t') \in \{\text{abstain, vote for } i\}$. By fact 2, for almost all types $t' = (j, c', d') \in g(t)$, $s(t') \in \{\text{abstain, vote for } j\}$. Therefore, for almost all types $t' = (j, c', d') \in g(t)$, $s(t') \in \{\text{abstain}\}$.

■

The next result shows that the restriction of $e^* \in E(t)$ to types in $g(t)$ with the same preferences as t, $\tilde{g}(t)$, defines a rule $\tilde{e}$ that is best (for t) given the behavior of types outside $\tilde{g}(t)$.

Fact 4 Consider a type $t \in T$, $t = (i, c, d)$. Let e^* be the rule defined in part 1 of the definition of consistency. Let $\tilde{g}(t) \subset g(t)$ be the subset of all types in $g(t)$ with preferences i. Let $\tilde{e} : \tilde{g}(t) \to A$ be the restriction of e^* to $\tilde{g}(t)$. Then,

$$R_i(\tilde{e} \mid s_{-\tilde{g}(t)}) \geqslant R_i(e \mid s_{-\tilde{g}(t)}) \text{ for all functions } e : \tilde{g}(t) \to A.$$

Proof: Assume, by contradiction that there exists a function $e_1 : \tilde{g}(t) \to A$ such that

$$R_i(\tilde{e}| \ s_{-\tilde{g}(t)}) < R_i(e_1| \ s_{-\tilde{g}(t)}).$$

Let $e_2 : g(t) \to A$ be a function defined by

$$e_2(t') = \begin{array}{cc} e_1(t') & \text{if } t' \in \tilde{g}(t); \\ \text{abstain} & \text{if } t' \in g(t), \ t' \notin \tilde{g}(t). \end{array}$$

By definition, $s^{e_1}(t') = s^{e_2}(t')$ for all $t' \in \tilde{g}(t)$. By fact 3, for almost all $t' \in g(t)$, $t' \notin \tilde{g}(t)$,

$$s_{-\tilde{g}(t)}(t') = s^{e_2}(t') = \{\text{abstain}\}.$$

Therefore, the behavior profile $\left(s^{e_1}, \ s_{-\tilde{g}(t)}\right)$ is, for almost all types, identical to the behavior profile $\left(s^{e_2}, \ s_{-g(t)}\right)$. So, $R_i(e_1 \mid s_{-\tilde{g}(t)}) = R_i(e_2 \mid s_{-g(t)})$.

By part 2 of the definition of consistency and the fact that $\tilde{e}$ is a restriction of e^* to $\tilde{g}(t)$, the behavior profile $\left(s^{e^*}, s_{-g(t)}\right)$ is identical to the behavior profile $(s^{\tilde{e}}, s_{-\tilde{g}(t)})$, i.e., they both equal to s. So,

$$R_i(\tilde{e}|\ s_{-\tilde{g}(t)}) = R_i\left(e^*|\ s_{-g(t)}\right).$$

Therefore, $R_i\left(e^*, s_{-g(t)}\right) < R_i(e_2 \mid s_{-g(t)})$. This is a contradiction with part 1 of the definition of consistency. ■

We are now ready to demonstrate proposition 2. Assume that the behavior profile s is consistent with an group function g that satisfies assumption 1. Let e^* be the rule defined in part 1 of the definition of consistency. Given a type $t = (i, c, d)$, let $\tilde{g}(t) \subset g(t)$ be the subset of all types in $g(t)$ with preferences i. Let $\tilde{E}(t)$ be the set of all functions $e : \tilde{g}(t) \to A$. Let $\tilde{e} \in \tilde{E}(t)$ be the restriction of e^* to $\tilde{g}(t)$. By fact 4,

$$R_i(\tilde{e}|\ s_{-\tilde{g}(t)}) \geqslant R_i(e|\ s_{-\tilde{g}(t)}) \text{ for all } e \in \tilde{E}(t).$$

By part 2 of the definition of consistency, $s_{\tilde{g}(t)} = s^{\tilde{e}}$. ■

References

[1] Abrahmson, P., J. Aldrich, P apolino and D., Rhode (1992). ""Sophisticated" Voting in the 1988 Presidential Primaries." *American Political Science Review,* Vol. 86, No. 1. pp. 55-69.

[2] Aldrich, J. “Rational Choice and Turnout,” *American Journal of Political Science,* 37-1, 246-278.

[3] Andreoni, J. (1989) “Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence,” *Journal of Political Economy*, 97-6, 1447-1458.

[4] Becker, G. (1976) “Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology,” *Journal of Economic Literature,* 14-3, 817-826.

[5] Becker, G. (1977) “Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology,” *Journal of Economic Literature*” 14-3, 817-826

[6] Bergstrom, T. (1995) “On the Evolution of Altruistic Ethical Rules for Siblings,” *American Economic Review*, 85-1, 58-81.

[7] Bernheim, D. and O. Stark (1988) "Altruism within the Family Reconsidered: Do Nice Guys Finish Last?" *American Economic Review*, 78-5, 1034-1045.

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

[9] Borgers, T. (2004) "Costly Voting," *The American Economic Review*, 94(1), pp. 57-66.

[10] Campbell, C. (1999) "Large Electorates and Decisive Minorities," *The Journal of Political Economy*, 107(6), pp. 1199-1217.

[11] Coate, S. and M. Conlin (2005) "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence." *The American Economic Review*, 94(5), pp. 1476-1504.

[12] Cox, G. (1997). *Making Votes Count.* Cambridge: Cambridge University Press.

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

[14] Feddersen, T. (2004) "Rational Choice Theory and the Paradox of Voting." *Journal of Economic Perspectives.* Vol 18, Winter.

[15] Feddersen, T. and W. Pesendorfer. 1996. "The Swing Voter's Curse." *American Economic Review*, 86(3): 408-24.

[16] Feddersen, T. and W. Pesendorfer. 1997. "Voting Behavior and Information Aggregation in Elections with Private Information." *Econometrica* 65(5): 1029-58.

[17] Feddersen T. and W. Pesendorfer. 1999. "Abstention in Elections with Asymmetric Information and Diverse Preferences." *American Political Science Review*, Vol. 93, No. 2., pp. 381-398.

[18] Feddersen, T. and W. Pesendorfer (1999) "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica,* 65-5, 1029-1058.

[19] Feddersen, T., and A. Sandroni (2005). "A Theory of Participation in Elections." Typescript.

[20] Franklin, M, R. Niemi and G. Whitten (1994) "Two Faces of Tactical Voting," *British Journal of Political Science*, 24, 549-57.

[21] Harsanyi, J. (1977) "Morality and the Theory of Rational Behavior." *Social Research* 44: 623-656.

[22] Harsanyi, J. (1980) "Rule Utilitarianism, Rights, Obligations and The Theory of Rational Behavior," *Theory and Decision,* 12, 115-33.

[23] Harsanyi, J. (1992) "Game and Decision Theory in Ethics," *Handbook of Game Theory,* vol 1, Edited by R. Aumann and S. Hart.

[24] Hirshleifer, J. (1985) "The Expanding Domain of Economics," *American Economic Review,* 75-6, 53-68.

[25] Krasa, S. and M. Polburn (2004) "Is Mandatory Voting Better than Voluntary Voting?"

[26] Monroe, K. (1994)) "A Fat Lady in a Corset: Altruism and Social Theory," *American Journal of Political Science*, 38- 4, 861-893.

[27] Morton, R. (1991) "Groups in Rational Turnout Models," *American Journal of Political Science,* 35, 758-76.

[28] Nalebuff, N. and R. Shachar (1999) "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review,* 89-3, 525-47.

[29] Palfrey, T., and H. Rosenthal (1983) "A Strategic Calculus of Voting," *Public Choice*, 41-1, 7-53.

[30] Palfrey, T., and H. Rosenthal (1985) "Voter Participation and Strategic Uncertainty," *American Political Science Review,* 79-1, 62-78.

[31] Rabin, M. (1993) "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281-1302.

[32] Riker, W., and P. Ordeshook (1968) "A Theory of the Calculus of Voting," *American Political Science Review,* 62, 25-42.

[33] Roemer, J. (1996) *Theories of Distributive Justice.* Harvard University Press.

[34] Roemer, J. (2003) "Political equilibrium with private and/or public campaign finance: A Comparison of Institutions." Working paper.

[35] Samuelson, P. (1993) "Altruism as a Problem Involving Group versus Individual Selection in Economics and Biology," *American Economic Review*, Vol. 83, No. 2, Papers and Proceedings of the Hundred and Fifth Annual Meeting of the American Economic Association, 143-148.

[36] Sen, A. (1977) "Social Choice Theory: A Re-Examination," *Econometrica*, 45-1, 53-89.

[37] Simon, H. (1955) "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, Vol. LXIX, February.

[38] Sugden, R. (1984) "Reciprocity: The Supply of Public Goods Through Voluntary Contributions," *The Economic Journal*, 94, 772-787.

[39] Taylor, C. and H. Yildiram (2005) "Public Information and Electoral Bias." Unpublished paper..

[40] Teixeira, R. *Why Americans Don't Vote: Turnout Decline in the United States: 1960-1984.* New York Greenwood Press, 1987.

[41] Uhlaner, C. (1989) "Rational Turnout: The Neglected Role of Groups," *American Journal of Political Science,* 33-390-422.