Learning in Elections and Voter Turnout

Stefano Demichelis* and Amrita Dhillon†

January 2002

Abstract

We analyze voter learning in two candidate elections with costly voting in which voters may or may not know the costs of other voters. We show that the only "robust" equilibria, i.e. those consistent with reasonable models of voter learning, are those with low turnout in this stylised model. Increases in costs of voting affect turnout adversely but there may be persistence of turnout levels between elections even though costs and other parameters change. Increases in the size of the electorate are shown to decrease turnout, in line with intuition.

JEL Classification: C72, D72

Keywords: Voter Participation, Voter Learning, Asymptotically Stable Equilibrium, Markov chain, long run equilibria

*CORE, 34 Voie du Roman Pays, 1348 Louvain la Neuve, Belgium, email: demichelis@core.ucl.ac.be

†Address for Correspondence: Amrita Dhillon, Department of Economics, University of Warwick, Coventry CV4 7AL, UK. Email:A.Dhillon@warwick.ac.uk. We would like to thank Alan Beggs, Thomas Palfrey, Matthew Jackson, V.Bhaskar, Myrna Wooders, Jonathan Cave, Michel Le Breton, Rabah Amir and seminar participants at Ischia and the universities of Cambridge, Exeter, Essex, Warwick Mathematics and Toulouse, for comments.

1 Introduction

There has been a lot of recent literature looking at the problem of voter turnout. It is well recognised that a game theoretic model is the correct one to use since it captures the simultaneity of the voting process with the probability of being pivotal (Palfrey and Rosenthal 1983). Such models of voter behaviour, however, typically do not lend themselves to easy analysis, not least because of the problems of multiple Nash equilibria that are pervasive in such models. Multiple equilibria become a serious problem when one wants to test the several different theories about voting (e.g. instrumental versus expressive see Mueller (1989), Dhillon and Peralta (2001)). For this a correct estimation of the probability of being pivotal is vital, but, as Fischer (1999) remarks; "Since it could not be known in advance which equilibrium would be attained in any particular situation, it is implied that the probability of being decisive is not well defined in this framework. It would only make sense if probabilities could be assigned to the occurence of each of the possible equilibria ...". This is exactly what our paper aims at: giving a plausible and coherent selection criterion that will enable us to check whether the predictions of voting games are consistent with reality.

The model we will start from in this paper is the one of Palfrey and Rosenthal (1983,1985) who examined the issue of voter turnout[1] in two cases, one with complete information and one with incomplete information. They found that already even in the simplest model with complete information (two candidates, two types of voters, symmetric equilibria) there was a problem of multiple equilibria. In particular, even in the simplest case with an equal number of the two types of voters in the population and restricting attention to symmetric mixed strategy equilibria they found two types of mixed strategy equilibria for plausible cost levels– one with low turnout but one with substantially high turnout. They say that this high turnout equilibrium "has the unappealing feature that there is another equilibrium with almost no one voting. Apparently the only reason the upper one can be sustained is that the two electorates are of the same size so that for the probability of voting very close to 1, the probability of a tied election is very high. Again, the result rests on the fact that in equilibrium there is essentially no strategic uncertainty." (Palfrey and Rosenthal, 1985). Of course, there is no reason why this argument should apply only to the high turnout equilibrium[2]. If we measure the degree of strategic uncer-

[1] Downs (1957) paradox states that if voters are rational and voting is instrumental, then, since voting is costly, they should not vote. Empirically however, levels of turnout are quite high.

[2] When analysing the general game where the size of the two electorates is different they

tainty by the standard deviation at the two equilibria, it is exactly the same in both high and low turnout equilibria. They show that moving to the corresponding incomplete information game gets rid of this high turnout equilibrium. Indeed, in their model of the incomplete information game, they show that with some assumptions on the type of uncertainty allowed, the only symmetric mixed strategy equilibrium that survives is the low turnout one.

Since the game theoretic models cannot explain the observed levels of turnout, it is accepted in the literature that the game theoretic model fails to capture some essential feature of reality.

We have two remarks to make: first, we agree that the model fails to capture an important feature of real elections: the information aggregation devices like pre-election polls that exist, and that can coordinate people's beliefs and that are crucial in predicting the right equilibria[3]. Second, we believe that such models cannot hope to predict exactly observed outcomes as they are simple stylised models[4]. Still it should be remarked that, the Palfrey and Rosenthal model with uncertainty and reasonable parameters can give realistic predictions in turnout: the actual problem seems to lie more in the predicted outcome of the election. This issue is discussed in section 5. In any case what should be asked from this type of models is to make predictions using comparative statics or make other qualitative predictions which *can* be tested. The essential features of the game theoretic model are that voters are strategic in their decisions, they care about the probability of affecting the outcome and about the costs of voting.

We believe that our model offers both these features: first by providing a role for information aggregating devices like polls and second, by isolating a unique equilibrium we can actually do some comparative statics at least at a qualitative level.[5].

While we agree with Palfrey and Rosenthal (1985) that the low turnout equilibrium is the more appealing one, we argue that the reason they cite may not be the most compelling one. As we said before, there is a sense in which the high turnout equilibria are not robust – they require precise beliefs about what other voters are doing in equilibrium. The aim of this paper is to attempt to make this claim more precise. We show that removing some of the assumptions

still get some "quasi-symmetric mixed-pure strategy equilibria" which have high turnout even as the size of the electorate increases. But all symmetric totally mixed strategy equilibria have the property that as the size of the electorate increases the turnout decreases.

[3]See Fey (1997)for a similar idea but in the context of costless voting and multiple candidate elections.

[4]As was pointed out by Ledyard (1981).

[5]Feddersen and Sandroni (2001) independently stress the importance of comparative statics in their model.

Palfrey Rosenthal (1985) make on the type of voter uncertainty, the incomplete information model may still lead to multiple equilibria (at least for small populations). Indeed, some realistic probability distributions on the cost of voting produce *more* equilibria than in the complete information case. We show that introducing voter learning about equilibrium (i.e. considering the dynamics of reaching a Nash equilibrium by boundedly rational agents) would lead in a very simple and natural way to the low turnout equilibria in their example. @@ The introduction of learning is justified both on theoretical and empirical grounds in the discussion at the beginning of section 3. Finally let us point to the connection with the problem of the provision of public goods (see Bagnoli and Lipman (1989)): the voting decision can be seen as a two sided version of the usual problem, wher the 'public' good is the fact of having one's candidate winning and the contribution is the cost of voting. Indeed both cases present similar problems (freeriding, instability of the equilibrium in which everybody contributes as the number of players increases etc.)[6].
The framework of this paper is as follows: We introduce the model of voter learning in Section 2, then we consider the complete information game in Section 3 and the incomplete information game in Section 4. Section 5 concludes.

2 The Model

The model is due to Palfrey and Rosenthal (1983, 1985) (henceforth PR). There are two candidates (or two alternatives): 1 and 2. The voting rule is Simple Majority Rule: in case of tie either 1 is chosen or a coin toss takes place. There are N voters in the population. There are two groups of voters: T_1 (with N_1 voters who prefer 1) and T_2 (with N_2 voters who prefer 2) and all voters belong to one of these groups. Voting is costly and the cost of voting is the same for all voters. Voters have two pure strategies: vote (participate) or abstain – if they vote they always vote sincerely (i.e. for their best candidate).
Let R represent the expected net benefit from voting , p the probability of being pivotal, C the cost of voting, B the benefit from voting i.e. the difference between the benefits of i's more preferred alternative winning as opposed to the less preferred one, and D a fixed benefit from the act of voting (civic duty). Let n_r^i, $r = 1, 2$ denote the number of voters excluding voter i who are expected by voter i to turn out to vote for candidate r. Let p_1 denote the probability that $n_1^i = n_2^i$, p_2 the probability that $n_1^i = n_2^i - 1$, q_1 the probability that voter i is

[6] We thank Michel Le Breton for pointing this out to us.

not pivotal and candidate 1 leads, while q_2 denotes the probability that voter i is not pivotal and candidate 2 leads.
Thus the expected utility from voting for voter i is given by:

$$p_1 B + p_2 \frac{B}{2} + q_1 B + q_2 0 - C + D \tag{1}$$

The expected utility of voter i from abstaining is denoted by:

$$p_1 \frac{B}{2} + p_2 0 + q_1 B + q_2 0 \tag{2}$$

Therefore R is the difference between 1 and 2, and has the following expression:

$$R = (p_1 + p_2)\frac{B}{2} - C + D \tag{3}$$

Let $c = C - D$ represent the net cost of voting. We can normalise by dividing throughout by B so that only the ratio of costs to benefit matters. W.l.o.g set $B = 1$. Let $p = p_1 + p_2$. Thus a voter i will vote if and only if $p - 2c \geq 0$.
We consider Nash equilibria of this game.
Let the probability for player i to choose to vote be q_i. Then $(q_1, q_2, ..., q_n)$ is a mixed strategy Nash equilibrium, if for all i voting and non-voting give the same expected payoff, given the mixed strategies of other players. As PR (1983) shows there are many Nash equilibria to this game: they can be divided into three categories: the first are pure strategy Nash equilibria but these do not exist in general, but only in special cases. The second category where all voter's probabilities of voting are strictly between 0 and 1 is called a Totally Mixed Strategy Equilibrium (TMSE). The equilibria in this category have the property that as electorates become large, the probability of voting becomes smaller. The other category has all voters in one group using a mixed strategy while in the other group, voters are divided into two subgroups, one in which voters definitely abstain and the other in which voters definitely vote – these are called "Mixed-pure" equilibria by PR (1983). This does not have the property that turnout becomes smaller as the size of the electorate becomes larger.

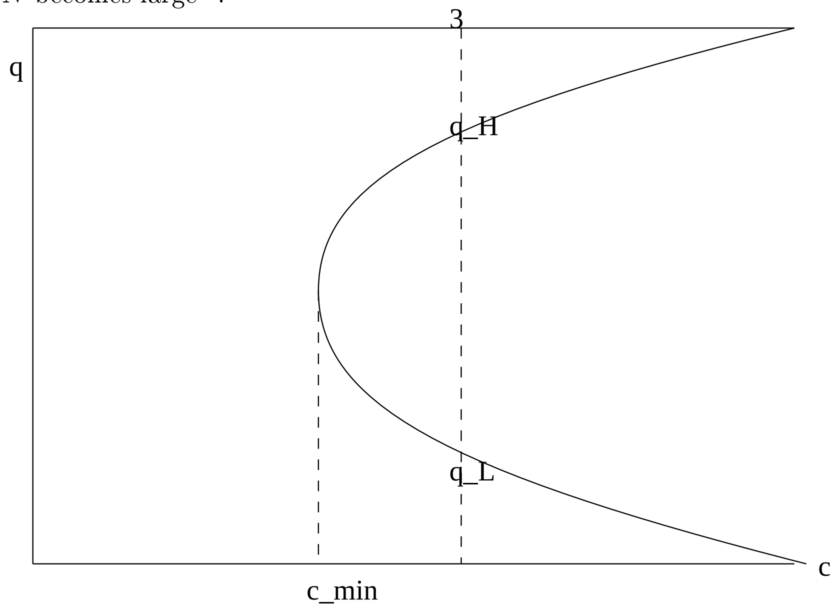
In the rest of the paper, unless otherwise stated, we we will limit our discussion to the case of equal number of voters in each party, i.e. $N_1 = N_2$, this case exhibits the same problems of multiplicity of equilibria as the general case, but allows us to make several technical simplifications in the solution that can be discussed at a very intuitive level.

2.1 The case with $N_1 = N_2$:

The symmetric mixed strategy equilibrium is denoted q which satisfies the following equation:

$$\begin{aligned} 2c &= \sum_{k=0,\ldots,N_1-1} C_{N_1-1,k} C_{N_1,k} q^{2k} (1-q)^{2N_1-2k-1} \\ &+ \sum_{k=0,\ldots,N_1-1} C_{N_1-1,k} C_{N_1,k+1} q^{2k+1} (1-q)^{2N_1-2k-2} \end{aligned} \quad (4)$$

If $0 < c < 1/2$ then this equation has either no solution, 1 solution or two. We can plot a graph (as in PR, 1985) to see the equilibria and how they change as N becomes large [7]:



The example shows that as the cost increases beyond c_{min} there are two types of mixed strategy equilibria: one where almost everyone votes (denoted q_H^*) and one with almost no one voting (denoted q_L^*). In addition there is a pure equilibrium in which everybody votes! Let us analyze them separately: in the pure strategy equilibrium everybody goes to vote believing that everybody else will go, such a situation may arise with a little number of voters, but seems unlikely for electorates of large size, note also that it relies heavily on complete information on the number of voters[8]. The high turnout equilibrium q_H^*

[7] In the current litterature it is usually conjectured and assumed that expression 3) is decreasing in q if $0 \leq q \leq 1/2$ and increasing for $1/2 \leq q \leq 1$, an argument showing how this is at least approximately true is available from the authors

[8] The assumption needed here is merely knowledge of the number of voters in each group,

is counterintuitive, particularly if one considers comparative statics: it predicts an increase in turnout when the cost of voting increases. The low turnout equilibrium q_L^*, has good comparative statics properties and, apart from predicting very low turnout, seems the one that should appear for large N, note also that it Pareto dominates the other two.

PR, 1985 state that the high turnout equilibrium is not robust in that they arise because of the fact that there is almost no strategic uncertainty in the complete information model.

We claim that this is not the main problem with the high turnout equilibrium. In fact the problem of multiple equilibria remains in the incomplete information model: it will be shown below (see Section 4) that at least in small populations, unless the amount of uncertainty is quite large (probably too large to be observed in concrete applications), not only does one still get three equilibria near the original ones but, in certain cases, many more may appear.

We claim, instead, that the main problem with the equilibria other than the low turnout one is that they are inconsistent with reasonable models of voter learning such as fictitious play. We suggest, in this paper an alternative way to select the voting equilibria that gives the "good" prediction even in the case of an equal number of voters in the two electorate. There is a very intuitive reason why the high turnout equilibrium is not a robust one, and this is true regardless of whether the game is of complete information or not. This is because of its stability properties with respect to learning dynamics. But first we will describe our model of voter learning.

3 Complete Information

In this section we will discuss learning in the PR model with complete information. Its primary purpose is to present the learning dynamics and show its use in selecting away "bad" equilibria, the artificial assumption of fixed cost for everybody will be lifted in section 4.

3.1 Dynamics: Single Elections

Before discussing technicalities, the first question one probably should answer is: "Why should an adaptive process be a good refinement criterion in this case?" We think there are two types of reasons for believing this: methodological and

not equality: if $N_1 \leq N_2$ there is still a (mixed-pure) equilibrium in which everybody in group 1 votes based on similar beliefs.

empirical. From the former point of view: we are studying a game with a large number of players who have very little possibilities of communicating and who have little a priori information on what the others will do. It is not reasonable to assume that they will divine the Nash equilibria and that they will assume that everyone else will do. This applies 'a fortiori' to all the selection criteria that are used to refine Nash equilibria (it is easy to see that these have no cutting edge in our problem any way). It seems more sensible to assume that during the period that precedes the election players will collect the information that is available to them (statistics from previous elections, polls etc.), and then find what is their optimal behaviour in a process of trial and error in which they will assume that everybody else will follow a similar path of learning. It is interesting to see that the outcomes of such processes not only are Nash equilibria, but satisfy the strongest known rationality criteria (although the converse is not true – see Demichelis and Ritzberger (2000) for details). On the empirical side there is good evidence (see Cooper et al (1990) and van Huyck et al (1990)) that players do not apply a rationalistic selection criterion when they select strategies to play , for instance weakly dominated strategies play a role in their choice, and again adaptive considerations seem to play a strong role in guiding their behaviour.

Learning can be modeled in several different ways, see Fudenberg and Levine (1998) for an overview of the literature. A good process should satisfy the, possibly conflicting, requirements of assuming as little rationality as possible from the players and of being efficient enough to give acceptable outcomes (we want the voters to learn something before election day comes...). Typical models are the best reply dynamics , that goes back to Cournot, and its improvement (discrete) fictitious play (Brown (1951)) . These assume rather myopical players but have the defect that sometimes they forecast players behaving in a rather silly way, overshooting their beliefs and strategies so as to cycle for ever around an equilibrium without ever coordinating on it (see Fudenberg and Kreps (1993)).

The problem can be circumvented, without having to assume a more sophisticated behaviour of the players, using continuous fictitious play, a model in which they gradually adapt their strategy instead of jumping suddenly from one another (Demichelis and Germano(2000)).

We now describe our model of learning in detail: Let $N = N_1 = N_2$ and as before let q denote the probability that a voter participates in the election and c the cost of voting. Let $f(q, N)$ denote the probability of being pivotal as a function of the equilibrium probability of voting of all voters (computed above in (5) with the coin tossing tie breaking rule). Then, as before the best reply is to vote if

$$f(q) - 2c > 0 \tag{5}$$

and abstain in the opposite case with equality corresponding to indifference between the two strategies.

A voter starts by observing a q, e.g. the share of the population who voted on the last election or the result given by a poll (we assume N is not too small so that the empirical observation of this quantities is a reasonable estimator of q) and checks whether the inequality 5 tells her to go to vote (or to abstain). She realizes that everybody will do the same and so expects the correct q to be higher (lower). Once she has adjusted q she checks again what is her best reply and so on. To simplify notation, we assume that everyone begins with the same q and everyone adjusts in the same way[9]. This is consistent with PR's focus on symmetric equilibria. If this process converges to some point q^*, she will apply the corresponding mixed strategy[10].

The dynamics can be described by the differential equation[11]:

$$\frac{dq}{dt} = K(q) \tag{6}$$

and sign $K(q)$ =sign$(f(q) - 2c)$ This is the *Monotonicity property* assumed in Kandori-Mailath-Rob (93)(henceforth KMR)[12]. We assume that the function $K(q)$ satisfies Lipschitz continuity and so there exists a unique solution for any initial q_0, for all $t \in R$. The solution of this equation for any initial condition, $q(q_0, t)$ is continuous[13].

Now we examine the behaviour of the dynamics on the strategy space. The basic intuition is that the outcomes of the learning process that are likely to be seen in concrete cases are the limit points of $q(t)$ as t goes to infinity. To be more precise we will introduce some definitions adapted from Weibull (1995): Let $q(q_0, t)$ be the solution of the differential equation (6) with initial condition

[9]the reader may check that the latter assumption is innocuous since voters are ex-amnte identical and we focus on symmetric equilibria. The hypothesis of a common initial q is not very drastic, too; see the discussion at the end of this subsection.

[10]We may also think of voters using pure strategies and q describes the proportion of the population that votes – then our assumption is that population shares move continuously with changes in the expected benefit of voting.

[11]if the adjustment steps are small enough, taking a discrete adjustment process would give essentially the same results

[12]The functional form of $K(q)$ depends on the particular speed of learning, different individuals could have different K, as we said before our discussion and our results depend only on the assumptions on its sign.

[13]See Weibull (1995) Appendix.

$q(0) = q_0$. Let $q(q_0, t) \in X = [0,1], \forall t \in R$, i.e. the state variable q is a symmetric mixed strategy (the same for all players).

Definition 1: A state $q^* \in X$ is said to be Lyapunov stable if every neighborhood B of q^* contains a neighborhood B^0 of q such that $q(q_0, t) \in B$ for all $q_0 \in B^0 \cap X$ and $t \geq 0$.

Intuitively a state is Lyapunov stable, or just stable, if no small perturbation away from it induces a movement away from it.

Definition 2: A state q^* is asymptotically stable if it is Lyapunov stable and exists a neighborhood B^* such that the following holds for all $q_0 \in B^* \cap X$:

$$Lima_{t\to\infty} q(q_0, t) = q^* \tag{7}$$

A point that is not stable will be called unstable. While stability requires that there be no pull away from the state, asymptotic stability requires in addition that there be a local pull towards it as well.

Definition 3: Basin of attraction of state q^* : is the set of points $q_0 \in C$: $q(q_0, t)_{t\to\infty} \to q^*$.

Intuitively, the basin of attraction of q^* is the set of initial conjectures $q_0 \in C$ that, with learning, will lead to q^*.

Recall that equilibrium 1 in Figure 1 (also Figure 2 below) is the low turnout mixed strategy equilibrium q_L, 2 is the high turnout mixed strategy equilibrium, q_H, and 3 is the pure strategy full turnout equilibrium. Now we can state Proposition 1:

Proposition 1: For any learning dynamics of type (6) Equilibria 1 and 3 will be asymptotically stable, while equilibrium 2 will always be unstable.

We refer to the appendix for the (elementary) proof, here is an informal discussion: If $c > 1$ or $c < c_{min}$ any trajectory trivially converges to the unique equilibrium which is the zero turnout or the full turnout equilibria respectively. When $c_{min} < c < 1$ the qualitative behaviour of the dynamics is as in Figure 2 below:

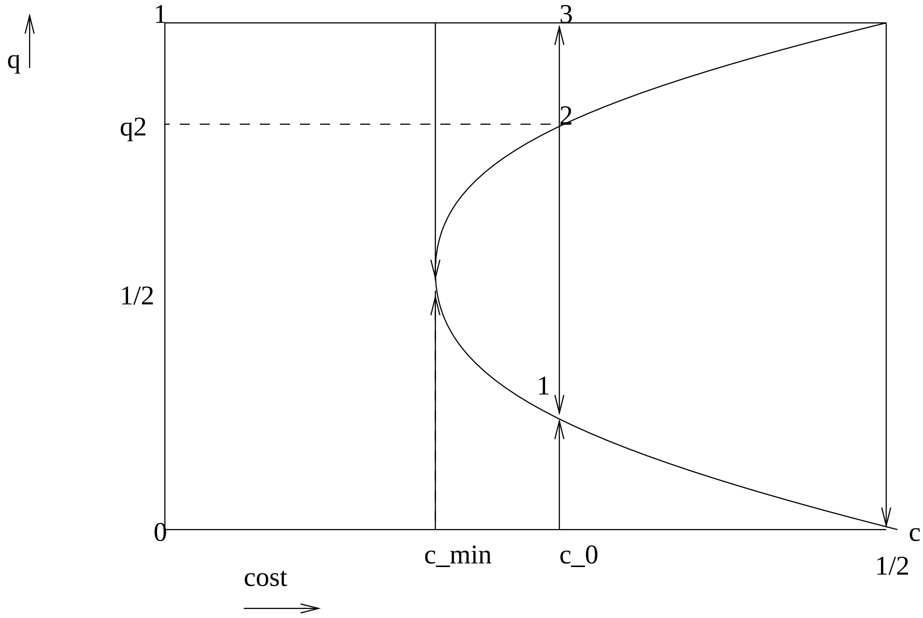


FIGURE 2

Any trajectory starting in the interval $[0, q_2)$ (its basin of attraction) converges to equilibrium 1, therefore it is stable ; equilibrium 2 is unstable, no trajectory leads to it; equilibrium 3 (the pure strategy equilibrium where everyone turns out) is stable with basin of attraction $(q_2, 1]$. Moreover, the basin of attraction of 1 is larger than that of 3 for any cost higher than c_{min}, and as c increases the basin of attraction for equilibrium 3 shrinks.

Considerations about the size of the basin of attraction allow us to improve the predictions of the model: for a stable equilibrium to have a large basin of attraction means having many initial conditions leading to it and so has a high probability of being observed. Note also that we don't have to assume that players start the learning process at the same point, provided all the points are in the same basin of attraction.

So the prediction of the model in case of a single election is that equilibrium 2 will never be observed, equilibrium 1 will be observed with high probability and equilibrium 3 has a smaller chance to appear (if N is moderately large this probability goes to zero very fast).

3.2 Repeated Elections

We mentioned earlier that the low turnout equilibrium has a large basin of attraction. This remark becomes crucial in the following extension of the model: suppose that elections are repeated regularly: we can index elections with $i = 0, 1, 2, 3, ...$, all other elements of the model being the same as before. Now, at election i, voters begin the learning process at q_0^i. This depends on the turnout in the preceding election q_∞^{i-1} in the following (non-deterministic) way: q_0^i is a random variable uniformly distributed on the interval $[\underline{q}; \overline{q}]$ with $\underline{q} = max\{0; q_\infty^{i-1} - \delta\}$ and $\overline{q} = min\{1, q_\infty^{i-1} + \delta\}$; δ is a small positive number that gives a measure of the possible mistakes in ascertaining the turnout. In this way we get a random dynamical system, a Markov chain since the system is time independent, whose states are the stable Nash equilibria, 1 and 3. The behaviour is given by proposition 2:

Proposition 2: If the number of voters is larger than $N_0(\delta)$, the limit distribution of the outcomes q_∞^i is concentrated on equilibrium 1.

Proof in Appendix.

Note that the precise form of the probability distribution of the q_0^i is, to a large extent, irrelevant to the result, provided the distribution is concentrated around q_∞^{i-1}. As before we give an informal discussion of the result: using a terminology borrowed from mechanics, equilibrium 3 is called "metastable". Intuitively the fact that the basin of attraction has positive measure but is very small means that if there are random disturbances, the equilibrium will be stable for a while but after a sufficiently long time we should observe a jump out of it towards equilibrium 1. Equilibrium 1 is "stochastically stable" in the sense of KMR (1993).

A consequence of metastability is that, if elections are repeated, equilibrium 3 tends to jump to equilibrium 1 after a long sequence if the cost is below 1 but higher than a certain value $1/2 > c_{min}$. If the cost is not in this range there is only one equilibrium[14].

Note that, in all cases, equilibria with positive probability predict a nondecreasing q when c decreases, as intuition suggests. It is interesting to investigate further what happens when the cost, or the interest in the outcome, changes from one election to the other. In this case it is natural to ask that the fictitious play dynamics starts at the percentage of voting of the last election. As before,

[14]equilibrium 3 tends eventually to jump to equilibrium 1 even if the number of voters N is small, the only change is in the average time it may take; such a behaviour is typical of risk dominated equilibria in the sense of (Harsanyi and Selten, 1988). See also discussion in the introduction on public goods.

to simplify the discussion and isolate the different effects we will assume that there are no mistakes , i.e. that δ is zero. This allows us to see how q varies as a function of c. Suppose at some time we are given $c = c_0$ and we are in equilibrium 1. Then suppose that the introduction of electronic voting,for instance, causes c to decrease so that the predicted effect is a little increase of turnout (see Figure 2). But when c decreases below c_{min}, the mixed equilibria disappear and we suddenly fall in the basin of attraction of equilibrium 3, so q suddenly jumps up, i.e. at $c = c_{min}$ and below, equilibrium 3 is the only stable equilibrium. The equilibrium will tend to persist for a while even if cost increases again until we reach the point where $c = 1/2$ and the only equilibrium possible is the pure strategy one, where nobody votes; alternatively even if c stays below 1 but over 1/2, in the "very long run" we will see q jumping down. This is phenomenon is known as hysteresis or "memory" of the system, and explains why the same values of parameters can cause the emergence of different equilibria, depending on the initial state.[15]

Again we refer to the Appendix for rigorous statements and proofs. Thus, we should observe phenomena of this type: in countries where there has been a large turnout in preceding elections one expects large turnout in the next election too even if cost has (moderately) increased or interest for the candidates has diminished. When a critical cost level is reached, or after a sequence of many elections, turnout will suddenly jump down and stay low even when cost decreases back to the original one. See section 4.1 for more details in the case of incomplete information. Regarding the predictions for changes in population size: since the probability of being pivotal decreases as population size increases (the function $f(q)$ shifts towards the vertical axis so that for any q the probability of being pivotal is lower for a higher N), thus for any fixed level of cost we see lower turnout as population size increases.

4 Incomplete Information

We capture uncertainty by a model of incomplete information about costs, exactly as in the PR (1985) model[16] Each voter i has a cost of voting c_i which is private information to him. Let the cumulative distribution of costs be denoted $F(c)$ and for simplicity we assume the distribution to be the same between the two groups. We look for the Bayesian Nash equilibria of this game (as in PR

[15]We thank Jonathan Cave for pointing out this interesting feature of the model.

[16]PR (1985) also have a section on incomplete information about preferences but this has similar results.

(1985)). Each voter then has a decision rule that specifies whether to vote or not as a function of his own cost c_i. It is easy to see that in any symmetric Bayesian equilibrium a voter votes if his cost is below a certain threshold level, c^*, so in this case a learning process will be a dynamic on the c^* to choose. Thus a mixed (symmetric) Bayesian equilibrium is a cost level c^* such that $2c* = f(q(c^*))$, with the corresponding $q^* = F(c*)$. This corresponds to the equilibrium outcome in the game of complete information where all voters have cost c^* and vote with probability $q^* = F(c^*)$. Note that, although it is natural to assume that players choose the cost level c^*, this is equivalent, for symmetric equilibria, to choosing q^*. All dynamics in terms of one variable can be easily translated in dynamics in terms of the other. So let $C(q)$ represent the inverse of $q(c) = F(c)$. Then we need $2C(q^*) = f(q^*)$. In the graph below (Figure 3), this equilibrium is given by the point where the distribution function intersects the curve $f(q)/2$, which shows the probability of being pivotal (as in Figures 1 and 2 above).

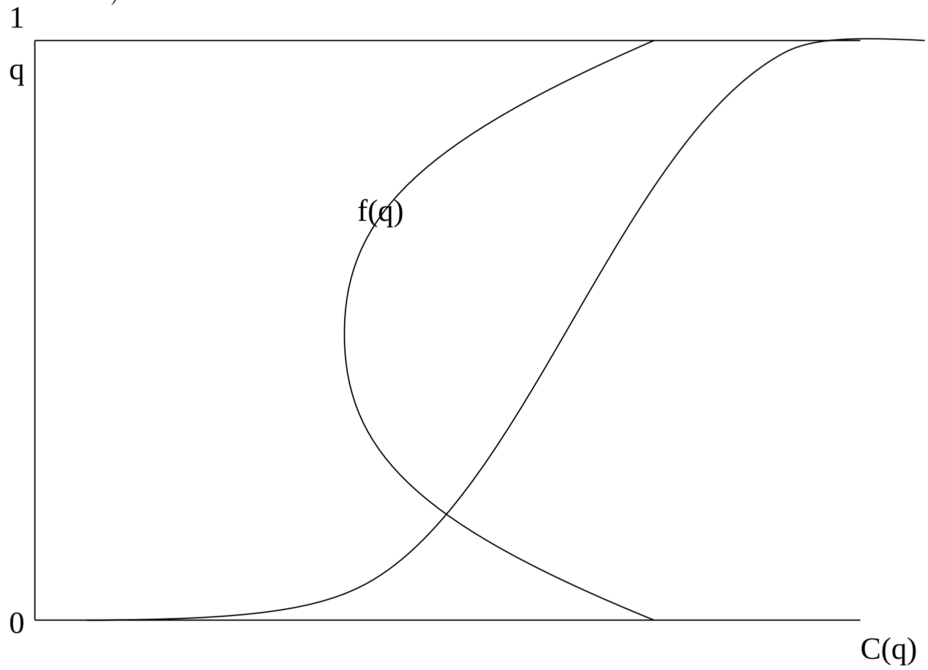


FIGURE 3

PR (1985) use assumptions under which there is only one intersection and the intersection converges to the point $q = 0$ as N becomes large. As is evident from the graph, however, everything depends on the shape of the distribution function $F(c)$ (or its inverse $C(q)$. The assumptions they use are: (1) $F(c)$ is continuous on $(-\infty, \infty)$, (2) $F(0) > 0$ and (3) $F(1) < 1$.

The first assumption is rather natural and corresponds to assuming that the probability distribution of c has no atoms. Assumption 2 is quite realistic also, i.e. that there is a positive probability that cost will be negative (civic sense will prompt some people to vote regardless of their assessment about being pivotal.) However, the last assumption is stronger: it implies that there is a positive probability that a voter would not show up even if he were sure to be pivotal. It is also not innocuous and PR (1985) has an example where relaxing this assumption takes us back to the problem of multiple equilibria in the complete information case (see Figure 3). Moreover, there seems to be an implicit assumption that $F(c)$ is not too wiggly: Figure shows a case where the curvature of $F(c)$ can change quite fast so that many additional equilibria are introduced. Note that such a multimodal probability distribution is not so pathological: it could model a population made of different groups each with different costs and with small variance within a group. Thus, incomplete information does not solve the problem of multiple equilibria, sometimes it even introduces more of them, some of which have high turnout and our intuition suggests that they should be "non-robust" in some sense. However we should mention that for sufficiently large N, the PR (1985) conclusion is still valid.
To address this we now consider learning as before: $\frac{dq}{dt} = K(q)$. We can isolate the stable equilibria as being the ones where $C(q)$ intersects $f(q)$ from below. These are the equilibria 1,3 and 5 in Figure 4.

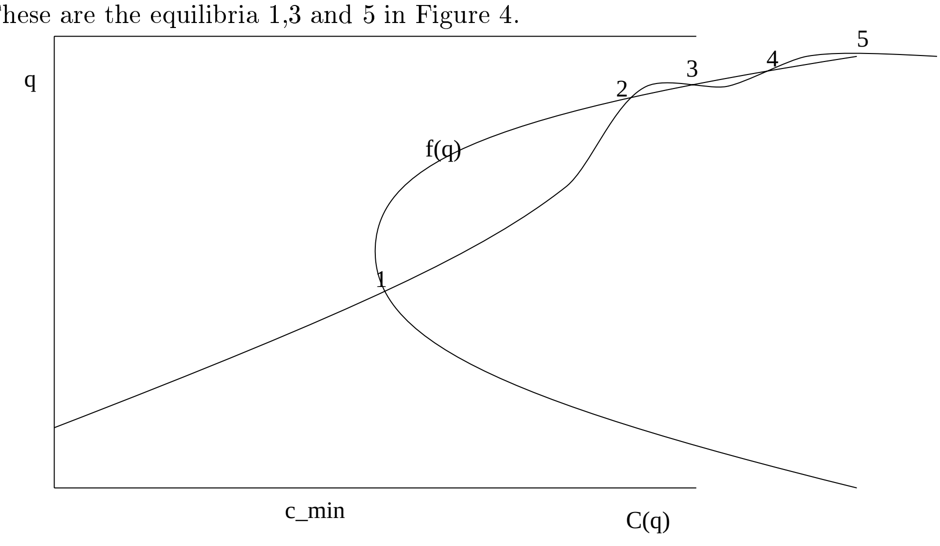


FIGURE 4

More formally we state this in the next Proposition:

Proposition 3: Let the graphs of $C(q)$ and $f(q)$ be in a generic position (this means that they intersect transversely i.e. $2C(q) - f(q) = 0 \Rightarrow K'(q) \neq 0$), then the asymptotically stable points are those such that $d/dq(2C(q) - f(q)) \geq 0$.

The proof is the same as for Proposition 1.

In the same way as in proposition 2, it is not hard to see that if N is large the only equilibrium with a large basin of attraction, containing at least the interval $[0, 1/2]$, is 1 – the low turnout equilibrium. All the others are either unstable, i.e. with zero probability, or metastable, i.e. with a small basin of attraction, whose size goes to zero when N goes to infinity. In the next section we do a similar analysis of the behaviour of equilibria when the parameters of the model are allowed to vary.

4.1 Comparative Statics

We now investigate how equilibria change when the distribution of c varies; For simplicity we shall assume that the c are uniformly distributed on the interval $[\overline{c} - s; \overline{c} + s]$, so that we have two parameters the average $\overline{c}$ and a measure of the dispersion s. Other distributions, such as the Gaussian, can be discussed in the same way and give qualitatively similar results (see the discussion at the end of the section). A uniform distribution corresponds to an F whose graph is shaped as in Figure 5.

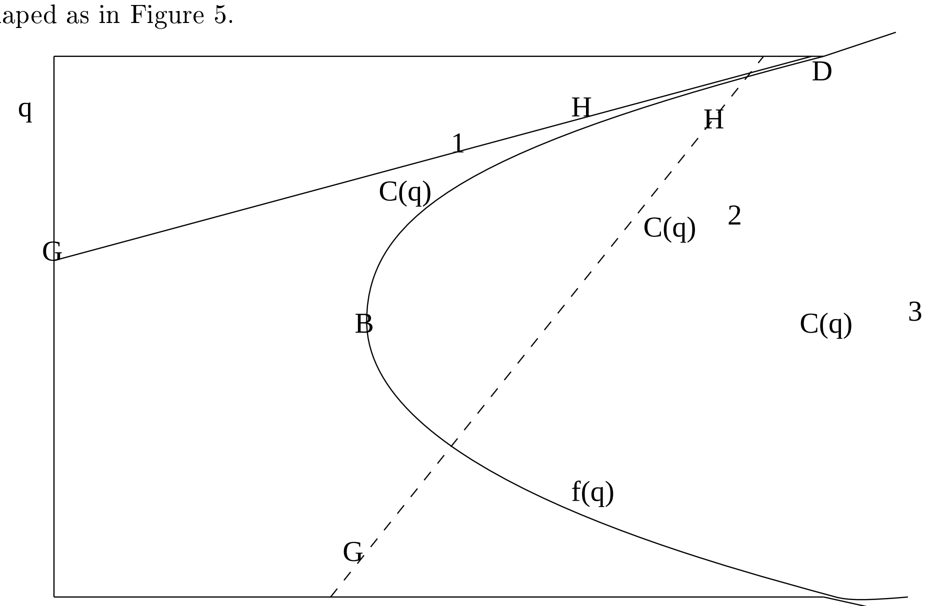
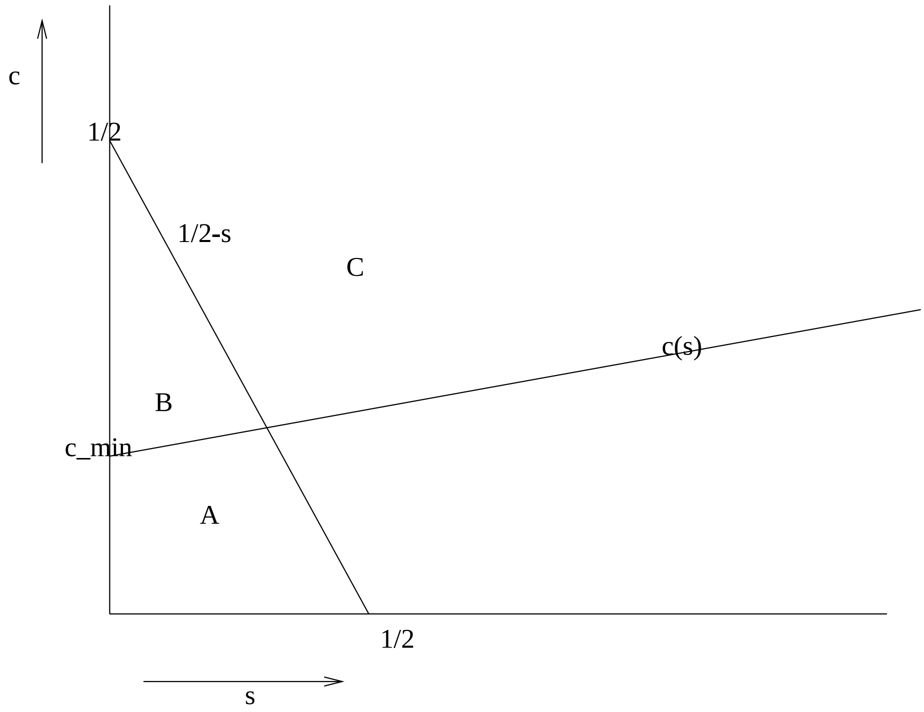


FIGURE 5

The intersections with the curve studied before give the equilibria. It should be geometrically intuitive, and it is easily proved, that if s is large enough so that the slope of the line GH is less than the slope of the arc BD at D, there is only one equilibrium, (see Figure 5). This slope can be computed explicitly: it is $(\frac{N-1}{2})$. This tells us that the condition for such a behaviour is that $s \geq (N-1)/2$. Note that this is a very large value for s, for plausible values of N. In this range of s the unique equilibrium corresponds to nobody voting if $\overline{c} - s > 1/2$, everybody voting if $\overline{c} + s < 1/2$ and a percentage of voters that is a smoothly decreasing function of $\overline{c}$ in the cases in between, in good agreement with intuition. This corresponds to the case studied by PR (1985). The more realistic case of small (or rather not enormous), s , i.e. $s < (N-1)/2$, is more interesting and presents several analogies with the case of complete information , that corresponds to $s = 0$. For any s, let $c(s)$ be the value of $\overline{c}$ such that the line GH is tangent to the curve BD. Note that $c(s)$ is equal to c_{min} for $s = 0$ and increases monotonically to $1/2$ when $s = (N-1)/2$. It is also clear from the picture that $\overline{c} < 1/2 - s$. We now have, as in the complete information case, three cases: 1) For $\overline{c} < c(s)$ there is one equilibrium with $q = 1$ (everybody votes) (shown by area A in Figure 6) 2) For $c(s) < \overline{c} < 1 - s$ there are three equilibria: the previous one, that is stable, an unstable one with large q and a stable one with a low q, with a large basin of attraction, (shown by area B in Figure 6) which makes the $q = 1$ equilibrium stable at first and then makes it disappear when 3)$\overline{c} + s > 1/2$ and there is only one low turnout equilibrium (shown by area C in figure 6) . The domain in the $\overline{c}, s$ plane corresponding to the three cases is shown in Figure 6.



Most of our discussion applies to other distributions as well provided they are regular enough. Note however that the tangency between the F curve and the q curve may not be unique in particular cases, such as costs concentrated around a finite number of values, this would make the jump in the hysteresis cycle split into the composition of several smaller ones.

Our model, and in particular hysteresis, can be used to see what happens when the cost c is changed by introduction or removal of voting laws. For example: abstention is high in the U.S.A. and has been significantly lower in countries such as Belgium and Italy, even though there is no reason to expect significant differences in cost of voting or interest in the elections. An explanation of this fact might be as follows: in the past Belgium and Italy had laws against abstention that made c quite low and possibly negative, so equilibria were high turnout equilibria. The abolition, or lack of enforcement, of such laws has moved the state to the segment of higher cost but with persistence of large q; in U.S.A. where there have never been such laws the more stable low turnout equilibria are observed.

In the incomplete information case, the low turnout equilibrium has a large basin of attraction that increases with $\overline{c}$ making first the $q = 1$ equilibrium metastable and then making it disappear when $\overline{c} + s > 1$ when there is only one, low turnout, equilibrium left. The domain in the $\overline{c}, s$ plane corresponding to the three cases are shown in Figure 6.

Metastability gives an explanation of another phenomenon that should arise in small electorates : with a long history of voting, and if c does not change too much abstention is often lower in first elections than in subsequent elections. This may not stem from the closeness of the electoral platforms (or high cost of voting) or indifference of the voters about the issues, which is a common (and a little tautological) explanation. Rather one may suppose that voters first coordinate on the high turnout equilibrium (and here the assumption of relatively small electorate is important, but quite reasonable in that the franchise was limited in most countries to begin with) and then random fluctuations make it jump to the lower as in the process explained in section 3.2. One does not expect a similar behaviour in elections for higher electorates, because in that case the size of the basin of attraction of the high turnout equilibrium is too small to give it a reasonable probability of being chosen. See more on this in the conclusions of section 6.
One would see the same phenomena of hysteresis as in the complete information case when one moves back and forth on a line crossing the regions as AB does.

5 Instrumental or expessive voting?

An important question in voting theory is what makes people vote, it can actually be split in two: Is there a rational or rather consistent reason that makes them vote and if there is what is it?

More formally the question is whether the sign of the expression (3) triggers the decion to vote and, if it does, what are the relevant terms in it.

As for the first question empirical studies (see Mueller (1989) Chapter 18). seem to give evidence that voters are rational although they often contradict each other in the details.

As for the relative importance of the various factors: the so called instrumental hypothesis claim that voters expect their vote to have an appreciable effect on the outcome of the election i.e. the term pB is relevant, while the expressive hypothesis claims that voters vote because they enjoy it or that they are pushed by a sense of civic duty that overcomes the cost they incur in voting, i.e. what matters is $D - C$.

To defend the expressive theory it is usually claimed that $pB \leq C$ because p is negligible so that almost nobody would vote unless D were comparable and significantly larger than C.

Getting a reliable measure of p from statistics is problematic: basing themselves on official reports Mulligan and Hunter (2000) compute that one out of 100,000 votes matters, however they note that margin specific election procedures may have significantly lowered the observed p in comparison to the p voters use to make their choice. We claim that, using the PR model, one can get at least orders of magnitude of the quantities pB etc that correspond to observed behaviour and support the instrumental voter hypothesis. We first begin by estimating B, : in most electoral campaigns candidates promise advantages in terms of tax cuts, subsidies, cash tranfers from a group to another etc that correspond to an increase of income for their supporters amounting to several thousands euro (or dollars) per year. A legislation usually lasts 4 or 5 years. A conservative estimate (discounting some suspicion about electoral promises) gives more than 10.000 euro overall. The usual methods to evaluate p come from decision theory. The drawback is that the result strongly depends on the method used: Beck (1975) has a model in which the probability r with which a candidate is voted is exogenous. Unless r is EXACTLY equal to $1/2$ this gives a p that is exponentially decreasing with twice the size of the population. For example if we take $r = 0.51$ and a population of just million voters p becomes less than e^{-200} i.e. completely negligible. Good and Mayer (1975), have a more refined model in which r is a random variable extimated via an opinion poll on a small sample of the population, this gives a much larger p , roughly of the order of $1/N$, that is nevertheless too small to give an appreciable incentive to vote,since pB would be less than a cent.

Let us now turn to the PR model with uncertainty: in PR (1983) it is proved that the value of c_{min} in figure 1 is $2/\sqrt{\pi}(2N-1)$ and, with incomplete information, the model predicts that at least everybody whose cost is below this figure will go to vote. Estimating $2N$ is easy, voting populations are of the order of about 40 millions for large European states and 200 millions for USA (http://www.IDEA). By substitution in the formula one obtains an expected benefit from voting pB of about 1 to 2 euros, and more for small countries. It is reasonable to believe that a non trivial part of the population will bother to cash a check of that amount e.g. on their way to shopping. So it seems that we don't have to postulate a significant D to make sense of appreciable turnout. Our argument can be made stronger: it is not clear what should be the right N to take because people may get utility from their candidate winning just in their province or state or even their electoral college, i.e in much smaller units than the whole country. Another reason that may alter the size of N as the players perceive it is given by the tendency people have of thinking of the electorate as being composed of groups of individuals of the same size (e.g. women voters, ethnic minorities etc) whose electoral behaviour coincide, so that in this case one should think of N as the number of these types[17]. In the latter case it would be the expected closeness of the election that mattered rather than the probability of being pivotal – a much easier interpretation.

It should also be remarked , as as often been done in the literature, that there is a tendency to overestimate very small probabilities, which could contribute to make the pB term relevant.

Let us note, however that the weakness of the PR model is in the outcomes it predicts: a simple estimation of the binomial distribution in the case of type symmetric equilibria with a cost of voting of the order of some euro it gives a 50-50 outcome with a standard deviation of the order of less than $1/\sqrt{N}$, that means less than 0.1 percent! It can be proved that this holds also for unequal populations: i.e. for N1 dif from N2. In that case the only equilibria giving outcomes with an appreciable spread must be sustained by very low cost of voting. So, in our opinion, the model seems to give a consistent picture of the phaenomenon of abstension while it fails to capture the behaviour of voters once they decide to go to the polls, in any case this is not what the model was devised for.@@@

6 Conclusions

In this paper we showed how considerations based on learning dynamics and stability can select the intuitive equilibria in the PR (1985) model without having to resort to ad hoc arguments. In this way it is also possible to give the model some predictive value as we saw in the last section.

We would like to make a remark on the "paradox" of voting (Downs 1957). Clearly this stylised model misses many features of real elections: usually there are more parties, voters are not homogenous – they have different costs and benefits, and there exist many other ways to influence policy than voting, such as lobby groups and other forms of direct action. Moreover parameters such as the cost of voting or the measure of the interest in a candidate are not directly measurable with reasonable confidence, so that it seems hard to go beyond the rough estimates we made in the last section.

[17]This type of reasoning (if I chose strategy X, it means that all people in my group or my co-player will probably do the same) is based on a wrong application of Bayes rule but seems to be common in human behaviour, it probably lies at the root of irrational behaviours observed in Newcomb's paradox or the prisoner's dilemma.

It should be obvious that in such cases asking for precise quantitative predictions is too much – there are too many factors that can affect the predicted turnout and that are hard to quantify exactly. So what can we hope to achieve from such simple models? The essence of the PR (1983,1985) models is that voters are strategic – the probability of being pivotal matters to them, as do costs and benefits of voting. Thus the power of the model comes from predictions about comparative statics and other qualitative predictions. Since our model gives some sharp qualitative predictions (jumps , hysteresis, long time drifting away from metastable equilibria) that are very robust with respect to the parameters involved, it makes the PR model more apt to be tested in this way.

As an example take the case of repeated elections. It is commonly believed that there is a significant drop in turnout after the first election.

The prediction of our model, described in section 4 ,is that for small number of voters, a sudden drop in turnout after the first few elections is possible, but that, if the electorate is large, the sequence is much more likely to begin at the stable low turnout equilibrium and then to have small oscillations due to variable costs of voting, spread of different platforms etc. This seems to be confirmed by empirical data from www.idea.int/.

With more refined data, and better estimates on the perceived benefits from one's candidate election, one could even deduce from a sequence of electoral behaviours something about the shape of the distribution of the $c's$, for instance a case in which a gradual change of c would introduce several severe jumps in the turnout should point to the existence of a multimodal distribution, as described at the end of the last section, while a unique jump would be evidence of a more homogeneous electorate as far as costs are concerned.

One objection to this approach that we anticipate is that the model is not really suited to large populations because for large populations the probability of being pivotal is very small, and hence the game theoretic model is more suited to smaller groups e.g. committees. A partial answer has been provided above, anyway our model is quite applicable to voting in committees as well, except for the assumption that voters are myopic in their learning behaviour. We plan to extend this model allowing for more sophisticated learning on the part of voters.

The basic model of PR (1983) did not allow for uncertainty about candidates on the part of voters. Feddersen and Pesendorfer (1996) show that if there are common values among a group of informed and uninformed voters about the candidate (in a two candidate setting as in PR (1983)), then uninformed voters might prefer to abstain even when voting is costless, to ensure that informed voters decide the election. This can happen even with partisans in the population. It seems that many of these equilibria suffer from a similar problem in the sense that they require extreme co-ordination on the part of uninformed voters. It would be worthwhile to check if these equilibria satisfy our requirements of stability.

References

1. Bagnoli and Lipman, 1989, " Provision of Public Goods : Fully implementing the Core through Private Contributions", Review of Economic Studies, Vol 56, pp. 583-601.

2. Beck, N. "A Note on the Probability of a Tied Election", Public Choice 23, Fall 1975, pp. 75–80.

3. Brown, G.W., 1951, "Iterative solutions of games by fictitous play", in "Activity analysis of production and allocation", ed. T.C. Koopmans, New York: Wiley.

4. Cooper, R.W., D. DeJong, R. Forsythe, T.W.Ross,1990," Selection criteria in coordination games: some experimental results", American Economic Review ,March 1990, 80, pp.218-233.

5. Demichelis, S. and F.Germano, 2000, "On knots and dynamics in games", CORE Discussion Paper 00/10.

6. Demichelis, S. and Klaus Ritzberger, 2000, "From evolutionary to Strategic Stability", CORE Discussion Paper 00/59.

7. Downs, A., 1957, "An Economic Theory of Democracy", Harper, New York.

8. Dhillon, A. and Susanna Peralta, "Economic Theories of Voter Turnout", mimeo, University of Warwick.

9. Feddersen, T. J. and W. Pesendorfer, 1996, "The swing voter's curse", The American Economic Review 86, pp. 408-424.

10. Feddersen, T. J. and A.Sandroni, 2001, "A Theory of Ethics and Participation in Elections", http://www.stanford.edu/group/SITE/sandroni.pdf.

11. Fey, M. (1997), "Stability and Coordination in Duverger's Law: A Formal Model of Preelection Polls and Strategic Voting", American Political Science Review.

12. Fischer, A.J. 1999, "The Probability of being Decisive", Public Choice 101, pp. 267-283.

13. Fudenberg, D. and D. Kreps, 1993, "Learning Mixed Equilibria", Games and Economic Behaviour,5, pp.320-367 .

14. Fudenberg, D and David K.Levine, (1998) "The theory of learning in Games", MIT Press.

15. Good, I.J. and Lawrence S.Mayer,1975 "Estimating the Efficacy of a Vote", Behavioural Science, 20(1), January, pp.25-33.

16. Harsanyi, J. and R. Selten, 1988, "A General Theory of Equilibrium Sel;ection in Games", MIT Press.

17. Kandori, M., G.J.Mailath, and R.Rob (1993), Learning, Mutations and Long Run Equilibria in Games, Econometrica, Vol.61, No.1, pp.29–56.

18. Ledyard, J. O., 1981, "The paradox of not voting and candidate competition: a general equilibrium ananysis", in G. Horwich and J. Quirck (eds.), *Essays in contemporary fields of economics*, Purdue.

19. Mueller, Dennis C., 1989, " Public Choice II", CUP.

20. Mulligan, Casey B. and C.Hunter, 2000, "The Empirical Frequency of a Pivotal Vote", University of Chicago Discussion Paper 00-25.

21. Palfrey,T.R. and H. Rosenthal (1983), A strategic calculus of voting, *Public Choice*, 41, pp.7–53.

22. Palfrey,T.R. and H. Rosenthal (1985), Voter Participation and Strategic Uncertainty, *The American Political Science Review*, Vol.79, pp.62–78.

23. van Huyck,R.B, R.C. Battalio, R.O.Beil, 1990, " Tacit Coordination games, Strategic uncertainty and Coordination failure", American Economic Review, March 1990, 80 ,pp. 234-248

24. Weibull, J.(1995), "Evolutionary Game Theory", MIT Press.

Appendix

Proof of Proposition 1: Let q_1 and q_2 represent the equilibria 1 and 2 respectively. We show that any trajectory beginning in the interval $[0, q_2]$ converges to q_1, thus $[0, q_2)$ is the required neighbourhood B^* for equilibrium 1, i.e. q_1, and any trajectory beginning in $(q_2, 1]$ converges to equilibrium 3 , i.e. to $q = 1$, hence the required neighbourhood for equilibrium 3 is $(q_2, 1]$.

Consider first a path starting in the interval $[0, q_2)$, i.e. $q(q_0, t) \in [0, q_2)$. By equation (6), $q(., t)$ is an increasing continuous function of t in this part of the domain, and remains so upto q_1. Hence $q(.)$ must converge to q_1. The other direction is the same. *square*

Proof of Proposition 2: Since any point in $[0, q_2)$ converges to equilibrium 1 and any point in $(q_2, 1]$ converges to equilibrium 3, the transition matrix of the Markov Chain is given by:

$$\begin{bmatrix} p_{11} & p_{13} \\ p_{31} & p_{33} \end{bmatrix}$$

where p_{11} is the probability that $q_0^i \in [0, q_2)$, the basin of attraction for equilibrium 1, if $q_\infty^{i-1} = q_1$, p_{31} is the probability that $q_0^i \in (q_2, 1]$, the basin of attraction of equilibrium 3 if $q_\infty^{i-1} = q_1$, p_{13} is the probability that $q_0^i \in [0, q_2)$, the basin of attraction for equilibrium 1, when $q_\infty^{i-1} = 3$ and p_{33} is the probability that $q_0^i \in (q_2, 1]$, the basin of attraction for equilibrium 3, when $q_\infty^{i-1} = 3$.

Note that when $N \to \infty$, $q_2 \to 1$. So given the distribution we have: $p_{1i} \to 1, p_{3i} \to 0, \quad i = 1, 3$. This conclusion is true as well with a more general distribution of q_0 as long as it is absolutely continuous with respect to the Lebesgue measure. In both cases it is easy to see that the invariant measure is concentration on state 1 in the first case or converges to a measure on 1 in the general case.