Evolution & Voting:

How Nature Makes us Public Spirited

John P. Conley[1]
University of Illinois
jpconley@uiuc.edu

Ali Toossi
University of Illinois
toossi@students.uiuc.edu

Myrna Wooders
University of Warwick
ecsdz@warwick.ac.uk

Revised: November 2001

Abstract

We reconsider the classic puzzle of why election turnouts are persistently so high even though formal analysis strongly suggests that rational agents should not vote. If we assume that voters are not making systematic mistakes, the most plausible explanation seems to be that agents receive benefits from the act of voting itself. This is very close to assuming the answer, however, and immediately begs the question of why agents feel a warm glow from participating in the electoral process. In this paper, we approach the question from an evolutionary standpoint. We show for a range of situations, that public-spirited agents have an evolutionary advantage over those who are not as public-spirited. We also explore conditions under which this kind of altruistic behavior is disadvantageous to agents. The details depend on the costs of voting, the degree to which different types of agents have different preferences over public policies and the relative proportions of various preference types in the population, but we conclude that evolution may often be a force that causes agents to internalize the benefits their actions confer on others.

1. Introduction

Any rational voter in a large society should realize that the probability that his vote will have an effect on the outcome of an election is negligible. Many classical writers in voting theory, Downs (1957) and Tullock (1968) for example, have argued that it simply does not pay to a citizen to show up at the polls. Even if a voter cares passionately about the outcome, the odds that his vote will be pivotal are so small that the expected benefit of casting a ballot would always be offset by even minor costs of voting. It is difficult to reconcile this with the fact that more than one hundred million Americans voted in the most recent presidential election.

Not surprisingly, there have been many attempts to provide a theory of voting that comports with actual observations. Ferejohn and Fiorina (1974), for example, suggested that voters might not be fully informed and so may not be able to calculate the probability that their vote would make a difference. They noted that this precludes voters from making an expected utility calculation and proposed instead that voters might be using minimax strategies. Since having voted when an agent ends up not being pivotal involves only a small regret (the cost of voting) but not having voted when an agent would have been pivotal may involve very large regret, minimax agents will generally choose to participate in elections.

Ferejohn and Fiorina's argument has the virtue that it does provide a foundation for rational voting. It is open to criticism, however on at least two grounds. Most obviously, it calls for agents to choose strategies in an extremely conservative and probably unrealistic way. For example, a minimax agent should never cross a street because it is possible that he might be hit by a car. More fundamentally, Ferejohn and Fiorina ignore the fact that the benefit of voting to any given agent depends on the actions of all of the other agents. While the expected utility approach can also be criticized for taking the probability a voter will be pivotal as exogenous and not depending on strategic interaction among voters, Ferejohn and

Fiorina go one step further. In suggesting that agents follow a minimax strategy, they are asserting that voters give no consideration at all to the strategic choices of others. It may be possible to justify this as an approximation for large economies, but it is at least a bit troubling to build a theory of voting on a foundation of strategic myopia.

More recently, several authors have reformulated the problem of why people vote to allow for strategic interaction between voters. For example, Palfrey and Rosenthal (1983) consider a model in which agents are completely informed about costs of voting and preferences of other voters. These agents play a noncooperative game in which they can either vote or abstain. They show that in some equilibria, there are substantial turnouts even for large economies. Unfortunately, these high turnout equilibria seem to be fragile, and as Palfrey and Rosenthal point out, the assumption of complete information is rather strong for large populations. The work of Palfrey and Rosenthal is partly based on the pioneering work of Ledyard (1981). There, and in a 1984 paper, Ledyard first explores the idea of strategic interaction among voters. In contrast to Palfrey and Rosenthal, Ledyard considers the case of voters who have incomplete information about the voting costs and preferences of their fellow citizens. Ledyard's key result is that equilibria with positive turnouts exist. Unfortunately, Palfrey and Rosenthal (1985) were able to show that when the electorate gets large, the cost of voting would again be the dominant factor for rational voters in Ledyard's model and so turnouts would be low. These results are reinforced by the recent work of DeMichelis and Dhillon (2001) in the context of a complete information learning model.

To summarize, although the game theoretic approach taken by Ledyard, Palfrey and Rosenthal do suggest that turnouts will be positive in many cases, their approach still does not explain what we actually observe. What is missing is a model in which agents have incomplete information and which at the same time exhibits robust equilibria in which the turnouts are substantial, even for large populations.

Riker and Ordeshook (1968) propose quite a different explanation of why it could be rational to vote. They suggest that agents might actually get utility from the act of voting itself. They show that if agents feel a sense of civic duty that is satisfied by going to polls, then large positive turnouts are not at all surprising regardless of the size of the electorate. This seems quite plausible, and the recent literature provides both empirical and experimental evidence that agents do indeed feel a "warm glow" from public-spirited activity. See Andreoni (1995), and references therein.[2] Despite the intuitive appeal of the civic duty explanation, it is somewhat disappointing from a theoretical standpoint. Saying that agents vote because they like to vote is essentially assuming the answer. As Andreoni (1990) points out in a somewhat different context, making such an assumption tends to rob the theory of its predictive power.

This provides the starting point for the current paper. Our main objective is to address the question of why agent might indeed have such a sense of civic duty. Is there some sense in which public-spiritedness in the context of voting is beneficial to agents? If so, to what degree of altruism is optimal? Fundamentally, we are asking how agents might come to have preferences that incorporate the welfare of others. We approach this question from an evolutionary standpoint. We show, for a range of situations in a voting game, that public-spirited agents have an evolutionary advantage over those who are not as public-spirited. We also explore when this kind of altruistic behavior is disadvantageous to agents. In general, we find that agents who like to vote will have an evolutionary advantage when voting is not too costly compared to the potential benefits of winning elections, and when the population of like-minded voters is large enough that winning an election is

---

[2] In an interesting paper, Kan and Yang (2001) explore an alterative explanation. They argue that agents get utility from voting because it allows them the pleasure of expressing themselves. they support this view with evidence for the 1988 US presidential elections. If "expressive voting" is in fact the reason that agents choose to turn up at the polls, our results would still make sense, but would need to be slightly reinterpreted. We would simply conclude that wanting to express one's opinion confers an evolutionary advantage rather than being public spirited *per se*.

a realistic possibility. The broader message is that evolution may be a force that causes agents to internalize the benefits their actions confer on others, at least to the extent that they all share a common set of preferences.

The plan of this paper is the following. In section 2 we describe the model. In section 3, we explore how the cost of voting, the size of the opposition and the degree to which preferences over public policies differ between groups affect the evolutionary benefits of voting. In section 4, we connect these results to the literature on evolution and altruism more generally and discuss possible extensions. Section 5 concludes.

2. The Model

We consider a dynamic economy with continuum of agents uniformly distributed on the interval [0,1]. Agent are divided into two types which we will designate H and L for "high" and "low" type voters, respectively. Two factors distinguish these types: preferences over public policies and propensities to vote. We denote the *share* of each type in the population by S_j. Since the population is divided between these two types we have

$$S_j \in [0,1] \text{ for } j = H, L, \text{ and } S_H = 1 - S_L.$$

Each period, agents vote on a randomly generated public proposal.[3] These proposals produce a cost or benefit for each type of agent that is uniformly distributed on the interval $[-1, 1]$. We assume that all agents of the same type have the same preferences over proposals, but that preferences between the types differ. Formally,

---

[3] Much of this model is a modification of Conley and Temimi (2001).

4

we denote the *benefit* that agents of type H receive from a proposal in any given period as a random variable B_H where:

$$B_H \sim \mathcal{U}(-1, 1).$$

We will allow the preferences of the two types to vary between perfectly correlated and perfectly independent.[4] Thus, we denote the benefit that agents of type L receive from a proposal in any given period as a random variable B_L where

$$B_L = \alpha B_H + (1 - \alpha) U_I$$

U_I is an independent uniform distribution on the interval $[-1, 1]$, and $\alpha \in [0, 1]$ is the *preference correlation parameter*. More generally, the correlation coefficient between B_H and B_L is:

$$Corr(B_H, B_L) = \frac{\alpha}{\sqrt{1 - 2\alpha + 2\alpha^2}}.$$

We denote the *propensity to vote* of the two types by V_H and V_L where

$$V_j \in (0, 1) \text{ for } j = H, L.$$

We also define the *relative public-spiritedness* of the two types as:

$$\beta = \frac{V_H}{V_L},$$

where $\beta \geq 1$. We shall assume that the likelihood of an agent choosing to vote for a proposal depends both on his innate propensity to vote (V_j) and the benefits that passage of a given realization the proposal will produce for him. More formally,

---

[4] In a previous version of this paper we also considered the case where the preferences are negatively correlated. The results do not differ radically and may be obtained from the author upon request. We have omitted them from the current paper in the interests of space.

we shall assume for any realization of the public proposal b_j, that $v_j \mid b_j \mid$ is the probability that a voter of type j will cast a ballot. This implies that the net turnout of voters of type j in any given election is a random variable given by:

$$TO_j = S_j V_j B_j.$$

Note that this number can be positive or negative. We will use the convention that a negative turnout measures the number of "No" votes while a positive one measures the number of "Yes" votes. Putting both types together implies that the total net turnout is a random variable given by:

$$TO_{net} = TO_H + TO_L = S_H V_H B_H + S_L V_L B_L,$$

We denote the *cost of casting a ballot* by $C > 0$ and assume it is the same for all agents. Since the voters show up at the polls with probabilities less than one, the realized voting cost to a voter of type j in any given election is also a random variable:

$$C_j = V_j \mid B_j \mid \ C.$$

Note that it is the probability of voting that affects the expected cost and not whether the vote was positive or negative, and this explains the absolute value term in the expression above.

From an algebraic standpoint, the expected payoff that members of each type receive in each period is rather complicated. It requires calculating the net turnout for any given realization of a proposal, and then integrating over all the proposals that pass, while subtracting the expected voting cost in each case. The net turnout, in turn, depends on the share of each type of agent in the population, the relative public-spiritedness of the types and the preference correlation parameter. We relegate both the expression and the derivation to the appendix. We shall, however, denote the *expected payoff* to agents of type j by:

6

$$\bar{\pi}_j, \text{ for } j = H, L$$

To model the evolution of the shares of each voter type over time we use standard replicator dynamics. According to this dynamic, the growth rate of the proportion of each type in the population is determined by the difference between its expected payoff and the population average payoff. Any type whose expected payoff is greater than average increases its share of the total population. Formally, the average payoff is:

$$\bar{\pi} = S_H \bar{\pi}_H + S_L \bar{\pi}_L.$$

In the interest of simplifying the model, we will treat the dynamics as taking place in continuous time. Since we will mainly be interested in showing how the parameters of the models and initial conditions of the economy influence which steady state the system converges to, this is innocuous. On the other hand, if we wanted to calculate the actual dynamic path we would have to explicitly take into account the fact the proposals are distinct and arrive at discrete points in time. This would introduce an degree on uncertainly in the paths since a particular set of initial conditions could lead to different steady states depending what specific proposals happened to randomly appear. We can think our choice to look at the continuous time version of the problem as moving the focus to "average" dynamics instead of exploring the entire distribution of possible paths. Thus, we shall assume that population shares evolve according to the following dynamic:

$$\dot{S}_j = S_j(\bar{\pi}_j - \bar{\pi})$$

where $\dot{S}_j$ is derivative of S_j with respect to time. The state of system at time t is given by the currently population shares:

$$S^t = (S_H^t, S_L^t).$$

7

We close this section with a remark. In the introduction, we told a story about evolution taking place over preferences and nature selecting for agents who felt a "warm glow" from altruistic actions. In the model above, however, altruistic actions appear to be programmed into behavior and evidently do not relate preferences at all. It would have been possible to derive the behavioral voting parameter (V_j) indirectly from an altruism parameter in preferences. Since the preference and behavioral parameters would be completely correlated in this case, we do not believe that much is to be gained from looking at these microfoundations. We therefore consider a reduced form in which V_j serves as a proxy for altruism in preferences. We do not believe any loss of generality results. In the appendix we provide an example of a utility function from which the behavior rule we describe can be derived. Further discussion on this point may also be found there.

3. Results

In this section, we focus on the steady states of the game. We will be most interested in showing how the parameters of voting game determine the population shares in the steady state to which the system converges.

The literature on evolution in economics is most concerned with the evolutionarily stable states (ESS). Testing for the stability of a steady state requires that the strategies agents play be shown to survive the introduction of small proportions of "mutant" strategies in the sense that they yield higher average payoffs. We will take up the question of how the presence of mutant players affects our equilibria in section 4. In this section, however, we will concentrate on finding the steady states themselves and will also study the likelihood a particular steady state will emerge as the outcome of the dynamic process. With this in mind, we shall say that the *type*

j wins the evolutionary game if the parameters and initial conditions are such that the economy converges over time to a stable steady state in which type j makes up the entire population ($S_j = 1$). We do this to simply our discussion, however, and not to assert that this necessarily is a compelling definition of evolutionary success.

We begin by the showing that steady states will always exist, and that there are three distinct possible dynamic situations for the economy.

Theorem 1. *Depending on values of parameters α, β and C there are three possible outcomes for the system:*[5]

1. **High type wins:** *The system has two steady states $S_H = 0$ and $S_H = 1$ where $S_H = 1$ is globally stable and $S_H = 0$ is unstable.*

2. **Large population wins:** *The system has three steady states, $S_H = 0$, $S_H = 1$ and $S_H = S_H^* \in (0,1)$ where $S_H = 0$ and $S_H = 1$ are asymptotically stable and their basins of attraction are $[0, S_H^*)$ and $(S_H^*, 1]$ respectively, and $S_H = S_H^*$ is unstable.*

3. **Low type wins:** *The system has two steady states $S_H = 0$ and $S_H = 1$ where $S_H = 0$ is globally stable and $S_H = 1$ is unstable.*

Figure 1 illustrates the three cases given in Theorem 1. What this result says is that sometimes, the high voter types will increase their share of the population until they make up the entire society regardless of how small their numbers are to begin with. This case is shown in Figure 1a. For other parameters, the low types will come to dominate the population regardless of their initial share. Figure 1c illustrates this. Both of these situations, however, are just limiting cases of what

---

[5] Let $F_t(S_0)$ be the value assumed by the state variable at time t when the initial condition at time 0 is S_0. A steady state S^* is stable if for every neighborhood U of S^* there is a neighborhood U_1 of S^* in U such that if $S_0 \in U_1$, $F_t(S_0) \in U_1$, $t > 0$. A steady state is asymptotically stable if it is stable and in addition if $S_0 \in U_1$, then $\lim_{t \to \infty} F_t(S_0) = S^*$. The basin of attraction of an asymptotically stable steady state is the set of all points S_0 such that $\lim_{t \to \infty} F_t(S_0) = S^*$. If there is a unique steady state with basin equal to the entire state space it is called globally stable.

we think of as the more typical case in which the initial population shares matter. In general, there will be two stable steady states and one unstable steady state that divides the basins of attraction. Figure 1b illustrates this. We will call this unstable steady state the *tipping point* and denote it S_H^*.

We now turn to the question of when being public-spirited is more likely to lead to evolutionary success. We begin by considering what happens as the cost of voting increases.

Theorem 2. *Assume that the parameters of the game are such that there are three steady states. Then all else equal, the higher the cost of voting C, the less likely the high voter types will win the evolutionary game.*

Proof/

See appendix.

To be slightly more formal, Theorem 2 says that if the parameters of the system are such that we are not in one of the two degenerate cases, then all else equal, as C increases, S_H^* approaches one and the basin of attraction of $S_H = 0$ expands. This means that as the cost of voting increases, the high voter type has to have a larger initial population share to prevent themselves from being squeezed out by the low voter type. Of course, this makes sense. If voting is costly, then the act of voting conveys that much less net increase in payoff to the high voter types. If voting is extremely costly, voting is a net loss, even to the group collectively. In this case, it is better to have a low voting parameter and we end in up in case 3 with the only stable steady state being $S_H = 0$ and the tipping point forced all the way up to $S_H = 1$.

Next we consider how the parameter of relative public-spiritedness β affects the evolutionary advantage of voting. It turns out that the cost of voting and the degree of preference correlation (which in turn affects the degree of free riding that the low types get to enjoy from the costly voting activity of high types) have an
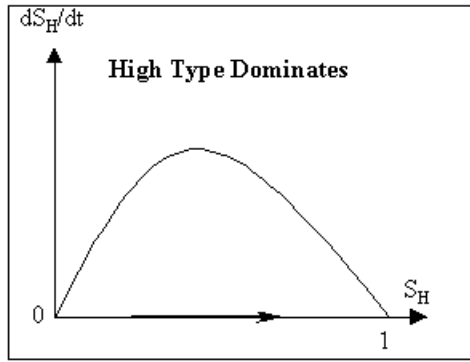
Figure 1a.
No matter what the initial population share, the high voting type will eventually make up the entire population. Thus, $S_H = 1$ is a globally absorbing state

Figure 1b.
The larger its initial population share, the more likely a type is to win the evolutionary game. The $S_H = S^*$ is an unstable steady state that divides the basins of attraction for the two stable steady states $S_H = 1$ and $S_H = 0$.

Figure 1c.
No matter what the initial population share, the low voting type will eventually make up the entire population. Thus, $S_H = 0$ is a globally absorbing state
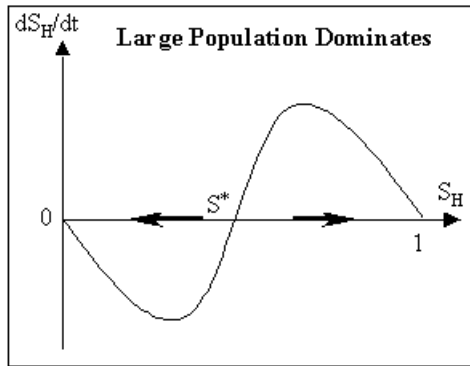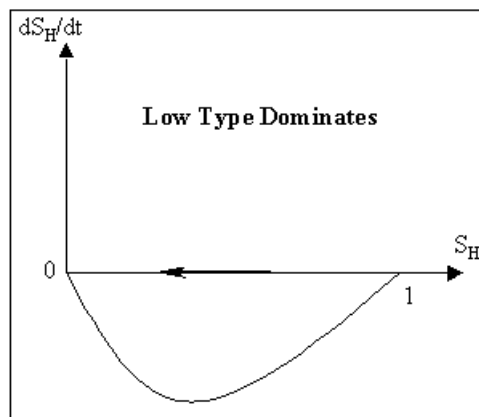
effect. As a consequence, more public spiritedness does not alway benefit a type. The next theorem shows this for the case of when voting is very costly.

Theorem 3. *If voting is too costly, then voting does not convey an evolutionary advantage.*

Proof/

The benefit and losses that voters get from each period from the public proposals that happen to pass by lie in the interval $[-1, 1]$ by assumption. On the other hand, voters must pay $V_i \, | \, B_i \, | \, C$ each period for voting. Thus, if C is high enough, the expected voting cost the high types pay compared to the low types each period (which grows without bound in C) will be larger than the expected difference the benefits from public projects. It follows that for large enough C, the low voter types get a higher expected payoff and so will always win the evolutionary game.

■

When voting is costless, a symmetric result is holds: public-spiritedness is always an advantage.

Theorem 4. *If voting is costless and preferences are not perfectly correlated, then voting conveys an evolutionary advantage.*

Proof/

If the agents with the high voter propensity increase their propensity to vote even more, the expected payoff from public projects relative to that received by the low voter type cannot decrease. This is easy to see. For any particular realization of a public proposal, the additional votes contributed by the high type voters either do or do not affect the outcome of the election. If the outcome is not affected, the relative payoff is not affected. If the outcome is affected it can only be because a proposal favored by the high type that would have failed passes instead (or the inverse). In either case the payoff to the high type goes up relative to the low type.

12

Given this, and since voting is costless, there is nothing on the negative side to offset these gains, and so the relative gains of the high voter type compared to the low increase as the V_H increases. A symmetric argument holds for the low types.

∎

We now consider the effects of the preference correlation parameter α. The question is: Is public-spiritedness more or less of an advantage for a group when they have preferences that are similar to the remaining population? Again it depends on the details of the economic parameters, but we are able to show an important result for the limiting case.

Theorem 5. *If preferences are perfectly correlated, then the low voter always win the evolutionary game.*

Proof/

Note that the payoff each type of agent gets from public proposals is identical in this case. Thus, if voting cost is positive, the type that votes more often gets a lower per capita payoff. The higher voting type therefore loses the evolutionary game.

∎

The fact that in the extreme case of perfect correlation, the high types are always supplanted by the low types in the steady state regardless of the initial population shares will turn out to have significant implications for the interpretation of our steady states as Evolutionary Stable Equilibria. We explore this more in the next section.

The previous two theorems explore only for extreme values of the parameters of the game. One might wonder whether voting conveys an evolutionary advantage in a more general case. We close by showing that for a range of parameters voting is beneficial to groups of agents.

Theorem 6. *If voting is sufficiently cheap and preferences are sufficiently uncorrelated, then the more public-spirited the low voter types are compared to the high voter types, the more likely they are to win the evolutionary game.*

Proof/

See appendix.

More formally, this says that the tipping point S^H moves in a way that favors the low type voters when they vote with higher frequency. This means that, all else equal they can win the evolutionary game with a lower initial share of the population. (A symmetric result holds when high type voters increase their voting propensity.) For this to be true, however, it must be the case voting is not too costly $C < \frac{(1-\alpha)}{2V_H}$. Otherwise, voting may be self-defeating. In addition, the preferences of voters must not be too highly correlated ($\alpha < 1/5$). Otherwise the free-riding benefits that the other type of agent gets from the costly voting efforts of the first may more than offset the advantages of winning a higher number of elections.

4. Evolution and Altruism.

The literature on evolution in economics is very large, and it is not our intention to survey it here. Instead, we shall concentrate on a discussion of how the model we present agrees with and differs from the existing literature.

Evolutionary game theory is typically used to explain who how agents might choose strategies in an arational way. Thus, evolution takes place over strategic choices. See Taylor and Jonker 1978, Friedman 1991, or more recently Lagunoff 2000, among many others. In contrast, we propose that evolution takes place over the underlying preferences of agents and those in turn determine their strategic

14

choices.[6] In this, we follow such authors as Becker (1976), Hirshleifer (1978) more recently Bergstrom and Stark (1993) and Robson (1996). (See Robson 2000 for a more complete survey.)

This raises an interesting question regarding whether or not our story can be reconciled with the traditional view in economics which seems to take evolution as metaphor for learning or imitation in strategic situations. See Kandori, Mailath and Rob (1993) or Fudenberg and Levine (1998) chapter 3, for example. We take a somewhat neutral view on this. Whether preferences come from nature (no learning) or nurture (passive learning) does not really matter for the results in our model. In either case, the actions of the parents are passed on through preferences to the children. What our model does not allow is a kind of active learning in which agents might somehow choose to undertake actions to shape their preferences as in Reiter (2001), for example. All in all, the major difference that evolving over preferences rather than strategies makes in interpretation is that the agents in our model are fully rational and behave in a strictly optimal way at all points.

The literature most closely related to the current paper relates to the evolutionary viability of altruism. In their seminal piece, Bergstrom and Stark (1993) consider a number of models but focus on one in which benefits of altruistic actions are felt amongst groups of siblings. Selfish siblings are at an advantage over altruistic ones in the same family, but pass on their selfish genes to their children. Since groups of altruistic siblings are at an advantage over groups of selfish siblings, the momentary benefit of exploiting one's own altruistic sibling is out-weighed by the evolutionary disadvantage of having a set of completely selfish children. The altruistic genes end up being successful. Eshel, Samuelson and Shaked (1998) pick up on another model described in Bergstrom and Stark in which agents are arranged in

---

[6] Recall that the voting propensity parameter (V_j) is a behavioral expression that reflects optimal altruistic actions of public-spirited agents. Thus, V_j is not a strategy, but rather a consequence of optimal voter choice given their preference for altruism. Of course, we treat the reduced form of the model and focus on providing an explanation for the presence of these altruistic preferences.

a circle and experience positive externalities when their direct neighbors choose to undertake costly altruistic actions. Agents choose a strategy each period by adopting the highest yielding action that they can directly observe. They show that for a correctly parameterized model, altruistic behavior survives and is stable against the introduction of mutations. Bester and Guth (1999) propose a model of externality producing duopolists. They show that if the production of one duopolist lowers the marginal cost of production for the other firm, then production choices are strategic complements. This means that when an altruistic firm chooses a higher than privately optimal production level, the other firm responses with its own higher production level, and this in turn benefits the first firm. Clearly, it is better to be selfish when paired with an altruist. Altruists, however, do much better when they happen to be paired with other altruists while egoists do much worse when they are paired with other egoists. As a result, altruists do better on the average, and are more successful from an evolutionary standpoint. (See also the comments of Bolle 2000 and Possajennikov 2000.)

There is common thread in all of these papers: local interaction. Eshel, Samuelson and Shaked's externalities extend only to adjacent neighbors, Bergstrom and Stark's only to groups of siblings, and Bester and Guth's only to pairs of duopolists. It is doubtful that any of these results could be generalized to more widespread externalities. What allows altruism to survive is that the altruist gene is able to recapture some part of the external benefit of its behavior.[7] In Eshel, Samuelson and Shaked's case, it is through teaching one's neighbors to be altruist, for Bergstrom and Stark it is by producing kids who have an evolutionary advantage, and for Bester and Guth it is though the strategic complementarity. It may appear that the model we describe breaks with this thread and does indeed allow for widespread

---

[7] To be a bit more precise, recapturing benefits of altruism only needs to take place in a relative sense. For example, recapture happens if egoists are benefit less from the acts of altruists that do other altruists.

externalities. This is only partly true. Our two groups of voters each consist of a continuum of agents, and when a proposal passes, the costs and benefits that result are purely public in nature. In this sense, the externalities are widespread. Notice, however, that preferences and voting propensity are completely linked by construction in our model. Thus, while benefits of proposals that pass are spread across many individuals, they are in a sense localized within a given genotype. We conclude that the gene recaptures much of the externality even though the individuals themselves do not get an advantage from voting.

Although the mechanism that allows altruism to survive in our group selection setting is similar the one at work for local interaction models described above, there remains the key question of the robustness of the steady states to the occurrence of mutations. Unless the steady states we find can survive strategic experimentation and random genetic drift, there is little reason to believe that we would ever observe them as the outcome of any evolutionary process.

As it turns out, the steady states in which the high voter types prevail are robust to the introduction almost all types of mutants. To see this, suppose we are in a steady state in which the high voter type makes up the entire population. Now introduce a small fraction of mutants with tastes that differ from the dominant type. Because the mutants make up such a small fraction of the economy, they have a negligible effect on elections and the proposals most favored by the dominant type will continue to pass. Thus, *provided that the tastes of the mutant are sufficiently different* from the dominant type, they will get a systematically smaller payoff than the dominant type regardless of their propensity to vote and will not upset the steady state. On the other hand, if the mutants have the same (or at least very similar) tastes for public proposals as the dominant type, then they can successfully free-ride on their voting efforts. Thus, a mutant with the same tastes, but a lower voting parameter can upset the steady state and will eventually supplant the original dominant type. Observe, however, that the free riding mutants are in turn

17

vulnerable to even less public-spirited mutants who otherwise share their tastes.

At first glance, this may seem like bad news. This analysis suggests that no steady state with agents who have any positive voting propensity is an ESS. There is a kind of Gresham's Law at work in which bad citizens force out good ones. We believe that the news is not so bad, however, and there are at least two possible ways to address the fact that the steady states we discover in our model are not ESS.

First, notice that the mutants we are worried about must have the same tastes but different voting propensity as the dominant type. There are reasonable arguments for why this may be an unlikely scenario in the real world. To the extent that preferences are literally based on genes, for example, it might be impossible to inherit a love of high levels of public spending without also having the public-spiritedness to vote. Both may be driven by the same "empathy" or "responsibility" gene, for example. To the extent that preferences are learned from the environment, the same argument might apply. Parents may teach their children to be empathetic and socially responsible and this would inform both the children's voting behavior and preferences over public proposals. If a child rejects his parent's teaching or gets a truly mutant gene he would necessarily find himself equipped both with preferences over public proposals and voting propensities that differ from those of his parents. Even if such mixed mutations were possible, it might be that there exist social sanctions to keep it from taking over. In other words, suppose a free rider arises. If this is detectable, the dominant group may protect itself by refusing to provide a mate for this mutant. After all, who wants a child to marry a selfish person? At a less extreme level, it may be that smaller social sanctions imposed by the dominant group more than offset the gain the free rider receives from not voting.[8] Thus, even though having the high voter types win the evolutionary game

---

[8] See Harbaugh (1996) for some interesting evidence that social sanctions and rewards do play a role in getting people to vote, and that people even try to lie about their voting behavior to receive

is not an ESS in a strict sense, there are still reasons to believe that this steady state my arise in real world settings.

Second, let us put aside the arguments just given, and suppose that free riding mutants with high type tastes can arise. The remarkable thing is that provided that mutation happens slowly enough, this actually improves the welfare of the dominant type and in a sense does not threaten its evolutionary success. Consider the following dynamic story: Initially we have two types of agents, high and low, and a small leavening of all possible types of mutants. Suppose that we are in a situation that converges to the high types dominating the population. If the mutants are small enough in number, their presence is not enough to prevent the high type for forcing the low voter types close to extinction. The only agent type that manages to increase its population proportion is the free riding mutant with high type tastes. Eventually, of course, enough time passes that the free riding mutant replaces the high type. This mutant in turn is eventually replaced by an even less public-spirited mutants with the same taste as the original high type and so on until public-spiritedness converges to zero. Thus, tastes of the original high types are evolutionary stable, but the altruism is not in the long run.

Notice, however, that in the initial state, there is a compelling social reason for the high types to vote. If there are many low voter types in the population with different tastes over public policies, voting by the high types is needed to assure that the public proposals favored by the high voter types win. As the low voter types begin to disappear, however, the high voters could win the elections even if they were less public-spirited since there are fewer of the low types to oppose them. Thus, in the steady state, *continuing to vote is socially wasteful because all of the opposition has been vanquished.* At this point, not only the individuals, but also the species itself benefits from having a lower voting parameter. In this modified

these rewards.

19

environment, free riders can thrive without threatening the survival of their type.

We think of this as a kind of decline and fall of the Roman empire story. Initially, for Rome to thrive, its citizens must be vigilant and willing to make sacrifices for the common good. If the neighboring cities contain less public-spirited citizens, they will be conquered and added to the empire. Eventually, however, Rome will have vanquished all of its enemies, and then it is better for everyone to spend public money on bread and circuses instead of a large standing army. Public-spirited sacrifice ceases to serve a useful purpose and it is time for Romans to rest on their laurels. The key, however, is to make sure that all of Rome's enemies have been destroyed before this decline into decadence. If the decline happens before all the Gauls have been pacified, decline turns into fall.

5. Conclusion

A feature of our model which may be open to criticism is that we find that only one type of agent can survive in the steady state. In reality, however, we seldom observe a completely homogeneous society. An interesting extension of our model might be to assume that agents experience diminishing marginal utility in public projects. In this case, the benefits accruing to whichever type of voter makes up the winning coalition would decline, while the prospective benefits to the opposition group of winning an election would remain high. This would suppress the winning coalition's turnout, make it more likely the opposition would begin to win elections, and slow the winning coalition's rate of growth even if they continued to win. It might be possible to find a stable interior solution in which both types of agents persist for such a model. Another interesting generalization would be allow more than two types. Simulation results suggest that if the groups are equally numerous,

20

preferences over public policies are uncorrelated and voting is cheap enough, then the type with the highest voting propensity will prevail in the evolutionary game. It is harder to prove theorems about this case, however, as the initial conditions (especially the covariance of tastes between agent types) can vary widely, and it is not immediately clear which are the most compelling benchmark cases.

Our work is motivated by our interpretation of the literature as suggesting that it is difficult to explain observed voting behavior on the basis of rational choice unless one assumes that agents get utility from the act of voting itself. In this paper we have attempted to provide a foundation for the warm glow associated with behaving in a public-spirited manner using evolutionary game theory. The basic result is that being public-spirited can confer an evolutionary advantage. Having a high propensity to vote is more advantageous when voting is less costly, when your group's preferences over public project differ sharply from those of competing groups, and when the competing group is less public-spirited or less numerous. We conclude that evolutionary forces may indeed play a role in causing agents to internalize the benefits their actions confer on their fellow agents, at least to the degree that they share a common set of preferences.

Appendix

Derivation of Payoff Functions

We begin with some preliminary that will simplify our calculations. First we define an the following:

$$\theta = -\frac{(1-\alpha)(1-S_H)}{\alpha + S_H(\beta - \alpha)}.$$

Denote the probability that a given proposal passes by P. This is calculated as follows:

$$P = Prob(S_H V_H B_H + S_L V_L B_L > 0 \mid B_H = b_H)$$
$$= Prob(S_H V_H b_H + (1 - S_H) V_L (\alpha b_H + (1 - \alpha) U_I) > 0)$$
$$= Prob\left(U_I > -\frac{\alpha + S_H(\beta - \alpha)}{(1 - \alpha)(1 - S_H)} b_H\right)$$
$$= Prob\left(U_I > \frac{b_H}{\theta}\right) = 1 - Prob\left(U_I < \frac{b_H}{\theta}\right)$$
$$= \begin{cases} 1 & \text{for } \frac{b_H}{\theta} \leq -1 \\ \frac{1}{2} - \frac{b_H}{2\theta} & \text{for } -1 < \frac{b_H}{\theta} < 1 \\ 0 & \text{for } \frac{b_H}{\theta} \geq 1. \end{cases}$$

In the calculations below, it will be more convenient to express this as follows:

$$P = \begin{cases} 1 & \text{for } -\theta \leq b_H \leq 1 \\ \frac{1}{2} - \frac{b_H}{2\theta} & \text{for } \theta < b_H < -\theta \\ 0 & \text{for } -1 \leq b_H \leq \theta \end{cases}.$$

The payoff that a high voting parameter agent can expect for a given proposal as follows:

$$E(\pi_H \mid B_H = b_H) = P(b_H - C_H) + (1 - P)(-C_H)$$
$$= P b_H - C_H.$$

Therefore the average payoff of a high voting type agent over all possible values of b_H is:

$$\bar{\pi}_H = E_{b_H}[E(\pi_H \mid B_H = b_H)]$$
$$= \int_{-1}^{1} \frac{(P b_H - C_H)}{2} db_H.$$

Substitution for P in the above integral gives:

$$\bar{\pi}_H = \int_{b_H=-1}^{b_H=\theta} \frac{-V_H C \mid b_H \mid}{2} db_H + \int_{b_H=\theta}^{b_H=-\theta} \left(\frac{b_H}{2}(\frac{1}{2} - \frac{b_H}{2\theta}) - \frac{V_H C \mid b_H \mid}{2}\right) db_H +$$
$$\int_{b_H=-\theta}^{b_H=1} \left(\frac{b_H}{2} - \frac{V_H C \mid b_H \mid}{2}\right) db_H$$
$$= .25 - \frac{\theta^2}{12} - \frac{V_H C}{2}.$$

Recall that b_H is constrained to lie in the interval $[-1, 1]$. Therefore, the calculation above is valid only if θ takes a value which keeps the limits of integration

within these bounds. It is immediate that $\theta \leq 0$. Thus, the calculation above is correct if and only if $\theta \geq -1$. We will therefore need to distinguish this case. It is easy to verify the following:

Case A: $-1 \leq \theta \leq 0$ if one of the following is true:
    i. $\frac{1}{2} \leq \alpha \leq 1$
    ii. $0 \leq \alpha \leq \frac{1}{2}$ and $S_1 \equiv \frac{1-2\alpha}{(1-2\alpha)+\beta} \leq S_H \leq 1$.

Case B: $\theta < -1$ if the following is true:
    i. $0 \leq \alpha \leq \frac{1}{2}$ and $0 \leq S_H < \frac{1-2\alpha}{(1-2\alpha)+\beta} \equiv S_1$.

Note that these two cases are exhaustive.

Clearly, if case B holds, it can never be true that $-1 \leq b_H \leq \theta$ or that $-\theta \leq b_H \leq 1$. Therefore the probability that a proposal passes is always given by the middle case: $\frac{1}{2} - \frac{b_H}{2\theta}$. This gives the following equation:

$$
\bar{\pi}_H = \int_{b_H=-1}^{b_H=1} \left(\frac{b_H}{2}(\frac{1}{2} - \frac{b_H}{2\theta}) - \frac{V_H C \mid b_H \mid}{2} \right) db_H
$$

$$
= -\frac{1}{6\theta} - \frac{V_H C}{2}.
$$

For the calculation of the low voting parameter agent payoff, we take a different route. Recall from the calculation of P that for values of $U_I > \frac{b_H}{\theta}$, the proposal passes, and otherwise it fails. Therefore we can calculate the payoff a low voting type can expect for a given proposal, $B_H = b_H$ as:

$$
E(\pi_L \mid B_H = b_H) = \int_{-1}^{\frac{b_H}{\theta}} \frac{-C_L}{2} du_I + \int_{\frac{b_H}{\theta}}^{1} \frac{(b_L - C_L)}{2} du_I.
$$

After a change of variable and taking expectation over all possible values of b_H we will have:

$$
\bar{\pi}_L = E_{b_H}[E(\pi_L \mid B_H = b_H)]
$$

$$
= \frac{1}{4(1-\alpha)} \left(\int_{-1}^{1} \int_{\frac{b_H(\alpha\theta+1-\alpha)}{\theta}}^{\alpha b_H+(1-\alpha)} b_L db_L db_H - \int_{-1}^{1} \int_{\alpha b_H-(1-\alpha)}^{\alpha b_H+(1-\alpha)} C_L db_L db_H \right).
$$

To make the presentation of the calculations easier we separate the above integration into two and substitute for C_L. We get:

$$
M \equiv \frac{1}{4(1-\alpha)} \int_{-1}^{1} \int_{\frac{b_H(\alpha\theta+1-\alpha)}{\theta}}^{\alpha b_H+(1-\alpha)} b_L db_L db_H
$$

23

$$N \equiv \frac{V_L C}{4(1-\alpha)} \int_{-1}^{1} \int_{\alpha b_H - (1-\alpha)}^{\alpha b_H + (1-\alpha)} |b_L| db_L db_H$$

This means that $\bar{\pi}_L = M - N$.

Not surprisingly, we run into similar problems regarding limits of integration. For different cases the calculation are as follows:

Case A.i:

$$M = \frac{1}{4(1-\alpha)} \left(\int_{\theta}^{-\theta} \int_{\frac{b_H(\alpha\theta+1-\alpha)}{\theta}}^{\alpha b_H + (1-\alpha)} b_L db_L db_H + \int_{-\theta}^{1} \int_{\alpha b_H - (1-\alpha)}^{\alpha b_H + (1-\alpha)} b_L db_L db_H \right)$$

$$= -\frac{\alpha\theta^2}{12} - \frac{\theta}{6} + \frac{\alpha\theta}{6} + \frac{\alpha}{4}$$

$$N = \frac{V_L C}{4(1-\alpha)} \left(\int_{-1}^{-\frac{(1-\alpha)}{\alpha}} -2\alpha(1-\alpha b_H) db_h + \int_{-\frac{(1-\alpha)}{\alpha}}^{\frac{(1-\alpha)}{\alpha}} \alpha^2 b_H^2 + (1-\alpha)^2 db_h + \right.$$

$$\left. \int_{\frac{(1-\alpha)}{\alpha}}^{1} (2\alpha(1-\alpha)b_H) db_h = \frac{(4\alpha^2 - 2\alpha + 1)V_L C}{6\alpha} \right.$$

Case A.ii:

In this case the calculation of the M is the same as in case A.i, but the calculation of N is as follows:

$$N = \frac{V_L C}{4(1-\alpha)} \int_{-1}^{1} (\alpha^2 b_H^2 + (1-\alpha)^2) db_h = \frac{(4\alpha^2 - 6\alpha + 3)V_L C}{6(1-\alpha)}$$

Case B:

In this case the calculation of the N is the same as in case A.ii, but the calculation of M is as follows:

$$M = \frac{1}{4(1-\alpha)} \int_{-1}^{1} \int_{\frac{b_H(\alpha\theta+1-\alpha)}{\theta}}^{\alpha b_H + 1 - \alpha} b_L db_L db_H = -\frac{\alpha}{6\theta} - \frac{(1-\alpha)}{12\theta^2} + \frac{(1-\alpha)}{4}$$

To summarize all of these results, the value of payoff functions for high and low voting types is the following:

Case	$\bar{\pi}_H$	$\bar{\pi}_L$
A. i.	$.25 - \frac{\theta^2}{12} - \frac{V_H C}{2}$	$-\frac{\alpha\theta^2}{12} - \frac{\theta}{6} + \frac{\alpha\theta}{6} + \frac{\alpha}{4} - V_L C d_1$
A. ii.	$.25 - \frac{\theta^2}{12} - \frac{V_H C}{2}$	$-\frac{\alpha\theta^2}{12} - \frac{\theta}{6} + \frac{\alpha\theta}{6} + \frac{\alpha}{4} - V_L C d_2$
B.	$-\frac{1}{6\theta} - \frac{V_H C}{2}$	$-\frac{\alpha}{6\theta} - \frac{(1-\alpha)}{12\theta^2} + \frac{(1-\alpha)}{4} - V_L C d_2$

where $d_1 = \frac{4\alpha^2 - 2\alpha + 1}{6\alpha}$ and $d_2 = \frac{4\alpha^2 - 6\alpha + 3}{6(1-\alpha)}$. Note that $\bar{\pi}_H$ and $\bar{\pi}_L$ are continuous and well behaved functions in S_H.

Proofs of Theorems

Theorem 1. *Depending on values of parameters α, β and C there are three possible outcomes for the system:*

1. **High type wins:** *The system has two steady states $S_H = 0$ and $S_H = 1$ where $S_H = 1$ is globally stable and $S_H = 0$ is unstable.*
2. **Large population wins:** *The system has three steady states, $S_H = 0$, $S_H = 1$ and $S_H = S_H^* \in (0,1)$ where $S_H = 0$ and $S_H = 1$ are asymptotically stable and their basins of attraction are $[0, S_H^*)$ and $(S_H^*, 1]$ respectively, and $S_H = S_H^*$ is unstable.*
3. **Low type wins:** *The system has two steady states $S_H = 0$ and $S_H = 1$ where $S_H = 0$ is globally stable and $S_H = 1$ is unstable.*

Proof/

The steady states are solution to $\dot{S}_H = 0$. The replicator dynamics can be written as follows:

$$\dot{S}_H = S_H(\bar{\pi}_H - \bar{\pi}) = S_H(1 - S_H)(\bar{\pi}_H - \bar{\pi}_L).$$

It is immediate that $S_H = 0$ and $S_H = 1$ are always steady states. The other steady state, if it exists, is the solution to $\bar{\pi}_H - \bar{\pi}_L = 0$. Calculating the roots of this equation is tedious, but the results are straightforward to verify. We show the calculations in detail for different cases.

Case A.i:

Substituting the values of payoff functions for this case into $\bar{\pi}_H - \bar{\pi}_L = 0$ gives us:

$$\Gamma^1 \equiv -(1-\alpha)\theta^2 + 2(1-\alpha)\theta - 6V_L C(\beta - 2d_1) + 3(1-\alpha) = 0.$$

We need some preliminary observations that makes the proof easier to understand. Note that the second derivative of Γ^1 with respect to θ is negative (for every θ). Thus Γ^1 is a concave function for all its range. Assuming $\alpha \neq 0$, a little algebra shows that both roots of equation $\Gamma^1 = 0$ are real if and only if $C \leq \frac{2(1-\alpha)}{3V_L(\beta - 2d_1)} \equiv C_1^*$. Thus for all values of $C < C_1^*$ the equation has two real roots. To simplify the equation define the constant term as follows: $K_1 \equiv 6V_L C(\beta - 2d_1) - 3(1-\alpha)$. Thus, the above equation becomes:

$$\Gamma^1 \equiv -(1-\alpha)\theta^2 + 2(1-\alpha)\theta - K_1 = 0$$

We call the two roots of this equation θ_1^+ and θ_1^- where:

$$\theta_1^+ = \frac{(1-\alpha) + \sqrt{(1-\alpha)^2 - K_1(1-\alpha)}}{(1-\alpha)}$$

$$\theta_1^- = \frac{(1-\alpha) - \sqrt{(1-\alpha)^2 - K_1(1-\alpha)}}{(1-\alpha)}$$

The fact that Γ^1 is concave implies the following:

A. $\theta < \theta_1^-$ or $\theta > \theta_1^+ \Rightarrow \Gamma^1 < 0 \Rightarrow \dot{S}_H < 0$.

B. $\theta_1^- < \theta < \theta_1^+ \Rightarrow \Gamma^1 > 0 \Rightarrow \dot{S}_H > 0$

Also recall that θ is a function of S_H and other variables. Solving for S_H in terms of θ gives the following:

$$S_H = \frac{\alpha\theta - \alpha + 1}{\alpha\theta - \alpha + 1 - \beta\theta}$$

Therefore, by substituting any valid roots, we can obtain the other steady state(s) of the system. The solution of the equation $\Gamma^1 = 0$ depends on the value of α.

1. First consider the case where $\alpha = 1$ (which implies the two types preferences are perfectly correlated). In this case $\bar{\pi}_H - \bar{\pi}_L = 0$ if and only if $6V_L C(\beta - 1) = 0$, which in is turn is true if and only if $\beta = 1$ (which implies there is no difference in voting behavior between the two types). For all $\beta > 1$, we have $\bar{\pi}_H < \bar{\pi}_L$. Thus, the only steady states in this case are $S_H = 0$ and $S_H = 1$. For all other values of S_H, $\dot{S}_H < 0$. This means that $S_H = 0$ is globally stable while $S_H = 1$ is globally unstable. Therefore, if preferences are perfectly positively correlated, no matter what the cost of voting is, the low voting type will be the winner. This means that case (3) of theorem obtains.

2. Next suppose $\frac{1}{2} \leq \alpha < 1$.

   As we saw above in this case, if $C < C_1^*$, the equation $\Gamma^1 = 0$ has two roots θ_1^+ and θ_1^-. However, θ_1^+ cannot be a solution. This is because $\theta_1^+ > 0$ and therefore either $S_H^* > 1$ (for $0 < \theta_1^+ < \frac{1-\alpha}{\beta-\alpha}$) or $S_H^* < 0$ (for $\theta_1^+ \geq \frac{1-\alpha}{\beta-\alpha}$).

   Now consider the other root, θ_1^-. As we mentioned above, for a root to give a valid solution, the associated steady state must satisfy the following: $0 \leq S_H^* \leq 1$. A little algebra shows that this implies that:

$$-\frac{(1-\alpha)}{\alpha} \leq \theta_1^- \leq 0.$$

26

Substituting the solution for θ_1^- given above and solving for allowable values of C gives us the following:

$$\frac{-4\alpha^3 + 4\alpha^2 + \alpha - 1}{6\alpha^2 V_L(\beta - 2d_1)} \equiv C_1^{min} \leq C \leq C_1^{max} \equiv \frac{(1-\alpha)}{2V_L(\beta - 2d_1)}$$

It is easy to verify that $C_1^{min} < C_1^{max} < C_1^*$.

Our final step is to determine the number and nature of the steady states as C varies.

a. If $C \leq C_1^{min}$ then from the above, we know $\theta_1^- < -\frac{(1-\alpha)}{\alpha}$. This in turn implies that $-\frac{\alpha}{\beta-\alpha} < S_H^* \leq 0$. There is no interior steady state in this case, only the boundaries, $S_H = 0$ and $S_H = 1$, remain. For stability properties of the steady states we find the sign of $\dot{S}_H$ for all values of $0 \leq S_H \leq 1$. For this note that the values of $0 \leq S_H \leq 1$ correspond to $\theta_1^- < -\frac{(1-\alpha)}{\alpha} \leq \theta \leq 0 < \theta_1^+$. As we saw above for these value of θ we have $\Gamma^1 > 0$ which implies $\dot{S}_H > 0$. This means that $S_H = 0$ is globally unstable while $S_H = 1$ is globally stable. Thus, case (1) of the theorem obtains.

b. If $C_1^{min} < C < C_1^{max}$ then $-\frac{(1-\alpha)}{\alpha} < \theta_1^- < 0$. This in turn means that $0 < S_H^* < 1$ and so we also have an interior $S_H = S_H^*$ in addition to the two at the boundaries. For determining the stability properties of steady states note that for values of $0 < S_H < S_H^*$, which correspond to $-\frac{(1-\alpha)}{\alpha} \leq \theta < \theta_1^-$ we have $\Gamma^1 < 0$ which means $\dot{S}_H < 0$. Also for values of $S_H^* < S_H < 1$ which correspond to $\theta_1^- < \theta < 0$ we have $\Gamma^1 > 0$ which means $\dot{S}_H > 0$. Therefore we have single interior steady state which is not stable. In addition, since $\dot{S}_H < 0$ for S_H close to zero, $S_H = 0$ is stable, and since $\dot{S}_H > 0$ for S_H close to one, $S_H = 1$ is stable. Thus, case (2) of the theorem obtains.

c. If $C_1^{max} \leq C < C_1^*$ then $\theta^- > 0$. As in the case of the positive root discussed above, this implies either $S_H^* > 1$ or $S_H^* < 0$. Thus, there is no interior steady state. For any interior value of the share of the high type ($0 \leq S_H \leq 1$ which implies $\theta < \theta_1^-$) we have $\Gamma^1 < 0$ which means $\dot{S}_H < 0$. This means that $S_H = 0$ is globally stable while $S_H = 1$ is globally unstable. Thus, case (3) of theorem obtains.

d. Finally, if $C \geq C_1^*$, then equation $\Gamma^1 = 0$ will not have any roots and since Γ^1 is concave, it will always be negative. Again, this means that there will be only two steady states $S_H = 0$ and $S_H = 1$ and for values of $0 \leq S_H \leq 1$ Γ^1 is negative, which means $\dot{S}_H < 0$. Again, case (3) of theorem obtains.

Case A.ii:

Substituting the values of payoff functions for this case into $\bar{\pi}_H - \bar{\pi}_L = 0$ gives us:
$$\Gamma^2 \equiv -(1-\alpha)\theta^2 + 2(1-\alpha)\theta - 6V_L C(\beta - 2d_2) + 3(1-\alpha) = 0.$$

Note that Γ^2 is also a concave function. Both roots of equation $\Gamma^2 = 0$ are real if and only if $C \leq C_2^* \equiv \frac{2(1-\alpha)}{3V_L(\beta-2d_2)}$. Thus for all values of $C < C_2^*$ the equation has two real roots. The same argument about the relationship of the location of θ relative to the roots of the equation and the sign of Γ^2 holds as in the previous case.

To simplify the equation define the constant term as follows: $K_2 \equiv 6V_L C(\beta - 2d_2) - 3(1-\alpha)$. Thus, the above equation becomes:

$$\Gamma^2 \equiv -(1-\alpha)\theta^2 + 2(1-\alpha)\theta - K_2 = 0$$

We call the two roots of this equation θ_2^+ and θ_2^- where:

$$\theta_2^+ = \frac{(1-\alpha) + \sqrt{(1-\alpha)^2 - K_2(1-\alpha)}}{(1-\alpha)}$$

$$\theta_2^- = \frac{(1-\alpha) - \sqrt{(1-\alpha)^2 - K_2(1-\alpha)}}{(1-\alpha)}$$

However, θ_2^+ cannot be a solution. This is because $\theta_2^+ > 0$ and therefore either $S_H^* > 1$ or $S_H^* < 0$. Now consider the other root, θ_2^-. As we mentioned above, for a root to give a valid solution, the associated steady state must satisfy the following: $0 \leq S_H^* \leq 1$. In this case this implies that:

$$-1 \leq \theta_2^- \leq 0.$$

Substituting the solution for θ_2^- given above and solving for allowable values of C gives us the following:

$$0 \leq C \leq C_2^{max} \equiv \frac{(1-\alpha)}{2V_L(\beta - 2d_2)}$$

It is easy to verify that $C_2^{max} < C_2^*$.

Our final step is to determine the number and nature of the steady states as C varies.

a. If $C \leq C_2^{max}$ then from the above, we know $-1 \leq \theta_2^- \leq 0$. This in turn implies that $S_1 \leq S_H^* \leq 1$. There is an interior steady state in this case in addition to the boundary solution $S_H = 1$. For stability properties of the steady states we find the sign of $\dot{S}_H$ for all values of $S_1 \leq S_H \leq 1$. Note that the values

28

$S_1 \leq S_H \leq S_H^*$, correspond to $-1 \leq \theta < \theta_2^-$. For these value of θ we have $\Gamma^2 < 0$ which implies $\dot{S}_H < 0$. Also the values $S_H^* < S_H \leq 1$, correspond to $\theta_2^- < \theta \leq 0$. For these value of θ we have $\Gamma^2 > 0$ which implies $\dot{S}_H > 0$. This means that $S_H = 1$ is stable while $S_H = S_H^*$ is unstable. We will show in case B that irrespective of value of C, there is no other interior solution between 0 and S_1 and there is only a boundary solution $S_H = 0$, which is stable. Thus, case (2) of the theorem obtains.

b. If $C_2^{max} < C \leq C_2^*$ then $\theta_2^- > 0$. Therefore the same argument for θ_2^+ applies here and we don't have any interior solution. Thus the only steady state is $S_H = 1$. For determining the stability properties of steady state note that for values of $S_1 < S_H < 1$, which correspond to $-1 \leq \theta < \theta_2^-$ we have $\Gamma^2 < 0$ which means $\dot{S}_H < 0$. As we will show in case B, irrespective of value of C, there is no other interior solution between 0 and S_1 and there is only a boundary solution $S_H = 0$, which is unstable. This means that $S_H = 0$ is globally stable and $S_H = 1$ is unstable. Thus, case (3) of the theorem obtains.

c. Finally, if $C > C_2^*$, then equation $\Gamma^2 = 0$ will not have any roots and since Γ^2 is concave, it will always be negative. Again, this means that there will be only one steady states $S_H = 1$. For values of $S_1 \leq S_H \leq 1$ Γ^2 is negative, which means $\dot{S}_H < 0$. Considering our results in case B, again, case (3) of theorem obtains.

Case B:

Substituting the values of payoff functions for this case into $\bar{\pi}_H - \bar{\pi}_L = 0$ gives us:

$$\Gamma^3 \equiv -\theta^2 \left(6V_L C(\beta - 2d_2) + 3(1 - \alpha)\right) - 2(1 - \alpha)\theta + (1 - \alpha) = 0.$$

To simplify the equation define the constant term as follows: $K_3 \equiv 6V_L C(\beta - 2d_2) + 3(1 - \alpha)$. Thus, the above equation becomes:

$$-K_3\theta^2 - 2(1 - \alpha)\theta + (1 - \alpha) = 0.$$

Note that $K^3 > 0$ and that Γ^3 is also a concave function. The same argument about the relationship of the location of θ relative to the roots of the equation and the sign of Γ^3 holds as in the previous cases.

We call the two roots of this equation θ_3^+ and θ_3^- where:

$$\theta_3^+ = \frac{-(1 - \alpha) + \sqrt{(1 - \alpha)^2 + (1 - \alpha)K_3}}{K_3}$$

$$\theta_3^- = \frac{-(1 - \alpha) - \sqrt{(1 - \alpha)^2 + (1 - \alpha)K_3}}{K_3}$$

29

Note that here since $(1-\alpha)^2 + (1-\alpha)K_3 > 0$ the two real roots always exist. For a root to give a valid solution, the associated steady state must satisfy the following: $0 \leq S_H^* \leq S_1$. In this case this implies that the roots must be between $-\frac{1-\alpha}{\alpha}$ and -1. The first root, i.e. θ_3^+ is positive. So there is no interior steady state corresponding to this root. It is also easy to check that $-1 < \theta_3^- < 0$. Therefore there are no interior steady states. Hence in this case for any value of $C > 0$ the steady states are $S_H = 0$ and $S_H = 1$. As to the stability property of the steady state, we note that for values of $0 \leq S_H^* \leq S_1$ which correspond to values of $\theta < \theta_3^-$, we have $\Gamma^3 < 0$ which means that $\dot{S}_H < 0$. This means that $S_H = 0$ is stable.

∎

Theorem 2. *All else equal, the higher the cost of voting C, the less likely the high voter types will win the evolutionary game.*

Proof/
    We assume that there is an interior steady state S_H^*. Therefore we should only consider cases A.i and A.ii. Since the two cases are very similar we will provide a proof only for case A. i. since the other cases are essentially repetitions of the same argument. As we argue above, $S_H^* = \frac{\alpha\theta^- + 1 - \alpha}{\alpha\theta^- + 1 - \alpha - \beta\theta^-}$ where $\theta^- = \frac{(1-\alpha) - \sqrt{(1-\alpha)^2 - K_1(1-\alpha)}}{(1-\alpha)}$, $K_1 = 6V_L C(\beta - 2d_1) - 3(1-\alpha)$, and $d_1 = \frac{4\alpha^2 - 2\alpha + 1}{6\alpha}$. Since we are considering case A. i., we know that $\frac{1}{2} \leq \alpha \leq 1$. It is easy to verify that this implies that $\frac{1}{3} \leq d_1 \leq \frac{1}{2}$ and since $\beta > 1$

$$\frac{\partial K^1}{\partial C} = 6V_L(\beta - 2d_1) > 0$$

It is also the case that $\frac{\partial \theta^-}{\partial K_1} > 0$, and that

$$\frac{\partial S_H^*}{\partial \theta^-} = \frac{(1-\alpha)\beta}{(\alpha\theta^- + 1 - \alpha - \beta\theta^-)^2} > 0.$$

Putting this altogether we get $\frac{\partial S_H^*}{\partial C} > 0$. Therefore as C increases S_H^* will move towards 1. This means that the high voter types must make up a larger share of the initial population if they are to win the evolutionary game. Thus, as C increases it is less likely that the high voter types will win.

∎

Theorem 6. *If voting is sufficiently cheap and preferences are sufficiently uncorrelated, then the more public spirited the low voter types are compared to the high voter types, the more likely they are to win the evolutionary game.*

Proof/

More formally, we shall demonstrate that if $C < \frac{(1-\alpha)}{2V_H}$ and $\alpha < 1/5$ that $\frac{\partial S_H^*}{\partial V_L} > 0$, that is, the basin of attraction for $S_H = 0$ expands.

Note that we are now considering a result for case (A. ii) By arguments similar to those given for theorem 1, we can establish that

$$S_H^* = \frac{\alpha\theta_2^- + 1 - \alpha}{\alpha\theta_2^- + 1 - \alpha - \beta\theta_2^-}$$

where $\theta_2^- = \frac{(1-\alpha)-\sqrt{(1-\alpha)^2-(1-\alpha)K_2}}{(1-\alpha)}$. and $K_2 = 6V_L C(\beta - 2d_2) - 3(1-\alpha)$

To prove our result, we need to take the derivative of S_H^* with respect to V_L. The algebra is dense, but after simplification we get the following:

$$\frac{\partial S_H^*}{\partial V_L} = \frac{\frac{V_H}{2(1-\theta_2^-)}[\theta_2^{-2}(2\alpha\theta_2^- + 1 - 3\alpha) - 6CV_H + 3(1-\alpha)]}{[(\alpha - 1 - \alpha\theta_2^-)V_L + \theta_2^- V_H]^2}.$$

The denominator is positive. Since we are considering case (A. ii.), $-1 < \theta_2^- < 0$ and therefore $1-\theta_2^- > 0$ and so $\frac{V_H}{2(1-\theta_2^-)} > 0$. We focus on the rest of the numerator.

1. First consider $2\alpha\theta_2^- + 1 - 3\alpha$. We know $-1 < \theta_2^- < 0$. We multiply by 2α and then add $(1 - 3\alpha)$ to get:

$$-2\alpha + 1 - 3\alpha < 2\alpha\theta_2^- + 1 - 3\alpha < 1 - 3\alpha.$$

Simplifying gives:
$$1 - 5\alpha < 2\alpha\theta_2^- + 1 - 3\alpha < 1 - 3\alpha.$$

Since $\alpha > 1/5$ by assumption, $1 - 5\alpha$ is positive and so is the expression under consideration.

2. Now consider $-6CV_H + 3(1-\alpha)$. Very directly, since $C < \frac{(1-\alpha)}{2V_H}$ by assumption the expression is positive. Therefore, $\frac{\partial S_H^*}{\partial V_L} > 0$ and so higher voting propensity conveys an evolutionary advantage on the low type voters. A similar result is true for the high type voters in this case.

■

Derivation of the voting behavioral rule for a rational agent

Suppose that an agent has the following utility function where

B - *ex anti per capita* cost or benefit of a the public proposal to agents of his type
B_r - *ex post* cost of benefit received by an agent (may be zero if the proposal fails)
C - cost of voting

p -probability of voting
x -private good consumption

$$U(B, B_r, C, p, x) = x + B_r + (B + C)p - (\frac{1}{2V})p^2.$$

The budget constraint is $\omega = x + pC$. Substituting this in to the utility function and maximizing this with respect to p gives the following first order condition:

$$\frac{\partial U}{\partial p} = -C + B + C - \frac{1}{V}p = 0,$$

Which gives a solution $p = vB$. This is the linear behavior rule we explore in above.

This utility function warrants some discussion. The idea we are attempting to capture is that the agent is an altruist who enjoys voting in proportion to how much benefit the proposal would covey to his type and how much effort he has to go to in order to vote. The second part of this may seem strange at first as it says that the more the agent has to exert himself to vote, the happier he is, at least as far as his altruistic feelings go. While we would not want to argue that this is always the case, it seems reasonable that in some cases agents get a warm glow from working hard to help their fellow man. (Note, however, that cost of voting is still a negative in that it affects the budget constraint.) If we were to remove this term, the behavioral rule would get more complicated in that agents would choose not to vote when the per capita benefits of voting were lower than the expected costs of voting. This would introduce discontinuities into the behavior and would substantially complicate the proof of the results. Since the proofs are already algebraically dense, we do not pursue this further.

References

Aldrich, John H. (1997): "When is it rational to vote," in *Perspectives on Public Choice,* by Dennis C. Muller Cambridge University Press, chapter 17.

Andreoni, James (1990): "Impure Altruism and Donations to Public Choice: A Theory of Warm-Glow Giving," *The Economic Journal,* 100: 464-477.

Andreoni, James (1995): "warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments," *The Quarterly Journal Of Economics,* pp. 1-21.

Becker, Gary (1976): "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology," *Journal of Economic Literature,* V. 14, pp. 817-26.

Bergstrom, Theodore C. and Stark Oded (1993): "How can Altruism prevail in an Evolutionary Environment," *American Economic Review,* v.83 pp. 149-155.

Bester, Helmut and Guth, Werner (1998): "Is altruism evolutionarily stable?," *Journal Of Economic Behavior & Organization,* pp. 193-209.

Bolle, Fried el (2000): "Is altruism evolutionarily stable? And envy and malevolence? Remarks on Bester and Guth," *Journal Of Economic Behavior & Organization,* pp. 131-133.

Conley, John and Akram Temimi (2001): "Endogenous Enfranchisement when Groups' Preferences Conflict," *Journal of Political Economy,* V. 109, pp. 79-102.

DeMichelis, Stefano, and Amrita Dhillon (2001): "Learning in Elections and Voter Turnout Equilibria," .

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

Eshel, Ilan ; Samuelson, Larry and Shaked Avner (1998): "Altruists, Egoists, and Hooligans in a Local Interaction Model," *The American Economic Review,* v.88 pp. 157-179.

Ferejohn, John A. , and Morris P. Fiorina (1974): "The Paradox of Not Voting: A Decision Theoretic Analysis," *American Political Science Review,* 68:525-36.

Friedman, Daniel (1991): "Evolutionary Games in Economics," *Econometrica,* pp. 637-666.

Fudenberg, Drew, and David K. Levine (1998): *The Theory of Learning in Games.* The MIT press.

Green, Donald P. , and Ian Shapiro (1994): *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science.* Yale University Press.

Hirshleifer, Jack (1977): "Economics from a Biological Viewpoint," *Journal of Law and Economics,* V. 20, pp 1-52.

Hirshleifer, Jack (1978): "Natural Economics Versus Political Economy," *Journal of Social and Biological Structures.,* V. 20, pp 1-52.

Harbaugh, William (1996): "If people vote because they like to, then why do so many of them lie?," *Public Choice,* V. 89, pp 63-76.

Lagunoff, Roger (2000): "On the Evolution of Pareto Optimal Behavior in Repeated Coordination Problems," *International Economic Review,* v. 41 pp. 273-93.

Ledyard, John O. (1981): "The Paradox of Voting and Candidate Competition: A general Equilibrium Analysis," in *Essays in Contemporary Fields of Economics,* by G. Horwich and J. Quirk West Lafayette: Purdue University Press.

Ledyard, John O. (1984): "The Pure Theory of Large Two-Candidate Elections," *Public Choice,* 44:7-41.

Kan, Kamhon and Cheng-Chen Yang (2001): "On Expressive Voting: Evidence from the 1988 US Presidential Election," *Public Choice,* Forthcoming.

Muller, Dennis C. (1989): *Public Choice II: A revised Edition of Public Choice.* Cambridge University Press.

Palfrey, Thomas R. , and Howard Rosenthal (1983): "A Strategic Calculus of Voting," *Public Choice,* 41:7-53.

Palfrey, Thomas R. , and Howard Rosenthal (1985): "Voter Participation and Voter Uncertainty," *American Political Science Review,* 79:62-78.

Possajennikov Alex (2000): "On the evolutionary stability of altruistic and spiteful preferences," *Journal Of Economic Behavior & Organization,* pp. 125-129.

Riker, William H. , and Peter C. Ordeshook (1968): "A Theory of the Calculus of Voting," *American Political Science Review,* 62:25-43.

Smith, J. Maynard and Price G. R. (1973): "The logic of animal conflict," *Nature,* v. 246, pp. 15-18.

Reiter, Stanley (2001): "Interdependent Preferences and Groups of Agents," *Journal of Economic Theory,* v. 3 pp. 27-69.

Robson, A (1996): "A Biological Basis for Expected and Non-Expected Utility Theory," *Journal of Economic Theory,* V. 68, pp. 397-424.

Robson, A (2000): "The Biological Basis for Economic Behavior," Manuscript.

Taylor, Peter D. and Jonker, Leo B. (1978): "Evolutionary Stable Strategies and Game Dynamics," *Mathematical Bioscience,* pp. 145-156.

Tullock, G. (1967): *Towards a Mathematics of Politics.* Ann Arbor: University of Michigan Press.

Weibull, Jörgen W. (1997): *Evolutionary Game Theory.* The MIT Press.