MEMETICS & VOTING:
HOW NATURE MAY MAKE US PUBLIC SPIRITED

by

John P. Conley and Myrna Wooders

Working Paper No. 05-W14

April 2005

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Memetics & Voting:

How Nature May Make us Public Spirited

John P. Conley[1]
Vanderbilt University
j.p.conley@vanderbilt.edu

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

Myrna Wooders
Vanderbilt University
Myrna.Wooders@vanderbilt.edu

Revised: April 2005

[1] We wish to thank Jim Andreoni, Robert Aumann, Larry Samuelson and the participants of Games 2000 in Bilbao, Spain for their helpful comments. All errors are the responsibility of the authors.

Abstract

We consider the classic puzzle of why people turn out for elections in substantial numbers even though formal analysis strongly suggests that rational agents would not vote. If one assumes that voters do not make systematic mistakes, the most plausible explanation seems to be that agents receive a warm glow from the act of voting itself. However, this begs the question of why agents feel a warm glow from participating in the electoral process in the first place. We approach this question from an memetic standpoint. More specifically, we consider a model in which social norms, ideas, values, or more generally, "memes" influence the behavior of groups of agents, and in turn, induce a kind of competition between value systems. We show for a range of situations that groups with a more public-spirited social norm have an advantage over groups that are not as public-spirited. We also explore conditions under which the altruistic behavior resulting from public-spiritedness is disadvantageous. The details depend on the costs of voting, the extent to which different types of citizens agree or disagree over the benefits of various public policies, and the relative proportions of various preference types in the population. We conclude that memetic evolution over social norms may be a force that causes individuals to internalize the benefits that their actions confer on others.

1. Introduction

Any rational voter in a large population should realize that the probability 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 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 American presidential election.

Not surprisingly, there have been many attempts to provide a theory of voting that agrees with actual observations. Ferejohn and Fiorina (1974), for example, suggest that voters might not be fully informed and so might not be able to calculate the probability that their votes would make a difference. They note that this precludes voters from maximizing expected utility and propose instead that voters might be using minimax strategies. Since having voted when not pivotal involves only a small regret (the cost of voting) while not having voted when pivotal may induce very large regret, minimax agents usually choose to participate in elections.

Ferejohn and Fiorina's insightful argument has the virtue that it provides 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 perhaps unrealistic way. For example, a minimax agent would 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 to any given citizen of voting depends on the actions of all the other citizens. 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 voters follow a minimax strategy, they are asserting that voters give no consideration to the strategic choices of others. It may be possible to justify this as an approximation for large societies, but it would be preferable to build a game theoretic approach to voting on a foundation that did not assume this type of strategy 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 voters are completely informed about costs of voting and preferences of other voters. These voters play a noncooperative game in which an individual can either vote or abstain. Palfrey and Rosenthal show that even for large societies, there are some equilibria with substantial turnouts. Unfortunately, these high turnout equilibria seem to be fragile, and as Palfrey and Rosenthal point out, the assumption of complete information appears to be 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 explores the idea of strategic interactions among voters. In contrast to Palfrey and Rosenthal, Ledyard considers the case of voters who have incomplete information about 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 in Ledyard's model that when the electorate gets large, the cost of voting would again be the dominant factor for rational voters 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 approaches taken by Ledyard and Palfrey and Rosenthal suggest that turnouts will be positive in many cases, it is still not clear that they reflect what we actually observe. What seems to be missing is a model with robust equilibria in which turnouts are substantial, even for large societies.

Riker and Ordeshook (1968) propose quite a different explanation of why it might be rational to vote. They suggest that voters may actually get utility from the act of voting itself. They show that if voters feel a sense of civic duty that is satisfied by going to the 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] While we feel civic duty/warm glow is the clearly the best explanation available for voting behavior, saying that voters vote because they like to vote has the somewhat troubling flavor of assuming the answer. As Andreoni (1990) points out in a somewhat different context, making such an assumption risks robbing the theory of its predictive power.

We take a mememetic approach to addressing this question.[3] Memes are similar to genes and to other replicators such as computer viruses. Memes can be transmitted between individuals and among groups in minutes. Thus, replication

---

[2] In an interesting paper, Kan and Yang (2001) explore an alternative 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*.

[3] Richard Dawkins introduced this word in the 1976 book "The Selfish Gene". There he writes: We need a name for the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of imitation. 'Mimeme' comes from a suitable Greek root, but I want a monosyllable that sounds a bit like 'gene'. I hope my classicist friends will forgive me if I abbreviate mimeme to meme. If it is any consolation, it could alternatively be thought of as being related to 'memory', or to the French word *même*. It should be pronounced to rhyme with 'cream'.
Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passed it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. As my colleague N.K. Humphrey neatly summed up an earlier draft of this chapter: '... memes should be regarded as living structures, not just metaphorically but technically.(3) When you plant a fertile meme in my mind you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell. And this isn't just a way of talking – the meme for, say, "belief in life after death" is actually realized physically, millions of times over, as a structure in the nervous systems of individual men the world over.'

of memes can be extremely rapid compared to genes. On the other hand, copy-fidelity of memes may be much lower, and the within population variation, higher, than for genes. An example of a meme in the animal world is bird songs. While it was once thought that bird songs evolved genetically, it was recently found that birds have begun imitating cell phone ring tones. In human society, almost may aspect of a culture can be seen as a meme: religion, language, fashion, music, scientific theories and concepts, social conventions, traditions, etc. We refer the reader to Blackmore (1999) for an interesting and wide ranging speculative discussion of memes in human societies.

This provides the starting point for the current paper. Our main objective is to address the more basic question of how populations might develop a social norm for civic duty. Is there some sense in which public-spiritedness in the context of voting is beneficial? If so, what degree of altruism is optimal? Fundamentally, we ask how memes for warm-glow/civic duty might successfully be encoded into altruistic preferences that incorporate the welfare of others and in turn lead individuals to choose to undertake public-spirited actions such as voting.

We show for a range of voting game situations that groups of public-spirited citizens have an advantage over those that are not as public-spirited. We also explore when this kind of altruistic behavior is disadvantageous. In general, we find that groups with a stronger social norm for civic duty will have an 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 so that winning an election is a realistic possibility. In these circumstances, societies that possess the more altruistic meme are better off and this meme tends to replicate more quickly as a result. We model this as an evolutionary process in which the fittest memes tend to spread and force out the less fit memes.

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 benefits of voting and in turn the population dynamics of a society. 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 a 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, citizens vote on a randomly generated public proposal[4] that produces a cost or benefit for each type of agent. We assume that all agents of a given type have the same preferences over proposals, but that preferences between the types differ.

We think of high types as social leaders of society and assume that the costs and benefits of proposals are uniformly distributed on the interval $[-1, 1]$. Formally, the *benefit* that agents of type H receive from a proposal in any given period is a uniformly distributed random variable denoted B_H:

[4] This part of model is similar to Conley and Temimi (2001) who study why current voters agree to extend the franchise to new voters as opposed to warm glow, preference formation and evolution as in the current work.

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

We think of low types as social followers in the sense that they experience only a partial spillover of the costs and benefits the mainstream gets from a project. They also experience a separate uncorrelated impact from policy proposals. We consider situations in which the relative weight on this uncorrelated benefit low types varies between unity (making benefits between high and low types zero) and zero (making benefits perfectly correlated)[5] Formally, 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*. This implies that the correlation coefficient between B_H andB_L is:

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

Note that this means that while the distribution of costs and benefits for the high types is uniform, the distribution for the low types is uniform when benefits are either perfectly correlated or perfectly uncorrelated, but becomes more less uniform and less variable when benefits are partially correlated. An interpretation of this is that groups who are partial integrated into the mainstream of society have, in a sense, a diversified portfolio. While they may not fully share in the benefits the broader society gets, they don't suffer all of the costs the mainstream might incur

[5] 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 authors upon request. We have omitted them from the current paper in the interest of space.

either. There are initiatives and an economic structure that exist within the less enfranchised group that yield a stream of costs and benefits that are unrelated to those received by the mainstream in the face of policy changes.[6]

We denote the *propensity to vote* for each 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 of the proposal will produce for him. More formally, 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 *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 depending on the sign of B_j. 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. If we add together the turnouts of low and high type we get a measure of "total net turnout". Note that if total net turnout is positive, a proposal receives the majority of votes cast.

$$TO_T = TO_H + TO_L = S_H V_H B_H + S_L V_L B_L.$$

[6] We thank an anonymous referee for pointing out some of these details.

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:

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

where $\mid \bullet \mid$ denotes the absolute value. Note that it is the probability of voting that affects the expected cost and not whether the vote was positive or negative; 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. Solving for the expected payoff 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 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:

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

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 be interested mainly in showing how the parameters of the models and initial conditions of the economy influence steady states to which the system converges, this is innocuous. Note that 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 a degree of uncertainty in the paths because a particular set of initial conditions could lead to different steady states depending on what specific proposals happened to randomly appear. Our choice to look at the continuous time version of the problem moves the focus to "average" dynamics as opposed to an exploration of the entire distribution of possible paths. Thus, we 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 the system at time t is given by the current population shares:

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

We close this section with an important remark. In the introduction we told a story about evolution taking place over behavioral memes and nature selecting for types of people who felt a more of "warm glow" from public-sprinted actions. The model above, however, is really a reduced form in which the social norm for altruistic actions appear to be programmed into behavior directly and evidently does not relate to preferences at all. Thus, so far we have described a kind of behavioral model in which some actions do not derive for rational maximization of preferences.

While this modeling approach is sufficient to explore how societies evolve behavioral norms in competition with one another, it is not entirely persuasive to claim that agents choose social behaviors though some ad hoc, a-rational process, but employ a rational process in the rest of their decision making. Fortunately this dichotomy is not necessary. We refer the reader to the second appendix in which we provide a utility function from which the reduced form behavioral rule we describe in the body of the paper can be derived as optimal, rational behavior. Thus, agents as described in the appendix go to the polls because, given their altruistic tastes, it is *privately rational to vote*. Agents, however, realize that the probability that they will be the pivotal voter is zero, and so the possibility of affecting an election outcome plays no role in this decision. To put this another way, agents rationally choose to vote because of the public-spiritedness encoded in their preferences. It turns out, however, that the preference parameter used in the appendix and the and behavioral parameter used in the paper are completely correlated. As a result not much is gained from looking at these more complicated microfoundations. We therefore consider a reduced form in which V_j serves as a proxy for altruism in preferences.

3. Characterization of the Stable States

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

The literature on evolution in economics tends to focus on evolutionarily stable strategies (ESS). Testing for the evolutionary stability of a strategy requires that the strategies agents play be shown to survive the introduction of small proportions of "mutant" strategies in the sense that steady state strategies yield higher average payoffs. We will examine how the presence of players who have adapted mutant social norms affects our equilibria in section 4. In this section, however, we concentrate on finding the steady states themselves and study the likelihood that a particular steady state will emerge as the stable outcome of the dynamic process. To simplify our discussion we shall say that *meme type j wins the evolutionary game* if the parameters and initial conditions are such that the population converges over time to a stable steady state in which type j makes up the entire population ($S_j = 1$).

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

Theorem 1. *Depending on the values of parameters α, β, and C, there are three possible outcomes for the system:*[7]

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 in some situations, regardless of how small their numbers are to begin with, the high voter types will increase their share of the population until they make up the entire society. This case is shown in Figure 1a. For other parameters, the low

[7] 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.

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 we think of as the more typical case in which 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 public-spiritedness is more likely to lead to evolutionary success.

> **Definition:** A change in the value of the parameters of an economy is said to *increase the probability of evolutionary success of a type* if the change results in an increase in size of the basin of attraction that favors that type (or equivalently, moves the tipping point in a way that implies that the type can win the evolutionary game with smaller and smaller initial shares of the population.)

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, a higher cost of voting C will decrease the probability of evolutionary success of the high voting type.*

Proof/

See appendix.

Theorem 2 says that if the parameters of the system are such that we are not in one of the two degenerate cases (cases 1 and 3 of Theorem 1), 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 is intuitive since as voting becomes more costly, the

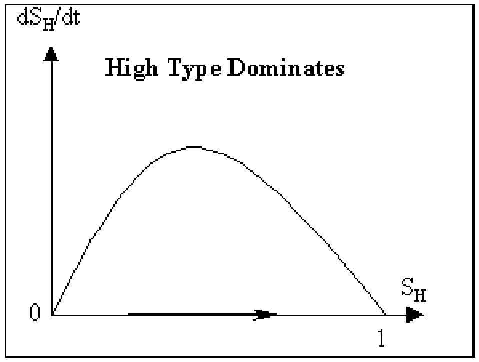

Figure 1a.
No matter what the initial population share, the high voting type will eventually make up the entire population. Thus, $S_H = 1$ is a globally absorbing state

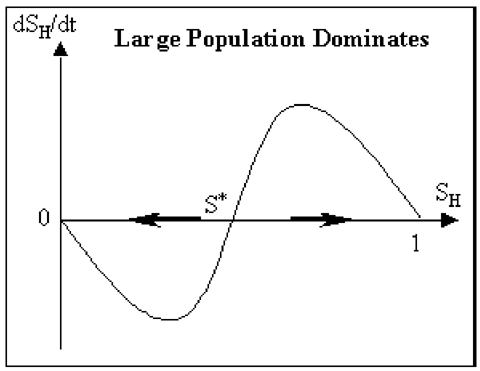

Figure 1b.
The larger its initial population share, the more likely a type is to win the evolutionary game. The $S_H = S^*$ is an unstable steady state that divides the basins of attraction for the two stable steady states $S_H = 1$ and $S_H = 0$.

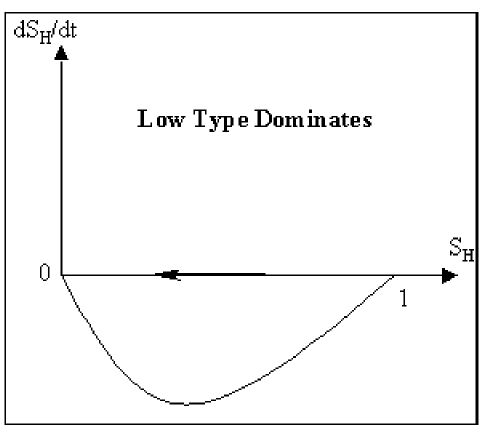

Figure 1c.
No matter what the initial population share, the low voting type will eventually make up the entire population. Thus, $S_H = 0$ is a globally absorbing state

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 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 whether a social norm for voting conveys an evolutionary advantage to a type. More precisely, we will say that this norm conveys an *evolutionary advantage* if an increase in the parameter of public spiritedness V_j increases the probability of evolutionary success of type j. 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 can enjoy from the costly voting activity of high types) both have an effect. As a consequence, more public spiritedness does not always benefit a type. The next theorem shows this for the case of when voting is very costly.

Theorem 3. *Assume that the parameters of the game are such that there are three steady states. In this case, if voting is too costly, then voting does not convey an evolutionary advantage.*

Proof/

By assumption, the benefits and losses that voters realize each period from the public proposals that happen to pass lie in the interval $[-1, 1]$. Recall that voters must pay $V_i \mid B_i \mid C$ each period for voting. Thus, if C is high enough, the expected per period voting costs the high types pay compared to the low types (which grow without bound in C) will be larger than the expected difference in benefits they receive from public projects. It follows that for large enough C, the low voter types have a higher expected payoff and so will win the evolutionary game.

■

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

Theorem 4. *Assume that the parameters of the game are such that there are three steady states. In this case, if voting is costless and preferences are not perfectly correlated, then voting conveys an evolutionary advantage.*

Proof/

If the agents with the high voting 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. 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 benefit correlation parameter α. The question is: Is public-spiritedness more or less of an advantage for a group when they experience benefits 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. *Assume that the parameters of the game are such that there are three steady states. In this case, if benefits are perfectly correlated, then the low voter type always wins the evolutionary game.*

Proof/

Note that in this case the payoff each type of agent gets from public proposals is identical. 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 of benefits the high type is always supplanted by the low type 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. In fact, in the strict sense, it implies that our steady states are not ESS since they cannot withstand the addition of these particular free riding mutants. We will argue in the next section, however, that this actually makes the population dynamics more natural and efficient.

The previous two theorems consider only 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. *Assume that the parameters of the game are such that there are three steady states. In this case, if voting is sufficiently low cost and benefits are sufficiently uncorrelated, then voting conveys an evolutionary advantage.*

Proof/

See appendix.

Theorem 6 says that the tipping point S^H moves in a way that favors the low voter type 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 voter types increase their voting propensity.)

For this to be true, however, it must be the case that 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. Memetics, 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 agrees with and differs from the existing literature.

Evolutionary game theory is typically used to explain how agents might choose strategies in an a-rational 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 choices.[8] In this, we follow such authors as Becker (1976), Hirshleifer (1978), and more recently, Bergstrom and Stark (1993) and Robson (1996). (See Robson 2001 for a more complete survey.) We further place this in a memetic context in that we describe agents' preferences as being formed by the social norms current in the social group into which they are born.

This raises an interesting question regarding whether our story can be reconciled with the traditional view in economics which seems to take evolution as a

[8] 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.

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 or imitation) 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 experienced 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 outweighed 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 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. Eshel, Samuelson and Shaked show that for an appropriately parameterized model, altruistic behavior survives and is stable against the introduction of mutations.

Bester and Guth (1998) 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 duopolist, then production choices are strategic complements. This means that when an altruistic firm chooses a higher than privately optimal production level, the other firm responds 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 a 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.[9] In Eshel, Samuelson and Shaked's case, it is through teaching one's neighbors to be altruists, for Bergstrom and Stark it is by producing kids who have an evolutionary advantage, and for Bester and Guth it is through the strategic complementarity. It may appear that the model we describe breaks with this thread and does indeed allow for widespread 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 by construction in our model preferences and voting propensity are completely linked. Thus, while benefits of proposals that pass are spread across many individuals, they are in a sense localized within a given

[9] 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 benefit less from the acts of altruists than do other altruists.

memotype. We conclude that the meme 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 to 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 can survive strategic experimentation and/or and random memetic 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 of almost any type of social mutant. 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 population, 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; thus, mutants will not upset the steady state. On the other hand, mutants who have the same (or at least very similar) tastes for public proposals as the dominant type, can successfully free-ride on their voting efforts. Thus, a mutant with the same tastes but a lower voting parameter as the dominant type can upset the steady state and will eventually supplant the original dominant type. Observe, however, that such free riding mutants are in turn 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. On the other hand, if our story holds and preferences are developed in a way that reflects the social norms of the society into which a child in born, the same argument might apply. Parents and society 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 these teaching or gets a truly mutant gene he would necessarily find himself equipped with preferences over public proposals and voting propensities *both of which* differ from those of his parents and society. Even if such mixed mutations were possible, it might be that there exist social sanctions to keep them from taking over. In other words, suppose a free riding mutant arises. Provided the mutation is detectable, the dominant group may protect itself by refusing to provide mates for such mutants. 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.[10] Thus, even though having the high voter types win the evolutionary game is not an ESS in a strict sense, there are still reasons to believe that this steady state may arise in real world settings.

Second, put the arguments above aside and suppose free riding mutants with

[10] 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 these rewards.

high type tastes do in fact 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 the initial situation is such that there has been a convergence to the high types being dominant in the population. If the mutants are small enough in number, their presence is not enough to prevent the high type from forcing the low type 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 mutant 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 in the long run, the altruism itself may not be.

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

We think of this as a kind of rise 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 Romans should 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 that accrue to whichever type of voter makes up the winning coalition decline while the prospective benefits to the opposition group of winning an election remain high. This suppresses the winning coalition's turnout and makes it more likely that the opposition would begin to win elections. The decline of benefits also slows the winning coalition's rate of growth even if they should continue to win the elections. For such a model, it might be possible to find a stable interior solution in which both types of agents persist. Another interesting generalization would be to allow more than two types of agents. Simulation results suggest that if the groups are equally numerous, preferences over public policies are uncorrelated and voting is sufficiently cheap, 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 memetic evolution. The basic result is a society with a public-spirited social norm may be at an advantage and thus the memes for public spiritedness may spread. Having agents with a high propensity to vote in a society is more advantageous when voting is less costly, when the high voting propensity group's preferences over public projects 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 when societies with different world views compete with one another.

Appendix

Derivation of Payoff Functions

We begin with some preliminary results that will simplify our calculations. First we define 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:

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

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:

$$
\begin{aligned}
E(\pi_H \mid B_H = b_H) =& P(b_H - C_H) + (1-P)(-C_H) \\
=& P b_H - C_H.
\end{aligned}
$$

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

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

Substitution for P in the above integral gives:

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

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:

$$\begin{aligned}\bar{\pi}_H &= \int\limits_{b_H=-1}^{b_H=1} \left(\frac{b_H}{2}(\frac{1}{2} - \frac{b_H}{2\theta}) - \frac{V_H C \mid b_H \mid}{2} \right) db_H \\ &= -\frac{1}{6\theta} - \frac{V_H C}{2}.\end{aligned}$$

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:

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

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

$$\begin{aligned}\bar{\pi}_L =& E_{b_H}[E(\pi_L \mid B_H = b_H)] \\ =& \frac{1}{4(1-\alpha)} \left(\int_{-1}^{1} \int_{\frac{b_H(\alpha\theta+1-\alpha)}{\theta}}^{\alpha b_H+(1-\alpha)} b_L db_L db_H - \int_{-1}^{1} \int_{\alpha b_H-(1-\alpha)}^{\alpha b_H+(1-\alpha)} C_L db_L db_H \right).\end{aligned}$$

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

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

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

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:

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

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

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 to make some preliminary observations that render 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 preference types 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 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 the 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:

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

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

$S_1 \leq S_H \leq S_H^*$, 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:

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

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

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 toward 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 ante 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 attempts to model an agent who feels civic duty. 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 convey to his type and how much effort he has put forth 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,* V: 100 pp. 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 Oded Stark (1993): "How can Altruism prevail in an Evolutionary Environment," *American Economic Review,* v.83, pp. 149-155.

Bester, Helmut and Werner Guth (1998): "Is altruism evolutionarily stable?," *Journal Of Economic Behavior & Organization,* v.34, pp. 193-209.

Blackmore, Susan (1999): *The meme Machine.* Oxford University Press.

Bolle, Friedel (2000): "Is altruism evolutionarily stable? And envy and malevolence? Remarks on Bester and Guth," *Journal Of Economic Behavior & Organization,* v.42, pp. 131-133.

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

Dawkins, Richard (1976): *The Selfish Gene.* Oxford University Press.

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

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

Eshel, Ilan, Larry Samuelson, and Avner Shaked (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,* v. 68, pp. 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.

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.

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.

Kan, Kamhon and Cheng-Chen Yang (2001): "On Expressive Voting: Evidence from the 1988 US Presidential Election," *Public Choice,* V, 108, pp. 295-312.

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,* v. 44, pp. 7-41.

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,* v. 41, pp. 7-53.

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

Possajennikov Alex (2000): "On the evolutionary stability of altruistic and spiteful preferences," *Journal Of Economic Behavior & Organization,* v.42, pp. 125-129.

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

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

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

Robson, A (2001): "The Biological Basis for Economic Behavior," *Journal of Economic Literature,* v. 39, pp. 11-33.

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

Taylor, Peter D. and Leo B. Jonker, (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.