A Behavioral Model of Turnout

Jonathan Bendor, Daniel Diermeier and Michael Ting

7th Annual Conference
October 20-21, 2000

W. ALLEN
WALLIS
Institute of
POLITICAL
ECONOMY

UNIVERSITY OF
ROCHESTER

A Behavioral Model of Turnout[1]

Jonathan Bendor
Graduate School of Business
Stanford University

Daniel Diermeier
Kellogg Graduate School of Management
Northwestern University

Michael Ting
Department of Political Science
University of North Carolina at Chapel Hill

September, 2000

[1]We would like to thank Steve Ansolabehere, Sorin Antohi, Glenn Ellison, Dedre Gentner, Sunil Kumar, David Laitin, Tze Lai, Arthur Lupia, Elijah Millgram, John Patty, Paul Pfleiderer, Adam Simon, Joel Sobel, Carole Uhlaner, and the participants of the Political Economics seminar at the GSB, the Stanford-CalTech workshop, the UCLA conference on Cognition, Emotion, and Rational Choice, panels at the 2000 Annual Meetings of the MPSA and APSA, and the Center for Advanced Study in the Behavioral Sciences seminar series for their helpful comments. This paper was revised while Bendor was a Fellow at the CASBS, and he is grateful for the Center's financial and intellectual support.

Abstract

The so-called "paradox of voting" is a major anomaly for rational choice theories of elections. If voting is costly and citizens are rational then in large electorates the expected turnout would be small, for if many people voted the chance of anyone being pivotal would be too small to make the act worthwhile. Yet many people *do* vote, even in large national elections. To address this puzzle we construct a model of *adaptive* rationality: citizens learn by simple trial-and-error, repeating satisfactory actions and avoiding unsatisfactory ones. (Their aspiration levels, which code current payoffs as satisfactory or unsatisfactory, are also endogenous, themselves adjusting to experience.) Our main result is that agents who adapt in this manner turn out in substantial numbers even in large electorates and even if voting is costly for everyone.

I. Introduction

"standard conceptions of rational behavior do not explain why anyone bothers to vote in a mass election... [Turnout is] the paradox that ate rational choice theory" (Fiorina 1990, p. 334).

Perhaps Fiorina's remark is too gloomy, but it does seem apparent that the phenomenon of substantial turnout in large-scale electorates is anomalous for rational choice theory, in either its decision theoretic or game theoretic guises. In a rough-and-ready sense, the problem is straightforward: in large electorates, the chance that any single voter will be pivotal is very small. Consequently, if voting imposes strictly positive costs, these will outweigh the expected gains from voting. Accordingly, rational citizens will not vote—contrary to the evidence. Hence, an anomaly.

This is the classical version of the turnout-problem as formulated by Downs (1957). Downs' formulation, however, is decision-theoretic. In strategic models of turnout, if the cost of participating is not too high then it will *not* be an equilibrium for *everyone* to stay home, for then one voter by herself could decide the election. The key insight of strategic models is that the probability of being pivotal is endogenous. That is, if citizens are rational, the voting decisions and pivot probabilities are determined simultaneously.

Strategic theories usually model turnout as a *large team game* (Palfrey and Rosenthal 1984, 1985; Myerson 1996). There are typically two alternatives (e.g., candidates) and two types of citizens (call them Democrats and Republicans), where each type or team has identical preferences. Preferences of each team are diametrically opposed. Each person can either vote, for either candidate, or stay home (shirk). Elections are decided by a simple plurality with some tie-breaking rule, usually a coin toss. All members of the winning faction earn a payoff for winning (whether or not they voted); losers get nothing. Independent of the outcome there is an additive and private cost of voting.

Strategic theories then solve for the equilibria of such team games. Since voting for the non-preferred candidate is strictly dominated for each voter, the relevant problem reduces to a *participation game* which simply involves the binary decision of whether to vote or stay home. The results of game theoretic models can be summarized as follows (Palfrey and Rosenthal 1984, 1985; Myerson 1995):

(1) There are no pure strategy equilibria except in degenerate cases. For example, "Everybody votes" is an equilibrium only if $c = 0$ or if the teams are of exactly the same size.

(2) There are many equilibria with positive turnout. Except in degenerate cases, these equilibria involve the use of mixed strategies by at least some voters.

(3) Equilibria with non-trivial turnout are asymmetric. That is, some voters of the same type use different strategies. For example, some of them use pure strategies, while others mix.

(4) High turnout equilibria are not robust to the introduction of uncertainty over either preferences and costs (Palfrey and Rosenthal 1985) or the number of players (Myerson 1995). The robust equilibria have vanishing turnout.

Thus, as Palfrey and Rosenthal (1985, p.64) pointed out "We have come full circle and are once again beset by the paradox of not voting." The initial rough-and-ready intuition is still close to the mark: as long as all voters have strictly positive costs of voting, the expected turnout will be "small" relative to the size of the electorate.

A common response to this anomaly has been to modify the payoff assumptions. For example, suppose citizens have been socialized into having a sense of obligation toward democratic institutions, i.e., toward voting (Riker and Ordeshook 1968). Then the costs of voting might be negative, at least for some citizens.[1]

Empirically this may well be true, and we suspect that it does explain at least some turnout. But there are obvious methodological concerns associated with completely dispelling an anomaly in this manner. To answer the question, "Why do people do x?", by saying that people have a taste for x seems theoretically shallow.

In this paper we take a completely different tack. For the most part, we leave the game structure, and hence payoffs, alone: i.e., we assume that voting is costly, as in the classical models. Instead, we modify the rationality assumptions. Specifically, we assume that voters are *adaptively* rational: they learn to vote or to stay home. Their learning is a simple form of trial-and-error behavior which is consistent with basic axioms of reinforcement learning (Bush and Mosteller 1955): actions that are successful today are more likely to be produced tomorrow; unsuccessful actions are less likely. This reinforcement learning is married to an aspiration level, a threshold that partitions all possible current payoffs into successful and unsuccessful ones, hence indicating which actions are coded as successes (and so worthy of reinforcement) and which as failures (and so inhibited). A voter's aspiration level itself adjusts to experience, reflecting prior payoffs.

[1] See also Ferejohn and Fiorina (1974), who assume that voters are regret minimizers. Their approach met considerable resistance on theoretical and empirical grounds. For example, it is inconsistent with the common observation of higher voting rates in closely contested elections (Palfrey and Rosenthal 1985).

Our model thus proposes a synthesis between two competing research programs in political science: (neo-)institutionalism and behavioralism. As in institutionalism, we explicitly model the strategic interdependence inherent in electoral participation, but instead of assuming fully rational voters with common knowledge about the game form, our players adapt their behavior over time as a consequence of past individual pay-offs and aspirations.[2]

Because of the model's complexity, we have turned to simulation to obtain many of our main results.[3] The simulation results indicate that even if *all* the voters experience strictly positive costs of voting, turnout is substantial. Perhaps most strikingly, *turnout remains non-negligible even when we increase the size of the electorate to one million voters.*

The paper is organized as follows. Section II describes the model. The main simulation result—the emergence of substantial turnout in large electorates—is reported in section III. The computational output is supplemented by several analytical results that both help to explain why the computer model generates non-negligible participation and to generalize the simulation findings. Section IV explains, via several analytical results, why participation falls well short of being universal. Section V presents some findings on comparative statics and dynamics (e.g., how varying the cost of voting affects turnout). Extensions and a few concluding observations follow in section VI.

II. The Model

Each agent has two choices, to vote or to stay home (shirk). We assume that the electorate is divided into two clear blocs or factions, of n_D Democrats and n_R Republicans. Whichever side turns out more voters wins the election. (Candidates and their behavior are suppressed in this model.)[4] Members of the winning faction earn a payoff of $b > 0$ for winning, whether or not they voted; losers get nothing. (In the unlikely event of a tie, everyone gets $\frac{b}{2}$.) The private cost of voting is $c \geq 0$, where $c < b$. (In some variants we allow costs to be negative, to reflect the duty to vote.) Payoffs are additive in the benefits and costs. Thus winning voters get $b - c$; winning shirkers get b. Losing voters get $-c$; losing shirkers get 0. Unless stated otherwise we normalize b

[2]For a more detailed discussion of these issues see Bendor, Diermeier and Ting (1999).

[3]The simulation model bears a family resemblance to an analytical model of aspiration-based learning (Bendor, Mookherjee and Ray 1998).

[4]For simulation models of adaptively rational candidates, see Kollman, Miller and Page (1992, 1998). Also, the analytical results of Bendor, Mookherjee and Ray (1998) can be extended to apply to two-party races. All of this work complements the present manuscript in that in the models of Kollman et al. and of Bendor et al. voters are passive mechanisms: essentially, they are represented as vote shares (payoffs) in a two-person game in which the active players are the candidates. Here, voters are active while candidates or parties are in the background.

to 1.

Propensities, Aspirations and Adaptation. The heart of the model is the learning behavior of each agent. As indicated earlier, the adaptive behavior combines reinforcement learning and endogenous aspirations.[5] Thus in the beginning, at $t = 1$, every citizen begins with an initial propensity to vote; call this $p_{i,1}(v) \in [0, 1]$ for citizen i.[6] Each citizen is also endowed with an initial aspiration level, call it $a_{i,1}$ for agent i. (At this point we can only distinguish between Democrats and Republicans, so voters in different blocs can have different initial propensities or different initial aspirations, but voters in the same bloc are initially homogeneous in these and all other respects.) Given the vector of propensities, actions are realized and the outcome is determined, as are payoffs.

As noted earlier, any action that is coded as a "success" is reinforced; any action coded as a "failure" is inhibited. Thus if a citizen votes and her payoff is greater than or equal to her aspiration, then her propensity to vote in the next period rises. The equation is the classical linear Bush-Mosteller equation:

$$p_{i,t+1}(v) = p_{i,t}(v) + \alpha(1 - p_{i,t}(v)) \tag{1}$$

where $\alpha \in (0, 1)$ represents the speed of learning or adaptation, given a successful outcome. Similarly, if her payoff is less than her aspiration and she had voted in t, then her propensity to vote falls:

$$p_{i,t+1}(v) = p_{i,t}(v) - \alpha \cdot p_{i,t}(v) \tag{2}$$

Thus, because α represents the speed of learning in the face of failures as well as in the face of success, we are assuming that the rate of adaptation is symmetric in the two directions. (This can be easily changed in the simulation, but it is reasonable as a benchmark case and furthermore turns out to be useful analytically.) The adjustment equations that hold when the agent shirked in t have exactly the same form and the same value of α.

Aspirations also adjust (probabilistically) to experience: when they adjust, tomorrow's aspirations are a weighted average of today's level and today's payoffs:

$$a_{i,t+1} = \lambda \cdot a_{i,t} + (1 - \lambda)\pi_{i,t} \tag{3}$$

[5] For a more detailed presentation of this model of aspiration-based learning, see Bendor, Diermeier and Ting (1999).

[6] We use the term 'propensity' to underscore the psychological nature of the theory. Mathematically, of course, it is a probability.

where $\pi_{i,t}$ denotes citizen i's payoff at date t and λ is in $(0, 1)$.

Because aspirations adjust to experience, eventually aspirations and actions must be consistent with each other. For example, if a citizen settles down into a propensity of always voting, then in a steady state his aspiration cannot exceed $b - c$. And more generally, all aspirations will be drawn toward the $[-c, b]$ interval, which includes all feasible payoffs.

A full cycle of learning is depicted in figure 1.

[figure 1 about here]

Trembles. Actors may tremble, i.e., occasionally do the opposite of what they had intended. A tremble (should it happen) occurs after an agent's propensity has produced an intended action. That is, suppose a citizen has a propensity of .25 of voting in some period. That produces an intended action, say (in this case) of shirking. If the probability of a tremble is one in a hundred, then the would-be shirker winds up voting with a chance of .01. In general, if the probability of trembling is ϵ, then the chance that voter i actually winds up voting in period t equals $p_{i,t}(v) \cdot (1 - \epsilon) + (1 - p_{i,t}(v)) \cdot \epsilon$. We assume throughout that $0 < \epsilon < \frac{1}{2}$.[7]

We introduce trembles in order to ensure that the model has empirical content, i.e., to avoid vacuity. Without trembles, *any* outcome could be predicted, given suitable initial conditions. For example, suppose that initially Democrats had aspirations of $b - c$ and were completely disposed to vote, while Republicans had aspirations of zero and were completely disposed to staying home. Without trembles, all Democrats would vote and all Republicans would abstain. Thus everyone would get payoffs that equalled their aspirations, and so their propensities (and aspirations) would remain unchanged. Or more extremely, in a tremble-free model of aspiration-based learning in the two-person prisoner's dilemma, the outcome in which player A is continually suckered could be generated, with suitable initial conditions. (A would have to start with aspirations that are no higher than the sucker's payoff, and so on.) By producing experimentation, trembles eliminate such possibilities, and so restore empirical content to the model.[8]

Though simple, the model poses some severe tractability issues. Because closed-form solutions are so difficult to derive, our results are generated by simulation and then reinforced by some complementary analytical results. We describe the general simulation model of electoral participation,

[7] The assumption that the tremble probability is less than one half is very mild. If it equalled one half, then behavior would be random and consequently unrelated to intentions. If ϵ exceeded one half then actions and intentions would be perversely connected.

[8] See Bendor, Mookherjee, and Ray (1998) for a more extensive discussion of the role of trembles in this kind of model.

upon which the present findings are based, in the Appendix.

III. The Main Results

A. Simulation Results

The main question obviously is, will these adaptively rational agents learn to vote in significant numbers? Initial results, with a small electorate and equal sized factions ($n_D = n_R = 50$), are encouraging (run 1).

[run 1 about here]

Run one shows that the percentage of citizens who voted seemed to quickly stabilize at about 50 percent.[9] Although not shown here, the turnout percentage remains stable throughout the run, indicating that the system has probably reached its steady state. Thus turnout has evidently stabilized at quite a substantial level.

A natural question is whether turnout will collapse in much larger electorates. The answer to this question is *no*: with equal faction size the turnout in an electorate of one million citizens remains substantial (run 2). Of course, this is not yet the number involved in a Presidential election, but we are already well past Congressional races.

[run 2 about here]

Since this is probably the paper's central result, and one which may puzzle some readers, it is worthwhile to pause at this point in order to try to understand it. In an electorate of one million people, the chance that any one person will cast a pivotal vote is miniscule. Why, then, do so many people learn to vote?

Perhaps the best way to understand this outcome is to conduct the following thought experiment. Imagine that everyone in a population has settled into a complete propensity of shirking, and this has lasted long enough for aspirations to adjust to current payoffs (of $\frac{b}{2}$, for everyone). What will happen thereafter? Of course, in a tremble-free world this could indeed be stable, but that is trivial, since without trembles anything can stabilize, given aspiration-based reinforcement learning. So the thought experiment must allow for trembles (as do all of the runs reported here).

[9] The parameter values for run 1 are displayed below the histogram. The reader can assume that these are the default parameter values in all runs covered in this paper, i.e., unless we explicitly say otherwise, a particular parameter (say, initial vote probabilities) will have the values of this run (0.50, in the case of this parameter). This will permit us to focus attention on those parameters which are different for a run in question.

Suppose, then, that in period t voter D1 trembles and turns out. With $c = 0.25$, this accident reinforces her propensity to vote (she now gets a payoff of $b - c = 0.75 > \frac{b}{2} = 0.5$)—but the slothful behavior of her comrades, who are now enjoying a free-riding payoff of 1.0, is also reinforced. So this is not the place to look for an explanation. The place to look is the effect that D1's tremble has *on her opponents*. In the hypothetical all-shirk state, the Republicans were enjoying a payoff of 0.5, with aspirations to match. With D1's voting, suddenly the Republicans are stuck with a payoff of zero. Thus, D1's tremble imposes negative reinforcement on the Republicans' shirking. Hence in the following period all of the Republicans will put positive probability on voting. We call this *loser-driven mobilization*.

The story is not over. In $t+1$ the Republicans will almost certainly win. The effect on their propensity to vote is complicated. All Republicans who actually voted will be positively reinforced for doing so; but all of their freeriding comrades will have *that* action supported as well. Thus once again focusing on the winners does not explain why the system will wind up at a much higher turnout level; once again we must look at the losers—in this period, the Democrats. In $t+1$ virtually all of the Democrats stayed home—and got a payoff of zero. With aspirations adjusting slowly, and hence still close to one half, the players will code a payoff of zero as a failure. So now the Democrats's shirking is inhibited. Hence more of them turn out in period $t+2$, and loser-driven mobilization continues. The mobilization of one side begets counter-mobilization, in a typically pluralist fashion.[10]

Having a computer at our disposal allows us to make this more than a thought experiment. In run 3a all citizens start out completely disposed to stay home, and initial aspirations are consistent with those propensities. Yet by period 50 average turnout has risen to almost half the citizens, where it stabilizes (runs 3a and 3b).

[runs 3a and 3b about here]

This seems to confirm at least part of the preceding argument: starting from a state of universal shirking, a process of mobilization and counter-mobilization is unleashed.

B. Explaining the "Breakout of Participation": Analytical Results

We will now make this preceding argument more rigorous, in order to explain the mobilization process more tightly and completely. Some "mini-results", analytically derived, will provide the

[10] This process has an arms race aspect, in that the Pareto-optimal symmetric outcome is for everyone to stay home and divide the electoral pie in half with zero effort. Thus the mobilization and counter-mobilization amounts to an escalating arms race of effort that is collectively inefficient.

foundation for this explanation. We will show in a step-by-step fashion why the mobilization observed in run 3a occurs. Indeed, we will also show why much of it does not depend on the specific form of adaptation, the Bush-Mosteller mechanism, used in the simulation but is instead driven by much more general properties of trial-and-error learning.

The first result is exceedingly simple but it supplies the basis for what follows. Observation 1 reports a simple property about aspirations and electoral payoffs. It presumes (as do all subsequent results) that the election in date t is *conclusive*, i.e., not a tie. We call a player's aspirations "not too high" if they are less than $b - c$ and "not too low" if they exceed 0. We also say that an actor is "satisfied" if her action is coded a success, and "unsatisfied" if it is coded a failure.

Observation 1: If in t the aspirations of people in the winning faction are not too high and the aspirations of people on the losing side are not too low, then all winners are satisfied by the outcome in t while all losers are dissatisfied.

This makes intuitive sense. Winners get payoffs of either $b - c$ or b, depending on whether they voted or stayed home. So if all winners have aspirations below $b - c$ then they are all content with the outcome. Similarly, losers get payoffs of either $-c$ or 0, depending on their individual choices. So if the aspirations of all losers exceed zero, then losing is unacceptable.

Though simple, this condition is very important both substantively and analytically. Substantively, it identifies situations in which everyone's satisfaction is determined exclusively by the *collective* outcome: if your side wins, you're happy; if it loses, you're unhappy (figure 2).

[figure 2 about here]

An important special case of this condition is when all citizens have "middling" aspirations, in the interval $(0, b - c)$. (In that case the conclusion of observation 1 holds if *either* party wins.) A natural interpretation of this special case is that people identify with the well-being of their factions: personal satisfaction is driven completely by collective outcomes.

Now given the conditions described by observation 1 (which are satisfied by the initial conditions of run 3a), which citizens will mobilize after the election of period one? That is, if observation 1 applies to period one, which citizens will increase their propensity to vote in period two?

In order to answer that question one must make some assumptions about how propensities are adjusted. Of course, we could use the Bush-Mosteller mechanism employed in the simulation. Happily, however, the next result depends only on general *qualitative* properties of the Bush-

Mosteller rule, not on its specific functional form. The essence of these qualitative properties is simply that the propensity to try satisfactory actions increases while the propensity to try unsatisfactory ones must fall.

Let us then call an *Aspiration-Based Adaptive Rule* (ABAR) any learning mechanism with the following properties. First, a citizen (say i) compares her current payoff to her current aspiration level. Second, her action propensities adjust so that successful actions are reinforced and unsuccessful ones, inhibited. Thus if, e.g., she voted in period t then her propensity to vote changes as follows (the analogous properties hold if she shirked in t):[11]

(1) if $\pi_{i,t} > a_{i,t}$ then $p_{i,t+1}(v) \geq p_{i,t}(v)$ (a strict inequality if $p_{i,t}(v) < 1$);

(2) if $\pi_{i,t} = a_{i,t}$ then $p_{i,t+1}(v) \geq p_{i,t}(v)$;

(3) if $\pi_{i,t} < a_{i,t}$ then $p_{i,t+1}(v) \leq p_{i,t}(v)$ (a strict inequality if $p_{i,t}(v) > 0$).

These ordinal properties are at the heart of aspiration-based trial-and-error learning.[12] Thus an ABAR is a function that takes as input an agent's current aspiration and propensity and realized action and payoff, and yields as an output a new propensity.

The next observation shows that these properties tell us what we need to know, i.e, which citizens become more inclined to vote and which become less inclined.

Observation 2: Suppose observation 1 holds. If adjustment is by any arbitrary mix of ABARs, then all winning voters and all losing shirkers become more disposed to vote after the election in t (or remain fully disposed to vote). The other citizens become less inclined to vote (or remain fully inclined to shirk).

It is obvious that winning voters increase their propensity to participate: they voted and were pleased with the outcome. A bit less obviously, *so will losing shirkers*, since all losers were displeased with the outcome.[13]

[11] Recall that $p_{i,t}(v)$ denotes i's propensity to vote in t, while $a_{i,t}$ is her aspiration and $\pi_{i,t}$ her payoff.

[12] It is worth noting that the concept of an ABAR does not presume that propensity-adjustment is *quantitatively* the same for different actions (here, voting versus shirking). Only the *direction* of change is stipulated by the definition. Thus it is permitted for, e.g., an agent to change more in response to feedback on voting than in response to feedback on shirking. Later we shall encounter contexts in which it is important to make the additional assumption that an ABAR is quantitatively the same for all feasible actions, a property that we call *action-invariance*.

[13] There is a good reason for focusing on these two sets of citizens: it is they who generally determine whether mobilization will occur. Given what we will later call *realistic* aspirations—those above the game's minimal payoff ($-c$) or below its maximal one (b)—winning shirkers and losing voters never become more likely to turn out: the former's apathy is reinforced while the latter's activism is inhibited. Hence to explain the "breakout of participation" we must turn our attention to winning voters and losing shirkers.

The next piece of the story involves political demography: how *many* citizens are either winning voters or losing shirkers?

Observation 3: If in t the Democrats win and $n_D \leq n_R$ or the Republicans win and $n_D \geq n_R$, then the number of winning voters plus losing shirkers is a majority of the electorate in t.

Since in our benchmark simulation the factions are of equal size, observation 3 applies to all elections in run 3a that aren't ties, i.e., nearly always.

Now let us combine observations 1, 2 and 3 in order to see how much of the electorate will, at the end of a given election, become more mobilized (increase their propensity to vote). It is important to note that the following result does *not* assume that everyone in the electorate uses the same type of adaptive rule. People may adapt in different ways and at different speeds, per observation 2. All that is required is that voters use *some* ABAR.

Proposition 1: Suppose the conditions of observations 1, 2 and 3 hold. Further, everyone in t has a vote-propensity of less than one. Then after the election in t a majority of citizens become more disposed to vote.

Beyond the conditions of the three observations, the further requirement of proposition 1—before the election nobody is completely disposed to vote—is very mild. It is obviously satisfied by run 3a, which presumes that initially everyone is completely inclined to *shirk*. More importantly, it will hold in many contexts of low mobilization. Thus proposition 1 makes clear that outbreaks of participation are to be expected, if citizens learn via this large class of adaptive rules.[14] Further, it is important to note that this result *does not depend in any way on the size of the electorate.*[15] In particular, proposition 1 does *not* require that the electorate be small. This gives us confidence that the results of run 2—stable, substantial amounts of turnout in a district with one million citizens—will generalize to even larger jurisdictions.

The notion of "loser-driven mobilization" that we discussed earlier falls naturally out of special cases of proposition 1. For example, consider circumstances in which initially everyone is fully disposed to shirk, as in run 3a. Since that run satisfies all of the conditions of proposition 1, we know (analytically, now) that after any conclusive election in period one over half of the community

[14] It is worth noting that proposition 1 holds even if shirking by all citizens forms a Nash equilibrium. (In our game, universal pure shirking is Nash if $c > \frac{b}{2}$.) We will explain in more detail later why and how this can occur. For now we merely point out to readers that the logic of observations 1-3 and proposition 1 does not depend in any way on whether the system's initial state (of universal shirking) is a Nash equilibrium.

[15] Nor do any of our other analytical results, as we will see shortly.

will become more inclined to vote. However, since no one initially has any inclination to vote and the benchmark tremble value is small (0.01), few people actually turn out in period one. Since together winning voters and losing shirkers are over half the community but the odds are overwhelming that only a few percent of the electorate participate, *most of the newly mobilized must be losing shirkers*. Thus, in the early goings mobilization is loser-driven.[16]

Proposition 1 establishes mobilization in a demographic or head-counting sense: when its assumptions are satisfied, more citizens will increase their propensity to participate than will decrease that tendency. This does not necessarily imply, however, that the electorate's *average* propensity to turn out rises. Whether that follows depends on *how much* those increasing their vote-propensity boost their participation tendencies versus how much the decreasers reduce theirs.

Because this point is very important both in the simulation and, we suspect, in any adaptive model of participation, we provide two numerical examples to flesh it out. In the first, citizens are initially highly disposed to participate while in the second they are inclined to shirk. We will see that although proposition 1 holds for both of them, and so both experience further demographic mobilization, the electorate's average propensity to vote rises only in the second example. In the first example, ceiling effects prevent this.

In both cases we use the adjustment rule of the simulation, symmetric Bush-Mosteller, and presume that the speed of learning (α) is 0.1, and all of the relevant conditions of proposition 1 hold.

Example 1: high status quo mobilization. Suppose that in period t everyone's propensity to vote is 0.90. Since proposition 1 holds, we know that a majority of citizens in $t+1$ will be more predisposed to participate than they were in t. But this need not imply an increase in the average propensity to vote. To see this, suppose that the factions are equally big and that ninety percent of the Democrats turn out but only eighty percent of the Republicans do so. Then 55 percent of the electorate (the Democratic voters plus the Republican shirkers) will increase their participation-propensity while 45 percent decrease. So more people are mobilized. But with an α of 0.1 the increasers boost their participation-propensity only to 0.91 while the decreasers fall to 0.81. Thus

[16] That loser-driven mobilization occurs even when universal shirking is Nash (i.e., when $c > \frac{b}{2}$) indicates it is not a *protected* Nash, as Bendor, Mookherjee and Ray define the term (1998). In a protected Nash, if one player deviates the other players are not hurt by this deviation. (E.g., mutual defection in the two-person prisoner's dilemma is a protected Nash.) But if a Democrat deviates from all-shirk, then many other players—the Republicans—*are* hurt. In general, whether or not a Nash equilibrium is protected greatly affects its stability, given adaptive learning. As Bendor, Mookherjee and Ray show, protected Nash equilibria are stable against random shocks: if player A trembles, B's propensity to play her Nash action is undisturbed. Unprotected Nash lack this stabilizing property. (See proposition 2 of Bendor et al., 1998.)

the community's average propensity *falls*, to $(0.91){\cdot}(0.55) + (0.81){\cdot}(0.45) = 0.865$. Thus under symmetric Bush-Mosteller and many other adaptive rules, very high rates of participation tend to be self-limiting: in these circumstances the winning voters cannot increase their vote-propensities much while the losing voters have a long way to fall; similarly, the losing shirkers cannot become much less inclined to shirk, while winning shirkers can become much more slothful.

Example 2: low status quo mobilization. Here suppose that in period t the Democrats' propensity to vote is .3 while the Republicans' is .2. There are 100 in each faction. In the election 30 Democrats and 20 Republicans turn out. Now observe the effects. Democratic (winning) voters increase their propensity to participate by a lot, to .37. But (losing) Republican voters drop only a bit, to .18. On the other hand, Republican shirkers, whose propensity to shirk had been .8, become significantly less inclined to stay home—it falls to .72—so their tendency to vote rises by the same large amount of .08. Meanwhile (winning) Democratic shirkers increase their propensity to shirk from .7 to .73, so their inclination to vote falls only a bit. Since winning voters outnumber losing voters *and* change more, per capita, and since losing shirkers outnumber winning shirkers *and* change more, per capita, it should come as no surprise that the electorate's average propensity to vote rises, from .25 to .28.

Thus we see that although demographic mobilization occurs in both cases, it is the extent of per capita change that determines whether the district's average propensity to vote (denoted by $\overline{p}_t(v)$) rises. In example 2, not only are there more winning voters than losing ones, but in addition the former change more per capita (in the "right" direction) than the latter change per capita in the "wrong" direction. Similarly, not only are there more losing shirkers than winning ones, but in addition the former change more per capita, in the right direction, than the latter change per capita in the wrong direction.

It should be apparent, then, that once demographic mobilization is assured, e.g., via proposition 1, these per capita quantities constitute a sufficient condition for the electorate's average participation-propensity to rise. Two features of the symmetric Bush-Mosteller rule are crucial here, so we isolate and identify them. First, propensity-adjustments under (any) Bush-Mosteller are *monotonic* in both directions. Increases in the propensity to use any action x, if it was tried and succeeded, are strictly decreasing in the current propensity to use x. Similarly, decreases in the propensity to use x, given that it was tried and failed, are strictly increasing in the current propensity to use x.

Second, the symmetric Bush-Mosteller is, of course, symmetric: the degree of propensity-adjustment is the same in the face of failure and of success. More precisely, let us call an ABAR *symmetric* if, for any propensity in $[0, 1]$, the increase in propensity to try an action x, given a successful use of x and a current propensity of $p_t(x)$, must equal the decrease in the propensity to try x, given a failed use of x and a current propensity of $q_t(x) = 1 - p_t(x)$. Thus, *given the same amount of adjustment room*, symmetric ABARs adjust equally in response to failure and to success.[17]

Adaptive rules that are monotonic or symmetric or both are displayed in figure 3.

[figure 3 about here]

As figure 3 illustrates, an ABAR that is both monotonic and symmetric must display ceiling effects. Thus, if citizens use a monotonic and symmetric ABAR, then when citizens' participation propensities are already low the tendency to shirk cannot rise much more, whereas feedback that is favorable to voting can lead to big changes. And so the next result shows that the breakout of participation illustrated by run 3a holds for a rather large class of adaptive rules, under a broad range of parameter values. In particular, this generalization away from Bush-Mosteller rules shows that their linearity (i.e., that adjustment magnitudes are linear in the status quo propensity) plays no essential role in the outcome of example 2. Thus generalizing beyond the Bush-Mosteller is useful both as an isolating abstraction—to identify the properties that drive a result—and to show that the breakout of participation is not an artifact of one particular adaptive rule.[18]

Proposition 2: Suppose that observations 1-3 hold. Further, everyone uses the same ABAR, which is monotonic and symmetric. If $p_{i,t}(v) \leq \frac{1}{2}$ for all i, then $\overline{p}_{t+1}(v) > \overline{p}_t(v)$.

Although this result generalizes beyond Bush-Mosteller rules, it does make a rather restrictive assumption about the current amount of mobilization: no one has a propensity to participate over one half. In the simulation, however, we typically see the district's *average* propensity rising to around 50 percent, which inevitably means that some citizens' vote-propensities exceed one half. So proposition 2, while useful as a robustness check on the type of learning mechanism used in the simulation, does not fully explain the breakout of participation that we see in run 3a.

[17] Symmetry implies that $\alpha = \beta$ for Bush-Mosteller rules.

[18] Further, the proof of proposition 2 reveals that the result can be generalized—given a plausible turnout condition—to cover adjustment rules that are asymmetric, responding more to failure than to success. (There is some psychophysiological evidence for this "negativity bias" (Ito and Cacioppo 1999, p.474). An independent source of evidence is experimental support for the prospect theory premise that people are loss-averse, though see Kahneman (1999, p.18-19) for a cautious assessment of the asymmetry's magnitude.) The appendix provides a precise definition of failure-sensitive ABARs.

It is important to keep in mind, however, that proposition 2 describes a sufficient condition for increasing average propensity to participate, not a necessary one. So it is consistent with run 3a; i.e., it is consistent with the fact that mobilization continues in that run even after some $p_{i,t}(v)$'s exceed one half. However, the simulation uses a very specific rule, the symmetric Bush-Mosteller, which is also action-invariant. The next result shows that we can say something analytical if we exploit that specificity by assuming that everyone uses that adjustment rule.

Before doing so, however, we must define action-invariance (introduced informally in footnote 12) precisely. In this definition $p_{j,t}(s)$ is j's propensity to shirk.

Definition. Suppose $p_{i,t}(v) = p_{j,t}(s)$. In t citizen i votes and j shirks but they get the same feedback (i.e., either both are reinforced or both are inhibited). The ABAR is *action-invariant* if $p_{i,t+1}(v) = p_{j,t+1}(s)$.

This property essentially says that the propensities to vote and to shirk are adjusted identically.

Proposition 3: Suppose observations 1 and 3 hold, and everyone adjusts by the same symmetric and action-invariant Bush-Mosteller rule.

(i) If $\overline{p}_t(v) \leq \frac{1}{2}$, then $\overline{p}_t(v) < \overline{p}_{t+1}(v)$.

(ii) If $\overline{p}_t(v) \geq \frac{1}{2}$, then $\overline{p}_{t+1}(v) > \frac{1}{2}$.

This result gives up the generality regarding the class of adaptive processes covered by proposition 2. But in exchange we relax the assumption about status quo participation propensities: part (i) shows that the electorate's average propensity continues to rise so long as the *community's*—not all individuals'—current average propensity is under one half. Further, part (ii) shows that once the community's average propensity reaches one half, it will remain at least that high so long as observations 1 and 2 continue to hold. Hence proposition 3 provides analytical backbone to our simulation results. And by doing so it provides an efficient kind of sensitivity testing: it reduces the need to search laboriously over different parameter values in order to see whether the breakout of participation seen in run 3a holds for a wide range of parametric configurations (e.g., proposition 3 holds for any values of the benefits and costs of voting such that $b > c > 0$).

It is worth noting that if aspirations were fixed ($\lambda = 1$) then under the conditions of proposition 3, eventually any district would reach a (community-wide) average vote-propensity exceeding one half, *regardless of its initial distribution of vote-propensities*. Further, once this happened then the average propensity to participate would remain above one half thereafter. Thus, partitioning all levels of average participation-propensities into two sets—moderately low (less than or equal to one half) versus moderately high (above one half)—we see that for fixed and intermediate aspirations, moderately high average vote-propensities is the system's *unique absorbing state.*

Propensities are mental states and so are unobservable. Hence propositions 2 and 3 do not themselves offer testable predictions. Fortunately, they do generate implications about turnout, which is, of course, observable.

Corollary 1: If the assumptions of propositions 2 or 3 hold then expected turnout in $t+1$ exceeds the expected turnout in t.

Since corollary 1 is essentially the payoff to the sequence of analytical mini-results that began with observation 1, it is worth recapitulating the sequence to get a clear overview of the logic that leads to this conclusion. When observation 1 holds, aspirations are such that winners are happy and losers are sad. By observation 2, if everyone adjusts via some form of Aspiration-Based Adaptive Rule, then the winning voters and the losing shirkers will become more disposed to vote (whenever that is possible). By observation 3, these two groups are a majority of the electorate if (say) there are as many Democrats as Republicans. Hence it follows that if no citizen is fully disposed to vote in t and observations 1-3 hold, then a majority of the electorate becomes more inclined to participate at the end of period t (proposition 1). If the district's current distribution of propensities isn't so high as to evoke ceiling effects, then the demographic mobilization of proposition 1 in turn implies that the electorate's average propensity to vote rises (propositions 2 or 3). Finally, since propensities are meaningfully related to behavior ($\epsilon < \frac{1}{2}$), expected turnout must also rise (corollary 1).

In sum, we see that run 3a is neither a fluke nor a mysterious product of a mysterious program. Given a starting point of complete shirking, there are good analytical reasons for expecting to observe a breakout of participation. Hence the results of run 3a make sense: they are the product of a few simple mechanisms which are instantiated by the parametric setting of this run.

Of course, these simple mechanisms hold for many other parameter settings. Consequently, the analytical mini-results constitute a sweeping sensitivity test for that run. Instead of laboriously investigating a huge number of other parametric configurations, we can invoke the "if" part of a result and know that a finding stands up for all parameter values swept up by that clause. Deduction has its uses.

IV. Why Does Mobilization Stop?

We know from the simulations reported thus far that mobilization does not continue indefinitely: it appears to level out at about 50% turnout. Why?

Clearly the above analytical results do not directly explain why mobilization comes to an end. They cannot do so, since they describe sufficient but *not* necessary conditions for increasing participation. Thus, based on the analytical results reported thus far, we simply do not know what happens when one or another of the relevant conditions fails to hold. We turn now to that issue.

Since behavior in our model is mediated by aspiration levels, the distribution of citizens' aspirations is very important. It is therefore no accident that the cornerstone of our analytical results, observation 1, pertains to aspirations. The hypothesis of observation 1 is that winners' aspirations are low enough so that winning is gratifying even if one paid the costs of participating, and losers' aspirations are high enough so that losing is dissatisfying even if one avoided those costs. Let us take up the hint inherent in observation 1 by examining what happens when we go to the opposite extreme: the aspirations of winning voters are high (exceed $b - c$) while those of losing shirkers are low (are below 0). It is also convenient to assume throughout this subsection that all aspirations are *realistic*, i.e., exceed the stage game's lowest possible payoff of $-c$ and are strictly less than the game's highest possible payoff of b.[19]

Proposition 4: Suppose $p_{i,t}(v) > 0$ for all i. People adjust by any arbitrary mix of ABARs. If in t the winning voters' aspirations exceed $b - c$, the losing shirkers' aspirations are less than 0 and all aspirations are realistic, then after the election *all* citizens will become less inclined to vote.

Obviously, given that *everyone's* propensity to vote decreases, proposition 4 immediately implies that the electorate's average propensity to vote falls.[20] Indeed, with only a bit more structure we can say much more than that.

The following result uses the idea of one distribution being stochastically smaller than another. This concept is defined precisely in the appendix; here, it suffices to define it informally. We will say that distribution X is stochastically smaller than distribution Y if X puts more weight on lower

[19] It is easy to show, given the existence of trembles, that no matter what initial aspiration a citizen begins with, eventually her aspiration will become realistic with probability one. And clearly, any aspiration that is realistic in t remains realistic forever after.

[20] For this reason, when the hypotheses of proposition 4 are satisfied we don't need the analogues to propositions 2 or 3. Since the decline of participation is universal when proposition 4 obtains, merely assuming that people adapt via ABARs suffices; we don't need to make more specific assumptions about the nature of trial-and-error.

values than Y does, and Y puts more weight on higher values than X does. Thus the stochastically bigger distribution, Y, is "shifted to the right" compared to X.

Corollary 2: Suppose the conditions of proposition 4 hold. Then the distributions of vote-propensities and of turnout probabilities in $t+1$ are strictly stochastically smaller than the corresponding distributions in t.

Since the distribution of turnout becomes stochastically smaller, we get for free that the expected turnout in the next election also falls. The former property is much stronger than the latter.

Clearly, since everyone is becoming more likely to stay home, this situation is unstable. Hence this one-sided domination cannot be a long-run probabilistic equilibrium. Something must give. What will give, we believe, is that the dominating faction will become too complacent: too many will learn to free ride on their comrades' efforts. This will make the race competitive again.

Note that proposition 4 and its implications rest on very weak assumptions about the nature of trial-and-error learning: only the qualitative properties of ABARs were assumed. Thus, citizens may differ substantially in how they learn.

What will produce aspirations of the type assumed by proposition 4? The proximate cause is that recently one faction has been winning a lot. Since winning produces payoffs of either $b-c$ or b, a long enough string of victories will drive the winners' aspirations into the $(b-c, b)$ interval. Meanwhile, losers have been getting either zero or $-c$, so a long run of defeats sends their aspirations into $(-c, 0)$.

What produces such one-sided strings? There are two main possibilities. If one side is much larger than the other it will win many elections by virtue of size. (As we will see in section V, changing the *relative* sizes of the factions in the simulation seems to affect turnout as this reasoning indicates.) Alternatively, if they are relatively balanced one side might reel off a string of victories by chance. Thus, because aspirations adjust to experience and one side has been winning while the other has been losing, people in the dominant faction currently have high aspirations while members of the weaker party have low aspirations.

Proposition 4 makes the rather restrictive assumption that all winners have high aspirations and all losers have low aspirations. This allows us to use the very weak assumption that people adjust via any arbitrary mix of ABARs. If we make stronger assumptions about adjustment mechanisms, then we can weaken the assumption about the distribution of aspirations. The following results, 5 and 6, parallel propositions 2 and 3, respectively: like proposition 2, proposition 5 assumes that adaptation is via an ABAR that satisfies general properties such as monotonicity while 6 (like proposition 3) presumes that adaptation is via the symmetric Bush-Mosteller rule.

Proposition 5 reverses several key assumptions of proposition 2.[21] It also adds the useful property of action-invariance.

Proposition 5: Suppose everyone adapts via the same monotonic, symmetric, and action-invariant ABAR. All aspirations are realistic in t. Together, winning voters whose aspirations are less than or equal to $b - c$ and losing shirkers whose aspirations are at least 0 are a minority in t. If $p_{i,t}(v) \geq \frac{1}{2}$ for all i, then $\overline{p}_{t+1}(v) < \overline{p}_t(v)$.

Similarly, proposition 6 is essentially the opposite of proposition 3: now, there are more decreasers than increasers, and the average propensity to vote is already at least one half. The former condition obviously produces a tendency toward reduced mobilization, and the latter produces a ceiling effect on the per capita quantities of further increases in participation.

Proposition 6: Suppose everyone uses the same symmetric and action-invariant Bush-Mosteller rule. All aspirations are realistic in t. Further, winning voters whose aspirations are less than or equal to $b - c$ and losing shirkers whose aspirations are at least 0 together form a minority of the electorate in t. If $\overline{p}_t \geq \frac{1}{2}$ then $\overline{p}_{t+1}(v) < \overline{p}_t(v)$.

And once again, since propensities are meaningfully connected to actions (i.e., $\epsilon < \frac{1}{2}$), both proposition 5 and proposition 6 imply that expected turnout in $t + 1$ is less than that in period t.

These results help to explain not only why voter participation stops well short of universality, but also why it occasionally collapses altogether. Figure 4 shows the evolution of voter participation over the course of a single 5,000 period simulation. After a string of defeats starting near period 1750, voters from party D quickly become disaffected: their aspirations are negative, so shirking satisfices even in defeat, while occasional votes are nowhere near pivotal. For members of the victorious party R, however, voting still satisfices, and only slowly (through trembles) do these citizens learn to stay home. But for reasons that we now understand (proposition 3 and run 3a), the no-voting outcome that inevitably results (near period 3500) is unstable, and as expected the citizens mobilize once again. Such collapses occur intermittently throughout our simulation results, accounting for the slight bimodality apparent in our histograms of voting propensities. Since each histogram captures a "snapshot" of propensities at $t = 1,000$, typically a few of the

[21] Like proposition 2, it can be generalized to cover failure-sensitive ABARs, given a plausible turnout condition. See the appendix for the more general form of proposition 5 and its proof.

1,000 simulations in each run will be in the midst of a participation collapse.

[figure 4 about here]

Together, propositions 3 and 6 give us a rather good understanding of the dynamics of turnout in the simulation, given (for example) a start of universal pure shirking. Initially, demographic mobilization and the ceiling effect—per capita amounts of propensity-change—reinforce each other, as explained by observations 1-3: more people increase their propensity to vote than decrease it, and since they began with zero inclination to participate, increasers have plenty of adjustment room while decreasers have little. Eventually, however, mobilization is self-limiting because one or both of the underlying factors will reverse themselves. First, once the community's average propensity to vote exceeds one half, the ceiling effect favors shirking: there is now more room to decrease than to increase. Second, one side may run off a string of victories, which will send some of the winners' aspirations above $b - c$ and some of the losers below zero. If this happens to enough people, the demography of mobilization will turn around: now a majority of people will become less inclined to vote. And so turnout will start to fall.

Further, once the community's average turnout has fallen below one half, if a majority of people still are becoming less inclined to vote then the district's average participation will remain below one half, as the following corollary to propositions 3 and 6 shows.

Corollary 3: Assume the conditions of proposition 6 hold, except that $\overline{p}_t(v) \leq \frac{1}{2}$. Then $\overline{p}_{t+1}(v) < \frac{1}{2}$.

For any symmetric Bush-Mosteller rule, the ceiling effect flips at the midpoint of the scale—a vote-propensity of 0.5—so it is not surprising that in our benchmark runs the average turnout is around 50 percent.

Decreasing Turnout in Unstable Democracies. Thus far we have primarily focused on circumstances in which citizens have realistic aspirations. These are likely to occur in stable democracies, in which citizens have learned which payoffs are feasible and which are utopian. But in a new or unstable democracy people may not know what is a satisfactory outcome; they may expect a new regime to transform the society overnight.

Having unrealistically high aspirations—those exceeding the game's maximal payoff (here, b)—can only produce disappointment. And if in the excitement of building a new society these unreasonable aspirations are coupled to high rates of participation, then disappointment leads to disillusionment with politics, i.e., to decreased political participation, as the next result shows.

Proposition 7: Suppose everyone adapts by the same monotonic and action-invariant ABAR. If in t all aspirations are unrealistically high and $p_{i,t}(v) > \frac{1}{2}$ for all i, then $\overline{p}_t(v) > E[\overline{p}_{t+1}(v)]$. If further a majority of citizens vote in t, then $\overline{p}_t(v) > \overline{p}_{t+1}(v)$.

Hence disillusionment with the results of a fledgling democracy can lower the average propensity to participate, and (since $\epsilon < \frac{1}{2}$) expected turnout as well.[22]

V. Comparative Statics and Dynamics

Although, the main result of this paper is the emergence of substantial turnout in large electorates, this finding does not exhaust what the model can generate. We turn now to several other simulation results.

Variations in Initial Values. Our default value for the initial propensity to vote is one half, and this was also the value used in runs 1 and 2. Since in each run participation was on average about half the electorate, one might wonder whether the result was being driven by the initial propensity. Is the system sensitive to starting conditions in this manner?

To address this we have run the model with different starting values of the propensity to vote and of aspiration levels. Runs 3a and 3c record typical patterns. (We have already seen run 3a in another context.) They indicate that the starting value of the vote-propensity has no long term effect on the system.

[runs 3a and 3c about here]

Runs such as these indicate that the system reaches a unique limiting distribution, no matter what the starting aspiration levels and propensities were. (Of course, the limiting distribution *is* affected by different values in exogenous parameters such as payoffs, as it should be.)

Variations in the Costs of Voting. One of the most puzzling features of rational choice models of turnout is that while the models cannot explain why voters care to vote at all, they are consistent with many empirical regularities concerning the *variation* of turnout in response to changing exogenous features (e.g. Hansen, Palfrey, and Rosenthal 1985). One of these regularities—

[22]We would like to thank Carole Uhlaner for suggesting that we use our model to analyze the phenomenon of decreasing participation in new democracies.

with considerable evidence supporting it [citations]—is that voting falls as its costs rise (run 4a). This is an implication of our model, for large (symmetric) differences in costs (runs 4a-4c).

[runs 4a-4c about here]

Small variations in cost seem to have little effect on turnout. This is partly a matter of the system being "well-behaved": its endogenous variables appear to respond in a continuous manner to parametric variations. But there is also a substantive issue here regarding the effects of aspirations. In general, the effects of changing payoffs on behavior are muted by aspirational mechanisms (March 1994), and we see that phenomenon played out here. The reason is aspirations adjust to experience—here, payoffs. Hence an increase in the cost of voting is partly absorbed by lower aspirations in the steady state; thus voting stays higher than it would have had aspirations been exogenously fixed.

Similarly, if the costs of voting are decreased participation rises (run 4b), though perhaps not as much as one would expect. Again, the change in the cost of voting is partly absorbed—this time, by higher aspirations, which "eat up" some of the parametric shift. This mediating effect of aspirations can be strong enough to produce rather counterintuitive effects. Consider, for example, what happens when a sense of civic duty makes the costs of voting *negative* (run 4c). As expected, participation rates continue to climb, relative to runs in which voting is costly. But they do not climb as high as rational choice considerations would lead one to expect. In particular, if the costs of voting are negative then participating is a strictly dominant strategy. Classical game theory predicts that people will always play a strictly dominant strategy. Yet this is not what happens in run 4c: people do not become fully disposed to vote. This seems strange—until we recall the mediating effects of aspirations.[23] Note that the aspirations in run 4c are much higher than they were in the comparable benchmark run (with positive costs of voting). Hence losing-while-voting is still usually dissatisfying. Consequently losing voters often become less inclined to participate, which means that vote-propensities cannot stabilize at 1.0.

Variations in the Benefits of Voting. Turning now to variations in the value of winning the election, we see a similar pattern: turnout generally moves in the intuitive direction—higher benefits, more voting—but the effect on behavior is dampened by the mediating factor of changes in aspirations (see, e.g., runs 5a-5b).

[23] As Bendor, Diermeier and Ting (1999) show, ABARs frequently do not converge to Nash equilibria, or may even fail to select equilibria supported by dominant strategies. For example, in the Prisoners' Dilemma, agents using a Bush-Mosteller rule cooperate most of the time.

[runs 5a-5b about here]

However, changes in benefits can produce effects that are *qualitatively* counterintuitive, not merely quantitatively so. Consider run 5a, which is just like run 4c (which has negative costs of voting) except that in this latest run the benefits of winning have been sharply reduced. Surprisingly, participation is *higher* when the benefits of winning are *lower.* Indeed, when we reduce b still further, to the point where people do not care about winning at all (run 5b), then citizens become fully disposed to vote!

What is going on here? Are runs 5a and 5b bizarre? Do they indicate that our simulation model is fundamentally flawed?

Not so: far from being pathological, these runs illustrate a more general result. The counterintuitive effect that mobilization increases when the value of winning falls to zero is not peculiar either to the particular configuration of parameter values in runs 5a and 5b or to the functional forms of the simulation model. The next result will show analytically that if citizens adapt by processes that satisfy two general conditions, then we obtain patterns similar to those of the above runs.[24]

The first property is a generalization of how aspirations adjust over time in our simulation model. It isolates the essential property of that adaptation: aspirations adjust by moving toward current payoffs. For this reason we call dynamics of aspiration-adjustment that have this feature *payoff-responsive.*

Definition. Payoff-responsive aspiration adjustment:

(1) If $a_{i,t} < \pi_{i,t}$ then $a_{i,t} < a_{i,t+1} < \pi_{i,t}$;

(2) If $a_{i,t} = \pi_{i,t}$ then $a_{i,t} = a_{i,t+1} = \pi_{i,t}$;

(3) If $a_{i,t} > \pi_{i,t}$ then $a_{i,t} > a_{i,t+1} > \pi_{i,t}$.

The significance of payoff-responsiveness for the following result is that it ensures that once an aspiration becomes realistic, it stays realistic thereafter. Note also that the amount of adjustment need not be stationary. It can, e.g., decrease over time: an aspiration level that is anchored in much experience may respond little to one new outcome.[25] Finally, note that this property allows

[24] For a more detailed analysis of the counterintuitive effects of aspiration-mediated change, see Bendor, Diermeier and Ting (1999, p.18-20).

[25] Suppose that one's aspiration tomorrow is the simple average of all prior payoffs and one's initial aspiration: $a_{i,t+1} = \frac{(a_{i,0}+\pi_{i,1}+\cdots+\pi_{i,t})}{t+1}$. Then the impact of current payoffs will steadily fall over time.

aspiration-adjustment to be somewhat random, as long as current payoffs are not identical to current aspirations. (For example, if $a_{i,t} < \pi_{i,t}$ then tomorrow's aspiration could be randomly drawn from the interval $(a_{i,t}, \pi_{i,t})$.)

The second property is a "no-leapfrogging" condition. Informally, it says that if i's vote-propensity is no less than j's at the beginning of period t and i's propensity increases in t then j's propensity cannot leapfrog past i's, no matter what happens to j. This sounds eminently reasonable but it does have some bite: e.g., for the Bush-Mosteller it implies adjustment that is symmetric in the face of success and failure.

Definition. No leapfrogging:

Suppose i and j use the same ABAR, $p_{i,t}(v) \geq p_{j,t}(v)$ and in t either (1) i votes and is reinforced or (2) i shirks and is inhibited. We will say that the ABAR permits *no leapfrogging* if $p_{i,t+1}(v) \geq p_{j,t1}(v)$.

With these properties in hand we can turn to the next result. Note that the following result holds for *any* value of b larger than $-c$. Hence it does not matter how valuable winning is; the pure intrinsic motivation of civic duty leads to stochastically greater mobilization. Note also that the result generalizes run 4c considerably in that it allows for heterogeneous initial propensities and aspirations.

Proposition 8: Districts A and B are parametrically identical except that in A, $b = 0$, while in B, $b > -c$ ($c < 0$ in both). Everyone adjusts propensities by the same ABAR, which permits no leapfrogging. Aspirations are adjusted by any arbitrary mix of payoff-responsive dynamics. The initial distributions of vote-propensities are the same in the two districts. Everyone begins with realistic aspirations. Then for all $t > 1$, A's distribution of vote-propensities is stochastically bigger than B's.

Proposition 8 immediately implies that after period one, the expected vote-propensity in district A exceeds the expected propensity in B, at every date. The result is driven precisely by the fact that people in district A are motivated only by their sense of civic duty. Since they do not care about the collective outcome, there are only two payoffs for them: $-c > 0$ for voting and zero for shirking. Since everyone's aspirations are realistic (in the beginning and, by payoff-responsiveness, thereafter), in district A they must always be in $(0, -c)$. Hence for people in this district participating is always gratifying while shirking never is. Consequently voting is always reinforced and shirking is always inhibited, *regardless of the electoral outcome.* Thus, since everyone is using an ABAR, in district A *everyone's* propensity to turn out rises in every period (until they reach 1.0, where they will stay). In contrast, in district B people care about winning. This implies that losing while voting can be disappointing. If this occurs—and such an outcome is always possible—then the losing voters will become less inclined to vote.

Note that proposition 8 holds for an extreme comparison, between a district where winning is worthless and one where winning is worth more than the private satisfaction of doing one's civic duty—and yet, despite this, the district where winning is worthless has stochastically *more* turnout than the one where winning is not only worth something, it is the more important payoff. The remaining possibility that is not considered by this result is that $0 < b \leq -c$. To get the result here would be less counterintuitive, because in this case winning in district B is not worth very much.

Asymmetric Costs. When one faction has systematically higher costs of voting (due either to harassment by the other faction, as in the old South, or to more innocuous factors), one would expect that bloc to participate less in elections. That seems to be the case here (run 6).

[run 6 about here]

Run 6 spreads the factions' costs of voting apart from the common default value of 0.25. With the Democrats' cost of voting equal to 0.4, while the Republicans incur a cost of only 0.1, we see that by about period 800 only about one third of all Democrats are voting. It is more interesting to observe that Republican turnout is also falling below the values we saw in the benchmark cases of runs 1 and 2, even though their costs in run 6 are lower than in 1 and 2. Evidently this reflects an indirect influence of the higher Democrat costs: since Democratic participation is reduced, Republicans learn that they can stay home and still win. Generally in runs such as this the disadvantaged group will experience a rebound after their opponents stop voting. However, this recovery is brief as the advantaged group will quickly re-assert itself, thus producing again the pattern observed in run 6.

However, run 5a and proposition 8 suggest that if citizens had negative costs of voting but one faction was indifferent to winning ($b = 0$) while the other did care, then the faction with the lower b would turn out *more* than the side with the higher b. This expectation is confirmed by run 7.

[run 7 about here]

Variations in Relative Faction Size. The voters' behavior in the simulation does not respond to changes in the absolute size of the electorate. (This, as noted, is consistent with our analytical results, which do not depend on this parameter at all.) However, the system does respond in interesting ways to the *relative* sizes of the factions.

Consider first run 8a, in which there are 5,200 Democrats and 4,800 Republicans. Even with this modest partisan split, minority faction members (Republicans) in nearly half of the electorates sit it out almost completely by period 1,000. One would expect this to depress long-run Democratic turnout as well, and indeed their participation is substantially lower than in the benchmark cases.

[run 8a about here]

What causes this? Even relatively small differences in factional sizes can produce big differences in winning elections, at least initially. Hence Republicans quickly learn that voting doesn't pay: it yields a payoff of $-c$. This generates negative reinforcement, given realistic aspirations. As their propensities drop to zero, Democrats gradually learn that voting is unnecessary (as in Figure 4). But after Democratic turnout drops low enough, Republicans learn to vote again, winning briefly until the larger faction mobilizes once more, driving the minority faction into another prolonged slump.

This pattern is troublesome: empirically, turnout is not this sensitive to small departures from exact partisan balance. Surprisingly, however, a seemingly innocuous (and yet more realistic) assumption about payoffs produces far more reasonable turnout patterns. Suppose that electoral payoffs are stochastic rather than deterministic. Winners get $b + \theta_{i,t}$ and losers get zero plus $\theta_{i,t}$, where $\theta_{i,t} \sim U[-s, s]$ ($s > 0$) and is independent and identically distributed across voters and periods. Using $s = 0.1$, run 8b shows that with the same distribution of voters, turnout levels are both higher and more stable. In fact, the effect noted in run 8a is exactly reversed: voters in the minority faction are *more* likely to participate![26]

[run 8b about here]

These effects are the joint product of the size of the random shock, the size of the minority faction, and (once again) the mediating role of aspirations. To see how all this works, let us consider a minority faction that is hurting: it has lost repeatedly, and so both aspirations and vote-propensities have been depressed. Aspirations are now low enough so that losing-while-shirking is acceptable (i.e., they do not exceed zero, the payoff to a losing shirker), but they are still realistic,

[26] A similar effect can be achieved by varying the cost of voting. Further, this additional result does *not* require making voting costs negative. Instead, as with random electoral payoffs, random voting costs matter when the support of the shock is sufficiently large.

so they exceed the worst possible payoff, $-c$, which accrues to a losing voter. Further, their vote propensities are so low that they are almost sure to lose in the current period. And given their realistically low aspirations, a loss that yields deterministic electoral payoffs would further reduce the vote-propensities of *all* minority citizens: the losing voters will become still more disenchanted with turning out, while the losing shirkers accept their lot.

Now consider, however, randomly perturbed electoral payoffs. Since the deterministic payoff to losers is zero, every minority citizen will get a random draw of θ_t. Now comes the crucial point: if θ_t has a sufficiently big variance, then a substantial number of losing voters will be satisfied (since there will be an appreciable chance that $-c + \theta_{i,t}$ will exceed i's aspiration), and similarly a substantial number of losing shirkers will be *dis*satisfied (since there will be a substantial chance that $\theta_{i,t}$ will fall below i's aspiration). And both of these sets will become more inclined to vote. Thus, stochastic electoral payoffs can interrupt the vicious cycle that keeps the minority's voting propensity low.

Runs 8a and 8b have shown that as we increase the support of θ from 0 to 0.4 (i.e., $s = 0.2$), the participation level of the minority stabilizes at near 50%. Runs 8c and 8d show some results for even more lopsided distributions of voters. In 8c, the Democrats have 60% of voters, but still between 45 and 50 percent of the minority faction turn out. Minority participation has its limits, however. In run 8d, where only 25% of citizens are Republicans, the pattern more closely resembles that of run 8a. Significantly, the variation in b is quite small in runs 8b-8d, ranging from 0.9 and 1.1. Thus the instability in run 8a (where $s = 0$) appears to be more the exception than the rule.

[runs 8c and 8d about here]

A glance at the histograms reveals some interesting empirical implications. As the disparity in faction size increases, *aggregate* turnout drops. The minority turnout rate also exceeds the majority's, with the disparity increasing in the asymmetry in faction size. This disparity arises because it is harder for minority members to free-ride on their colleagues. As the runs show, this is reflected in the aspirations of the two parties. When factions are very unbalanced, majority citizens, who often win without voting, have high aspirations; minority voters aspire to little more than zero.

Our last analytical mini-result shows that the patterns exhibited by runs 8b-8d hold for many types of adjustment rules and many kinds of shocks. Proposition 9 establishes that minority vote propensities can increase even in the toughest of situations—i.e., even when (1) minority aspirations are realistically low, so as to make losing-while-shirking acceptable but losing-while-voting dissatisfying, and (2) the minority faction loses—*if* the random shock is sufficiently spread out.

This crucial assumption about the random shock is defined informally as follows. A random variable θ *thins out* if we can make the probability that θ is in any interval $[a, b]$ as small as we please. (A precise definition of thinning out is provided in the appendix.) Distributions that thin out include the normal and the uniform.

In the following result, $p_{m,t}(v)$ denotes the vote-propensity of a minority faction member in period t.

Proposition 9: Suppose the minority faction loses in t. Aspirations of all minority faction members are in $(-c, 0]$ in t; all their vote propensities are strictly less than $\frac{1}{2}$. Players can use any ABAR that is monotonic, action-invariant, and symmetric. The random payoff shock $\theta_{i,t}$ is independent and identically distributed across agents, it is continuously distributed with $\Pr(\theta < 0) = \Pr(\theta > 0) = \frac{1}{2}$, and it thins out. If θ_t is sufficiently thin then for every minority faction member $E[p_{m,t+1}(v)] > p_{m,t}(v)$.

The conclusion of the proposition is that, under the stated conditions, the expected propensity (in $t + 1$) of *everyone* in the minority faction rises.[27] Obviously, therefore, the expected vote-propensity of the overall minority faction must also rise. And since $\epsilon < \frac{1}{2}$, the minority's expected turnout also increases in the next period.

VI. Extensions and Conclusions

This work attempts to construct a mathematical model of adaptively rational electoral participation. As such it is quite preliminary. There are a number of modifications which would improve the model. We consider several of these here.

(1) Perhaps the most important change would be to allow the citizens to decide which party to vote for. Thus each person would have three choices: stay at home, vote Democratic, vote Republican. Due to the coarseness of the feedback and the simple nature of learning posited here, we expect this modification will result in voters often making mistakes: voter i might vote for the wrong party, but for other reasons (how *most* people voted) i's payoff exceeds his aspiration level, thus reinforcing his error. Hence it may take a long time for the weakly dominated strategy of

[27] Proposition 9 generalizes to asymmetric (failure-sensitive) ABARs. See the proof in the appendix.

voting for the wrong party to be extinguished (if indeed it ever is). However, we do not expect that this change will cause participation to plummet.

Alternatively, it would be equally consistent with the spirit of adaptive rationality to assume that voters satisficed by voting for the incumbent if and only if the incumbent's performance exceeded some critical threshold.

(2) Next on the agenda would be to introduce active (though adaptive) parties or candidates. Payoffs to voters would then no longer be exogenously fixed; instead, they would depend on the platform taken by the winning party. It would be interesting to see whether the candidates tend to converge to the median voter.[28] If this convergence does obtain, it might depress turnout, though again we suspect that it will not drive participation down to negligible levels.

(3) The rate of adjustment in learning probably should respond to how much current aspirations and current payoffs differ. Big differences should produce large changes in propensities.

The results reported here are quite encouraging. Reinforcement learning, mediated by endogenous aspirations, seems to lead naturally to substantial turnout under a wide array of parametric configurations. Indeed, we have had to work hard in order to *suppress* participation. Thus, we look forward to seeing the model subjected to systematic empirical testing.[29]

Finally, we think it is worth noting that we did not design this model in order to study the paradox of voting. Indeed, our original purpose was quite far removed from this: we wanted to study how adaptively rational agents would behave in two-person, mixed-motive games that represented various kinds of institutional situations, such as committee-floor relations (Bendor, Diermeier and Ting 1999). That the model thus far seems to do rather well in a different domain has come as a pleasant surprise. And given the often-articulated criterion in the philosophy of science (e.g., Laudan, Laudan and Donovan 1988) that theories should be able to handle new problems (i.e., problems which a theory was not designed to solve), our surprise may reflect more than merely a report on the psychology of the researchers.

[28] In an analytical model of adaptive parties in a canonical Downsian setting (Bendor, Mookherjee and Ray 1999), there is a unique stable outcome: the candidates converge to the median. However, as in Bendor, Mookherjee and Ray (1998), this model assumes static (though consistent) aspirations for vote-share by the parties, and it also assumes sincere and error-free voting by the citizens, with universal turnout. Whether parties converge when aspirations adjust dynamically and voters also behave adaptively remains to be seen. (We are trying to bring these two streams of research together, but this will take awhile.)

[29] See Kanazawa (1998) for a significant step forward on the empirical front.

APPENDIX

A. Notation, Definitions, and Some Basic Assumptions

We adopt the following terminology in describing the model and simulation. A *period* is a one-shot play of the turnout game. A *simulation* is a sequence of periods, for a specific set of voters. Finally, a *run* is a collection of simulations.

$p_{i,t}(x)$ $(x = v, s)$ denotes the propensity of citizen i to choose action x in period t.

Since all the results are couched in terms of the propensity to vote, for convenience we shorten $p_{i,t}(v)$ to $p_{i,t}$ in most of the appendix. Thus $p_{i,t}$ denotes i's propensity to vote in period t.

$\overline{p}_t$ denotes the electorate's average propensity to vote in period t. (This is a summary statistic of a particular sample path; it is not an expectation.)

$N > 1$ is the number of citizens in the district.

n_D is the number of Democrats; n_R, the number of Republicans. (We always assume that $n_D > 0$ and $n_R > 0$.)

$a_{i,t}$ is citizen i's aspiration level in period t.

An aspiration is *realistic* in t if it is strictly larger than the stage game's minimal payoff and strictly less than the stage game's maximal payoff.

$\pi_{i,t}$ is citizen i's payoff in period t.

b is the benefit of winning and c is the cost of voting. Unless specified otherwise (i.e., in proposition 8), we assume $b > 0$ and $c > 0$.

ϵ is the probability of a tremble, i.i.d. over time and over agents.

$\alpha \in (0, 1]$ is the extent to which propensities adjust toward one under reinforcement using a Bush-Mosteller rule.

$\beta \in (0, 1]$ is the extent to which propensities adjust toward zero under inhibition using a Bush-Mosteller rule.

$\lambda \in (0, 1)$ is the extent to which aspirations are weighted toward extant aspirations.

We call an election *conclusive* if it isn't a tie.

B. The General Model

Players and Payoffs. We consider a series of elections in a single "district." There are two types of citizens, D and R, whose populations satisfy $n_D + n_R = N$. In each period, each citizen may either vote ($x = v$) or stay home ($x = s$). The side which turns out more voters wins the election. Members of the winning side earn a payoff of $b + \theta_{i,t}$, while members of the losing faction get a payoff of $d + \theta_{i,t}$, where $\theta_{i,t} \sim U[-s, s]$ ($s \geq 0$) and is i.i.d. across voters and periods. (Unless otherwise stated—i.e., in runs 8a-8c and in proposition 9—$s = 0$, so the electoral payoffs are deterministic for most simulation runs and most of the analytical results.) In the event of a tie, everyone gets $\frac{b}{2}$. The private cost of voting is c. Payoffs are additive in the benefits and costs. Thus winning voters get an expected payoff of $b - c$; winning shirkers get an expected payoff of b. Losing voters get an expected payoff of $d - c$; that of losing shirkers is d. Except where noted, we assume that $b > c > 0 = d$.

At $t = 1$, each citizen i begins with an initial propensity to vote, $p_{i,1} \in [0, 1]$, and an initial aspiration level, $a_{i,1}$. Voters in the same bloc are assumed to be initially homogeneous in these and all other respects.

Propensities, Trembles and Actions. In each period, intended actions are determined randomly through each citizen's propensities. Citizens then "tremble" their moves with probability ϵ. A trembled move is the opposite of the intended move. Thus, the probability that voter i actually chooses action x in period t equals $p_{i,t}(x) \cdot (1 - \epsilon) + (1 - p_{i,t}(x)) \cdot \epsilon$.

The realized actions determine the election winner. Payoffs are then assigned as above.

Aspirations and Adaptation. A player who achieves a payoff of at least her aspiration ($\pi_{i,t} \geq a_{i,t}$) codes her last action as a "success," while if $\pi_{i,t} < a_{i,t}$ the action is coded as a "failure." Propensities for successful actions are reinforced according to the classical linear Bush-Mosteller equation:

$$p_{i,t+1}(x) = p_{i,t}(x) + \alpha(1 - p_{i,t}(x))$$

Likewise, propensities for unsuccessful actions are inhibited:

$$p_{i,t+1}(x) = p_{i,t}(x) - \beta p_{i,t}(x)$$

Obviously, the action not taken in period t is adjusted suitably downwards if x is reinforced, and upwards if x is inhibited. We assume that $\alpha = \beta$ for the simulation runs in this paper, i.e., we use a symmetric Bush-Mosteller rule. (Symmetry is relaxed in propositions 1, 2, 4, 5, and 7-9.)

Aspirations also adjust to experience. At the end of each period there is a payoff-weighted probability of aspirations updating for each citizen:

$$\Pr(update) = \left(\frac{|\pi_{i,t} - a_{i,t}|}{max\{\pi_i\} - min\{\pi_i\} + C} \right)^K$$

Here, $max\{\pi_i\}$ and $min\{\pi_i\}$ denote the maximum and minimum payoffs attainable in a single election for player i (typically, $max\{\pi_i\} - min\{\pi_i\} = b + c$). C and K are auxiliary parameters that may be used to fine-tune the probability of aspiration adjustment. In the runs conducted here, we assume $C = K = 1$.

When aspirations do adjust, tomorrow's aspirations are a weighted average of today's aspirations and today's payoffs:

$$a_{i,t+1} = \lambda \cdot a_{i,t} + (1 - \lambda)\pi_{i,t}$$

This process repeats until the user-specified number of election periods is reached.

C. Simulation Program

Because of the serious tractability issues involved, we have written a program in ANSI C to simulate the described adaptively rational behavior. The program is compatible with the GNU C compiler and most UNIX operating system configurations.

When started, the program allows the user to set the following parameters before each run:

— the number of simulations;

— the number of periods per simulation;

— the faction sizes, n_D, and n_R;

— the payoff parameters, b, c, and d;

— the tremble probability, ϵ;

— the reinforcement and inhibition rates, α and β;

— the aspiration updating parameters, λ, C, and K.

— each faction's vote propensity at $t = 0$, $(p_{i,0})$;

— each faction's aspiration level at $t = 0$, $(a_{i,0})$;

When a run begins, the program initializes a pseudorandom number generator with an integer representing the current time. This standard procedure effectively ensures that each run's random parameters are independent of those of other runs. The program then initializes a custom data structure that keeps track of state variables (i.e., propensities and aspirations, which may change over the course of a run) and statistics related to the history of play.

In each period, intended moves are determined and then trembled to produce actual moves in the manner described previously. After payoffs are revealed, the data structure is updated to reflect the changed propensities and aspirations. These variables revert back to their original values for each new simulation.

The data structure associated with each run can be used to recover statistics at different levels of a run. To reduce runtimes, only those statistics that are requested by the user are collected. In particular, the user may opt to view or save to disk any of the following:

— the moves, payoffs, and adjusted propensities and aspirations after each period;

— average and cumulative propensities and aspirations for each simulation;

— a histogram of propensities for each voter for each simulation;

— average propensities and aspirations across simulations for certain periods;

— a histogram of final-period propensity and aspiration levels across simulations.

For the games investigated in this paper, the final two items proved to be most useful. All of the run reports are composed of either time series of average propensities or histograms of final period propensities across simulations. While the program does not generate graphics on its own, the output files it creates are easily read into other programs for further processing.

The program is able to generate large samples of game play quickly, although naturally large electorates will slow run times considerably. Typical runtimes for most of the runs reported here (10,000 voters, 1,000 periods and 1,000 simulations) ranged between 5 and 30 hours, depending on hardware configurations. However, the invariance in both our simulation and analytical results with respect to population size (see run 2 and propositions 1-7) gives us high confidence in their robustness.

D. Proofs

Before we prove the results described in the text, it is convenient to establish some facts about ABARs. These will be used in propositions 2, 5, 7 and 9.

Useful Facts about ABARs

Notation: we use $\delta^+_{i,t}(v;p)$ to denote the increment in i's vote-propensity if i's current vote-propensity is p, i votes in t, and is satisfied. $\delta^+_{i,t}(s;1-p)$ is the analogous increase in i's shirk-propensity, if i's current shirk-propensity is $1-p$, i shirks in t, and is satisfied. $\delta^-_{i,t}(v;p)$ and $\delta^-_{i,t}(s;1-p)$ denote the analogous decrements (their absolute values, i.e.) in vote-propensity and shirk-propensity, conditional on voting and being dissatisfied or on shirking and being dissatisfied, respectively. (We will suppress the t subscript whenever the date is sufficiently clear. At other points we will simplify the notation to, e.g., $\delta^+_{i,t}(v)$ when the value of the vote-propensity or shirk-propensity is sufficiently clear.)

Fact 1: If all agents use the same ABAR, which is monotonic and action-invariant, and $p_{i,t} < (>)\ \frac{1}{2}$ $\forall i$, then for all i and j (where i may equal j or not),

(a) $\delta^+_i(v;p) > (<)\ \delta^+_j(s;1-p)$, and

(b) $\delta^-_i(v;p) < (>)\ \delta^-_j(s;1-p)$.

(If $p_{i,t} \leq (\geq) \frac{1}{2}$ $\forall i$, then the conclusions in (a) and (b) change to weak inequalities.)

Proof: We prove the case of $p_{i,t} < \frac{1}{2}$; the proof for $p_{i,t} > \frac{1}{2}$ is completely analogous.

(a) Given $p < 1-p$, monotonicity implies that $\delta^+_i(v;p) > \delta^+_i(v;1-p)$. In turn the latter equals $\delta^+_j(v;1-p)$, since everyone uses the same ABAR. By action-invariance, $\delta^+_j(v;1-p) = \delta^+_j(s;1-p)$. Then transitivity implies $\delta^+_i(v;p) > \delta^+_j(s;1-p)$.

(b) Again, given $p < 1-p$, monotonicity implies $\delta^-_i(s;1-p) > \delta^-_i(s;p)$. In turn the latter equals $\delta^-_j(s;p)$, since the ABAR is common. By action-invariance, $\delta^-_j(s;p) = \delta^-(v;p)$. Then transitivity implies $\delta^-_i(s;1-p) > \delta^-(v;p)$. QED.

Fact 2: If all agents i use the same ABAR, which is monotonic and symmetric, and $p_{i,t} < (>)\ \frac{1}{2}$ $\forall i$, then for all i and j (where i may equal j or not) and $\forall q < \frac{1}{2}$,

(a) $\delta^+_i(v;p) > (<)\ \delta^-_j(v;q)$, and

(b) $\delta^-_i(s;1-p) > (<)\ \delta^+_j(s;1-q)$.

(If $p_{i,t} \leq (\geq) \frac{1}{2}$, the conclusions change to weak inequalities.)

Proof: We prove the case of $p_{i,t} < \frac{1}{2}$; the proof for $p_{i,t} > \frac{1}{2}$ is analogous.

(a) Since $p < \frac{1}{2}$, monotonicity implies that $\delta_i^+(v;p) > \delta_i^+(v;\frac{1}{2})$. Symmetry yields $\delta_i^+(v;\frac{1}{2}) = \delta_i^-(v;\frac{1}{2})$. Then, for all $q < \frac{1}{2}$, monotonicity implies $\delta_i^-(v;\frac{1}{2}) > \delta_i^-(v;q)$. Because everyone uses the same ABAR, $\delta_i^-(v;q) = \delta_j^-(v;q)$. Finally, transitivity gives $\delta_i^+(v;p) > \delta_j^-(v;q)$.

(b) Since $1-p > \frac{1}{2}$, monotonicity implies that $\delta_i^-(s;1-p) > \delta_i^-(s;\frac{1}{2})$. Symmetry yields $\delta_i^-(s;\frac{1}{2}) = \delta_i^+(s;\frac{1}{2})$. Then, because $q < \frac{1}{2}$, monotonicity implies $\delta_i^+(s;\frac{1}{2}) > \delta_i^+(s;1-q)$. Because everyone uses the same ABAR, $\delta_i^+(s;1-q) = \delta_j^+(s;1-q)$. Finally, transitivity gives $\delta_i^-(s;1-p) > \delta_j^+(s;1-q)$. QED.

Fact 3: If all agents use the same ABAR which is monotonic, action-invariant and symmetric, and $p_{i,t} < (>) \frac{1}{2}$, then $\delta_i^-(s;1-p) = \delta_j^+(v;p) > (<) \delta_j^-(v;p) = \delta_i^+(s;1-p)$.
(If $p_{i,t} \leq (\geq) \frac{1}{2}$, then the conclusion changes to a weak inequality.)

Proof: This is proved for the case of $p < \frac{1}{2}$; the proof for $p > \frac{1}{2}$ is analogous.

By the definition of symmetric adjustment, $\delta_i^-(s;1-p) = \delta_i^+(s;p)$. Action invariance implies $\delta_i^+(s;p) = \delta_i^+(v;p)$, so $\delta_i^-(s;1-p) = \delta_i^+(v;p)$. Similarly, symmetric adjustment and action-invariance imply that $\delta_i^-(v;p) = \delta_i^+(s;1-p)$. Because everyone uses the same ABAR, $\delta_i^+(v;p) = \delta_j^+(v;p)$ and $\delta_i^-(v;p) = \delta_j^-(v;p)$; hence $\delta_i^-(s;1-p) = \delta_j^+(v;p)$ and $\delta_i^-(v;p) = \delta_j^+(s;1-p)$. Finally, Fact 2 shows that $\delta_i^+(v;p) > \delta_j^-(v;q)$ for $p < \frac{1}{2}$, and since this held for $j = i$ as well, this implies $\delta_j^+(v;p) > \delta_j^-(v;q)$. QED.

The next fact drops the assumption that the ABAR responds symmetrically to failures and successes, assuming instead that it is *failure-sensitive*, which we now define.

Definition. Suppose that agents i and j use the same ABAR. In t, $p_{i,t} = 1 - p_{j,t}$ and both agents vote, but i gets negative feedback while j gets positive feedback. The ABAR is *failure-sensitive* if

$$p_{i,t} - p_{i,t+1} \geq p_{j,t+1} - p_{j,t}$$

for all $p_{i,t} \in [0,1]$, and the inequality is strict for some values of $p_{i,t}$. (When both agents shirk the definition is analogous.)

Fact 4: If all agents i use the same ABAR, which is monotonic and failure-sensitive, and $p_{i,t} < \frac{1}{2}$ $\forall i$, then for all i and j (where i may equal j or not) and $\forall q < \frac{1}{2}$, $\delta_i^-(s;1-p) > \delta_j^+(s;1-q)$.

(If $p_{i,t} \leq \frac{1}{2}$, the conclusion changes to a weak inequality.)

Proof: Since $p < \frac{1}{2}$, monotonicity implies that $d_i^-(s; 1-p) > d_i^-(s; \frac{1}{2})$. Failure-sensitivity gives $d_i^-(s; \frac{1}{2}) \geq d_i^+(s; \frac{1}{2})$. Since $1-q > \frac{1}{2}$, monotonicity yields $d_i^+(s; \frac{1}{2}) > d_i^+(s; 1-q)$. Because everyone uses the same ABAR, $d_i^+(s; 1-q) = d_j^+(s; 1-q)$, and transitivity completes the proof by supplying $d_i^-(s; 1-p) > d_j^+(s; 1-q)$. QED.

Proofs of Observations and Propositions

Observation 1: If in t the aspirations of everyone in the winning faction are strictly less than $b - c$ and the aspirations of everyone on the losing side are strictly positive, then all winners are satisfied by the outcome in t while all losers are dissatisfied.

Proof: This follows immediately from the defined payoffs (winners get either b or $b - c$, losers get either 0 or $-c$) and the assumption about aspirations.

Observation 2: Under the conditions of observation 1, if adjustment is by any arbitrary mix of Aspiration-Based Adaptive Rules (ABARs), then all winning voters and all losing shirkers become more disposed to vote (if that is possible, i.e., for $p_{i,t} < 1$) after the election in t. The other citizens become less inclined (if that is possible, i.e., for $p_{i,t} > 0$).

Proof: This follows immediately from observation 1 and the definition of ABARs.

Observation 3: If in t the Democrats win and $n_D \leq n_R$ or the Republicans win and $n_D \geq n_R$, then the number of winning voters plus losing shirkers is a majority of the electorate in t.

Proof: Suppose the Republicans win. Of course, there must be more winning voters than losing voters. And since the number of Democratic shirkers in t must equal n_D minus the number of Democratic voters, while the number of Republican shirkers equals n_R minus the number of Republican voters, if $n_D \geq n_R$ while Republican voters outnumber Democratic voters, then obviously there must be more Democratic shirkers than Republican ones. The proof of the other case is identical. QED.

Proposition 1: Suppose the conditions of observations 1, 2 and 3 hold. Further, $p_{i,t} < 1$ for all i. Then after the election in t a majority of citizens become more disposed to vote.

Proof: Suppose the Republicans win. Given the assumed aspirations, a Republican victory produces satisfied Republicans and dissatisfied Democrats. Hence, since everyone adjusts by some kind of ABAR, Republican voters and Democratic shirkers will increase their propensity to vote. Given that Republicans do not outnumber Democrats but Republicans won anyway, observation 3 applies: Republican voters plus Democratic shirkers must exceed $\frac{N}{2}$. Exactly the same logic holds for the other case. QED.

Proposition 2: Suppose that observations 1-3 hold. Further, everyone uses the same monotonic ABAR, which is either (i) symmetric or (ii) failure-sensitive and also there are in t at least as many losing shirkers as losing voters. If $p_{i,t} \leq \frac{1}{2}$ for all i then $\overline{p}_{t+1} > \overline{p}_t$.

Proof: The proof is divided into parts (i) and (ii), depending on whether the ABAR is symmetric or failure-sensitive.

(i) Consider first winning voters, who increase their participation-propensities, versus losing voters, who decrease theirs. We know that there are more of the former. Fact 2 shows that none of these per-capita increases can be less than any of the decreases. Hence the net effect of the changes of these two groups must be to increase the average propensity to vote.

Now compare the losing shirkers, who decrease their propensity to *shirk*, with winning shirkers, who increase theirs. The former are more numerous, given the assumptions of the proposition. Hence we only need to check that their per capita decreases are at least as big as the per capita increases of the winning shirkers. This is also ensured by Fact 2. QED.

(ii) It is convenient to divide this into two cases.

Case 1: there are exactly as many losing shirkers as losing voters.

Fact 1 implies that no losing voter decreases his vote-propensity more than any losing shirker decreases her shirk-propensity. And since there are as many of the latter as of the former, the total change between these two sets cannot lower the average propensity to vote.

Fact 1 also implies that no winning shirker can increase his shirk-propensity more than any winning voter increases her vote-propensity. And winning voters must outnumber winning shirkers, since observation 3 holds and the losers are exactly split between voters and shirkers. Thus the total change between these two sets must increase the average propensity to vote.

Case 2: losing shirkers outnumber losing voters.

For the losing voters who can be matched up with losing shirkers, the comparison is as in case 1.

Take the excess (of losing shirkers) and match them up with winning shirkers. Fact 4 implies that no winning shirker can increase her propensity to shirk more than any losing shirker decreases hers.

The comparison of winning voters with winning shirkers is as in case 1.

Finally, since observation 3 holds, once every losing shirker has been matched up with a losing voter, the sum of winning voters plus remaining (excess) losing shirkers must exceed the winning shirkers, so the result holds. QED.

Proposition 3: Suppose observations 1 and 3 hold, and everyone adjusts by the same symmetric and action-invariant Bush-Mosteller rule.

(i) If $\overline{p}_t \leq \frac{1}{2}$, then $\overline{p}_t < \overline{p}_{t+1}$.

(ii) If $\overline{p}_t \geq \frac{1}{2}$, then $\overline{p}_{t+1} > \frac{1}{2}$.

Proof: (i) For the usual reasons we know that more people will become more inclined to vote (or hold steady at a propensity of one) than will become less inclined (or hold steady at a propensity of zero). Hence it would suffice to show that the average propensity to vote doesn't fall if these two groups are exactly equal.

Since we are given that $\overline{p}_t = \frac{\sum_{k=1}^{N} p_{k,t}}{N} \leq \frac{1}{2}$, it follows that

$$\sum_{k=1}^{N} p_{k,t} \leq \sum_{k=1}^{N} (1 - p_{k,t})$$

or, breaking up the N citizens into two equal-size groups with arbitrary labels,

$$\sum_{j=1}^{\frac{N}{2}} p_{j,t} + \sum_{i=\frac{N}{2}+1}^{N} p_{i,t} \leq \sum_{j=1}^{\frac{N}{2}} (1 - p_{j,t}) + \sum_{i=\frac{N}{2}+1}^{N} (1 - p_{i,t})$$

Rearranging we get

$$\sum_{j=1}^{\frac{N}{2}} p_{j,t} + \sum_{j=1}^{\frac{N}{2}} (1 - p_{j,t}) = \frac{N}{2} = \sum_{i=\frac{N}{2}+1}^{N} p_{i,t} + \sum_{i=\frac{N}{2}+1}^{N} (1 - p_{i,t})$$

Since switching $\sum_{j=1}^{\frac{N}{2}} (1 - p_{j,t})$ for $\sum_{i=\frac{N}{2}+1}^{N} p_{i,t}$ produced a strict equality out of what had been a weak inequality, it must be true that

$$\sum_{j=1}^{\frac{N}{2}} (1 - p_{j,t}) \geq \sum_{i=\frac{N}{2}+1}^{N} p_{i,t}$$

But then it immediately follows that

$$\sum_{j=1}^{\frac{N}{2}} p_{j,t} \leq \sum_{i=\frac{N}{2}+1}^{N} (1 - p_{i,t})$$

Since this was an arbitrary partition of the N citizens, it follows that $\sum_{k=1}^{N} p_{k,t} \leq \sum_{k=1}^{N}(1-p_{k,t})$ implies that for any arbitrary break-down of the N citizens into two groups of $\frac{N}{2}$ decreasers and $\frac{N}{2}$ increasers, $\sum_{j=1}^{\frac{N}{2}} p_{j,t} \leq \sum_{i=\frac{N}{2}+1}^{N}(1 - p_{i,t})$. But this implies that $\sum_{j=1}^{\frac{N}{2}} \alpha p_{j,t} \leq \sum_{i=\frac{N}{2}+1}^{N} \alpha(1 - p_{i,t})$, i.e., the total decrease in participation-propensity cannot exceed the total increase. Since this holds for the case of equally many decreasers as increasers, it follows that if the latter outnumber the former, $\overline{p}_{t+1} > \overline{p}_t$. QED.

(ii) Let $i = 1, \ldots, m$ index the citizens who after the election in t will become more disposed to participate (or remain at $p_{i,t} = 1$) and let $j = m+1, \ldots, N$ denote the citizens who will become less disposed (or remain at $p_{j,t} = 0$). Since the election is conclusive, $m > \frac{N}{2}$.

Since adjustment is by symmetric Bush-Mosteller, it is easy to show that the vote-propensities of both winning voters and losing shirkers adjust in the same way: in each case $p_{i,t+1} = p_{i,t} + \alpha(1 - p_{i,t})$. Similarly, the propensity adjustment for both winning shirkers and losing voters is $p_{j,t+1} = p_{j,t} - \alpha \cdot p_{j,t}$. Hence

$$\begin{aligned}
\overline{p}_{t+1} &= \frac{1}{N}\left\{\sum_{i=1}^{m} p_{i,t} + \alpha(1 - p_{i,t}) + \sum_{j=m+1}^{N} p_{j,t} - \alpha \cdot p_{j,t}\right\} \\
&= \frac{1}{N}\left\{\sum_{i=1}^{m} p_{i,t} + \sum_{j=m+1}^{N} p_{j,t}\right\} + \frac{\alpha}{N}\left\{\sum_{i=1}^{m}(1 - p_{i,t}) - \sum_{j=m+1}^{N} p_{j,t}\right\} \\
&= \overline{p}_t + \frac{\alpha}{N}\left\{\sum_{i=1}^{m}(1 - p_{i,t}) - \sum_{j=m+1}^{N} p_{j,t}\right\} \\
&= \overline{p}_t + \frac{\alpha}{N}\left\{\sum_{i=1}^{m} 1 - \sum_{i=1}^{m} p_{i,t} - \sum_{j=m+1}^{N} p_{j,t}\right\} \\
&= \overline{p}_t + \frac{\alpha}{N} \cdot m - \alpha\overline{p}_t \\
&= (1 - \alpha)\overline{p}_t + \alpha\frac{m}{N} > \frac{1}{2}
\end{aligned}$$

since by assumption $\overline{p}_t \geq \frac{1}{2}$ and $\frac{m}{N}$ must exceed $\frac{1}{2}$. QED.

Corollary 1: Under the assumptions of propositions 2 or 3, expected turnout in $t + 1$ exceeds the expected turnout in t.

Proof: Under both propositions 2 and 3, $\overline{p}_{t+1} > \overline{p}_t$. This implies that $\sum_{i=1}^N p_{i,t+1} > \sum_{i=1}^N p_{i,t}$.

Now expected turnout in period t equals $\sum_{i=1}^N (p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon)$. But since $\epsilon < \frac{1}{2}$, we have $1 - \epsilon > \epsilon$, so $\sum_{i=1}^N (p_{i,t+1}(1-\epsilon) + (1-p_{i,t+1})\epsilon) > \sum_{i=1}^N (p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon)$ by simple algebra. QED.

Proposition 4: Suppose for all i, $p_{i,t} > 0$ and $a_{i,t}$ is realistic. The election is conclusive in t. People adjust by any arbitrary mix of ABARs. If in t the winning voters' aspirations are strictly bigger than $b - c$ and the losers' aspirations are strictly negative, then $p_{i,t+1} < p_{i,t}$ for all i.

Proof: Suppose the Democrats win. Given the realized payoffs and the assumed aspiration levels, Democratic shirkers are satisfied but Democratic voters aren't. Similarly, Republican shirkers are content but Republican voters aren't. By the definition of a ABAR, satisfied shirkers increase their propensity to shirk (if possible) and dissatisfied voters decrease their propensity to vote (if possible). QED.

Definition: The distribution of vote-propensities in district A is *stochastically bigger* than that of district B if for some t, for all $n = 1, \ldots, N$, and for all $x \in [0,1]$, $\Pr(|\{i|p_{i,t} \geq x\}| \geq n$ in A) $\geq \Pr(|\{j|p_{j,t} \geq x\}| \geq n$ in B), and this inequality holds strictly for at least one value of n and at least one value of x.

The distribution of turnout probabilities is defined analogously.

Corollary 2: Suppose the conditions of proposition 4 hold. Then

(i) the distribution of propensities to vote in $t+1$ is strictly stochastically smaller than the distribution in t;

(ii) the distribution of turnout probabilities in $t+1$ is strictly stochastically smaller than the distribution in t.

Proof: (i) Given proposition 4, *every* citizen will decrease his/her propensity to vote, which immediately implies a stochastically smaller distribution of propensities.

(ii) For any citizen i, $\Pr(i$ votes in $t) = p_{i,t} \cdot (1-\epsilon) + (1-p_{i,t}) \cdot \epsilon$. This is obviously increasing in $p_{i,t}$, since $1 - \epsilon > \frac{1}{2}$. Thus, given that $p_{i,t+1} < p_{i,t}$ for all i, every individual is less likely to turn out in $t+1$, so the distribution of turnout is stochastically smaller. QED.

Proposition 5: Suppose everyone adapts via the same monotonic and action-invariant ABAR. Further, the ABAR is either (i) symmetric or (ii) failure-sensitive and also there are in t at least as many losing voters as losing shirkers. All aspirations are realistic in t. Together, winning voters whose aspirations are less than or equal to $b - c$ and losing shirkers whose aspirations are at least 0 are a minority in t. If $p_{i,t} \geq \frac{1}{2}$ for all i then $\overline{p}_{t+1} < \overline{p}_t$.

Proof: (i) Since all aspirations are realistic, the only people in t who will become more disposed to vote are winning voters for whom aspirations don't exceed $b - c$ and losing shirkers with non-negative aspirations. These are a minority, by assumption. Everyone else becomes less inclined to vote. (A technicality: note that the proposition assumes that winning voters with aspirations less than *or equal to* $b - c$ and losing shirkers with aspirations greater than *or equal to* 0 form a minority. Hence the aspirations of the remaining winning voters must strictly exceed $b - c$, so by the definition of an ABAR they must strictly decrease their vote-propensity; similarly, the aspirations of the remaining losing shirkers must be strictly less than zero, so they must strictly increase their propensity to shirk.)

Given these demographic facts, it suffices to show that all those decreasing their participation-propensities must reduce their propensity at least much as any increaser raises theirs. This is established by Fact 3.

(ii) This proof parallels the proof of part (ii) of proposition 2. In case 1 we assume that there are equally many losing voters as losing shirkers, and compare their per capita propensity changes. Then we compare the per capita propensity changes of winning voters and winning shirkers. All the per capita effects weakly reduce the average vote-propensity. The demographics (i.e., the assumption that the satisfied winners plus the dissatisfied losers are a minority) then tilt the net effect toward a loss in the average propensity to vote. In case 2 we assume that losing voters outnumber the losing shirkers, and then we use the excess losing voters in a manner analogous to how the excess losing shirkers were deployed in part (ii) of proposition 2. QED.

Proposition 6: Suppose everyone uses the same symmetric and action-invariant Bush-Mosteller rule. All aspirations are realistic in t. Together, winning voters whose aspirations are less than or equal to $b - c$ and losing shirkers whose aspirations are at least zero are a minority of the electorate in t. If $\overline{p}_t \geq \frac{1}{2}$ then $\overline{p}_{t+1} < \overline{p}_t$.

Proof: This proof parallels the proof of proposition 3. The latter's two key assumptions are reversed: here, there are more decreasers than increasers, and $\overline{p}_t \geq \frac{1}{2}$. The latter implies, by a simple reapplication of proposition 3's proof, that even if there were only equally many decreasers as increasers the total participation-propensity increase could not exceed the total propensity decrease.

Since there are actually more decreasers than increasers, the average propensity to vote must fall. QED.

Corollary 3: Assume the conditions of proposition 6 hold, except that $\overline{p}_t \leq \frac{1}{2}$. Then $\overline{p}_{t+1} < \frac{1}{2}$.

Proof: This parallels the proof of proposition 3, part (ii). In particular, re-examine the last line of that proof:

$$\overline{p}_{t+1} = (1-\alpha)\overline{p}_t + \alpha\frac{m}{N}$$

Here by assumption $m < \frac{N}{2}$, and also $\overline{p}_t \leq \frac{1}{2}$. Hence it is clear that $\overline{p}_{t+1} < \frac{1}{2}$. QED.

Proposition 7: Suppose everyone uses the same monotonic and action-invariant ABAR.

(i) If $a_{i,t} > b$ and $p_{i,t} > \frac{1}{2}$ $\forall i$, then $\overline{p}_t > E[\overline{p}_{t+1}]$.

(ii) If in addition a majority of citizens vote in t, then $\overline{p}_t > \overline{p}_{t+1}$.

Proof: (i) Let $\delta^-_{i,t}(v)$ equal the decrement in i's propensity to vote, given that i voted in t and was dissatisfied. Similarly, $\delta^-_{i,t}(s)$ is the decrement in i's propensity to shirk, given that i shirked and was dissatisfied. Then for any i,

$$\begin{aligned}
E[p_{i,t+1}] &= [p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon][p_{i,t} - \delta^-_{i,t}(v)] \\
&\quad + [(1-p_{i,t})(1-\epsilon) + p_{i,t}\cdot\epsilon]\Big[1 - [(1-p_{i,t}) - \delta^-_{i,t}(s)]\Big] \\
&= [p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon][p_{i,t} - \delta^-_{i,t}(v)] + [(1-p_{i,t})(1-\epsilon) + p_{i,t}\cdot\epsilon][p_{i,t} + \delta^-_{i,t}(s)] \\
&= p_{i,t} + [p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon](-\delta^-_{i,t}(v)) + [(1-p_{i,t})(1-\epsilon) + p_{i,t}\cdot\epsilon](\delta^-_{i,t}(s)).
\end{aligned}$$

Since $p_{i,t} > (1-p_{i,t})$ and $1-\epsilon > \epsilon$, $p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon > (1-p_{i,t})(1-\epsilon) + p_{i,t}\cdot\epsilon$. Further, since $p_{i,t} > 1 - p_{i,t}$ and the ABAR is monotonic and action-invariant, it follows from Fact 1 that $\delta^-_{i,t}(v) \geq \delta^-_{i,t}(s)$. Hence $[p_{i,t}(1-\epsilon) + (1-p_{i,t})\epsilon](-\delta^-_{i,t}(v)) + [(1-p_{i,t})(1-\epsilon) + p_{i,t}\cdot\epsilon](\delta^-_{i,t}(s)) < 0$, so $E[p_{i,t+1}] < p_{i,t}$. Since this inequality holds for every i, it follows immediately that $E[\overline{p}_{t+1}] < \overline{p}_t$.

(ii) Since all aspirations are unrealistically high and a majority of people voted, a majority of people must decrease their propensity to vote. And since $p_{i,t} > \frac{1}{2}$, by Fact 1 each person decreasing their vote propensity does so by at least as much as people decreasing their shirk propensity; i.e., $\delta^-_{i,t}(v) \geq \delta^-_{i,t}(s)$. Hence the average propensity to vote must fall. QED.

Proposition 8: Districts A and B are parametrically identical except that in A, $b = 0$, while in B, $b > -c$. In both, $-c > 0$. The initial distributions of vote-propensities are the same in the two districts. Everyone begins with realistic aspirations. Everyone adjusts propensities by the same ABAR, which permits no leapfrogging. Aspirations are governed by any payoff-responsive dynamic. Then for all $t > 1$, A's distribution of vote-propensities is stochastically bigger than B's.

Proof: Consider first district A. Since b equals zero in A, there are only two possible payoffs, $-c$ for voting and zero for shirking. Since the costs of voting are negative, $-c > 0$. Since all aspirations start off realistic, aspirations in district A must be in $(0, -c)$. By payoff-responsiveness, they must always remain in that interval. Consequently voting, in period one and thereafter, is always reinforced and shirking is always inhibited, regardless of the electoral outcome. Hence, since everyone is using an ABAR, with probability one *everyone's* propensity in A rises in every period, unless/until they reach 1.0, where they'll stay forever.

Recall that the initial distributions of propensities in A and B are the same. Focus then on a pair of citizens, i from A and j from B, who start out with the same vote-propensity: $p_{i,0} = p_{j,0}$. Since everyone is using the same ABAR, which precludes leapfrogging, and since i increases his vote-propensity (whenever possible), it follows that $p_{j,t}$ can never exceed $p_{i,t}$, in any date t. Note that everyone can be paired up in this manner, because the districts' initial propensity-distributions are identical. Hence for *every* i in A and j in B, if $p_{i,0} = p_{j,0}$ then $p_{i,t} \geq p_{j,t}$ for all t. Further, by transitivity and the no-leapfrogging property, if $p_{i,0} \geq p_{j,0}$ then $p_{i,t} \geq p_{j,t}$ for all t. This immediately implies that A's distribution of propensities weakly stochastically dominates B's, i.e., after $t = 1$, for $n = 1, \ldots, N$, Pr(at least n citizens in A have $p_{i,t}$'s $\geq x$) $\geq$ Pr(at least n citizens in B have $p_{j,t}$'s $\geq x$). So to show that this inequality must hold strictly for at least one n and one x, it suffices to establish that in every period, with positive probability at least one B *decreases* his propensity to vote (or, if that is impossible, that at least one B does not increase his vote-propensity).

Now consider two possible cases concerning the distribution of initial aspiration-values for the residents of B: either there exists an i such that $a_{i,0} > -c$ or no such i exists (i.e., $a_0 \leq -c$ for all i).

Case 1: $a_{i,0} > -c$ for at least one i.

Since $\frac{1}{2} > \epsilon > 0$, all outcomes occur with positive probability in every period, regardless of the distribution of propensities. For example, it is possible that in the first period the Democrats lose and some Democrats vote. Without loss of generality, then, suppose that there is one Democrat, i, whose aspiration exceeds $-c$, and that in period one i votes but the Democrats lose. The losing voters get $-c$ which in this case is less than $a_{i,0}$, so i reduces her propensity to vote. Moreover, at the end of period i's aspiration (and those of all winning voters) must exceed $-c$, by payoff-responsiveness. Hence on this sample path the assumption defining the case self-replicates, so the same argument can be re-applied to show that with positive probability at least one voter in the next period will also become less disposed to participate. And so on.

Case 2: $a_{i,0} \leq -c$ for all i.

With positive probability one side (say the Republicans) wins in period one even though some Republicans shirk. These shirkers get b which exceeds $-c \geq 0$. Hence they're satisfied with shirking, so their propensity to vote falls. Moreover, at the end of period one the aspirations of Democrats must still be less than or equal to $-c$, so in the next period repeat the argument, with the Democrats winning and some Democrats shirking. If someone's aspiration rises above $-c$ then switch to case 1. And so on.

Thus, since either case 1 or case 2 must hold, we can always insure, with positive probability, that at least one citizen in B decreases her vote-propensity in every period.

This in turn ensures that after $t = 1$, for $n = 1, \ldots, N$, the inequality of Pr(at least n citizens in A have $p_{i,t}$'s $\geq x$) $\geq$ Pr(at least n citizens in B have $p_{j,t}$'s $\geq x$) must hold strictly for at least one n and at least one x in $[0, 1]$. QED.

Definition. A random shock θ *thins out* if there exists a sequence of probability distributions $F_n(\theta)$ such that, for any $\epsilon > 0$ and $a < b$, there exists an n^* such that $F(b) - F(a) < \epsilon$ for all $n \geq n^*$.

In the last result, let $p_{m,t}$ denote the vote-propensity of a minority faction member in t.

Proposition 9: Suppose the minority faction loses in t. All their aspirations in t are in $(-c, 0]$; all their vote propensities are strictly less than $\frac{1}{2}$. Players can use any ABAR that is monotonic, action-invariant, and either symmetric or failure-sensitive. The random payoff shock $\theta_{i,t}$ is i.i.d. across agents, it is continuously distributed around zero with $\Pr(\theta < 0) = \Pr(\theta > 0)$, and it thins out. If $\theta_{i,t}$ is sufficiently thin then $E[p_{m,t+1}] > p_{m,t}$ for every minority citizen.

Proof: Since by assumption $\theta_{i,t}$ thins out, we can make the probability over any finite interval as small as we like. In particular, $\Pr(-c \leq \theta_{i,t} \leq 0)$ (or equivalently, $\Pr(0 \leq \theta_{i,t} \leq c)$) can be made arbitrarily small, thus driving $\Pr(\theta_{i,t} < -c)$ (or equivalently, $\Pr(\theta_{i,t} > c)$) up arbitrarily closely to $\frac{1}{2}$. Hence we can ensure that the conditional probability that any shirking minority citizen is dissatisfied, or the probability that any voting minority person satisfied, is arbitrarily close to one half.

Consider, therefore, the expected vote-propensity of a specific minority faction member, m, at the end of period t. (For convenience we let $v_{m,t}$ denote the probability that m votes in this period, which is $[p_{m,t}(1-\epsilon)+(1-p_{m,t})\epsilon]$. Further, we use $\delta^+_{m,t}(v)$ to denote the increment in m's vote-propensity if m votes in t and is satisfied. $\delta^+_{m,t}(s)$ is the analogous increase in m's shirk-propensity, if m shirks in t and is satisfied; $\delta^-_{m,t}(v)$ and $\delta^-_{m,t}(s)$ are the analogous decrements in vote- and shirk-propensity, respectively.)

$$\begin{aligned} E[p_{m,t+1}] &= p_{m,t} + v_{m,t}\Pr(-c+\theta_{i,t} \geq a_{m,t})\delta^+_{m,t}(v) + (1-v_{m,t})\Pr(\theta_{i,t} < a_{m,t})\delta^-_{m,t}(s) \\ &\quad + v_{m,t}\Pr(-c+\theta_{i,t} < a_{m,t})(-\delta^-_{m,t}(v)) + (1-v_{m,t})\Pr(\theta_{i,t} \geq a_{m,t})(-\delta^+_{m,t}(s)) \end{aligned}$$

So it would suffice to establish that

$$\begin{aligned} &v_{m,t}\Pr(-c+\theta_{i,t} \geq a_{m,t})\delta^+_{m,t}(v) + (1-v_{m,t})\Pr(\theta_{i,t} < a_{m,t})\delta^-_{m,t}(s) > \\ &v_{m,t}\Pr(-c+\theta_{i,t} < a_{m,t})\delta^-_{m,t}(v) + (1-v_{m,t})\Pr(\theta_{i,t} \geq a_{m,t})\delta^+_{m,t}(s) \end{aligned}$$

Case 1: m's ABAR is symmetric.

Fact 3 tells us that if the ABAR is monotonic, action-invariant and symmetric, then $p_{m,t} < \frac{1}{2}$ implies that $\delta^+_{m,t}(v) = \delta^s_{m,t}(s) > \delta^-_{m,t}(v) = \delta^+_{m,t}(s)$. Hence in this case we need only show that

$$\begin{aligned} &[v_{m,t}\Pr(-c+\theta_{i,t} \geq a_{m,t}) + (1-v_{m,t})\Pr(\theta_{i,t} < a_{m,t})]\delta^+_{m,t}(v) > \\ &[v_{m,t}\Pr(-c+\theta_{i,t} < a_{m,t}) + (1-v_{m,t})\Pr(\theta_{i,t} \geq a_{m,t})]\delta^-_{m,t}(v) \end{aligned}$$

Since $\delta^+_{m,t}(v)$ is strictly bigger than $\delta^-_{m,t}(v)$, it suffices to show that $v_{m,t}\Pr(-c+\theta_{i,t} \geq a_{m,t}) + (1-v_{m,t})\Pr(\theta_{i,t} < a_{m,t})$ can get sufficiently close to $v_{m,t}\Pr(-c+\theta_{i,t} < a_{m,t}) + (1-v_{m,t})\Pr(\theta_{i,t} \geq a_{m,t})$. This, in turn, is ensured by the thinning out property of θ, since all four probability terms converge to $\frac{1}{2}$ as θ becomes arbitrarily thin.

Case 2: m's ABAR is failure-sensitive.

Fact 1 tells us that since the ABAR is monotonic and action-invariant, with $p_{m,t} < .5$, $\delta^+_{m,t}(v) > \delta^+_{m,t}(s)$ and $\delta^-_{m,t}(s) > \delta^-_{m,t}(v)$. Further, since the ABAR is also failure-sensitive, $\delta^-_{m,t}(s) > \delta^+_{m,t}(s)$ by Fact 4. Thus the only relevant pair that is indeterminate is that of $\delta^+_{m,t}(v)$ and $\delta^-_{m,t}(v)$.

If $\delta^+_{m,t}(v) \geq \delta^-_{m,t}(v)$, then a trivial extension of case 1 would settle matters. So we only need to consider the case of $\delta^+_{m,t}(v) < \delta^-_{m,t}(v)$.

Given this, we obtain a complete ordering of all four propensity changes: $\delta^-_{m,t}(s) > \delta^-_{m,t}(v) > \delta^+_{m,t}(v) > \delta^+_{m,t}(s)$.

Since both $p_{m,t} < \frac{1}{2}$ and $\epsilon < \frac{1}{2}$, $v_{m,t} < \frac{1}{2}$ as well. This, combined with the above rank-ordering of the four propensity changes, yields $v_{m,t}\delta^+_{m,t}(v)+(1-v_{m,t})\delta^-_{m,t}(s) > v_{m,t}\delta^-_{m,t}(v)+(1-v_{m,t})\delta^+_{m,t}(s)$.

Again, since this inequality is strict, by making $\theta_{i,t}$ sufficiently thin we can drive $\Pr(-c+\theta_{i,t} \geq a_{m,t})$ and $\Pr(\theta_{i,t} < a_{m,t})$ sufficiently close to $\Pr(-c+\theta_{i,t} < a_{m,t})$ and $\Pr(\theta_{i,t} \geq a_{m,t})$ to ensure that $v_{m,t}\Pr(-c+\theta_{i,t} \geq a_{m,t})\delta^+_{m,t}(v)+(1-v_{m,t})\Pr(\theta_{i,t} < a_{m,t})\delta^-_{m,t}(s) > v_{m,t}\Pr(-c+\theta_{i,t} < a_{m,t})\delta^-_{m,t}(v)+(1-v_{m,t})\Pr(\theta_{i,t} \geq a_{m,t})\delta^+_{m,t}(s)$. QED.

REFERENCES

Bendor, Jonathan, Daniel Diermeier, and Michael Ting. 1999. "Adaptively Rational Strategic Behavior with Endogenous Aspirations." Unpublished manuscript. GSB: Stanford University.

Bendor, Jonathan, Dilip Mookherjee, and Debraj Ray. 1998. "Reinforcement Behavior in Repeated Games." Unpublished manuscript. Department of Economics: Boston University.

Bush, Robert, and Frederick Mosteller. 1955. *Stochastic Models of Learning.* New York: Wiley.

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

Ferejohn, John, and Morris Fiorina. 1974. "The Paradox of Not Voting." *American Political Science Review* 68:525-36.

Fiorina, Morris. 1990. "Information and Rationality in Elections." In John Ferejohn and James Kuklinski (eds.), *Information and Democratic Processes.* Urbana: University of Illinois Press.

Hansen, Steven, Thomas Palfrey, and Howard Rosenthal. 1987. "The Relationship between Constituency Size and Turnout: Using Game Theory to Estimate the Cost of Voting." *Public Choice* 52:15-33.

Ito, Tiffany, and John Cacioppo. 1999. "The Psychophysiology of Utility Appraisals." In Daniel Kahneman, Ed Diener, and Norbert Schwarz (eds.), *Well-Being: The Foundations of Hedonic Psychology.* New York: Russell Sage Foundation.

Kahneman, Daniel. 1999. "Objective Happiness." In Kahneman et al., *Well-Being.*

Kanazawa, Satoshi. 1998. "A Possible Solution to the Paradox of Voter Turnout." *Journal of Politics* 60:974-95.

Kollman, Ken, John Miller, and Scott Page. 1992. "Adaptive Parties in Spatial Elections." *American Political Science Review* 86:929-38.

Kollman, Ken, John Miller, and Scott Page. 1998. "Political Parties and Electoral Landscapes." *British Journal of Political Science* 28:139-58.

Laudan, Rachel, Larry Laudan, and Arthur Donovan. 1988. "Testing Theories of Scientific Change" in A. Donovan et al. (eds.), *Scrutinizing Science.* Dordrecht: Kluwer Academic.

March, James. 1994. *A Primer on Decision Making: How Decision Happen.* New York: Free Press.

Myerson, Roger. 1994 [1997]. "Population Uncertainty and Poisson Games." CMSEMS Discussion Paper No. 1120R. Northwestern University.

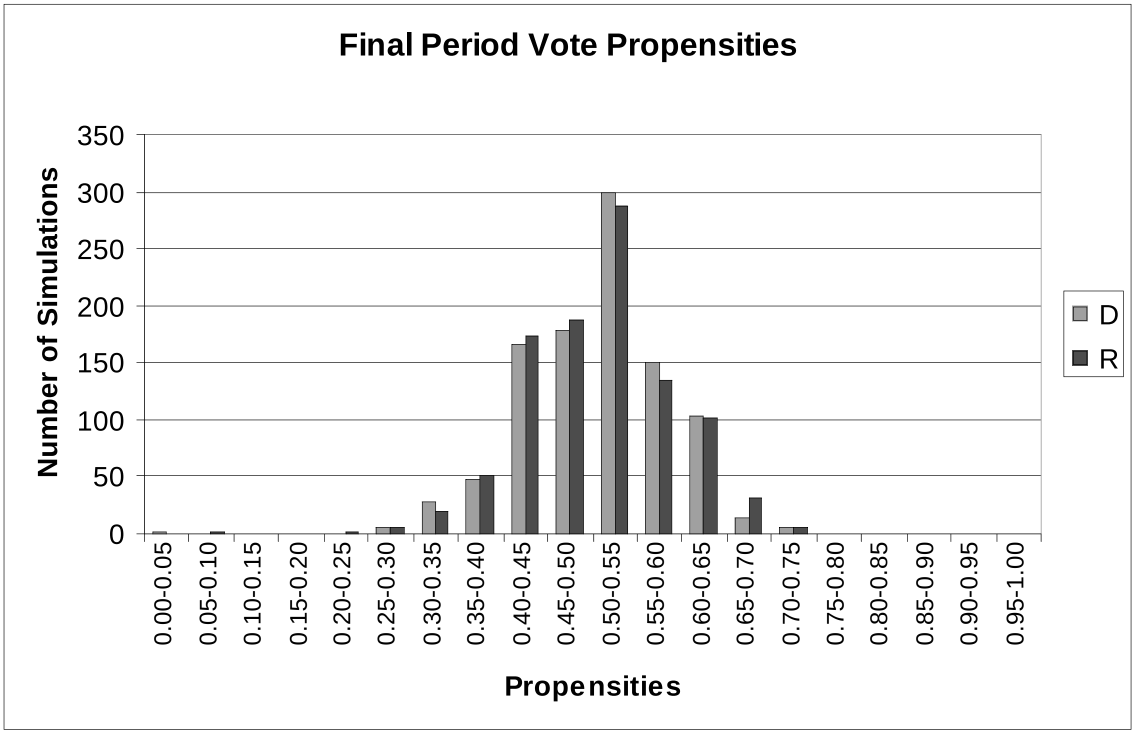
Myerson, Roger. 1996. "Economic Analysis of Political Institutions: An Introduction." CMSEMS Discussion Paper No. 1155. Northwestern University.

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

Palfrey, Thomas, and Howard Rosenthal. 1985. "Voter participation and Strategic Uncertainty." *American Political Science Review* 79:62-78.

Riker, William, and Peter Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62:25-42.

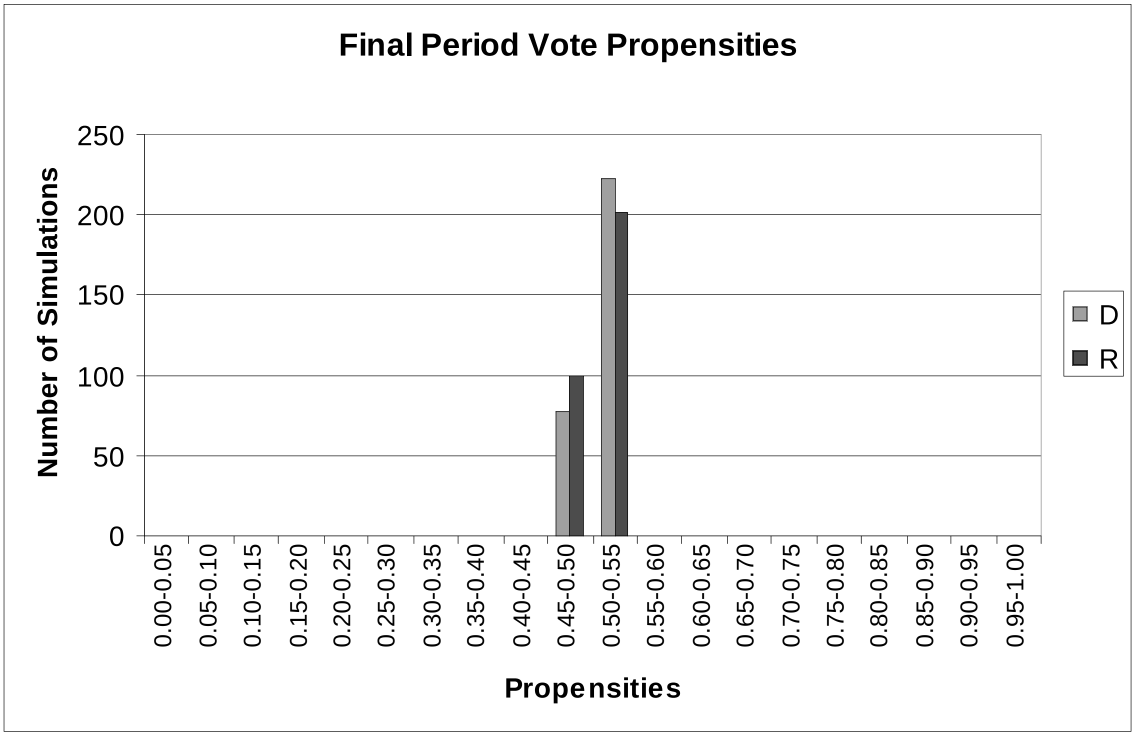
Run 1: Positive Turnout



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	50	50
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

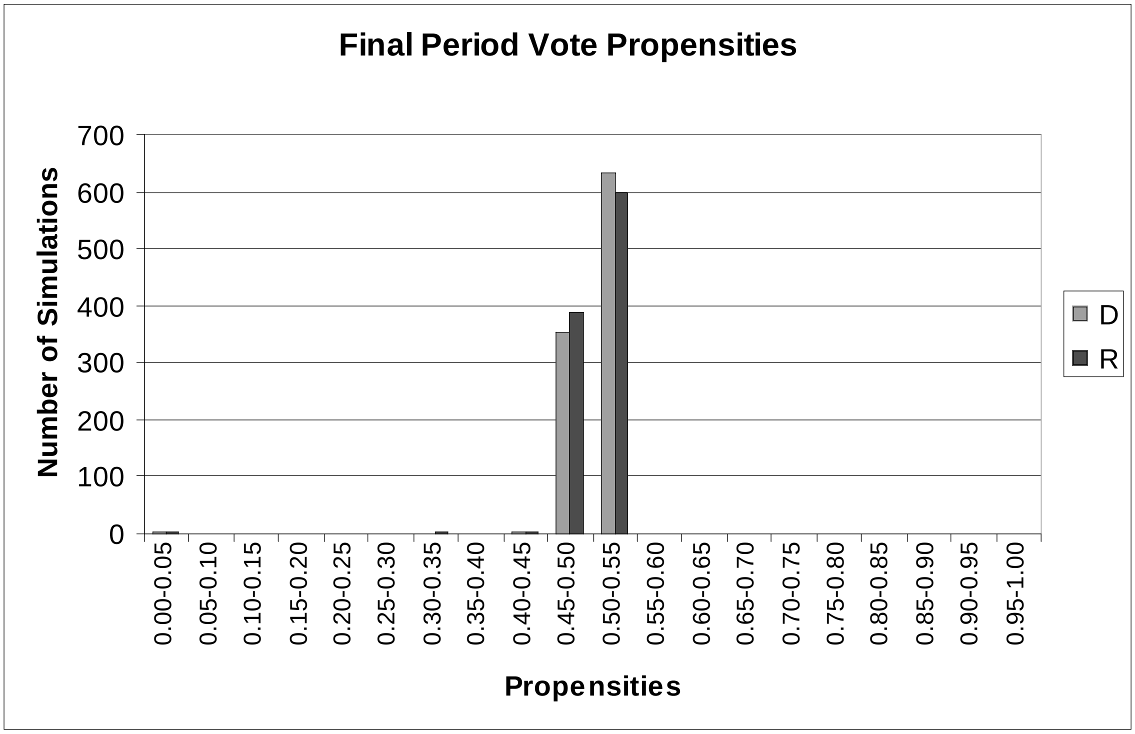
Run 2: Large Electorates



Starting Values: 1,000 Periods
300 Simulations

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	500,000	500,000
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

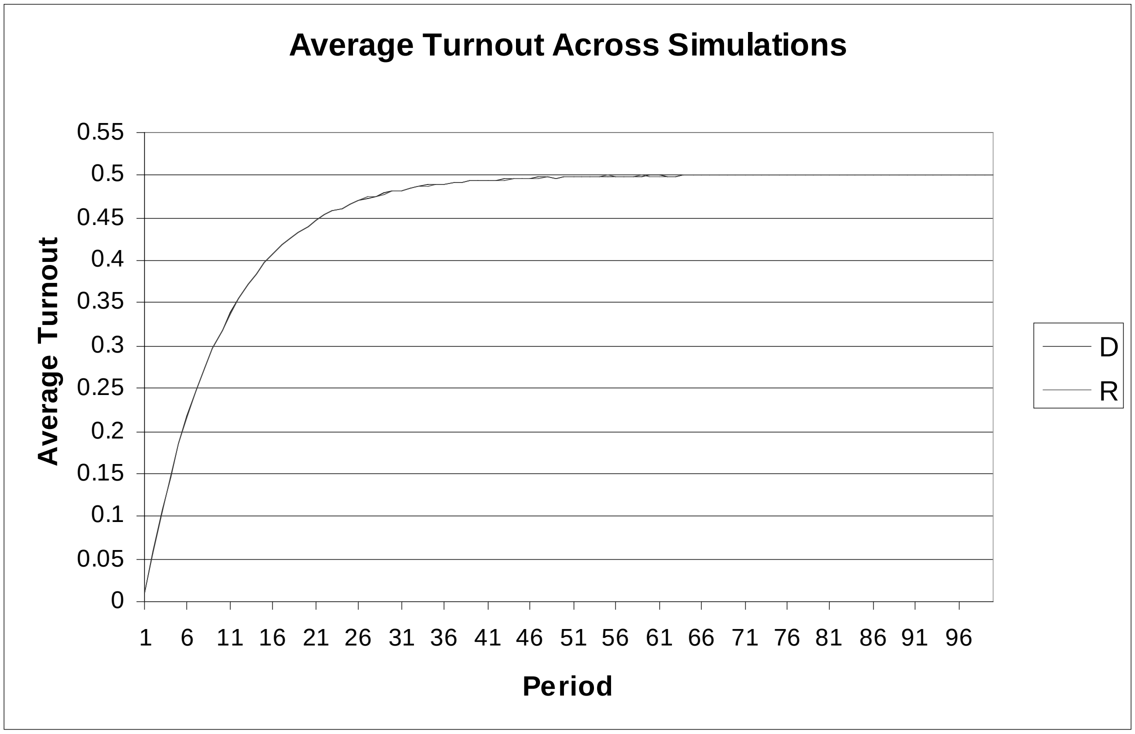
Run 3a: Low Initial Propensity - Breakout of Participation



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0	0

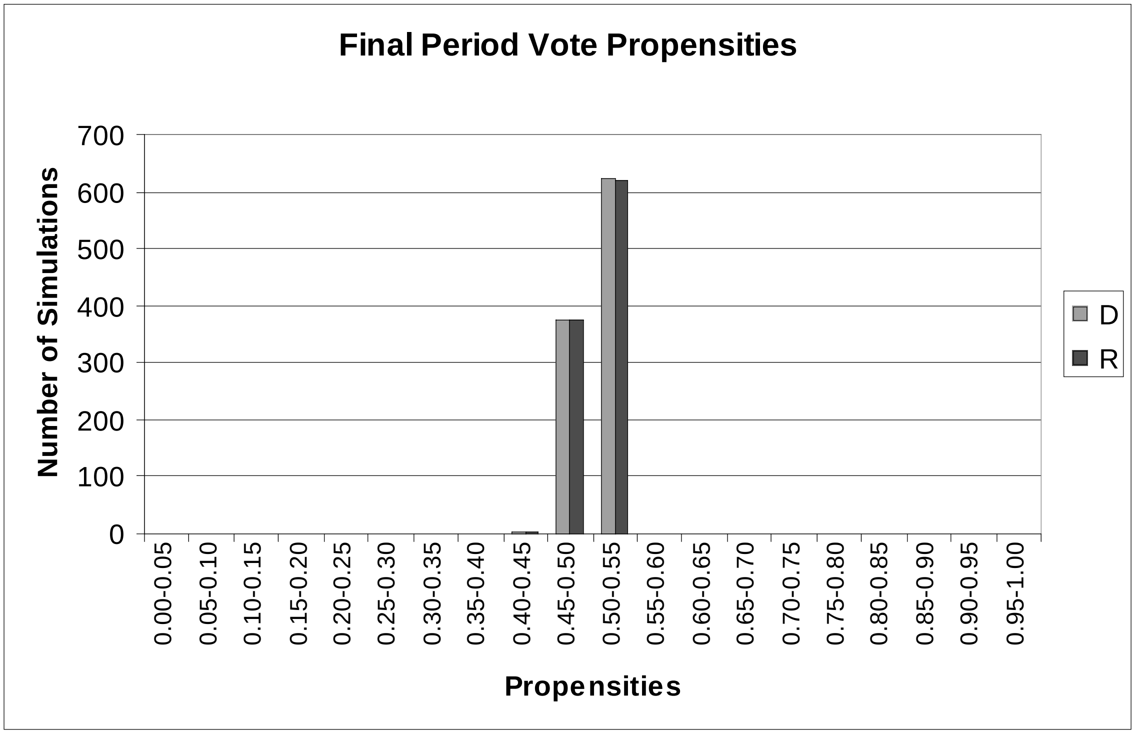
Run 3b: Breakout of Participation - Time Series



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0	0

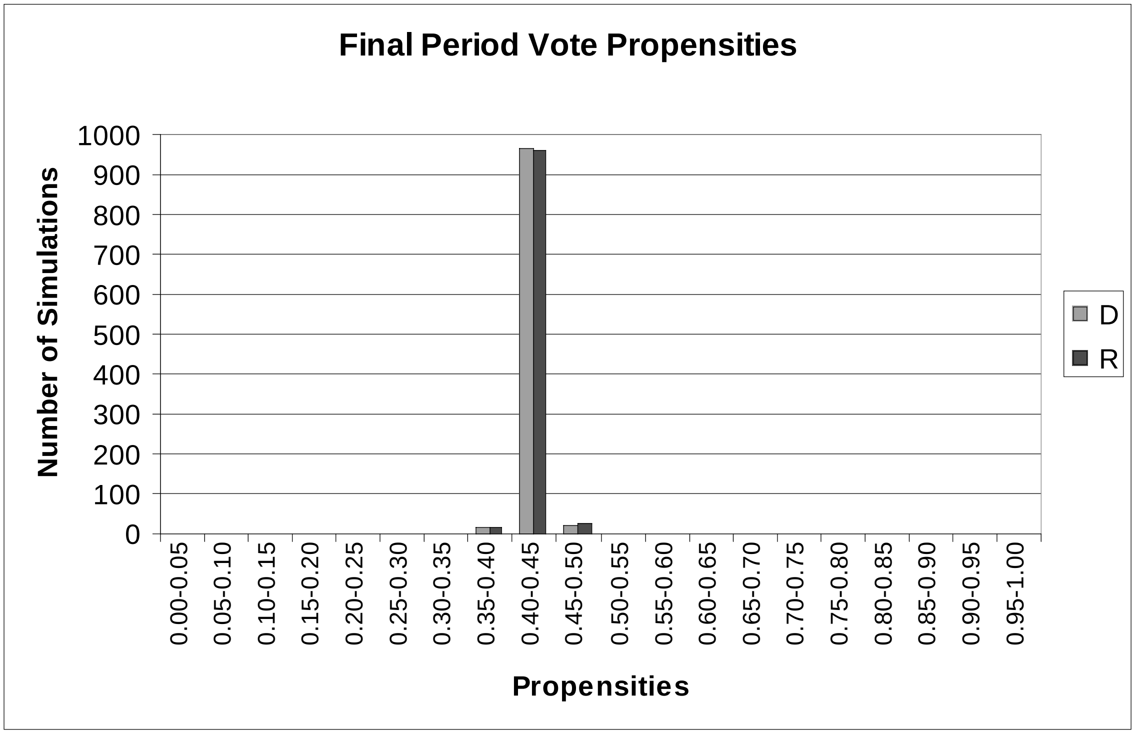
Run 3c: High Initial Propensity



Starting Values: 2,500 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	1.0	1.0
c	0.25	0.25
Aspirations	0.1	0.1
Vote Propensities	0.9	0.9

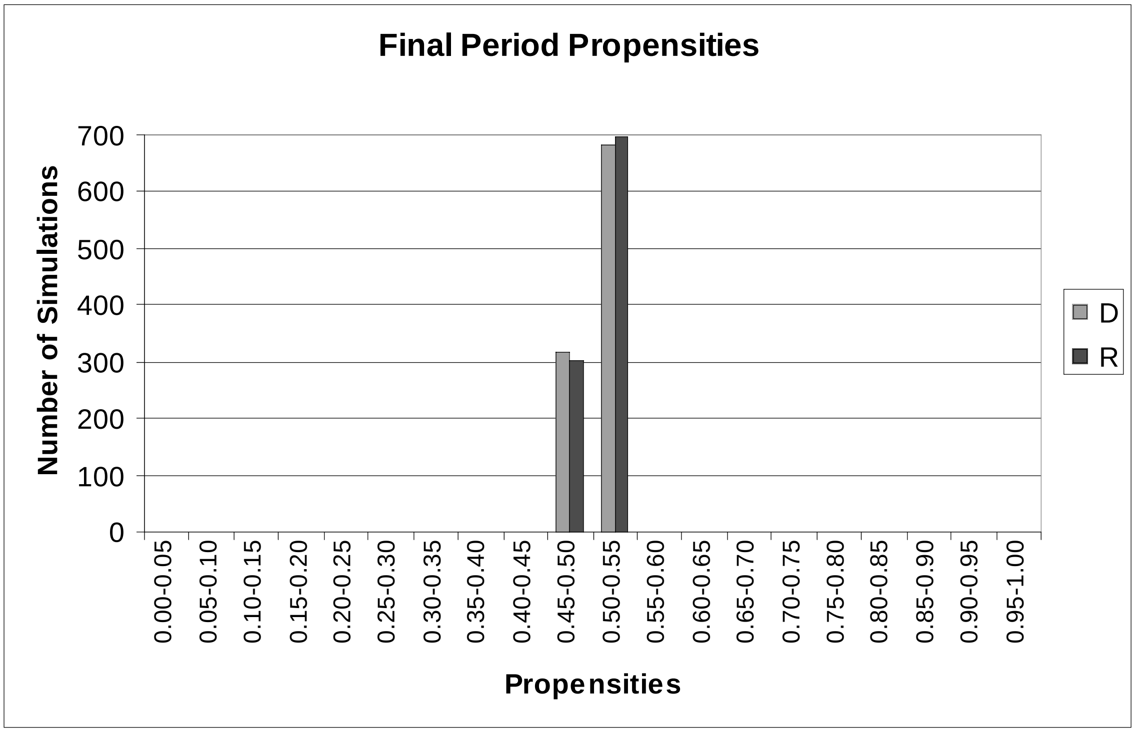
Run 4a: High Voting Costs



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	1.0	1.0
c	0.8	0.8
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

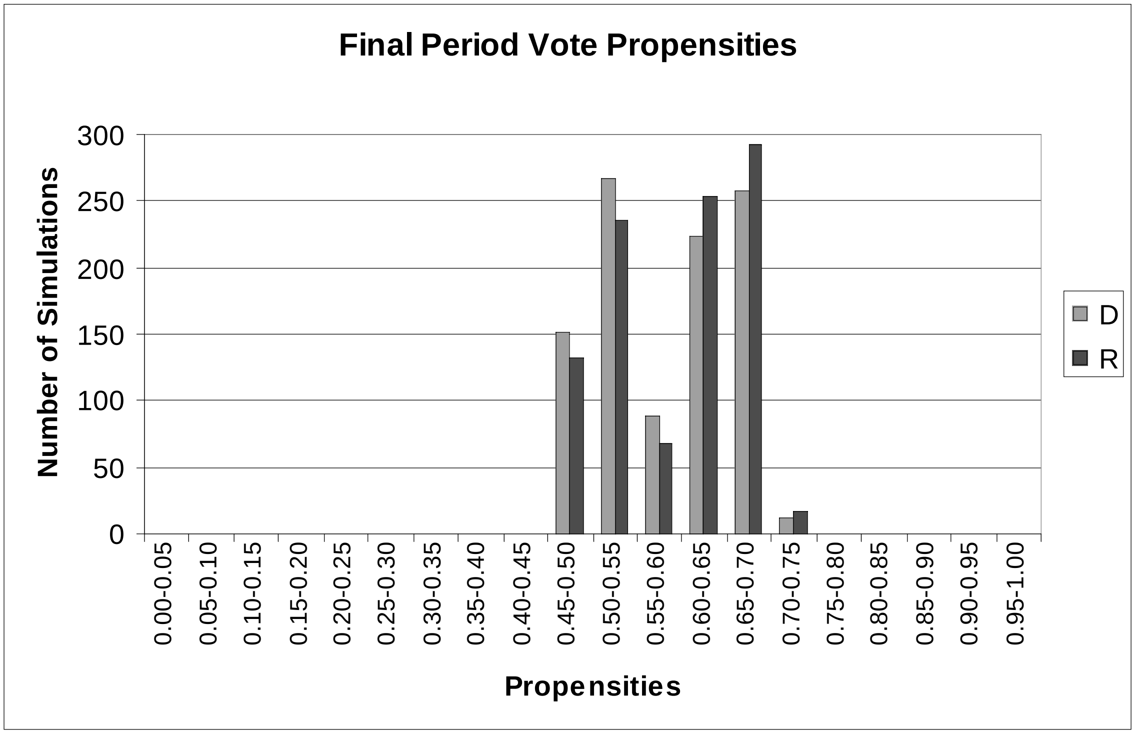
Run 4b: Low Voting Costs



Starting Values: 1,000 Periods
1,000 Simulations

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	5,000	5,000
b	1.0	1.0
c	0.05	0.05
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

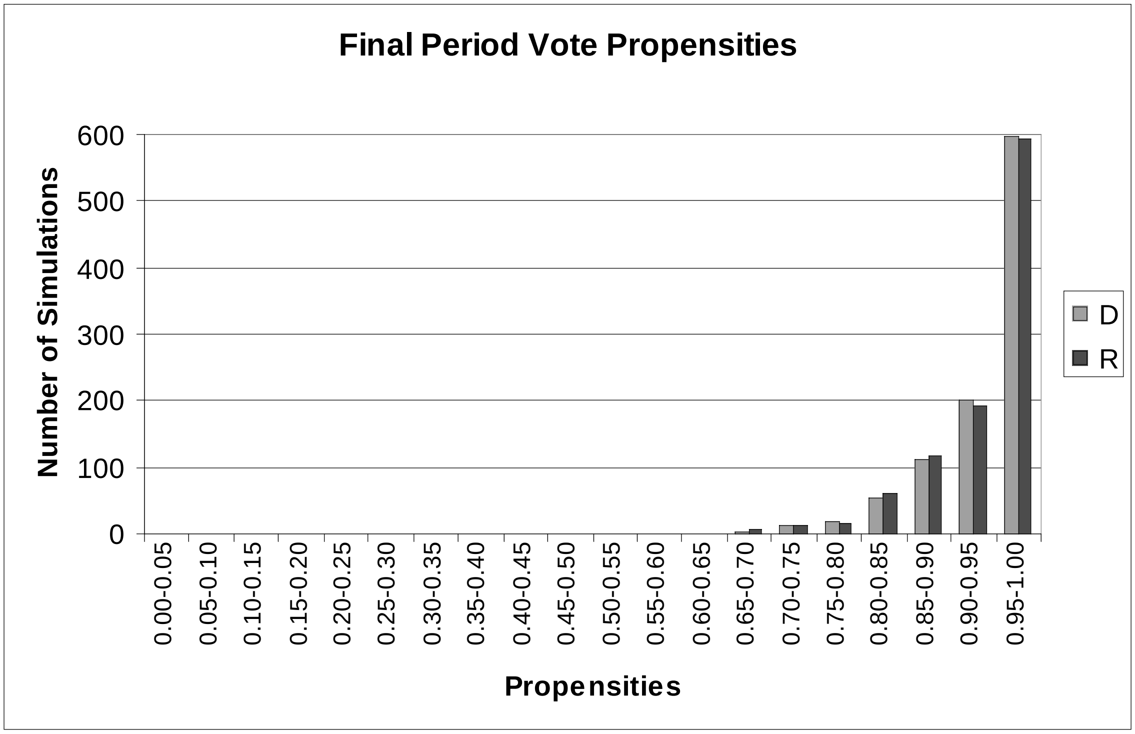
Run 4c: Negative Voting Costs



Starting Values: 1,000 Periods
1,000 Simulations

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	5,000	5,000
b	1.0	1.0
c	–0.25	–0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

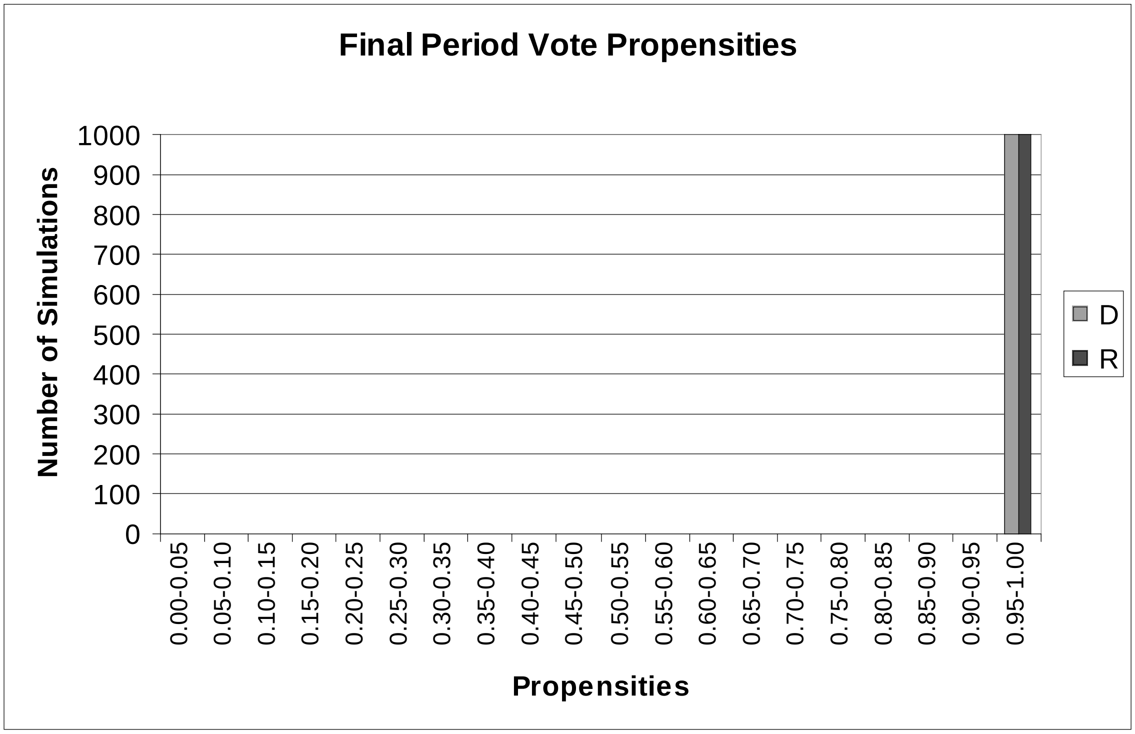
Run 5a: Negative Costs, Low Benefits



Starting Values: 1,000 Periods
1,000 Simulations

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	5,000	5,000
b	0.2	0.2
c	–0.25	–0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

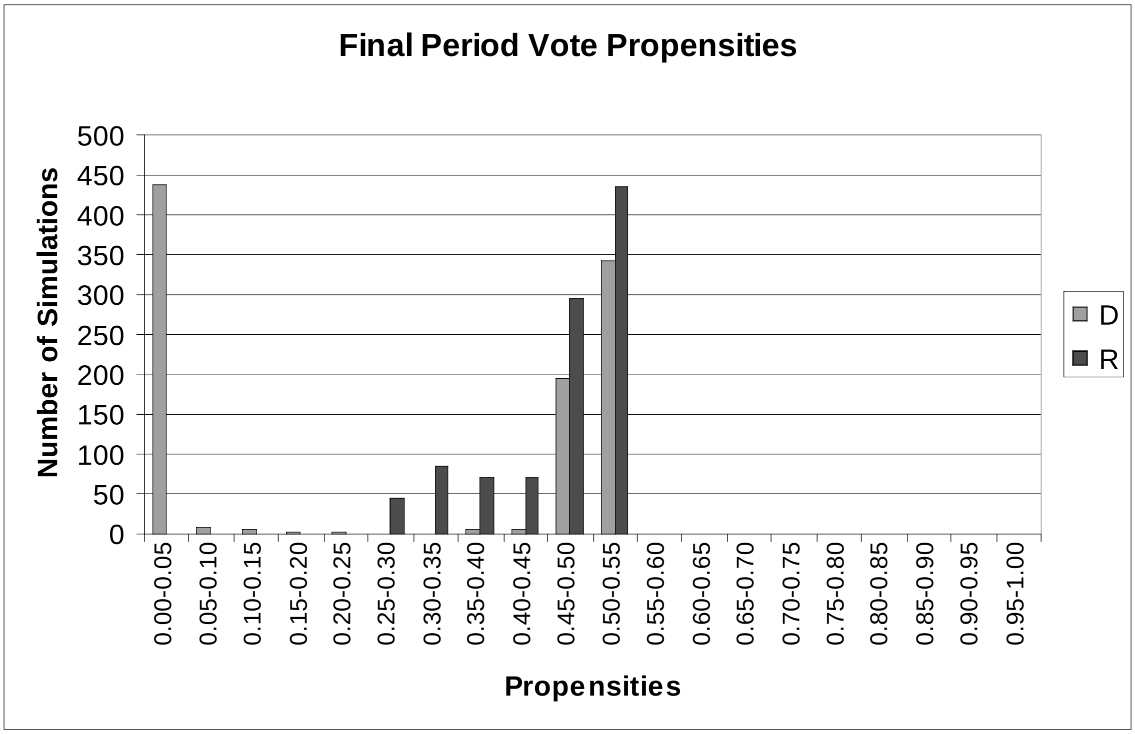
Run 5b: Negative Costs, Zero Benefits



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	0	0
c	–0.25	– 0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

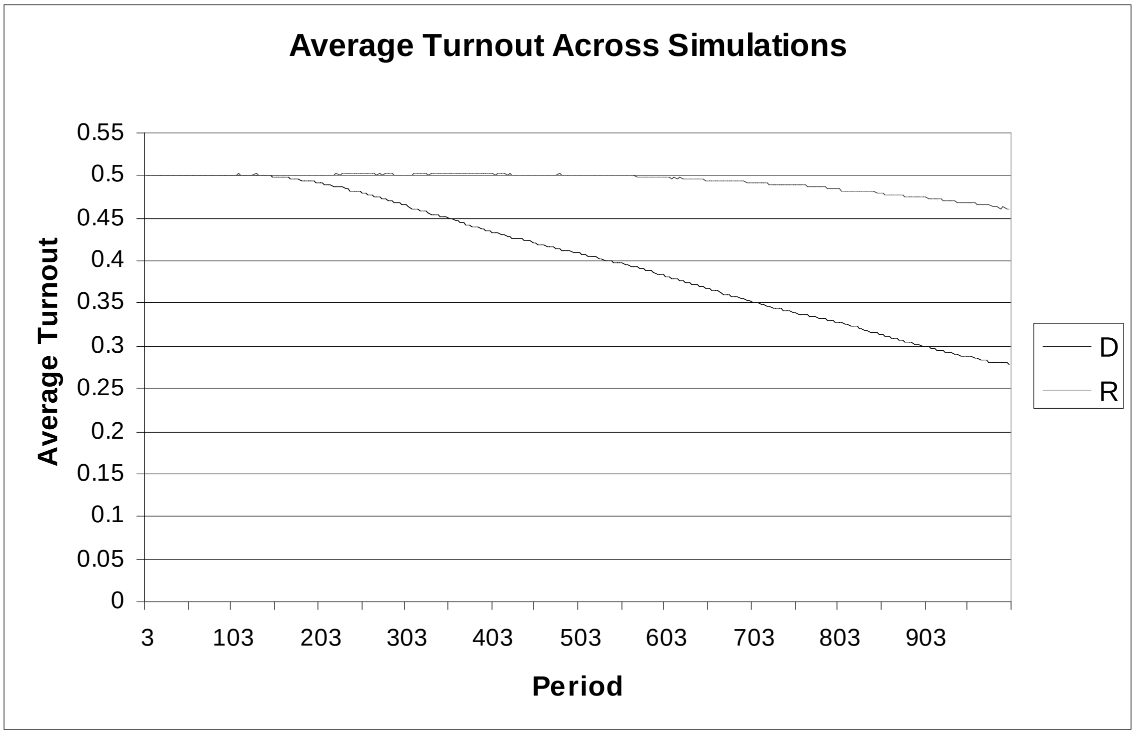
Run 6a: Asymmetric Costs



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	1.0	1.0
c	0.4	0.1
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

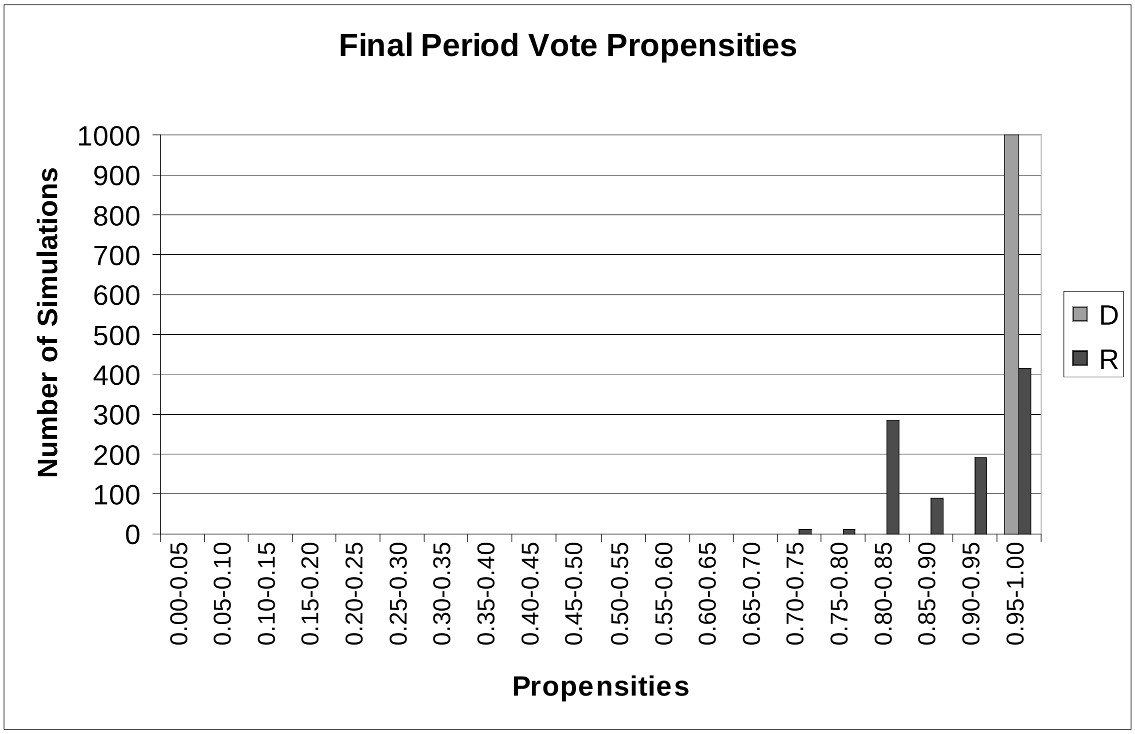
Run 6b: Asymmetric Costs - Time Series



Starting Values: 1,000 Periods
1,000 Simulations

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	5,000	5,000
b	1.0	1.0
c	0.4	0.1
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

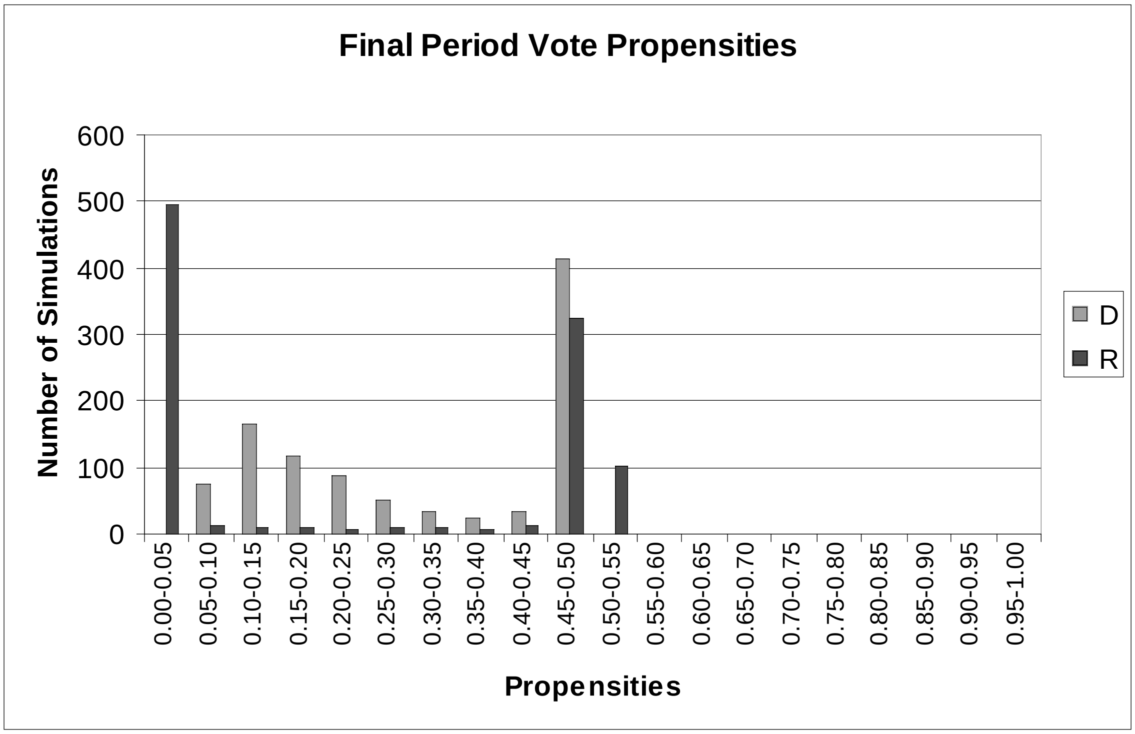
Run 7: Asymmetric Benefits, Negative Costs



Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,000	5,000
b	0	0.5
c	–0.1	–0.1
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

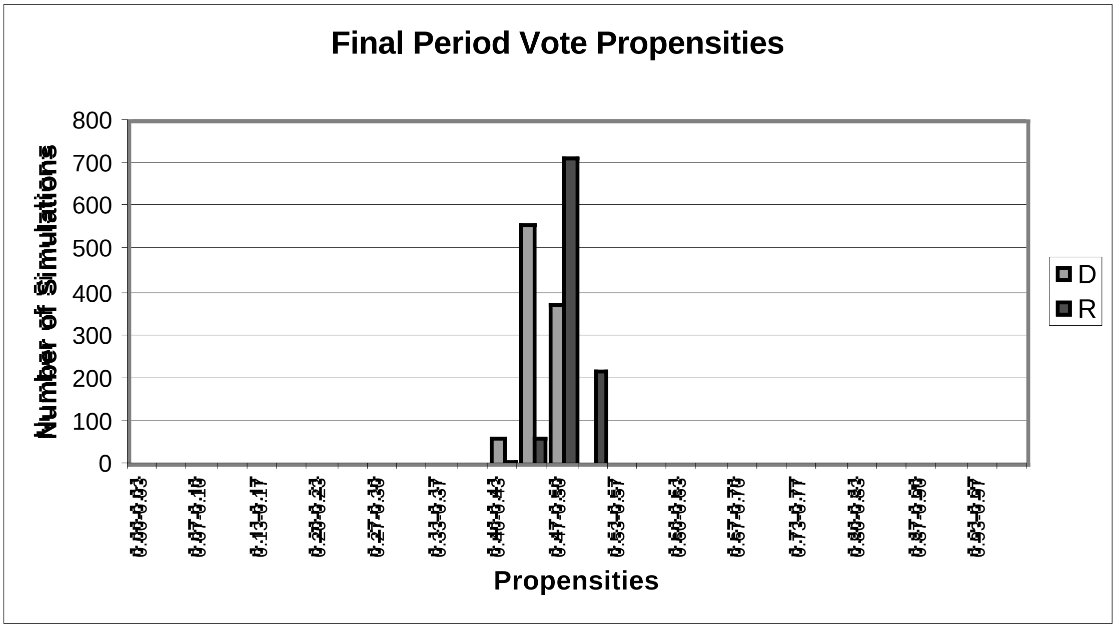
Run 8a: Asymmetric Factions

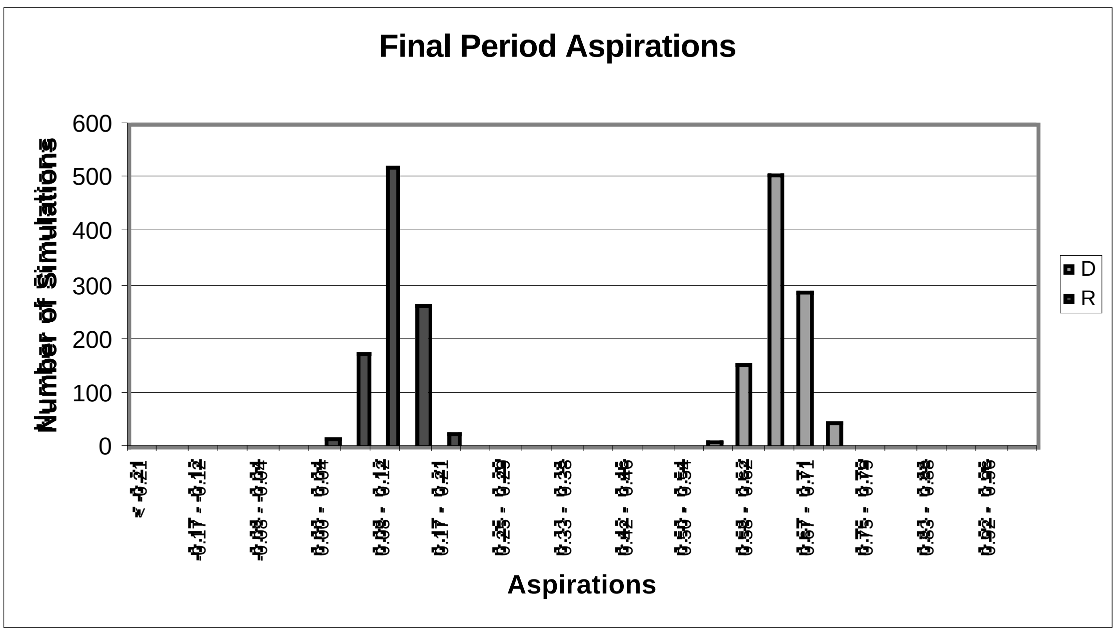


Starting Values: 1,000 Periods
1,000 Simulations

Faction	D	R
Population	5,200	4,800
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5

Run 8b: Asymmetric Factions

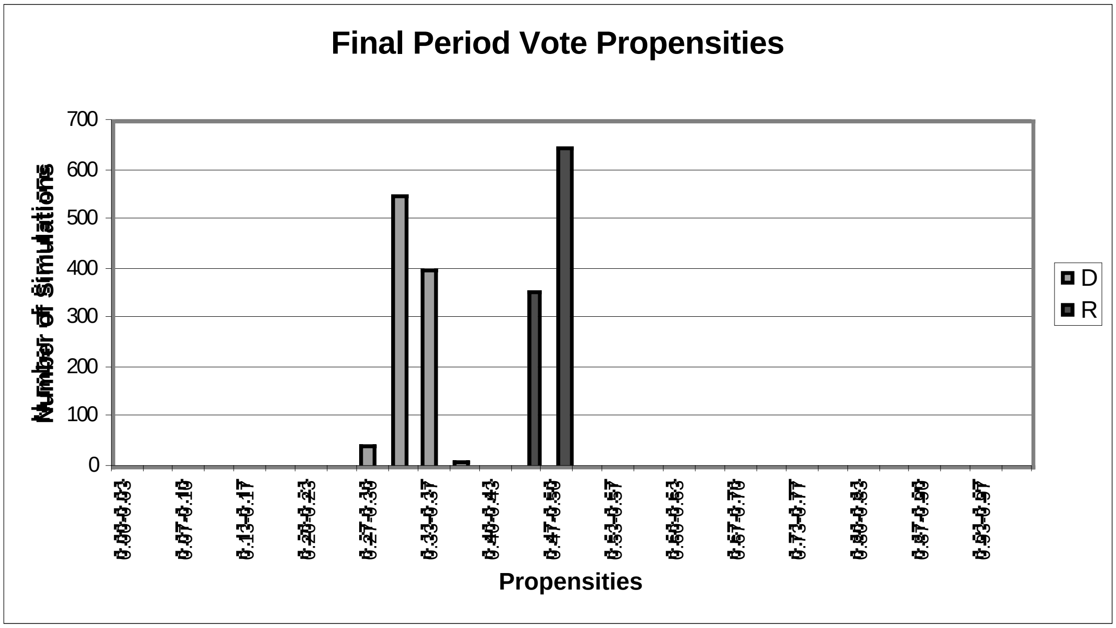


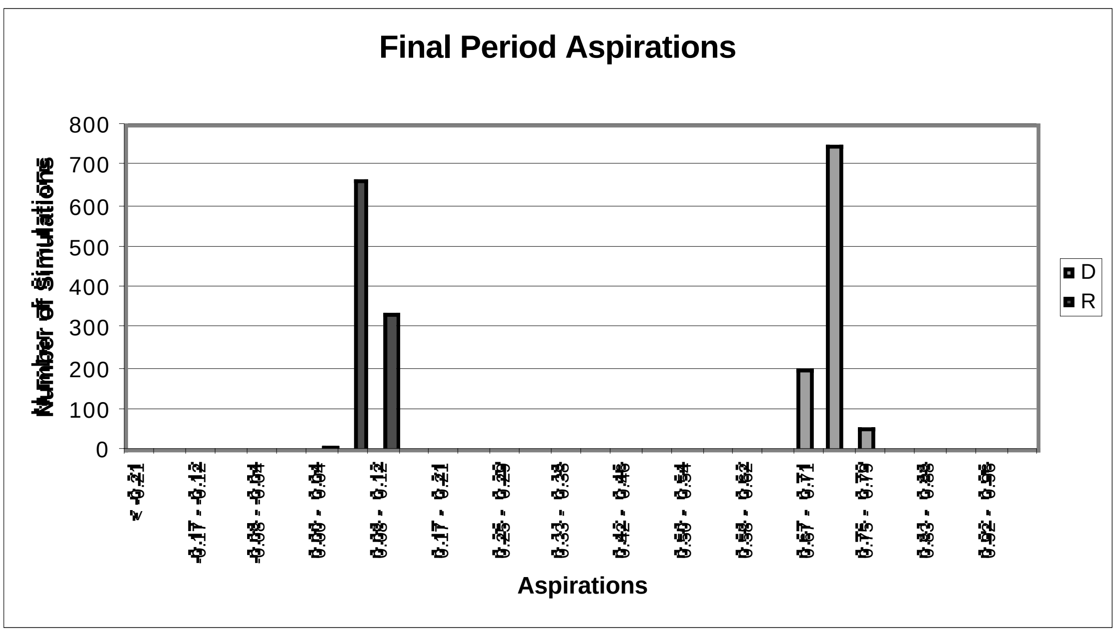


Starting Values: 1,000 Periods, 1,000 Simulations

Faction	Population	*b*	*c*	Aspirations	Propensities
D	5,200	1.0	0.25	0.5	0.5
R	4,800	1.0	0.25	0.5	0.5

Run 8c: Asymmetric Factions

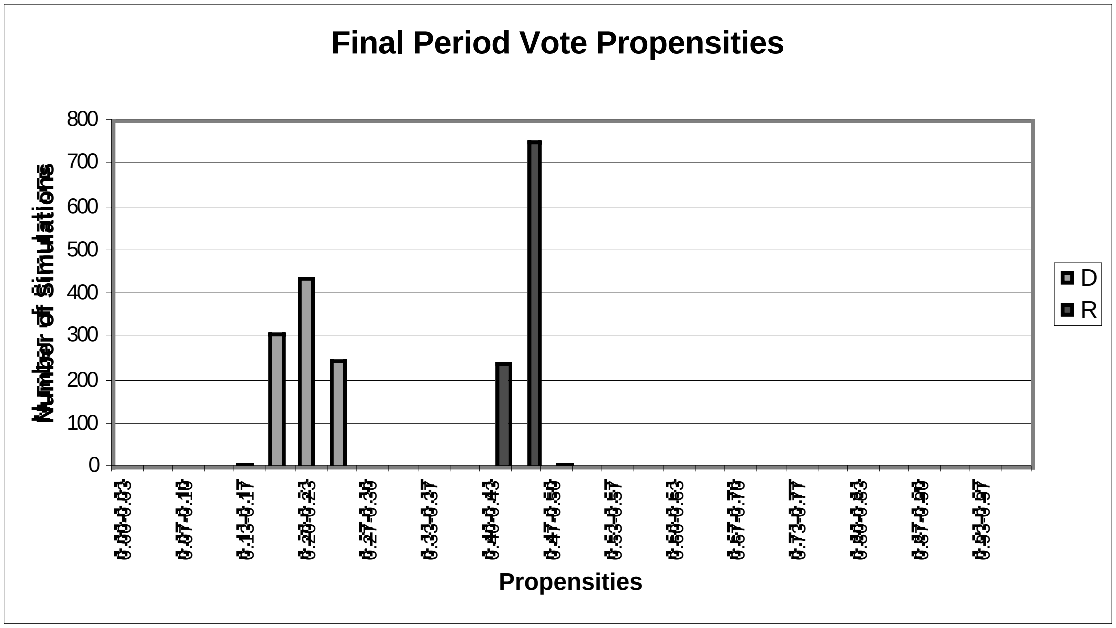


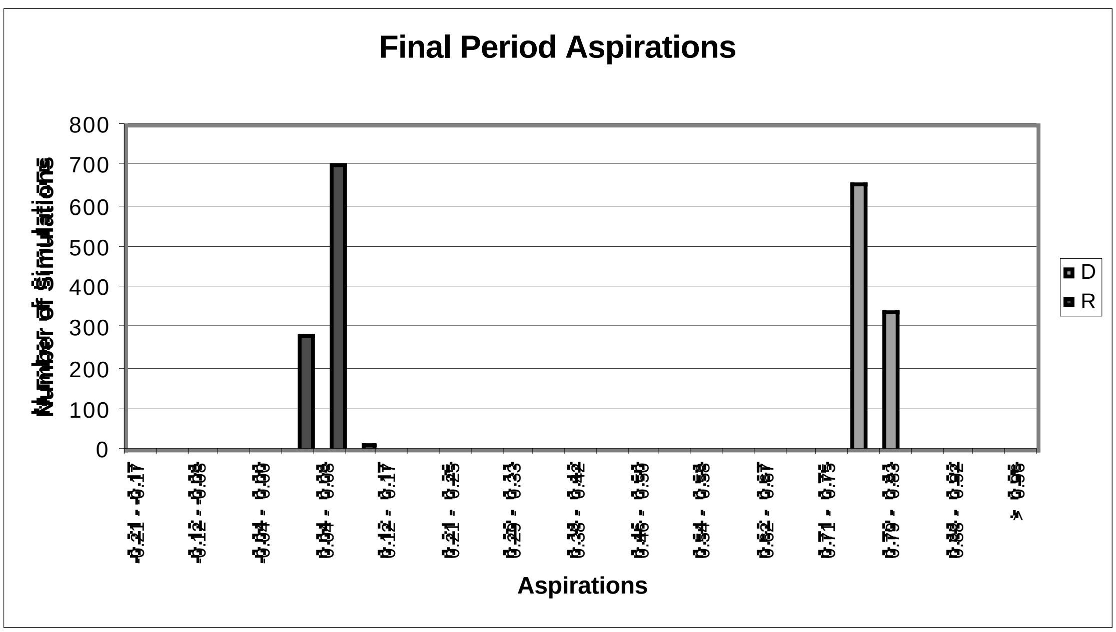


Starting Values: 1,000 Periods, 1,000 Simulations

Faction	Population	*b*	*c*	Aspirations	Propensities
D	6,000	1.0	0.25	0.5	0.5
R	4,000	1.0	0.25	0.5	0.5

Run 8d: Asymmetric Factions





Starting Values: 1,000 Periods, 1,000 Simulations

Faction	Population	*b*	*c*	Aspirations	Propensities
D	7,500	1.0	0.25	0.5	0.5
R	2,500	1.0	0.25	0.5	0.5

Figure 1: Basic Learning Cycle

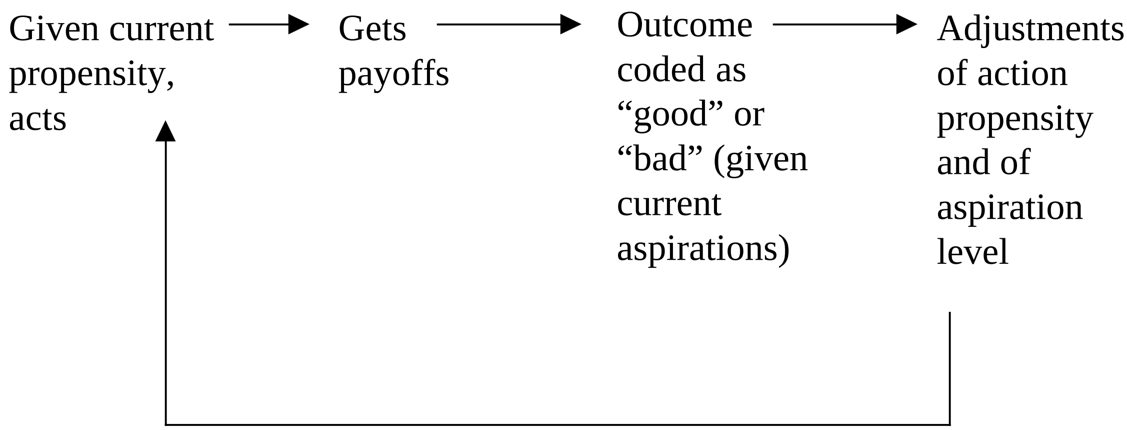
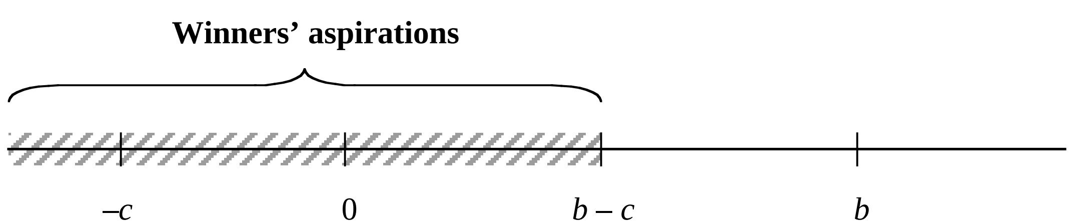
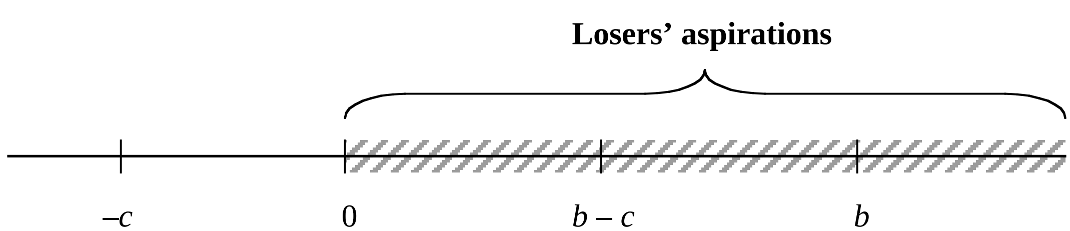


Figure 2: Aspiration Characteristics

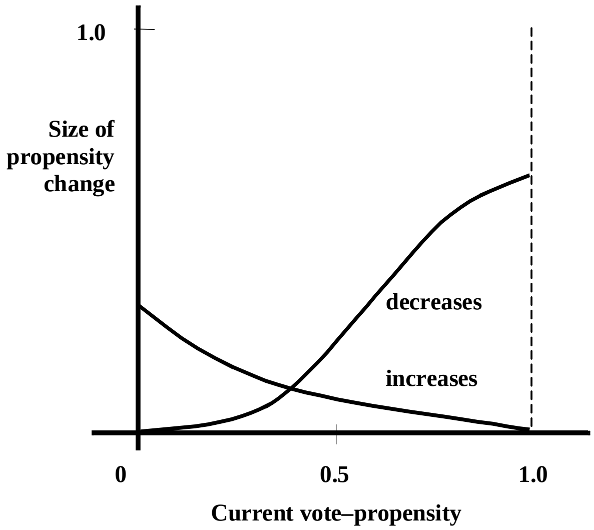


Winners' aspirations are not "too high," so winning is satisfying

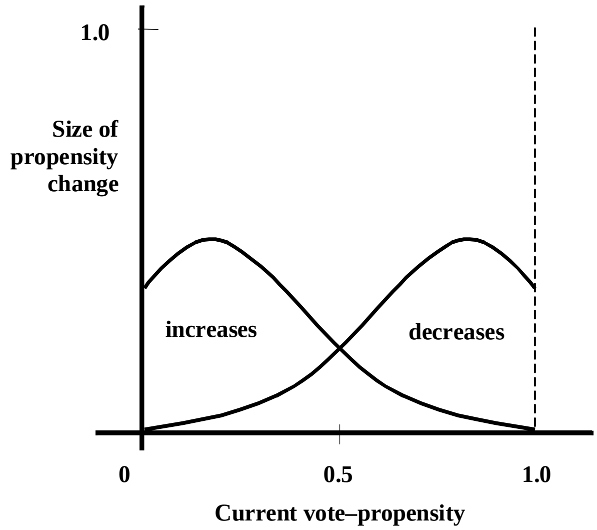


Losers' aspirations are not "too low," so losing is dissatisfying

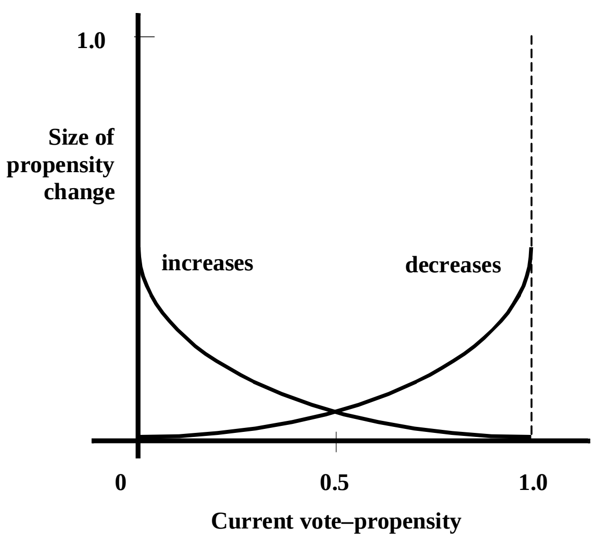
Figure 3: Types of Adjustment Rules



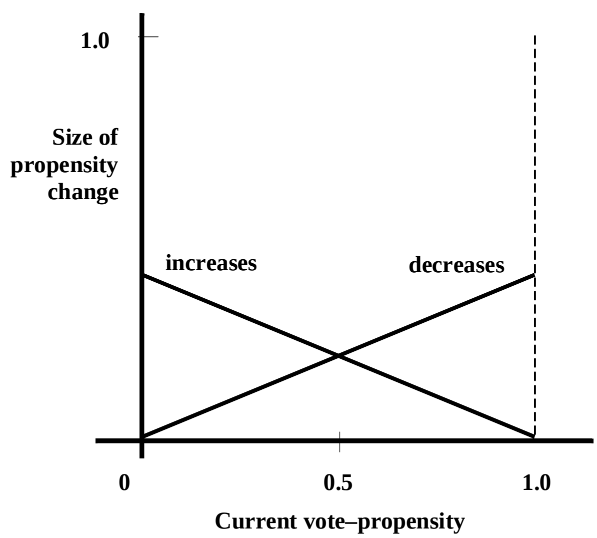
A monotonic but asymmetric rule



A symmetric but non-monotonic rule

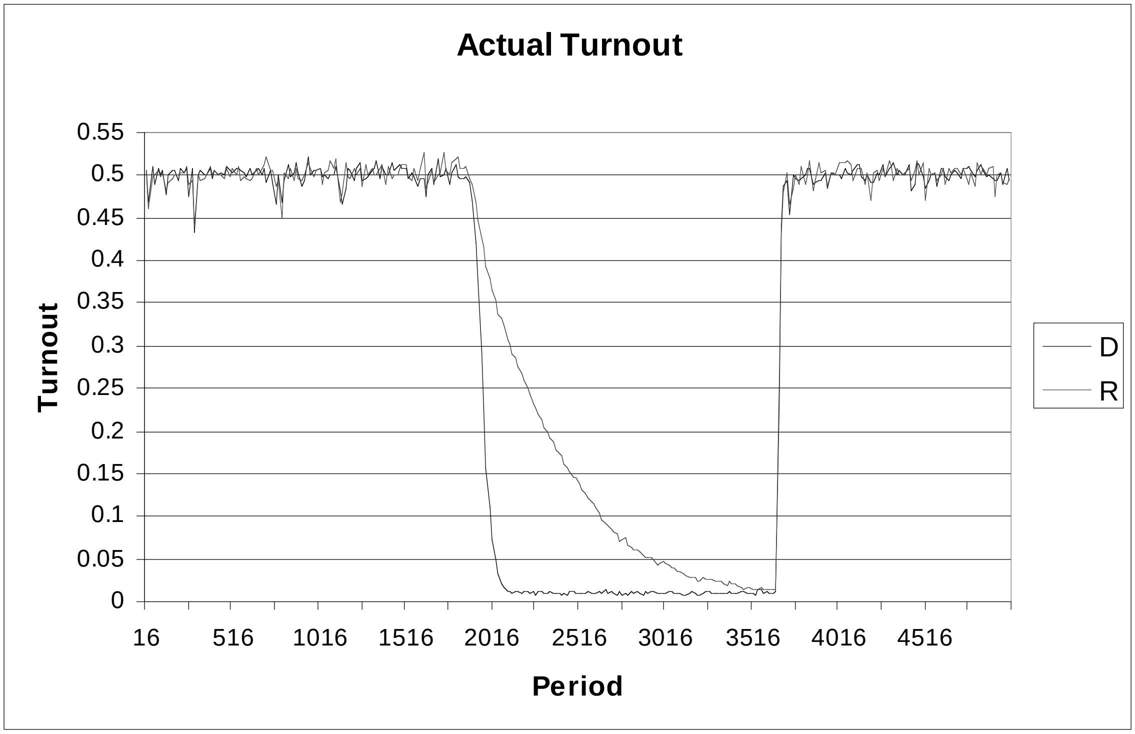


A monotonic and symmetric rule



A linearly monotonic and symmetric rule
(i.e., symmetric Bush-Mosteller)

Figure 4: Collapse of Participation



Starting Values: 5,000 Periods
1 Simulation

<u>Faction</u>	<u>D</u>	<u>R</u>
Population	5,000	5,000
b	1.0	1.0
c	0.25	0.25
Aspirations	0.5	0.5
Vote Propensities	0.5	0.5