V.O.Key Formalized:

Retrospective Voting as Adaptive Behavior

Jon Bendor Sunil Kumar David Siegel

Abstract

Since V. O. Key's seminal work, retrospective voting been regarded as a major component of voting theory, spawning a rich variety of models of voter choice utilizing Key's basic idea that the incumbent's performance influences citizens' votes. However, these models often assume that voters are fully rational and, for example, update their beliefs in accord with Bayes' rule. We suspect that Key had a less heroic view of voter cognition, and we propose a formalization of his verbal theory that we believe is closer to the spirit of his ideas. Our model is based on two fundamental axioms of aspiration-based retrospective voting: if an incumbent performed "well" (i.e., above voter A's aspiration level) then A's propensity to vote for the incumbent will rise; if an incumbent performed "poorly" (below A's aspiration level) then A's propensity to vote for the incumbent falls. These two assumptions, together with some postulates about how aspirations adjust, form the core of our model. We show analytically that citizens endogenously develop partisan voting tendencies, even though they lack overt political ideologies and instead simply vote retrospectively in the above manner. Further, this result is robust against perceptual errors (citizens evaluating an incumbent's performance incorrectly), given the conventional benchmark assumptions of independent and identically distributed errors. Lastly, we supplement this analytical model, which is spare in several respects, with a computational model that enables us to examine richer voting contexts.

I. Introduction

Two of the most robust findings about American voters is that few of them have coherent, detailed ideologies and few know much about politics. [1] Donald Kinder summarizes decades of survey research on ideology: "Precious few Americans make sophisticated use of political abstraction. Most are mystified by or at least indifferent to standard ideological concepts, and not many

[1]We thank Jim Fearon, Alan Gerber, Bob Luskin, Terry Moe and Ken Schultz for helpful comments on earlier versions of this paper. Bendor thanks the Fellows of the Center for Advanced Study in the Behavioral Sciences and CASBS itself for providing a wonderful work environment.

express consistently liberal, conservative, or centrist positions on government policy" (1999, p.796). Regarding information he reports that "the depth of ignorance demonstrated by modern mass publics can be quite breathtaking" and "the number of Americans who garble the most elementary points is... impressive" (p.785). Luskin's summary is harsher: most voters "know jaw-droppingly little about politics" (2002, p.282; see also Delli Carpini and Keeter 1996).

As is well-known, these empirical regularities contrast sharply with premises about voters in standard spatial models. In most Downsian formulations, citizens are assumed to have well-worked out ideologies in their heads—so, e.g., in unidimensional models a host of issue-positions are reduced in a consistent way to preferences over a single left-to-right spectrum—and to know a lot about politics: e.g., they know where candidates stand in the (commonly constructed) ideological space or at least to have unbiased estimates of these positions.

This gap between what we know empirically and what we assume theoretically was recognized long ago by Stokes: "[Downs'] model includes some cognitive postulates that need to be drastically qualified in view of what is known about the parties and electorates of actual political systems" (1963, p.369). And though the empirical critique *has* had some influence on theorizing (e.g., the work of Enelow and Hinich (1984) and Hinich and Munger (1994) clearly reflect concerns with standard spatial premises about voters), the debate has been hampered, we believe, by the failure of the critics to create their own models of electoral competition. Thus, the dialogue has mostly been confined to exchanges between rational choice theorists and empirically oriented critics, which has allowed the former to use the "you can't beat something with nothing" reply (e.g., Shepsle 1996, p.217).

However, nearly forty years ago V.O.Key (1966) sketched out an alternative: the retrospective theory of voting. Fiorina eloquently captured the heart of the idea.

> "citizens...typically have one comparatively hard bit of data: they know what life has been like during the incumbent's administration. They need *not* know the precise ecnomic or foreign policies of the incumbent administration in order to see or feel the *results* of those policies. And is it not reasonable to base voting decisions on results as well as on intentions [i.e., campaign promises]? In order to ascertain whether the incumbents have performed poorly or well, citizens need only calculate the changes in

their own welfare. If jobs have been lost in a recession, something is wrong. If sons have died in foreign rice paddies, something is wrong. If polluters foul food, water or air, something is wrong. And to the extent that citizens vote on the basis of such judgments, elections do not signal the direction in which society should move so much as they convey an evaluation of where society has been" (Fiorina 1981, p.5-6; emphasis in the original).

This is a plausible idea, but like many verbal theories it is somewhat vague and incomplete. In particular, how do voters evaluate governmental performance? How do they decide that an incumbent has "performed poorly or well"? And what are the effects of retrospective voting, either microscopic (e.g., the voting trajectories of individual citizens) or macroscopic (e.g., electoral outcomes)? We try to address these and related questions by developing a deductive model of retrospective voting. Formalizing Key's verbal theory has the usual benefits: it not only clarifies central notions, such as "good" governmental performance, it also allows us to extract testable implications.

Our model is designed to stay close to the general sense of Key's ideas. Since he clearly wanted to be realistic about voters and their capabilities, our model assumes that voters are not perfectly rational.[2] But in line with his famous remark that "voters are not fools" (1966, p.7), our voters are *adaptively* rational: they use sensible heuristics.[3]

The rest of the paper is organized as follows. Section II lays out the general ideas and defines

[2]There is a series of retrospective voting papers, pioneered by Ferejohn (1986), which assume completely rational voters. However, this line of work focuses on a different issue: not the capabilities of voters but whether politicians could credibly commit to implementing platforms advocated in campaigns. (Standard Downsian models assume that they could; in, e.g., Ferejohn (1986) words mean nothing; only the incumbent's actions matter.) Thus, Ferejohn-type models are explicitly principal-agent models, precisely as that term is used in economics. We completely agree with the assertion of Achen and Bartels that "rational retrospective voting is harder than it seems" (2004, p.36; see Wolfers [2002] for evidence on this point), and we suspect that Key would find the completely rational citizenry of these principal-agent models to be quite different from the American voters he studied empirically.

[3]For clarity and simplicity we examine voting heuristics that are purely retrospective: the citizens' votes are based totally on politicians' past performances. Fortunately, one can easily prove that most of our results are robust: if electoral choice is a weighted average of retrospective and prospective voting then they continue to hold if "most" (but not all) of the weight is on the past.

a class of adaptive voting rules. Section III presents an analytical model of adaptive retrospective voting. This model, by being starkly simple, allows us to deduce several important properties of adaptive voting rules. In particular, we show that (1) even if voters are initially completely ignorant of the conduct of political parties (i.e., they have no idea which party better serves their interests), over time they will learn to vote in partisan ways. Thus, party-based voting emerges endogenously (propositions 1 and 2). Further, the most informed voters are the most partisan (proposition 4). Section IV sketches out a richer model of retrospective voting which endogenizes aspirations, allowing them to adjust over time based on an individual citizen's experience (payoffs). We present some preliminary results of the corresponding computational model. Section V concludes.

II. General Ideas

Retrospective voting is based on voters evaluating the performance of an incumbent—either the party in power or a specific office-holder. This evaluation is largely instrumental: it is geared to the act of voting. Indeed, the heart of Key's theory of retrospective voting is that voters reward "good" performance by becoming more inclined to vote for the incumbent and punish "bad" performance by becoming less inclined to support the incumbent.

As noted above, the meaning of 'good' and 'bad' is somewhat opaque in Key's formulations. One way to make these notions more precise—a way that we think is close to the spirit of Key's argument—is to posit that voters have *aspirations*: internal evaluation-thresholds which code an incumbent's performance as good or bad, satisfactory or unsatisfactory. Once an incumbent's performance has been assessed in this manner, the *direction* of the voter's stance toward the incumbent official or party is determined: good performance is rewarded with increased support; bad, with reduced support.

We formalize these ideas as the two basic axioms of adaptive voting which we will use throughout this paper. For this formalization to make sense we must introduce the essentials of our model and some notation. We consider a sequence of elections in periods $1, 2, 3, \ldots$, with two parties, Democrats (D) and Republicans (R). In every period the voters get payoffs; these are due to an unspecified combination of the government's policies and factors (shark bites, terrorist attacks, and so forth) outside the government's control. Voter i compares his current payoff, $\pi_{i,t}$, to his prior aspiration level, $a_{i,t-1}$, and adjusts his propensity to vote for the incumbent accordingly. A new election is held, and the cycle repeats.

We assume that elections are held at the beginning of a period. Let W_{t-1} denote the winner of the election in $t-1$, who is then the incumbent in the election of period t, denoted by I_t: i.e., $I_t = W_{t-1}$. The citizen's propensity to vote for the incumbent is denoted by $p_{i,t}(I_t)$.

We can now state our two basic axioms of retrospective voting.

(A1) (positive feedback): If $\pi_{i,t} \geq a_{i,t-1}$ then with probability one $p_{i,t}(I_t) \geq p_{i,t-1}(W_{t-1})$, and this conclusion holds strictly if $\pi_{i,t} > a_{i,t-1}$ and $p_{i,t-1}(W_{t-1}) < 1$.

(A2) (negative feedback): If $\pi_{i,t} < a_{i,t-1}$ then with probability one $p_{i,t}(I_t) \leq p_{i,t-1}(W_{t-1})$, and this conclusion holds strictly if $p_{i,t-1}(W_{t-1}) > 0$.

Any rule that adjusts vote-propensities and that satisfies these two basic axioms belongs to a large set that we call Adaptive Voting Rules (AVoRs). We think that Key's notion of retrospective voting falls into this class. More importantly, we believe that whatever real voters are doing when they base their vote-choices (at least partly) on the performance of incumbents, it is plausible that they are doing so in ways consistent with (A1) and (A2). Behaving in accord with these postulates requires neither alot of political knowledge (fortunately, given how little most voters know) nor great powers of inference.[4]

(A1) and (A2) *do* presume, however, that people have aspirations and that these can change with experience. The latter claim is empirically plausible; indeed, it would be difficult to make sense of some important empirical patterns without positing that aspirations adjust.[5] The former claim—the sheer existence of aspirations—seems common-sensical, as we think it is, but nonetheless a significant theoretical issue is at stake here. (For example, some types of the (in)famous phenomenon of preference reversal can be explained by positing that decision makers have reference points—i.e., aspiration levels—which can be experimentally manipulated [Tversky, Slovic and Kahneman 1990, p.215].)

We now make our premises about aspiration-adjustment explicit.

[4]For evidence that voters are following rules consistent with (A1) and (A2) see Kinder (1999, p.839.

[5]For example, "rapid economic growth in countries such as Japan and France has been accompanied by a virtually flat line for SWB [subjective well-being]" (Diener and Suh 1999, p.441). This seems to reflect what psychologists now call the "hedonic treadmill": by adjusting to experience, higher aspirations reduce some of the potential psychic benefits of greater wealth.

(A3) Each agent i has an aspiration level, $a_{i,t}$, which is updated so that the following conditions hold for all i, t, and all histories leading up to t:

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

We have intentionally formulated (A1)-(A3) at a high level of generality. [6] More general theories are obviously preferable, ceteris paribus, to less general ones: "if x_1 then y" must rationally be more persuasive than "if x_1 and x_2 then y". Further, basic theoretical intuitions usually correspond to general qualitative properties, not to specific quantitative ones. Regarding voting, although we are confident that Key's verbal theory of retrospective voting pertains to processes that fall into our large class of AVoRs, it is far less obvious that he was contemplating a specific functional form (e.g., linear adjustment rules).

Readers who build formal models might guess that our three axioms are *so* general that deriving implications using only these premises must be difficult. This is indeed the case. Hence, throughout this paper we impose three assumptions on the kinds of AVoRs that we examine.

(1) We restrict attention to AVoRs that are *deterministic*: given a particular history and a current state of affairs—in particular, a voter's current vote-propensity—the AVoR that the citizen is using must determine a unique new vote-propensity, after the voter has evaluated the incumbent's performance. For example, if $I_t = D$, $\pi_{i,t} = h_t$ and $p_{i,t-1}(D) = 0.8$, then $p_{i,t}$ must, with probability one, be some unique propensity value in $(.8, 1]$.[7] We assume deterministic adjustment purely for tractability reasons: AVoRs that allow new propensities to take on multiple values (more precisely, that yield nondegenerate probability distributions over new vote-propensities) are harder to use. Allowing for nondeterministic adjustment is a natural candidate for future work.

(2) Because we do not want any aspects of party-oriented voting to be hardwired into the

[6] For simplicity we have made a specific modeling decision about what happens when payoffs exactly equal aspirations. Since this concerns a knife-edge circumstance, it isn't important.

[7] For example, Bush-Mosteller rules, which are often used in psychological learning theory, are deterministic in this sense. (Under Bush-Mosteller adaptation, if agent i has tried action x in t and the ensuing payoff exceeded aspirations, then $p_{i,t+1}(x) = p_{i,t}(x) + \alpha[1 - p_{i,t}(x)]$, where α, the rate of adjustment, is in $(0, 1]$. Thus Bush-Mosteller adjustment is linear as well as deterministic.)

adjustment rule itself (e.g. by allowing a citizen to use a rule that is biased toward D), we confine attention to AVoRs that are *party-neutral*. Defining party-neutrality is somewhat involved; the following example should give the flavor of the idea. Suppose citizens i and j, who are in different electorates, are using the same retrospective rule. Suppose that in t the incumbent in i's district is D; in j's it is R. If $p_{i,t-1}(D) = p_{j,t-1}(R)$ and $\pi_{i,t} = \pi_{j,t}$, then party-neutrality requires that i and j respond to D and R (respectively) the same way: $p_{i,t}(D) = p_{j,t}(R)$. Hence, we can attribute any difference between i's and j's lifetime of vote-propensities to either differences in their underlying interests (e.g., i is more likely to get an h-payoff from D than j is) or to macro-differences in the composition of their districts (e.g., i's community has a higher percentage of citizens with liberal interests than does j's).

(3) Most importantly, we examine only Markovian AVoRs: those in which adjustment of both voting propensities and aspirations in period t depend only the values of the state variables in that current period and on what happened in that period (e.g., what payoff a voter got).

It is important, especially for the benefit of readers who have heard that Markovian processes are "memoryless", to point out that Markovian rules can encode a lot of information about the past. For example, suppose a decision maker is trying to decide which of two slot machines to play. All she knows is that there is a fixed set of payoffs (e.g., nothing, 1 dollar, 2 dollars, etc.) and that the payoffs probabilities are stationary over time. Suppose she wants to try to maximize her expected payoff. Then one approach involves keeping track of each machine's average payoff and taking some action at every date based on this summary statistic. Because the average payoff at date t reflects the *entire* history of play, from date 1 through $t-1$, in an important sense this process is decidedly *not* memoryless. On the contrary: it is memory-intensive. It is Markovian because the action taken in t is based entirely on the values of the state variables—the two machines' performance indices—at t. (Specifically, if a second agent who used the same Markovian rule and who played the same two kinds of machines faced the same payoff averages at t then her decision-rule at t would be same, even though the first player and her clone reached those sets of performance indices via different sample paths.) Note further that this process is nonstationary: the effect of machine 1 delivering two dollars in trial 40 is less than the impact that payoff had in trial three. This feature allows still more history to affect today's decision. (E.g., suppose the rule is to stay forever with an alternative if its average payoff reaches some cutoff value. Then an unusually big payoff in period three might do the trick, while if it is first obtained in period 40 it might not.) [8]

It is important to point out that Markovian AVoRs can depend on time. (Consider, for example, a nonstationary Bush-Mosteller in which α_t, the amount of adjustment, is a decreasing function of t.) And, in fact, most of our results allow adaptation to be nonstationary. This is empirically important: it is likely that voters adjust less as they age [citations].

Because all the AVoRs examined in this paper (in both the analytical and the computational model) are deterministic, party-neutral and Markovian, we will not mention these properties as specific assumptions in the results that follow.

III. An Analytical Model of Retrospective Voting

We now present our analytical model of retrospective voting. In order to obtain closed-form results, it is insufficient to confine our attention to AVoRs that are deterministic and Markovian[9]: we must impose additional structure. To save on our own cognitive resources we construct a model in which we need not keep track of citizens' aspiration levels.[10] Nevertheless, aspiration levels are implicitly guiding voters' behavior. Because these two statements sound inconsistent, we must explain how we pull off this trick. To make our approach as clear as possible, we start with a decision theoretic example. Suppose a decision maker proceeds by some kind of trial-and-error consistent with (A1)-(A3): he tries an action, if it satisfies his current aspiration level then he becomes more disposed to try it again, and if he is dissatisfied then he becomes less inclined to try it. But his aspirations are themselves evolving with experience, which makes it harder for us to figure out what will happen.

[8] When payoffs are stationary, a Law of Large Numbers implies that eventually a performance index defined by the average payoff will eventually settle down, if the corresponding alternative is tried "long enough". Hence, Markovian AVoRs that are nonstationary might be one way to represent voting that becomes habitual. (For a cogent discussion of such "running tally" adaptive rules and their inertial properties, see Roth and Erev [20xx].)

[9] Some of our most important results do hold for party-biased rules, however. (Recall that we assume party neutrality throughout in order to not "rig" the results. Unlike the other two assumptions, party neutrality is not driven by tractability considerations.)

[10] Explicitly representing endogenous aspirations has been a formidable problem in the formal modeling of adaptive behavior.

Suppose, however, that unbeknownst to the agent, each action can yield only two payoffs: a high one, worth h, and a low one, l, where $l < h$. (Different actions generate h with different probabilities, so a genuine choice-problem exists.) Then *any* aspirational dynamic that is governed by (A3) will ensure that eventually, no matter what the agent initially aspired to, s/he will come to regard the h payoff as a 'success' and the l payoff as a 'failure'. That is, the sequence of aspirations $a_1, a_2, a_3, \ldots$ will with probability one get sucked into the (l, h) interval and then stay there. The following result, which we have proven elsewhere (Bendor, Kumar and Siegel 2004), says this more precisely.

Proposition 0: Consider a decision-theoretic problem in which the payoffs are either l or h, and every feasible action produces either payoff with positive probability. If aspirations adjust via (A3) then the following conclusions hold.

- (i) If $a_{t'} \in (l, h)$ then with probability one $a_t \in (l, h)$ for all $t > t'$.
- (ii) Suppose aspirations start outside (l, h): either $a_0 \leq l$ or $a_0 \geq h$. Then a_t moves monotonically toward (l, h) and is absorbed into that interval with probability one as $t \to \infty$.

Under these assumptions, decision makers whose aspirations adjust in accord with (A3) will eventually become dissatisfied with l and so will become less inclined to use an action that just delivered that payoff, and they will become more disposed to an action which has just produced the h-payoff. Thus, proposition 0 provides an analytical warrant for suppressing aspirations from models in which payoffs are binary.

Let us transfer this lesson to our analytical model of voting. Here we assume that governments generate binary payoffs for citizens. (It is natural to posit that a Republican administration is more likely to produce h's for conservative voters than for liberal ones, and vice versa for a Democratic administration. We will discuss these properties in detail shortly.) Proposition 0 implies that if citizens adjust aspirations in ways consistent with (A3) then eventually their aspirations will be in (l, h); hence, they will come to regard h's as satisfying and l's as dissatisfying. So if they adjust their vote-propensities via (A1) and (A2), their propensity to vote for the incumbent will rise if that incumbent produces an h-payoff, and it will fall if the incumbent gives them an l. Thus, binary payoffs and proposition 0 together enable us to have our analytical cake and eat it too: in these contexts it is reasonable to suppress aspirations from the model, yet it remains true that vote-propensities are modified as if aspirations explicitly guided propensity-changes.[11]

(Of course, we must acknowledge that *something* is lost by assuming binary payoffs and suppressing aspirations. The former assumption is often not the most natural way to represent choice situations in which a great many payoffs are possible.[12] The latter assumption makes it impossible for the present analytical model to represent aspirational dynamics or their effects. Both of these issues are addressed by the computational model presented in section IV.)

Because payoffs are binary, the relations between parties and voters are represented by the probability that a particular party gives the high payoff to different citizens. Thus, we assume the following properties (and also use the following terms and notation):

(1) A voter's "type", say i, is described by two parameters, $h_{i,D;t}$ and $h_{i,R;t}$, where $h_{i,D;t}$ denotes the probability that a voter of type i gets the h payoff from the D party in period t. (The meaning of $h_{i,R;t}$ is analogous.) When these probabilities are assumed to be stationary over time, we write them as $h_{i,D}$, etc. When they are nonstationary then the probabilities change according to some exogenous process. For example, disasters—hurricanes, 9-11's—can randomly impact the polity and lower the chance of an h-outcome, as can partly exogenous processes such as business cycles.[13]

(2) There are finitely many types of voters.[14] (Note: if there are only finitely many voters then

[11]Some readers may find it more natural to think of this section's model as having explicit aspirations; if so, then merely assume that $a_{i,t} \in (l, h)$ for all i, t. This is equivalent to the approach adopted here, i.e., suppressing aspirations and assuming that the propensity adjustments governed by (A1) and (A2) depend directly on payoffs alone.

[12]It is worth noting, however, that any set of preference ordering over a set of outcomes can be represented *either* by deterministic payoffs, one for each ordinal rank, *or* by only two payoffs, if the latter are stochastic. (Since the probability of getting the higher payoff can vary continuously, we have as many degrees of freedom as needed.)

[13]One could, e.g., think about the economic payoff that a Democratic incumbent gives a certain type of voter i as having a policy component plus an exogenous random shock. Both combine to make $h_{i,D;t}$ fluctuate over time. What the model requires is that either of the following hold: (1) the nonstationarity of $h_{i,D;t}$ is due to exogenous changes in the distribution of the random shock, or (2) there is some deterministic exogenous trend to $h_{\cdot,D;t}$ that affects all voters in a similar way.

[14]In some versions of this analytical model we will allow infinitely many *voters*, but even then there will be only finitely many types.

the model allows each voter to be a different type.)

(3) We say that a type i voter has "liberal interests" if $h_{i,D} > l_{i,D}$ and $h_{i,R} < l_{i,R}$; conservative interests are defined analogously. (We state the notion this way—voter i *has* certain kinds of interests—because we do not assume that a voter initially knows the underlying structure of his interests, as he would if he had a well worked-out political ideology.) When comparing voters, we say that the interests of a type i voter are *more* liberal than those of a type j voter if $h_{i,D} > h_{j,D}$ and $l_{i,R} > l_{j,R}$. Clearly, the former concept is stronger than the latter: if i's interests are liberal and j's are conservative then i's interests are more liberal than j's; but the converse does not hold, since i's interests could be *more* liberal than j's yet be conservative nonetheless.

The endogenous emergence of partisan affiliation

In this subsection we will show that even though citizens in this model lack political ideologies and only retrospectively evaluate governmental performance, they nonetheless tend to develop partisan voting tendencies. This tendency is probabilistic: for example, a citizen with underlying liberal interests may develop an affiliation for the Republican party if the GOP has done right by him or if the Democrats have governed badly. As Achen put it, "The voter's political history is the only causal variable" (1992, p.198). Hence we will be examining the probabilities that voters with different interests go down various sample paths, defined by voting propensities.

To fix ideas, let us consider an example. Suppose there are three voters in a district: i_1 and i_2 have the same liberal interests whereas j's are conservative. To show how partisan-voting can develop endogenously, let us start them off as without any party affiliations: $p_{i_1,0}(D) = p_{i_2,0}(D) = p_{j,0}(D) = \frac{1}{2}$. Let us assume that all of them use a stationary Bush-Mosteller rule with α, the rate of adjustment, equal to 0.2. (Since $p_{i,t}(R) = 1 - p_{i,t}(D)$, henceforth we drop the party subscript and let $p_{i,t}$ stand for $p_{i,t}(D)$.) Suppose the first incumbent is D and in period 1 citizen i_1 happens to get a high payoff whereas the others get low payoffs. Then i_1's inclination to vote Democratic rises to 0.6, whereas the other two become less disposed to vote for the incumbent, their propensities dropping to 0.4. So then the Republican challenger is likely (i.e., with probability .552) to win the election at the end of period 1. Suppose this happens, and in period 2 R delivers l to both i's and h to j. Then both i's will increase their propensity to vote for the Democratic challenger, to 0.68 and 0.52 for i_1 and i_2 respectively, whereas j will become more likely to vote for the Republican incumbent, so $p_{j,t}$ falls to 0.32. Now stop the process and inspect the citizens' biographies and voting affiliations in period 2. Citizen i_1 has gotten a good payoff from D and a bad one from R, so she is becoming a Democratic partisan, with a Democratic vote-propensity of 0.68. Citizen i_2 has had the happy experience of good payoffs from both parties, so he is nearly neutral: $p_{i_2,2} = 0.52$. And j, with a bad payoff from D and a good one from R, is becoming a Republican partisan: $p_{j,2}$ is only 0.32. Note that since i_2 has liberal interests, his experience of good payoffs from both parties was less likely than i_1's experience of h from D and then l from R. (The former occurs with a probability of $h_{i,D} \cdot l_{i,R}$ which exceeds i_2's combination of $h_{i,D} \cdot h_{i,R}$ since i_2 has liberal interests.) Nevertheless, because $h_{i,R}$ isn't zero this particular political biography of i_2 *can* happen.

Thus, this theory sees *no fundamental tension between party ID*, represented here as party-oriented vote-propensities, and *substantively attuned decision-making.* Instead, as Jackson (1975) and Achen (1992) have argued, citizens' current voting tendencies reflect their life-experiences with the two parties. To be sure, the substantive tuning generated by adaptive retrospective voting is crude. (Later we will show that many AVoRs fall short of optimal voting, even in the long-run.) Nevertheless, the connection between experience and partisan voting is real. And if normative theorists attend to the findings of 60+ years of voting behavior and adjust their aspirations about civic behavior accordingly, they may come to regard this substantive tuning as not unreasonable.[15]

Although the preceding example presumed that voters adjust their propensities via Bush-Mosteller, the example's *qualitative* features—e.g., that i_1 was more likely than j to emerge, after two elections, with a tendency to vote Democratic—do not depend on this precise functional form or its linearity. Indeed, our first main result, Proposition 1, will show that the endogenous emergence of party-oriented voting requires relatively little structure above and beyond the spare skeleton of (A1)-(A3) plus Markovianism. However, one adjustment-property possessed by Bush-Mosteller rules *is* important; fortunately, it is substantively quite intuitive. Roughly speaking, we will say that a (deterministic) AVoR is *order-preserving* if similar experiences in the current period maintain the propensity-ordering that obtained before those experiences. (For the precise definition see the Appendix.)

Our example reveals one rather abstract—yet important—property of an individual's vote-

[15] Of course, "not unreasonable" may be thought a pretty low standard. But that's what realistic aspirations imply.

propensities. Although i_1 and i_2 have identical underlying interests and started out with the same vote-propensity, by the second period their paths had diverged: i_1's probability of voting Democratic exceeded i_2's. We believe this illustrates a general pattern: because individual biographies always contain random shocks, what people learn—here, whom to vote for—can vary. Thus, the core of the theory is stochastic: instead of point predictions—an individual with such-and-such characteristics will vote Democratic—its predictions are probability distributions over vote-propensities. And because propensities are (mathematically speaking) just probabilities, the theory generates probability distributions over probabilities. This may sound strange but a close inspection of the example shows that it isn't. What happened to i_1 in the first two periods (h from D and then l from R) was merely one scenario; others *could* have happened. A realized political biography is just a sample path.

With this concept in hand, we are ready for our first major result.

Proposition 1: Suppose i and j, two citizens in the same electorate, use the same order-preserving AVoR. If i's interests are more liberal than j's and $p_{i,0} \geq p_{j,0}$, then i's probability distribution over her Democratic vote-propensities first-order stochastically dominates j's probability distribution, for all $t \geq 1$.

A few comments are in order. First, although our simple numerical example showed that different citizens can wander down different sample paths of political experience, a probabilistic pattern emerges. As long as people start adulthood with vote-propensities that are consistent with their underlying interests—which normal political socialization in stable democracies probably ensures—then any retrospective voting rule governed by (A1) and (A2) will produce party-oriented voting: under the hypotheses of Proposition 1, the unconditional probability that i will vote Democratic in t is higher than j's unconditional probability, for *all t*. [16] (Since this corollary of Proposition 1 is easier to state than is that result's conclusion, henceforth we will mostly refer to the former, not the latter. But it should be understood that Proposition 1's conclusion is stronger—it implies

[16] This is due to the law of total probability: one just multiplies a particular vote-propensity at date t with its associated probability, and then sums over all of that citizen's feasible vote-propensities for that election. Given that i's probability distribution over vote-propensities first-order stochastically dominates j's, it is easily shown that i's sum exceeds j's, whence i is more likely to vote Democratic than j in every election.

the corollary, but the converse doesn't hold: i's unconditional probability of voting Democratic could exceed j's, but that doesn't imply that i's probability over her Democratic vote-propensities stochastically dominates j's.)

Second, iteratively applying Proposition 1 shows that if i's interests are more liberal than j's which in turn are more liberal than k's and so forth all the way through some voter-type z, then we obtain a ranking of probabilistic party-oriented voting: in every election i is most likely to vote Democratic, j is next most likely, and so on to z, who is most likely to vote Republican.

Third, because the proposition doesn't require that the citizens use a stationary AVoR, the result allows for age-effects: e.g.. voters might change their vote-propensities less as they get older. (This seems to be the pattern in the U.S. [citation].)

Fourth, the result is quite general in two important respects: it holds for electorates of any size and any composition of interests. The generality along these important dimensions is our reward for suppressing aspirations.

Finally, note that proposition 1 compares the voting patterns of citizens *in the same electorate*. It does so for important substantive and theoretical reasons. Citizens who live in the same district have political experiences in common. Our model represents this common experience partly by the sequence of governing parties: at every date the same party is in power for both i and j. In a model whose basic mechanism is adaptation to experience, this is a powerful commonality. it implies, among other things, that we need to assume only that i's interests are *more* liberal than j's in order to derive the conclusion that i's probability of voting for D exceeds j's; we do not need to assume that i actually has liberal interests.

If, however, i and j lived in *different* districts then a mere ranking of interests would not suffice to yield this conclusion: i's interests could be more liberal than j's, yet j may be more inclined to vote Democratic. Consider the following example. Suppose i's district is composed of three voters: i, l_1 and l_2, and the other district is composed of j, c_1 and c_2. Voters l_1 and l_2 have completely liberal interests: they always get h's from D and l's from R. Similarly, c_1 and c_2 have completely conservative interests: they always get high payoffs from R and low payoffs from D. (Imagine that the D and R parties are dominated by l-types and c-types, respectively.) Both i and j are more moderate than their peers, though i's interests are more liberal than j's: $\{h_{i,D} = 0.45, h_{i,R} = 0.3\}$

and $\{h_{j,D} = 0.3, h_{j,R} = 0.4\}$. Suppose that all citizens use a very simple AVoR: vote for the incumbent if and only if you got a high payoff when that incumbent was in office. Then D always wins in i's district while R always wins in j's. And since $h_{i,D}$ equals 0.45, citizen i votes Democratic 45 percent of the time—which is *less* than j, who votes Democratic 60 percent of the time. Why does j, whose interests are more conservative than i's, vote Democratic more often?

The answer turns on two factors: the district's different interest-ecologies, and how this difference affects i and j. The former's district is dominated by fervent liberals who always elect Democrats; in j's, fervent conservatives always elect Republicans. But both parties gives low payoffs to more moderate voters (i or j) over half the time. This means, since i is always governed by Democrats and j always by Republicans, that i's dissatisfaction is reflected in a relatively low rate of voting Democratic while j's emerges as a relatively low rate of voting Republican. Different electorates, different experiences, different voting tendencies: hence, proposition 1's conclusion doesn't hold.

Yet a voter's environment—his electorate's ideological composition and the governance he experiences—is not completely decisive. It is clear, e.g., that in the above example that l_1 will always vote Democratic and c_1 will always vote Republican, regardless of where they live: their interests and the corresponding party alignments are so extreme that they are impervious to environmental impact. Since l_1 is the most extreme example of a citizen with "liberal interests" (per our preceding definition), just as c_1 most exemplifies someone with conservative interests, this provides a clue: perhaps proposition 1's conclusion holds for voters in different districts who have liberal versus conservative interests? The next result shows that this guess is correct, provided that the citizens use a certain type of AVoR: one that is *symmetric* as well as order-preserving. (If people use a symmetric AVoR then they respond similarly to positive and negative feedback. For example, suppose that i's propensity to vote Democratic in t is 0.8 and the Democratic incumbent gives her a high payoff, while in period s her propensity to vote Democratic is 0.2 and the Democratic incumbent gives her a low payoff. If i's AVoR is symmetric (and stationary), then she will increase her Democratic vote-propensity in t by the same amount she reduces it in s. For a formal definition of symmetry see the appendix.)

Proposition 2: Suppose i and j live in different electorates, and i has liberal interests while j's are conservative. Assume they use the same symmetric and order-preserving AVoR.

- **(i)** If $p_{i,0} \geq p_{j,0}$, then i's probability distribution over her Democratic vote-propensities first-order stochastically dominates j's probability distribution for all $t \geq 1$.

- **(ii)** If $p_{i,0} \geq \frac{1}{2}$ while $p_{j,0} \leq \frac{1}{2}$, then for all future periods, the *ex ante* probability of i's voting for D exceeds $\frac{1}{2}$ and the *ex ante* probability of j's voting for D is less than $\frac{1}{2}$ .for all $t \geq 1$ $p_{i,t} > \frac{1}{2}$ and $p_{j,t} < \frac{1}{2}$.

This proposition yields two kinds of testable implications: one is micro; the other, macro. The macro-implication is simple. Compare two districts, A and B, of equal size and composed of voters who satisfy the requirements of Proposition 2. If there are more liberal types in A than in B and the proposition's other assumptions hold, then we get the intuitively sensible prediction that the probability of a D-victory is higher in A than in B for all $t \geq 1$. (One could state this as a formal corollary of proposition 2; we prefer to state it as an empirical hypothesis.) This prediction holds for districts with multiple liberal types (ultra-liberal, liberal, moderate liberal, etc.) and multiple conservative types; with some computational effort it also extends to districts of different sizes. The micro-prediction is that if we can find observable characteristics which are correlated with voters' interests, then we will find that those voters whose characteristics correlate with liberal interests vote Democratic more often than do those with conservative interests, in either district.

Note that proposition 2, like its predecessor, allows for nonstationary voting: citizens could be adjusting their vote-propensities less as they age.

Do AVoRs lead to optimizal voting?

Although adaptive retrospective voting isn't crazy, it's not fully rational either. Some political scientists have recognized this (e.g., Achen 1992, p.199). Others, however, see little distinction between backward-looking adaptation and optimization. In particular, a simple kind of satisficing (vote for the incumbent if and only if today's payoff is satisfactory) belongs to the class of AVoRs, and some scholars have maintained that satisficing is "merely a shorthand label for optimzing under conditions of limited information and uncertainty" (Jackman 1993, p.282; see also Goodin 2000, p.63). Here we will show that, except under rather special circumstances, an important subset of AVoRs do *not* lead to optimizing behavior even in the ultra-long run (i.e., the limit). Thus, these retrospective rules remain forever distinct from optimization. Moreover, this holds even though voters get completely accurate feedback about their payoffs, as we've assumed thus far. (Later we will examine the effects of erroneous feedback.) Hence, the permanent divergence between these AVoRs and optimal behavior cannot be explained by informational errors. Instead, this divergence flows from basic properties of certain types of retrospective voting.[17]

To demonstrate how these AVoRs differ from optimization, we examine two simple but important demographic cases. In the first case, the electorate has just one voter; in the second, infinitely many. The first case is useful because in such electorates the voter is always pivotal; hence, we can study the suboptimality of AVoRs in splendid isolation, without the complexities arising from collective choice processes. The second case is useful because a Law of Large Numbers comes to our aid, taming the wildness of probabilistic voting and making the aggregate outcome—which party wins the election—deterministic, thus easier to analyze. In both cases we are interested in finding out whether citizens eventually learn to do what is best for them, i.e., to consistently vote for the party that delivers the high payoff with the higher probability.

Fact 1: Suppose $n = 1$ and the voter uses a stationary AVoR. If optimal voter-choice isn't trivially guaranteed—$h_D \neq h_R$—and parties are imperfect—$\max\{h_D, h_R\} < 1$—then the citizen's voting never becomes optimal.

We will understand AVoRs better if we pause long enough to understand why this important subset of retrospective voting rules fails the optimality test, despite the fact that the same choice problem recurs and feedback is perfect. (We suspect that few readers would find the assumption that parties are imperfect to be implausible!) Suppose that $h_D > h_R$ for the voter. Then it is optimal for the citizen to ensure that D is always elected. Since $n = 1$ the citizen is always pivotal; therefore, she can guarantee the optimal outcome by voting for D with certainty.

However, no stationary Markovian AVoR (recall that only Markovian rules are examined in this paper) will achieve this, even though the problem itself is stationary. Suppose that at some date t D is in power and the voter's propensity to vote Democratic has become very high, perhaps

[17] Since we do not presume a specific functional form such as Bush-Mosteller, the divergence between optimizing and retrospective voting doesn't arise from any properties of specific functional forms (such as the Bush-Mosteller's linearity). The fault lies elsewhere.

reaching one. But by the negative feedback axiom (A2), if the current payoff is low then the voter will reduce her propensity to vote for the incumbent. And since even the optimal party (D) is imperfect, eventually the voter will get l when D is in power. Hence locking permanently onto the optimal action of always voting for D is impossible.

The importance of the Markov property is evident: with a Markovian rule, the voter forgets about the (possibly wonderful) history that brought her to the present date, reacting instead only to today's poor performance by D. [18]

The significance of stationarity is more subtle. To understand why it matters, consider a nonstationary rule and see how it gives retrospective adaptation a shot at attaining optimality in the long-run.[19] Suppose that the adjustment is Bush-Mosteller, but the *size* of the adjustment decreases over time, say proportionally to $\frac{1}{t}$. Then in the limit the agent will become inertial, which is exactly what is required. (Of course, optimization also requires that her propensity to vote for D converges to 1. Becoming inertial is necessary but not sufficient for optimizing.)

We should recognize, however, constructing rules with just the right (i.e., optimal) amount of nonstationary inertia is a delicate affair. It is easy to overdo it and build in *too much* inertia. (For more on this problem see Bendor and Kumar (2005).) There is little reason to believe that ordinary citizens somehow stumbled on just the right amount; this flies in the face of everything that we know empirically about voters. Garden-variety AVoRs are probably suboptimal, even in the long-run.

But most important elections involve thousands or even millions of voters, and although any one of them might be using a suboptimal approach to voting, perhaps in the aggregate the associated errors cancel each other out. In short, perhaps Condorcet's Jury Theorem restores optimality.[20]

Condorcet's Theorem does have this benign effect if voters are prospective and compare the incumbent's *anticipated* performance *to the challenger's.* This comparison of alternatives *to each*

[18] Again, we caution the reader not to underestimate Markovian rules. As noted earlier in our example of a three-citizen district, a wonderful past *does* influence the voter's current behavior by making her current propensity to vote for the incumbent very high. Such rules do allow the past to matter, but *all* of its influence is embedded in the current value of the state variable $p_{i,t}$; none is reflected in today's *adjustment* to $p_{i,t}$.

[19] That such adaptive rules do, in fact, exist can be rigorously established along the lines of Kiefer and Wolfowitz (1952).

[20] For a thoughtful discussion of this topic, see Kinder (1998, p.797-800).

other is essential. If voters do this then their (independent) comparison or evaluation errors will indeed cancel each other out: in the limit—i.e., in an infinitely large electorate—the party that is better for the majority will win with probability one (Miller 1986). But in (pure) retrospective voting, the incumbent is compared not to the challenger but to an internal standard: a voter's aspiration level. Poor performance, as measured by that standard, weakens support for the incumbent. The promises of the challenger are ignored. This makes it harder for the Jury Theorem to help majority rule elections achieve the optimality standard. The next result reveals the problem in the simple context of a homogeneous electorate; most of the logic carries over to heterogeneous districts.

Part (iii) of this next result requires two new assumptions about how citizens adjust their vote-propensities. First, propensity-adjustment cannot become arbitrarily sluggish in the face of dissatisfaction; we call this (A2*), because it replaces the weaker negative feedback assumption of (A2). Second, part (iii) also requires that AVoRs involve *equal-adjustment*: from any given current vote-propensity in $(0, 1)$, positive and negative feedback produce changes of equal size (in absolute value).[21]

(A2*) (negative feedback): If $\pi_{i,t} < a_{i,t-1}$ then with probability one $p_{i,t}(I_t) \leq p_{i,t-1}(W_{t-1})$, and there exists an $\epsilon > 0$ such that for all t and for all histories leading up to t, if $p_{i,t-1}(W_t - 1) > 0$ then with probability one $p_{i,t}(I_t) \leq (1 - \epsilon)p_{i,t-1}(W_{t-1})$.

(A2*) says that when voters are dissatisfied, they will always decrease their propensity to vote for the incumbent by a fixed percentage that is bounded away from zero.

Fact 2: Suppose the standard "Condorcet conditions" hold and there are infinitely many voters. All the voters are the same type, and all use the same AVoR, which is symmetric and order-preserving. The start is unbiased: $p_{i,0} = \frac{1}{2}$ for all i.

- **(i)** If $h_D > h_R > \frac{1}{2}$ then whichever party wins the first election is elected thereafter, with probability one.

[21] A simple type of equal-adjustment AVoR is a one-step walk along a finite grid of equally-spaced propensities, e.g., $(0, 0.01, 0.02, \ldots, 0.99, 1.0)$. But in other equal-adjustment AVoRs, propensity changes taper off as one moves toward zero or one. And in nonstationary equal-adjustment rules the size of changes could decrease over time as a person ages.

- **(ii)** If $h_D > \frac{1}{2} > h_R$ then two paths can occur: (1) $W_0 = D$ and D is elected thereafter, or (2) $W_0 = R$, but then R is defeated in period 1 and never wins office again.
- **(iii)** If $\frac{1}{2} > h_D > h_R$ and the AVoR satisfies (A2*) and equal-adjustment then every incumbent eventually loses office, with probability one.

Thus, whether sheer size ensures that the optimal outcome—governance by D—is attained depends on payoff-probabilities. If both parties give the good outcome over half the time then a problem of excessive inertia can arise: the electorate can get permanently stuck with the weaker party because that one is "good enough", collectively speaking, and a Law of Large Numbers ensures that a majority of (satisfied) voters will have their way with probability one. At the other end of this parametric spectrum—both parties produce the *bad* outcome over half the time—the opposite problem of excessive restlessness appears. In such circumstances, over half the voters are dissatisfied with the government's performance before every election, so the incumbent is always kicked out of office eventually. Here the electorate is too demanding for its own good; to paraphrase Voltaire, "the perfect is the enemy of the optimal" (Bendor and Kumar 2005). The limit of Condorcet's solution—having infinitely many voters—does nothing to dispel this problem; instead, it ensures, via a Law of Large Numbers, that over half the electorate are displeased with every incumbent's performance with probability one.

Only when the choice context takes on the shape of (ii) is the match between these AVoRs and the problem environment so good that we get a result similar to Condorcet's Jury Theorem. When the better party pleases more than half the voters *and* the weaker one satisfies less than half *then* large numbers of voters secures the optimal outcome with certainty.

As noted earlier, the difference between (myopic) optimization (compare the two parties and vote for whichever you think will perform better) and adaptive retrospection (compare the incumbent's performance to your aspiration level and then raise or lower your vote-propensity accordingly) is crucial. This difference is revealed sharply in either the benign environment of case (1) or the malign environment of case (ii). In the former, if the district chooses R in the first election, a majority of the voters will compare R's performance to their aspirations and find it good. Their purely retrospective voting rule doesn't then pose the question of how D would fare; after all, why fix what isn't broken? In the malign environment, persistently comparing the incumbent's performance to an internal standard generates an unwillingness to stick with the lesser of two evils. (The question, "why vote for the lesser of two evils?", makes little sense in a normative theory of rational choice, yet it seems to describe how unhappy voters sometimes think about their alternatives.)

Thus we see that Achen was right: retrospective voting really *does* differ from optimization. Fact 1 shows that the difference persists even as time goes to infinity; fact 2 shows that it can persist even as the number of voters goes to infinity. These theoretical results are important not because we will observe time or populations increasing without limit; they matter because they demonstrate that optimization and backward-looking adaptation differ in fundamental ways.

Retrospective voting and the benefits of representative democracy

In some ways facts 1 and 2 are too pessimistic: they reflect how cognitively constrained voters can fall short of optimality, even in the limit, but they do not examine how the voters' *agent*—the politician—might improve matters. The next results will show that representative democracy can ameliorate problems caused by simple retrospective voting.

To see the value added by representation, let us analyze how retrospective voting fares under *direct* democracy. Under the latter institution, citizens vote directly on policies, as they do for state initiatives and referenda. We consider a single issue area. An initial policy is randomly selected to be the status quo in t_0. It generates high or low payoffs for citizens, who then adjust their propensities to vote for it. In the election in t_1, if the status quo policy wins majority approval then it continues in force for another period. If the voters reject, then a new policy becomes the status quo. (For what follows we do not need to specify how replacement works.)

Voters can use a variety of AVoRs, provided that they all satisfy (A2*) (Because we are now considering direct democracy it should be understood that $p_{i,t+1}(I_{t+1})$ denotes citizen i's propensity to vote for the incumbent *policy*, not the incumbent politician.)

We will say that a policy is *stable* if, once installed as the status quo, it wins approval from the electorate thereafter, with probability one. Hence a stable policy is an absorbing state in the corresponding stochastic process. We assume that policies belong to X, a compact and convex subset of R^n and voters have ideal points located in X. (Per Stokes' (1963) objections, we don't assume that voters can describe R^n or even that they think in those terms. A voter's ideal point is just our way of representing her "underlying interests" here.) The probability that policy x gives citizen i the high payoff in t, denoted $h_{i,x;t}$, is strictly decreasing in the eucldiean distance between policy x and i's ideal point. Voters continue to vote retrospectively, based on what they observe (their realized payoffs), using AVoRs to update their propensities to vote for the status quo policy.

Fact 3: Suppose that the above spatial framework holds. If the AVoRs satisfy (A2*), then policy x is stable if *and only if* $h_{i,x;t} = 1$ for a majority of citizens, for all $t \geq 1$.

Fact 3 immediately implies that if no policy can *guarantee* good outcomes to any citizen (which seems plausible) then no policy is stable, given (A2*). *This is true even if the electorate is completely homogeneous*: fully informed and completely rational citizens would unanimously agree that a particular policy is the unique social optimum.

Thus Achen and Bartels (2004) are right to be concerned about what they call "blind" retrospective voting: in an imperfect world, where no policy generates high payoffs with certainty, citizens using *any* AVoR that satisfies (A2*) cannot settle down on the social optimum, and will take up suboptimal policies infinitely often. The defect is intrinsic to retrospective voting (as formalized by (A2*)): since even the optimal policy is imperfect, eventually it will generate low payoffs for a majority of citizens, who will reject it with positive probability. Eventually, then, even the best policy will be shot down in a direct referendum.

Now let us see how *representative* democracy fares.

Fact 4: If the hypotheses of fact 3 hold then the following conclusions obtain.

(i) A fully informed incumbent who maximizes the probability of staying in office in the current election is stable if *and only if* there is a policy s/he can implement which generates high payoffs to a majority of citizens with certainty, at every date.

(ii) If no such policy exists but the electorate is homogenous, then no incumbent is stable but every incumbent implements the electorate's common ideal point.

In a realistically imperfect world, fact 4 gives a "shark bite" result: retrospective voters eventually become disgruntled with every incumbent, *even when they implement the collectively optimal policy* and voters are blaming incumbents for events beyond their control—i.e., for shark attacks.

Despite this Athenian pattern[22], office-oriented politicians will, in environments identified by fact 4's part (ii), protect the voters *from themselves*. They'll do so not out of benevolence but because pleasing the voters is the best way to continue enjoying the perks of office, per Adam Smith's famous quote: "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest" (quoted in Downs 1957, p.28). Thus, we see that a key part of the Smith-Downs argument does not require that voters be fully rational, at least not in all contexts.

Of course, the strongest form of this phenomenon, as stated in part (ii), depends on there being a unique social optimum and on incentives lining up correctly: the socially optimal policy—the common bliss point—coincides with the one that maximizes electoral support. But this result is not fragile: one can show that if the electorate is "nearly" homogeneous and our normative theory requires that socially optimal policies be pareto optimal, then a myopically support-maximizing politician will implement a policy that is "close" to a socially optimal policy, if such policies exist. [23] (If they don't exist—i.e., the relevant normative theory's criteria are inconsistent—then the matter is moot: no intelligible claims about democracy's failures can be made.) Thus, the Smithian condition of incentives lining up correctly is *not* a knife-edge property.

Misperceptions. Although the cognitive operations represented in reduced form by (A1) and (A2) could be quite simple, they depend on a vital input: a citizen's perception of her own payoffs. And since payoffs in a model of voting depend on public policies and the evaluation of the latter is far from easy—policy evaluation is a speciality in its own right—it is important to allow for the possibility that ordinary citizens may err in evaluating an incumbent's performance.

Further, it is essential to know whether payoff-misperceptions will affect party-oriented voting. It is rather obvious that *systematic* errors could have such effects: e.g., if a left-leaning citizen always perceived his payoff as being low when the incumbent was Republican, no matter what the R's performance was, then this bias will make induce him to vote for Democrats more often than he otherwise would. The more interesting question is whether *un*biased misperceptions can affect

[22]Ancient Athens was notorious for often kicking even successful leaders out of power (Knox 1985). We thank Joshua Ober for this reference.

[23]The text's statement implicitly allows "social optimality" to include properties in addition to pareto optimality. Hence, the socially optimal set is a (possibly improper) subset of the pareto set.

party-oriented voting.

Our first result on misperceptions shows that Proposition 1's conclusion is robust with respect to unbiased mistakes—fortunately, given that many voters are quite ignorant of governmental conduct. (Note that this and the other misperception results allow error-rates to change over time: e.g., citizens might evaluate incumbents more accurately as they acquire more information about politics over time.)

Proposition 3: Suppose the hypotheses of Proposition 1 hold. Further, citizens misperceive their payoffs with probability $\psi_t \in (0, \frac{1}{2})$, which is independently and identically distributed across voters and independently over time. Then the conclusion of Proposition 1 obtains.

It is well-established that citizens vary greatly in their political sophistication (e.g., Delli Carpini and Keeler 1989; Kinder 1999; Luskin 2002). Hence we should expect them to make errors of misperception at different rates. The next result examines an interesting effect of this variation in sophistication.

Proposition 4: Suppose there are 2 types, where i has liberal interests and j has conservative interests. At date t, sophisticated i's have error probability $\psi_{i,S;t}$ while unsophisticated i's make mistakes with probability $\psi_{i,U;t}$, with $\psi_{i,S;t} < \psi_{i,U;t} < \frac{1}{2}$ for all t. Similarly, sophisticated j's err in t with probability $\psi_{j,S;t}$ and unsophisticated ones do so with probability $\psi_{j,U;t}$, with $\psi_{j,S;t} < \psi_{j,U;t} < \frac{1}{2}$ for all t. Otherwise the hypotheses of proposition 2 hold. Then a sophisticated i's probability distribution over her Democratic vote-propensities first-order stochastically dominates an unsophisticated i's, which stochastically dominates an unsophisticated j's, which in turn dominates a sophisticated j's, for all $t \geq 1$.

Thus the more sophisticated voters—those who are better at evaluating governmental performance—are more partisan than less sophisticated ones. (This is consistent with the picture of "Independents" drawn in *The American Voter* [1960, p.143].) Further, Proposition 3 implies that voting tendencies converge as evaluation-errors become more common.

We include the following result regarding the effect of biased errors just to verify that our intuitions are correct. The result relies on a notion of misperceptions having a partisan bias. Informally, if a voter has a partisan bias then her errors are skewed in favor of one of the parties and against the other one. (The formal definition is in the appendix.) This bias could arise from a top-down processing of information: e.g., someone who thinks, perhaps unconsciously, "if D is in power, then my payoff must be good; but if R is in power then my payoff must be bad."

Proposition 5: Suppose the hypotheses of Proposition 1 hold, except that i_1's misperceptions have a more partisan bias than i_2's do. If they have the same underlying interests, then i_1's probability distribution over his Democratic vote-propensities first-order stochastically dominates i_2's probability distribution for all $t \geq 1$.

Thus, greater random error tends to push citizens toward each other (proposition 3); systematic bias tends to polarize the polity (proposition 5).

IV. Explicit and Endogenous Aspirations: The Computational Model

Throughout this paper we have referred to a computational model that can express the full complexity of endogenous aspirations and more-than-binary payoffs. While a full exploration of this model awaits future work, here we briefly use it to illustrate what can be derived from the very general assumptions of (A1)-(A3). In particular, we show that, while our analytical results indicate that it is difficult for individual voters to fix on any one candidate, a population of voters (all of whom begin with unbiased propensities of $\frac{1}{2}$) can become strongly polarized, endogenously generating party affiliations where none previously existed.

The pictures displayed below were generated via a simulation that instantiates the general axioms of (A1)-(A3). Just as in the analytical model, the computational one begins with an election. Then, payoffs are distributed to all voters, who subsequently adjust aspirations and vote propensities according to set rules. These propensity adjustment rules may take a variety of forms, including Bush-Mosteller, equal search along a grid, or, for what is displayed below, logistic adjustment, wherein individual propensities trace out a logistic curve as they update. (Of course, all these belong to the set of adjustment rules defined by axioms (A1) and (A2).) Under the logistic rule, voters with propensities near $\frac{1}{2}$ react strongly to feedback, while voters with extreme propensities are more sluggish, requiring many failures to sway them from their earlier inclinations. Aspirations adjust linearly, as in Cyert and March (1963); this is consistent with axiom (A3). Voters' ideal points are uniformly distributed over [-3,3]x[-3,3], the two parties are located at (-.5,

-.5) and (.5, .5), and payoffs consist of a quadratic loss function plus a normally distributed shock.

Figure 1 displays the electorate at time t=1. The locations of the rectangles signify voter bliss points; as one can see, they are scattered uniformly over two dimensions. The darker (blue) rectangles house voters who have voted Democratic in the last election, and the lighter (red) rectangles Republican. Intermediate colors (purple) indicate spots with voters of both types. The histogram is of propensities; all here are clumped near $\frac{1}{2}$, their initialization point.

[figure 1 about here]

By ten periods, shown in figure 2, the distribution of propensities has already begun sorting itself out. The upper right half of the main view is largely red, and the lower left largely blue. The propensity distribution is also starting to look polarized, with a strongly bimodal distribution.

[figure 2 about here]

By period thirty six[24], the system is in its steady state, as seen in figure 3. The propensity distribution exhibits extreme polarization, with the huge majority of voters voting for D or R with probability near one, and the main view displays an electorate split nearly evenly between red and blue along the line $y = -x$. Though the underlying dynamic remains constant throughout, our probabilistic voters have developed strong party identification. Future work will explore how well this can be maintained when parties are free to change their policies over time.

[figure 3 about here]

V. Conclusions

V. O. Key was right: citizens don't need detailed ideologies; they can get by with retrospective voting. Many of these backward-oriented rules do not guarantee optimal voting in the short- or even the long-run. But they are sensible heuristics: they produce behavior that is ordinally consistent with voters' underlying interests and they are well-adapted to the informational and cognitive constraints shouldered by most citizens.

We have focused here on the behavior of voters and have mostly assumed exogenously fixed parties. This is a common modeling approach (e.g., Palfrey and Rosenthal 1985; Bendor, Diermeier

[24] The exact number of periods before extreme polarization depends on the rate of logistic adjustment and the size of the payoff shock, and can be longer or shorter than indicated here. For our purposes, though, we are more interested in the system's qualitative patterns than in its quantitative ones.

and Ting 2003). But its justification is based on the cognitive constraints of modelers, not on verisimilitude. In related work (Bendor, Kumar and Siegel 2005) we "close the loop" by constructing more complex models in which parties are also active players.

The formulations in this paper are what might be called pure models of retrospective voting: the citizens rely completely on politicians' past performances as a voting guide. This is the opposite of pure Downsian models, in which citizens rely completely on forecasts of politicians' future conduct in order to make their electoral choices. The truth is surely somewhere in-between these polar types (Fiorina 1981). In subsequent work we will modify the present models by adding a forward-looking component to the voters' decision making: they will (imprecisely) compare the candidates' platforms and their vote-choices will combine both retrospective and prospective components. We believe that Key's major insights—as well as most of the present paper's results—will continue to hold in this more complex formulation.

References

Achen, Christopher. 1992. "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior* 14,3:195-211.

Achen, Christopher, and Larry Bartels. 2004. "Blind Retrospection: Electoral Responses to Drought, Flu, and Shark Attacks". Unpublished manuscript (January 27). Princeton University.

Campbell, Angus, Philip Converse, Warren Miller and Donald Stokes. 1960. *The American Voter.* New York: Wiley.

Delli Carpini, M., and S. Keeter. 1996. *What Americans Know about Politics and Why It Matters.* New Haven: Yale University Press.

Diener, Ed, and Eunkook Suh. 1999. "National Differences in Subjective Well-Being." In D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-Being: The Foundations of Hedonic Psychology.* New York: Russell Sage Foundation.

Enelow, James, and Melvin Hinich. 1984. *The Spatial Theory of Voting.* Cambridge: Cambridge University Press.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50:5-25.

Ferejohn, John. 1995. "The Development of the Spatial Theory of Elections." In James Farr and John Dryzek (eds.), *Political Science and History.* Cambridge: Cambridge University Press.

Fiorina, Morris. 1981. *Retrospective Voting in American National Elections.* New Haven: Yale University Press.

Hinich, Melvin, and Michael Munger. 1994. *Ideology and the Theory of Public Choice.* Ann Arbor: The University of Michigan Press.

Jackson, John. 1975. "Issues, Party Choices, and Presidential Votes." *American Journal of Political Science* 19(2):161-85.

Key, V.O. 1966. *The Responsible Electorate: Rationality in Presidential Voting 1936-1960.* Cambridge MA: Belknap Press.

Kiefer J. and J. Wolfowitz. 1952. "Stochastic approximation of a regression function,' *Annals of Mathematical Statistics,* 23: 462–466.

Kinder, Donald. 1998. "Opinion and Action in the Realm of Politics." In D.Gilbert, S.Fiske and G.Lindzey (eds.), *The Handbook of Social Psychology*, 4^{th} ed. Boston: McGraw-Hill.

Knox R. A. 1985. "So mischievous a beaste? The Athenian Demos and its treatment of its politicians." *Greece & Rome* 32:132-161.

Luskin, Robert. 2002. "From Denial to Extenuation (and Finally Beyond): Political Sophistication and Citizen Performance." In James Kuklinski (ed.), *Thinking about Political Psychology.* New York: Cambridge University Press.

Miller, Nicholas. 1986. "Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem." In B. Grofman and G. Owen (eds.), *Information Pooling and Group Decision Making.* Greenwich CT: JAI.

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

Shepsle, Kenneth. 1996. "Statistical Political Philosophy and Positive Political Theory." In Jeffrey Friedman (ed.), *The Rational Choice Controversy.* New Haven: Yale University Press.

Stokes, Donald. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57(2):368-77.

Tversky, Amos, Paul Slovic and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80:204-17.

Wittman, Donald. 1995. *The Myth of Democratic Failure: Why Political Institutions are Efficient.* Chicago: The University of Chicago Press.

Wolfers, Justin. 2002. "Are Voters Rational? Evidence from Gubernatorial Elections." Research Paper No. 1730, Graduate School of Business, Stanford University.

Appendix

Here we provide proofs for the propositions and facts stated in the text. In some cases we actually prove more general results than given there; for clarity we include the text of all propositions here. We begin with technical definitions of those terms defined verbally in the text and required for the proofs: order preservation, party neutrality, and symmetry.

As stated in the text, we focus our analytic results on the special case of two outcomes' arising from an incumbent's actions: success due to them, or failure. Since there are two parties, these create four possible combinations of outcomes: 1) success with a D incumbent, 2) failure with a D incumbent, 3) success with an R incumbent, and 4) failure with an R incumbent. As each election cycle consists of two linked stochastic events, we may combine them into a single probability of each outcome's occurring, contingent on the propensity set from the previous periods. If we condition only on the previous period's state space, our problem is Markovian, but this is not necessary for our results (though it does simplify notation greatly).

The most general specification of this model has a different transition rule for each person, for every outcome, for all times. We thus define the following continuous functions, listed in the same order as the outcomes above: 1) $\alpha_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})$, 2) $\beta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})$, 3) $\gamma_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})$, 4) $\delta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})$, with $\{p_{i,s}\}_{s=0}^{t-1}$ corresponding to the history of all previous prpensities. Though these functions (collectively denoted a propensity adjustment rule (PAR)) may take any form, for simplicity we assume that they are deterministic, and so uniquely specify the new propensity set (many of our results generalize to non-deterministic functions, with appropriate assumptions on distributions, but this adds to the complexity without offerring additional insights, and so we stick to a deterministic formulation here). We change notation slightly from the text, and denote each individual's probability of achieving a success with a D as $h_{i,t}^D$, and with an R as $h_{i,t}^R$.With these definitions, an individual's propensity to vote for D after period t:

$$\begin{aligned} p_{i,t} &= W(\mathbf{p}_{t-1})(h_{i,t}^D\alpha_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) + (1-h_{i,t}^D)\beta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})) \\ &\quad +(1-W(\mathbf{p}_{t-1}))(h_{i,t}^R\gamma_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) + (1-h_{i,t}^R)\delta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1})), \end{aligned} \tag{1}$$

where $W(\mathbf{p}_{t-1})$ is the probability that D wins in this period, given the previous period's propensity set (this function hides much of the complexity of the model).

This formulation is exceedingly general, and it should come as no surprise that at times we must make further assumptions on the structure of the PAR. We defined each of these intuitively in the text; here we offer technical definitions.

Definition: A set of adjustment rules is *order-preserving* if it obeys the following conditions at all times, for all $p_{i,t-1} \geq p_{j,t-1}$, and for all $j \in [1, N]$:

$$\alpha_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) \geq \alpha_{j,t}(\{p_{j,s}\}_{s=0}^{t-1}), \tag{2}$$

$$\beta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) \geq \beta_{j,t}(\{p_{j,s}\}_{s=0}^{t-1}), \tag{3}$$

$$\gamma_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) \geq \gamma_{j,t}(\{p_{j,s}\}_{s=0}^{t-1}), \tag{4}$$

$$\delta_{i,t}(\{p_{i,s}\}_{s=0}^{t-1}) \quad \geq \quad \delta_{j,t}(\{p_{j,s}\}_{s=0}^{t-1}). \tag{5}$$

Note that we assume in the definition only that $p_{i,t-1} \geq p_{j,t-1}$, so that the rest of a voter's past history is allowed to vary freely, as long as it satisfies the conditions. In doing so, we are effectively sweeping the non-Markovian aspects of the model under the rug; voters are allowed to condition on their past propensities before the most recent period, as long as the functions so generated obey order-preservation, which is a statement only about the change in vote propensities between periods. Accordingly, for notational simplicitly, we will henceforth write each function as dependent only on $p_{k,t-1}$ for $k = i, j$, while keeping in mind its more complicated dependence. Because we will generally be dealing with probability distributions over propsensities, we will reserve $p_{k,t-1}$ to refer to degenerate, single-valued propensities, and $f_{k,t-1}(p_{k,t-1})$ to refer to probability distribution functions over the same.

Though use of this assumption does constrain the set of possible adjustment rules, it does so in an intutively appealing way. Simply put, a rule is order-preserving if the same outcome applied to two different original propensities preserves the ordering between the two; i.e. a lower propensity cannot leapfrog a higher one, when the impetus to do so is the same in both cases. This applies both within a person's own functions, so that a success with D cannot take one's propensity from .7 to .8, but also from .6 to .9, and between people, so that one person's cannot be too much more reactive than another's. The latter is particular strong, as it forces each person's functions to be identical at each time due to the functions' continuity in propensities.[25] Thus, we drop the voter subscripts from the functions in all subsequent results.

Definition: A set of adjustment rules is *symmetric* if $\alpha(x) = 1-\beta(1-x)$ and $\delta(x) = 1-\gamma(1-x)$ for all x.

Definition: A set of adjustment rules is *party neutral* if $\alpha(x) = 1 - \gamma(1 - x)$ and $\beta(x) = 1 - \delta(1 - x)$ for all x.

Definition: A set of adjustment rules satisfies *equal adjustment* if $\alpha(x) - x = x - \beta(x)$ and $\delta(x) - x = x - \gamma(x)$ for all x.

[25] If they were not the same, then there must exist a point x at which one voter's function, say i's α, is greater than another's α, so that $\alpha_{i,t}(x) > \alpha_{j,t}(x)$. But then continuity implies $\alpha_{i,t}(x) > \alpha_{j,t}(x + \epsilon)$ for some $\epsilon > 0$, which violates order-preservation.

A symmetric PAR has all voters treating a decrease from propensity x the same as an increase from propensity $1-x$, and thus matches the typical definition of the term when viewed graphically. A party-neutral PAR has all voters treating a failure from R and a success from D as the same, and vice versa, and thus voters are not biased by the party of the incumbent. When both hold, $\alpha(x) = \delta(x)$ and $\beta(x) = \gamma(x)$, for all x, so that there are only two paths each voter's propensity can take in every period. Equal adjustment implies that adjustment up and down are the same at each point, though these can be different at different propensities. Whenever we assume equal adjustment, we will also make the subsidiary assumption that the PAR is defined for all propensities $x \in [0, 1]$.

With these, we can prove the texts claims. Note that the proof of the first proposition does not make use of party neutrality, though this is a maintained assumption in the text. Because we have done so to ensure that it is clear that we have not biased the outcomes, rather than for strictly mathematical reasons, we explicitly state whenever it is necessary here. We also prove the results from the text slightly out of order, for simplicity of explication.

Proposition 1: Let voter i and voter j be part of the same electorate, and assume order-preservation holds. If i is at least as liberal as j, so that $h^D_{i,t} \geq h^D_{j,t}$ and $h^R_{i,t} \leq h^R_{j,t}$, and i is at least as likely as j to vote for D in the previous period, so that $f_{i,t-1}(p_{i,t-1})$ first order stochastically dominates $f_{j,t-1}(p_{j,t-1})$, then the same relation holds in the following period; i.e. $f_{i,t}(p_{i,t})$ first order stochastically dominates $f_{j,t}(p_{j,t})$.

Proof: First we show that the conclusion holds when the previous period's distributions are degenerate, so that $p_{i,t-1} \geq p_{j,t-1}$, and then we extend this to the case of non-degenerate distibutions. One distribution first order stochastically dominates (fosd) another whenever the cumulative distribution function (cdf) of the dominant distribution never exceeds that of the greater at any point. Since the cdf of both propensity distributions here consists of a series of two step functions (one at each outcome possibility, given deterministic adjustment and the fact that the electoral uncertainty must resolve identically for each), a sufficient (though not strictly necessary) set of conditions for fosd to hold in this setting is for the first "step" of the dominant distribution to occur later than the other, and to have a lesser height. The heights here correspond to the probability of achieving each outcome, which are either $1 - h^D_{k,t}$ and $h^D_{k,t}$ if D wins, or $1 - h^R_{k,t}$ and $h^R_{k,t}$ if R wins, for

$k = i, j$. Consider first the case if D wins. By (A1) and (A2), $\alpha_t(x) \geq \beta_t(x)$ for all x, so the height of the first step is $1 - h^D_{i,t}$ for i, which is no more than $1 - h^D_{j,t}$ by assumption. Further, since $p_{i,t-1} \geq p_{j,t-1}$, $\beta_t(p_{i,t-1}) \geq \beta_t(p_{j,t-1})$ by order-preservation, j's step cannot occur second, thus proving fosd for this case. The case for R is identically shown, with the only change that the first step is $\gamma_t(x) \leq \delta_t(x)$. This proves the degenerate case, which is equivalent to the expression: $\int_0^x f_{i,t}(p_{i,t}|p_{i,t-1})dp_{i,t} \leq \int_0^x f_{j,t}(p_{j,t}|p_{j,t-1})dp_{j,t}$ for all x. To show the non-degenerate case, we just need to show that this expression holds for the unconditional distributions $f_{k,t}(p_{k,t})$ as well. These can be written as $f_{k,t}(p_{k,t}) = \int_0^1 f_{k,t}(p_{k,t}|p_{k,t-1})f_{k,t-1}(p_{k,t-1})dp_{k,t-1}$, an expression which calculates the unconditional distribution by weighting the conditional distributions for each $p_{k,t-1}$ by the distribution of $p_{k,t-1}$ and integrating over all possible values of $p_{k,t-1}$. If the set of conditional distributions for each voter were identical, identical distributions over $p_{k,t-1}$ would result in identical unconditional distributions. However, under our assumption that $f_{i,t-1}(p_{i,t-1})$ fosd the same for j, j's conditional distribution is weighted more heavily for low values of $p_{j,t-1}$ than i's is for $p_{i,t-1}$. Further, $f_{i,t}(p_{i,t}|p_{i,t-1})$ fosd $f_{j,t}(p_{j,t}|p_{j,t-1})$ for $p_{i,t-1} \geq p_{j,t-1}$, so relatively more weight is put for j than for i on conditional distributions that are biased toward low final propensity values. This implies that $f_{i,t}(p_{i,t})$ fosd $f_{j,t}(p_{j,t})$ as well, completing the proof. [26]

Proposition 2: Suppose voter i and voter j live in different electorates, and i has liberal interests while j's are conservative. Assume they each use the same order-preserving, symmetric, and party-neutral AVoR.

- (i) If i is at least as likely as j to vote for D in the previous period, so that $f_{i,t-1}(p_{i,t-1})$ first order stochastically dominates $f_{j,t-1}(p_{j,t-1})$, then the following period $f_{i,t}(p_{i,t})$ first order stochastically dominates $f_{j,t}(p_{j,t})$.
- (ii) If $p_{i,0} \geq \frac{1}{2}$ and $p_{j,0} \leq \frac{1}{2}$, then for all future periods, the *ex ante* probability of i's voting for D exceeds $\frac{1}{2}$ and the *ex ante* probability of j's voting for D is less than $\frac{1}{2}$.

[26] It helps to visualize an infinite array of cdf's here, each occurring with a probability given by the likelihood that a particular initial propensity obtains. By the first part of the proof, lower initial values lead to "higher" cdf's, in the sense that they put more probability on lower values of the final propensity. The unconditional cdf is the sum over all of these cdf's, weighted by their probabilities of occurring. If the weighting for j puts more weight than i's on small initial propensities, j's sum will skew "higher" in the sense above, and so be dominated by i's.

Proof: The manner of proof for (i) is the same as in Proposition 1, differing only in the definition of the two "steps." Thus we need prove the claim for degenerate initial propensities; the proof for moving to the non-degenerate case is identically accomplished. Symmetry and party neutrality imply $\alpha(x) = \delta(x)$ and $\beta(x) = \gamma(x)$, for all x; as $p_{i,t-1} \geq p_{j,t-1}$, order preservation implies that each step for j cannot occur after the corresponding step for i. So, as long as the height of the first step for j meets or exceeds that of the first for i, the claim holds. This amounts to the inequality $(1 - W_j(\mathbf{p}_t))h^R_{j,t} + W_i(\mathbf{p}_t)(1 - h^D_{i,t}) \geq (1 - W_i(\mathbf{p}_t))h^R_{i,t} + W_i(\mathbf{p}_t)(1 - h^D_{i,t})$. Because j has conservative interests, the left hand side is the convex combination of two numbers greater than a half, and so it must also be greater than a half. Similarly, because i has liberal interests, the right hand side is the convex combination of two numbers less than a half, and so it also must be less than a half. The combination of these two statements implies that the inequality holds always, and the claim is proved. (ii) follows from (i) and the law of total probability; since each subsequent period's distribution over vote propensities fosd's the one before, and since the voters have respectively liberal and conservative interests, the *ex ante* probabilities of voting for D are increasing and decreasing, respectively.

Proposition 3: Suppose the hypotheses of Proposition 1 (or 2) hold. Further, citizens misperceive their payoffs with probability $\psi_t \in (0, \frac{1}{2})$, which is independent and identically distributed across voters and independently over time. Then the conclusion of Proposition 1 (or 2) obtains.

Proof: This follows directly from the proofs to the relevant propositions; indeed, the perfect information case can be seen as a special case of this more general model. To see this, note that adding an error $\psi_t \in (0, \frac{1}{2})$ corresponds to changing the chance of success from $h^{D,R}_{i,t}$ to $(1-\psi_t)h^{D,R}_{i,t} + \psi_t(1-h^{D,R}_{i,t}) = \psi_t + h^{D,R}_{i,t}(1-2\psi_t)$. Clearly, if $h^{D,R}_{i,t} \geq h^{D,R}_{j,t}$, then $\psi_t + h^{D,R}_{i,t}(1-2\psi_t) \geq \psi_t + h^{D,R}_{j,t}(1 - 2\psi_t)$, and since the proofs of the previous two propositions depend on the former condition's holding, they still go through in the case of identical error.

If errors are not identical, so that some individuals are better informed than others, the the same logic holds. As $\epsilon + h^{D,R}_{i,t}(1 - 2\epsilon)$ is decreasing in ϵ for $h^{D,R}_{i,t} > \frac{1}{2}$ and increasing for $h^{D,R}_{i,t} < \frac{1}{2}$, increasing levels of error increasingly reduces the effects of holding extreme (i.e. towards 1 or 0) interests, thus biasing such voters toward moderation. We formalize this general result in the following proposition, which has the text's Proposition 4 as a special case.

Proposition 4': Within either the setting of Proposition 1 (for the single electorate case), or Proposition 2 (for the different electorates case), consider two voters i and j of identical types, with $h^D_{k,t} > \frac{1}{2} > h^R_{k,t}$ for $k = i, j$, save only that i apprehends success with error $0 \leq \psi_{i,t} \leq \psi_{j,t} \leq \frac{1}{2}$. Then if $f_{i,t-1}(p_{i,t-1})$ first order stochastically dominates $f_{j,t-1}(p_{j,t-1})$, the same relation holds in the following period; i.e. $f_{i,t}(p_{i,t})$ first order stochastically dominates $f_{j,t}(p_{j,t})$.

Proof: The proofs from proposition 1 or 2 apply directly, after replacing all $h^{D,R}_{k,t}$ by $\psi_{k,t}+h^{D,R}_{k,t}(1-2\psi_{k,t})$. The results follow since $\psi_{i,t} \leq \psi_{j,t}$ implies $\psi_{i,t} + h^D_{i,t}(1 - 2\psi_{i,t}) \geq \psi_{j,t} + h^D_{j,t}(1 - 2\psi_{j,t})$, and $\psi_{i,t} + h^R_{i,t}(1 - 2\psi_{i,t}) \leq \psi_{j,t} + h^R_{j,t}(1 - 2\psi_{j,t})$.

Proposition 4: Suppose there are 2 types, where i has liberal interests and j has conservative interests. At date t, sophisticated i's have error probability $\psi_{i,S;t}$ while unsophisticated i's make mistakes with probability $\psi_{i,U;t}$, with $\psi_{i,S;t} < \psi_{i,U;t} < \frac{1}{2}$ for all t. Similarly, sophisticated j's err in t with probability $\psi_{j,S;t}$ and unsophisticated ones do so with probability $\psi_{j,U;t}$, with $\psi_{j,S;t} < \psi_{j,U;t} < \frac{1}{2}$ for all t. Otherwise the hypotheses of proposition 2 hold. Then a sophisticated i's probability distribution over her Democratic vote-propensities first-order stochastically dominates an unsophisticated i's, which stochastically dominates an unsophisticated j's, which in turn dominates a sophisticated j's, for all $t \geq 1$.

Proof: The results within each sophisticated/unsophisticated pair follow from Proposition 4' applied to two different pairs of voters, with the i pair having $h^D_{i,t} > \frac{1}{2} > h^R_{i,t}$ and the j pair having $h^D_{j,t} < \frac{1}{2} < h^R_{j,t}$. The comparison between pairs follows from the fact that no error ever causes the effective chance of success $\psi_{k,t} + h^{D,R}_{k,t}(1 - 2\psi_{k,t})$ to cross $\frac{1}{2}$, so that Proposition 2's comparison between voters with liberal and conservative interests still holds for all sophistication levels of each type.

Proposition 5: Suppose the hypotheses of Proposition 1 hold, except that i_1's misperceptions have a more partisan bias than i_2's do. If they have the same underlying interests, then i_1's probability distribution over his Democratic vote-propensities first-order stochastically dominates i_2's probability distribution for all $t \geq 1$.

Proof: This follows from Proposition 1 and 3 directly; any bias in misperceptions may be absorbed into $h^{D,R}_{k,t}$, as this is a measure of perceived rather than absolute success, so that bias merely makes a voter more liberal or more conservative, leaving the rest of the assumptions unchanged.

Fact 1: Suppose $n = 1$ and the voter uses a stationary AVoR. If optimal voter-choice isn't trivially guaranteed—$h_D \neq h_R$ —and parties are imperfect—$\max\{h_D, h_R\} < 1$—then the citizen's voting never becomes optimal.

Proof: For $n = 1$, (1) becomes: $p_t = p_{t-1}(h_t^D \alpha_t(p_{t-1}) + (1 - h_t^D)\beta_t(p_{t-1})) + (1 - p_{t-1})(h_t^R \gamma_t(p_{t-1}) + (1 - h_t^R)\delta_t(p_{t-1}))$. Assume without loss of generality that $h_D > h_R$. If $p_{t-1} < 1$ then there is a non-zero probability that R wins in period t, and so voting is not optimal. If $p_{t-1} = 1$, then under the assumptions of the fact and (A1) and (A2) there is a non-zero probability that $p_t < 1$, and so there is a non-zero probability that R wins in period $t + 1$, proving the claim.

Fact 2: Suppose the standard "Condorcet conditions" hold and there are infinitely many voters. All the voters are the same type, and all use the same AVoR, which is symmetric, party neutral, and order preserving. The start is unbiased: $p_{i,0} = \frac{1}{2}$ for all i.

- **(i)** If $h_D > h_R > \frac{1}{2}$ then whichever party wins the first election is elected thereafter, with probability one.
- **(ii)** If $h_D > \frac{1}{2} > h_R$ then two paths can occur: (1) $W_0 = D$ and D is elected thereafter, or (2) $W_0 = R$, but then R is defeated in period 1 and never wins office again.
- **(iii)** If $\frac{1}{2} > h_D > h_R$ and the AVoR satisfies (A2*) and equal-adjustment then every incumbent eventually loses office, with probability one.

Proof: The proof for all three uses the same fundamental logic. Under (A1)-(A3), upon a success (failure) with D (R), voters must adjust their propensity to vote for D upwards, and similarly for the opposite case. Since there is an infinite population, success probabilities correspond directly into percentages of the population adjusting their propensities in this manner; symmetry implies that adjustments from symmetric locations around $\frac{1}{2}$ will be equal in both directions. If $h_D > h_R > \frac{1}{2}$, then over half the population will increase their likelihoods of voting for the winner of the first election, by the same amount that less than half the population decreases theirs, implying the average propensity will surpass 50% and the incumbent will win with certainty in the next period. Standard inductive logic implies that this will remain true in all subsequent periods. If $h_D > \frac{1}{2} > h_R$, then two things could happen. If D wins the first election, then the previous logic holds and it wins in every subsequent period. If R wins the first election, then by similar logic the average propensity to vote for D increases over 50%, and D wins in every period thereafter. Finally, if $\frac{1}{2} > h_D > h_R$, and (A2*) and equal adjustment hold, then we may split the population at any given time t into two infinite subgroups: one consisting of all voters with $p_i > 0$, and one with $p_i = 0$. By equal adjustment and (A2*), the former group must see its mean propensity decrease in every period, eventually passing below .5, while the latter group sees its mean weakly increase in every period. Because of the action of the latter group, it is possible that the overall population mean will increase in a given period. However, positive feedback cannot take $p = 0$ to a point higher than .5. To see this, recall that we have assumed for this case that the PAR is defined over all points in $[0, 1]$. As a result, by order preservation, the point $p = 0$ cannot go under positive feedback to a point higher than what its neighbor goes to, which cannot reach .5 due to equal adjustment, since if it did that would imply the existence of a propensity below zero, which is impossible. Thus, in every period, the population of voters consists of two subgroups: one with mean decreasing toward zero, and one with mean varying in [0,.5). The overall mean of the population must therefore decrease below .5 at some point, implying the incumbent eventually loses, with probability 1.

Fact 3: Suppose that the above spatial framework holds. If the AVoRs satisfy (A2*), then policy x is stable if *and only if* $h_{i,x;t} = 1$ for a majority of citizens, for all $t \geq 1$.

Proof: For x to be stable, there must be a strictly zero probability of voting x out. x wins whenever a majority vote to keep it; x wins with certainty if a majority votes with certainty to keep it. To achieve certainty in this setting, every member of a majority of the populace must always be satisfied with x; otherwise by (A1) and (A2*) there will be a nonzero probability that less than a majority of the population will vote for x with certainty in the following election, implying a nonzero chance that x will lose the next election. Thus, stability for x equates to having $h_{i,x;t} = 1$ for a majority of citizens, for all $t \geq 1$.

Fact 4: If the hypotheses of fact 3 hold then the following conclusions obtain.

(i) A fully informed incumbent who maximizes the probability of staying in office in the current election is stable if *and only if* there is a policy s/he can implement which generates high payoffs to a majority of citizens with certainty, at every date.

(ii) If no such policy exists but the electorate is homogenous, then no incumbent is stable but every incumbent implements the electorate's common ideal point.

Proof: By Fact 3, policy x is stable if *and only if* it generates high payoffs to a majority with certainty. Since x always wins if and only if case obtains, an incumbent who chooses policies wins always if and only if this case obtains, proving (i). If such a policy fails to obtain, the lack of stability stated in part (ii) follows directly from (i). The rest of (ii) follows from the fact a homogenous electorate's common ideal point maximizes the probability that the electorate will be satisfied with the policy chosen, which implies that this policy maximizes the probability of re-election for the incumbent who chooses it.

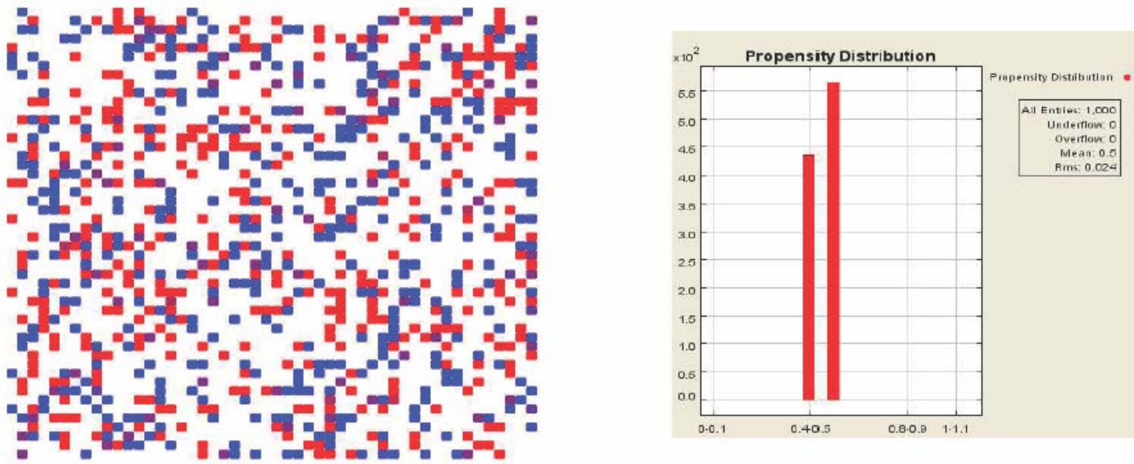

Figure 1: 1 Period

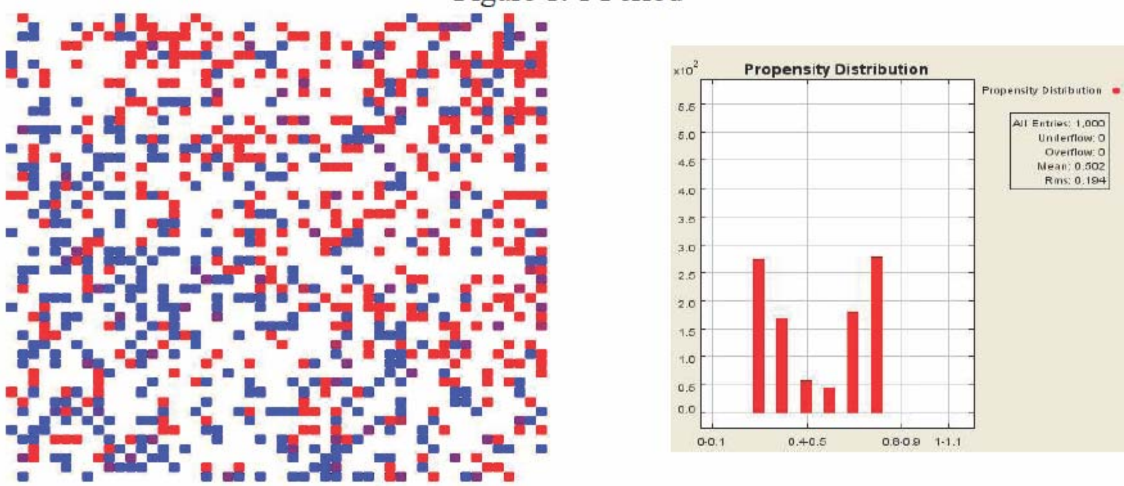

Figure 2: 10 Periods

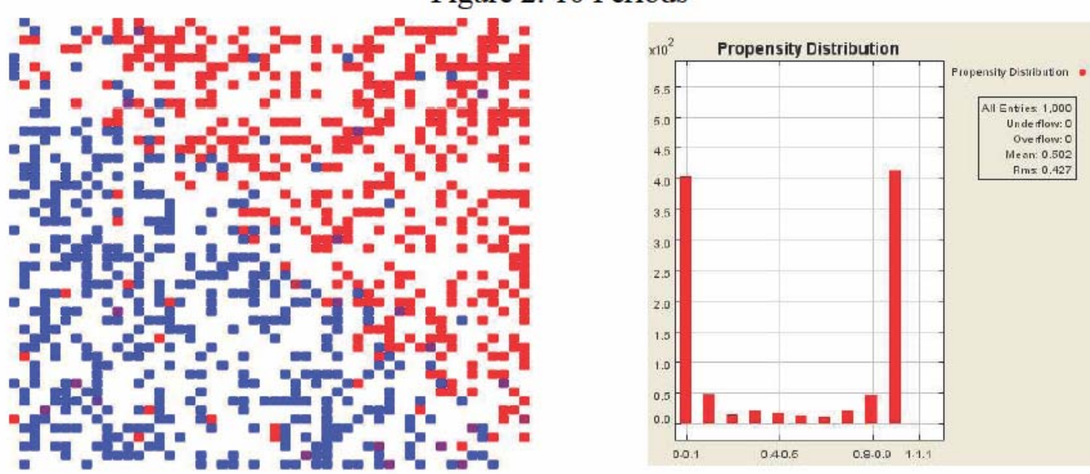

Figure 3: 36 Periods