

# Self-Enforcing Democracy

James D. Fearon  
Department of Political Science  
Stanford University  
DRAFT; Comments welcome\*

March 6, 2006

## Abstract

If democracy is to have any of the good effects said to justify it, it must be *self-enforcing*. Those who control the government must want to hold regular, competitive elections for the highest offices, and all parties must be willing to comply with the results. I consider a model in which citizens can always protest or rebel against the current ruler, but can unseat the ruler only if enough people rebel. When individuals privately observe a signal of the government's performance (e.g., their own welfare), they face a difficult problem of how to coordinate to pose a credible threat of rebellion necessary to keep the ruler from stealing. Further, if the signals are noisy, inefficient rebellions must occur in equilibrium to keep the ruler honest. Allowing for the possibility of elections makes for equilibria that eliminate both problems. The convention of holding elections at particular times provides a public signal for coordinating rebellion in the event that elections are suspended or blatantly rigged. The electoral results themselves aggregate private information about the ruler's performance, providing the ruler an incentive to stick to the terms. In the case of noisy signals of government performance, the electoral results act as a cheap talk signal that allows the public to commit to rebel if a losing ruler does not step down, avoiding the need for costly rebellions. These arguments pose an explanation for self-enforcing democracy, whereas the several models in the literature do not because they do not explain why anyone would want to use elections to allocate power.

## 1 Introduction

The standard justification for electoral democracy as a form of government is that competitive elections give leaders an incentive to provide public goods, and more generally to align public policy with citizens' preferences. In economists' jargon, democracy is viewed as a solution to a

---

\*Thanks to Nikolay Marinov, Robert Powell, Jim Robinson, and Barry Weingast for comments on an earlier, quite different version of this paper.

principal-agent problem with multiple principals (the citizens) and an agent (the government or ruler) who may face strong temptations, or moral hazard.

Work in political economy over the last 50 years has given us a plethora of models examining elections as an incentive system for producing public policies. The tradition of spatial models initiated by Downs (1957) studies how elections affect the policies chosen by office-motivated candidates. Within or beside this tradition, there is a literature focused on the moral hazard problems that arise if politicians are not purely office-motivated.<sup>1</sup> These models tend to support the intuition that elections can motivate politicians to do what voters want, but to a degree that depends on how well voters can link outcomes to politicians' actions (which may depend on the quality of the media, electoral rules, the structure of government, or the nature of different issues, among other things).

All such work assumes, however, that fair, competitive elections will be held. Elections are always taken as a given, a part of the structure of interactions being analyzed. But for electoral accountability to have any positive effects, incumbents must choose to hold free elections and all must choose to observe the results. If it is to work, democracy must be "self enforcing," as Przeworski (1991) put it, since at the level of a country there is no third party that can enforce a contract binding the ruler to hold elections.

Democracy might be rendered self-enforcing in a country if its potential rulers internalize a strong normative attachment to democratic "rules of the game." At least since Locke's *Second Treatise on Government*, however, liberal political theorists have not imagined that the problem could be solved by assuming (or creating) virtuous leaders. The temptations of power are too strong, and in any event, intense political competition could make almost any faction believe that it was doing the right thing by preventing the other side from taking over. Instead, the usual recourse is to suggest that for democracy to be self-enforcing, the public, or some significant part of it, must

---

<sup>1</sup>See Besley (2006) for an excellent survey and analysis.

be motivated *to protest or even rebel* if democracy is threatened.

People might have this motivation either because they have internalized democratic norms or culture – a commitment to the “rules of the game” – or because they expect that they will be materially worse off if dictatorship prevails. The post-World War II political science and sociological literatures on democratic transitions and stability stressed democratic norms and culture. More recently, some political scientists and economists have offered models in which people or out-of-power parties can have material motivations to defend democracy. In these studies an implicit threat of rebellion either leads to a transition to democracy, or ensures that the party in power will choose to hold elections and observe the results.<sup>2</sup>

In this paper I argue that existing accounts of self-enforcing democracy fail because they do not explain why elections would be needed to select and discipline leaders. In the models of Acemoglu and Robinson (2001), Boix (2003), and Przeworski (1991), competitive elections are completely unnecessary to implement the equilibrium outcomes they associate with democracy. In the approach of Acemoglu and Robinson and Boix, a dictatorship of the person with the median level of income will serve just as well as using elections. In Przeworski’s (1991) informal model, coin flips could replace elections, and a system in which an unelected ruler divides up the spoils in a certain proportion each period would be better still if the parties are risk averse.

The puzzle is this: If a leadership can be induced to hold elections by the implicit threat of protest or rebellion, then why do we need elections at all to get the leadership to provide public goods and choose policies in accord with public preferences? Why not just use the implicit threat of protest or rebellion? What value is added by competitive elections?

I argue that the institution of regular elections solves a coordination problem posed by imper-

---

<sup>2</sup>For norms or culture-based accounts, see in particular Almond and Verba (1963) and Lipset (1963). This idea also seems implicit in Barrington Moore’s influential *Social Origins of Democracy and Dictatorship* (1966), in which the bourgeoisie just carry around democracy and want to implement it. The idea of “self-enforcing democracy” as such comes from Przeworski (1991), who argued material motivations could be sufficient. See Persson and Tabellini (2005) for a model in which both motivations play a role.

fect monitoring. Individuals or out-of-power factions at best observe private signals about whether the ruler is complying with the terms of a distributive bargain. This means that to pose a credible threat of rebellion, they need some *public signal* that will allow them to coordinate opposition to a would-be dictator. In addition, they need a publicly observable measure of whether the ruler is keeping to the distributive bargain, in order to give the ruler an incentive to keep to it. Elections solve both problems at once. The convention of holding them at particular times – *not* the electoral outcome itself – provides a public signal for coordinating rebellion in the event that elections are suspended or blatantly rigged.<sup>3</sup> And the electoral results aggregate private information about the ruler’s performance, providing the ruler an incentive to stick to the terms.

I consider several models in which a large population of citizens and a ruler interact in successive periods. The ruler chooses each period how to distribute a fixed amount of resources (“public goods”) among the population, keeping whatever remains for herself. Citizens observe a signal of what the ruler did, and then have the option of protesting, which is costly. If enough citizens protest, the ruler is deposed.

I show that if signals are privately observed and there is a minimum size below which protests are not publicly visible, then the only equilibria of the game are “dictatorial” – the citizens get nothing from the ruler. Introducing the possibility of elections makes for equilibria in which the ruler distributes positive amounts, indeed, as much as they could get if the ruler’s actions were publicly observed.

If the citizens observe a noisy signal of the ruler’s policy choice – for example, if they observe their welfare which is a stochastic function of policy – then a new problem arises for control of the ruler in the absence of elections. With noisy signals, rebellions must sometimes occur if the ruler is to be kept honest, but this is costly for the citizens. I show that elections can resolve this problem by acting as a cheap talk signal that enables the public to commit to rebel if a losing incumbent

---

<sup>3</sup>I discuss the problem of partially rigged elections in the conclusion.

does not step down. In equilibrium, everyone knows that the electoral results “draw a line in the sand,” engaging a reputation that is valuable for the citizens when they cannot directly observe the government’s policy choices or their impact on public welfare. It is socially more efficient for rulers to exit under an implicit threat of rebellion than via actual rebellion.

There are relatively few papers in the literature in which electoral democracy is treated as, in effect, a choice variable. The contributions most relevant to this paper (by Przeworski and Acemoglu and Robinson) are discussed in the first section below. The analysis here owes a significant debt to Weingast (1997), who interprets a political constitution as an equilibrium in a repeated game in which two social groups agree to rebel collectively if the “sovereign” transgresses against the rights of either group. The models considered below are akin to an  $n$ -player, distributive politics version of Weingast’s game, but dropping the assumption that all players observe exactly what the sovereign does, and introducing elections and the problem of self-enforcing democracy into the analysis.

Ferejohn (1986), Baron and Ferejohn (1989), and Persson and Tabellini (2000, 238) consider distributive politics models in dynamic settings, although in these elections are exogenously given. In Ferejohn’s “nonhomogenous electorate” case, a large number of citizens are assumed to be able to announce and commit to thresholds such that if they get at least their threshold amount from the politician, they will vote to reelect. This sets up an auction-like situation that drives everyone’s threshold to zero (the same tendency appears in Persson and Tabellini’s model). The dynamic does not occur in the electoral versions of the games studied here because I do not assume either stationarity or that citizens can publicly commit to thresholds.

The next section reviews existing models of self-enforcing democracy in more detail. The third section presents a succession of models, beginning with the case in which the public observes perfectly what the ruler gives to everyone, and proceeding to the case of private and noisy signals with and without elections. The conclusion informally considers the problem of partially

fraudulent elections, and some empirical implications of the analysis.

## **2 Why use elections to allocate power?**

In Carl Schmitt's view, "democracy" means a government that rules according to the will of the people. Elections and parliaments are just one possible means of determining this will (Schmitt 1988[1923]). Thus, a regime like Hitler's could be a democracy if its policies express the popular will.

The Anglo-American tradition and current understanding takes a more procedural approach, defining democracy as a system of government that uses competitive elections based on a broad franchise to fill the highest political offices.<sup>4</sup> Given this definition, one does not have an explanation for how democracy can be a self-enforcing political system unless one can explain why people would want to use elections to select leaders. The several models of "self-enforcing democracy" in the literature do not provide such an explanation, or imply explanations that misunderstand the point of electoral democracy.

Przeworski (1991, ch. 1) introduced the idea of self-enforcing democracy, arguing that democracy could survive as a result of "self-interested strategic compliance" even in the absence of "normative commitments" to the system for its own sake. He sketches a model of a repeated interaction between two factions or parties that can try to grab power each period by violent means, or they can agree to hold an election, which is represented as a lottery. His main point is that both parties may prefer to hold elections and abide by the results if their electoral odds give them higher expected payoffs than they get by trying to grab power by force.

Suppose, for example, that the value of controlling government for one period is set at 1,

---

<sup>4</sup>A position forcefully developed by the Austrian Joseph Schumpeter (1975[1942]).

and that there are two parties, call them the rich and the poor.<sup>5</sup> Their probabilities of winning at the polls are  $p_R$  and  $p_P$  respectively, with  $p_R + p_P = 1$ . If both parties discount future payoffs by the factor  $\delta \in (0, 1)$ , then party  $i$ 's value for democracy is  $p_i/(1 - \delta)$ ,  $i \in \{R, P\}$ . Suppose that party  $i$  can use violence to grab power permanently with probability  $q_i = 1 - q_j$ , and that doing so reduces the size of the pie in subsequent periods by the factor  $\alpha \in (0, 1)$ . It is then easy to show that we can support an equilibrium in which the parties agree to hold elections and abide by the results provided that  $p_R \in [(1 - \alpha)q_R, (1 - \alpha)q_R + \alpha]$ . In words, this means that democracy is feasible and preferable provided that each party's chance of taking power by elections is similar enough to its chance of grabbing power by force. For example, if the rich are unlikely to win at the polls (low  $p_R$ ), but have control of the military and so can grab power easily in a coup (high  $q_R$ ), then democracy is not feasible if the costs of rebellion are not too large.

In this argument, the purpose of elections is to act as a sort of coin flip that allows the parties to share control of the government through time. Of course, a literal coin flip (or some randomizing device), or an alternation scheme, would be even better than elections, since these can be calibrated to the military odds rather being fixed at the electoral odds. Further, the whole argument depends on the assumption that the benefit of controlling government is indivisible and side-payments are not possible. This assumption implies that the only way to share the "pie" of rule is through time. By contrast, if the pie is divisible and the parties are risk averse, elections are unnecessary and undesirable. Both sides would prefer to let one manage the government, with the understanding that failure to observe the terms of the political bargain would lead to mutual attempts to grab power by force.<sup>6</sup> This account of self-enforcing democracy fails because it renders

---

<sup>5</sup>This model is a slightly more developed version of the model sketched in Przeworski (1991).

<sup>6</sup>Przeworski (2005, 270) briefly addresses this point, arguing that elections are needed because it is impossible to write a complete contract to cover all the issues that can be decided by the ruler. "Hence, to avoid violence, the conflicting political forces adopt the following device: agree over those issues that can be specified and allow the residual control to alternate according to specified probabilities." But elections are bizarrely elaborate and ineffectual if they are nothing more than a randomizing device.

elections puzzling rather than comprehensible.<sup>7</sup>

Acemoglu and Robinson (2000, 2001) examine models in which rich and poor players choose whether to hold elections, and have the option of rebelling if they are unhappy with the status quo. In their approach elections are explained as a device for implementing a particular distribution of income, via the choice of a linear tax rate.<sup>8</sup> In the 1999 paper, they argue that the rich may be unable to commit to implement a tax rate in the future that would make the poor prefer not to rebel in the present, due to random fluctuations in the ability of the poor to mobilize. In a period when the poor are strong, the rich may then prefer to allow democracy, in which elections will select the tax rate preferred by the median voter (who is presumed to be poor). Doing so avoids a revolution, which is even worse for the rich than democracy because it destroys resources.

Acemoglu and Robinson's analysis could be said to explain "democracy" in the sense of Schmitt – that is, it is a democracy if the policies accord with the will of the median income individual. But it clearly does not explain electoral democracy, since elections are completely unnecessary to implement the median income individual's preferred tax rate. These might as well be models explaining the dictatorship of the proletariat rather than electoral democracy.

Electoral democracy is not properly understood as a device for conflict resolution by randomizing who controls the state, or as a device for credibly committing to a particular income distribution. The next sections formalize the more traditional notion that the point of elections is to make public private information about the performance of the government, so giving the leadership an incentive to perform well. The novel feature of the models is that elections are a choice variable

---

<sup>7</sup>Przeworski (2005) presents a more complicated version of his 1991 model, in which the "pie" is divisible via the choice of a linear tax rate that determines the extent of income redistribution. A new assumption is introduced to make elections more attractive than dictatorship. Namely, without elections, all citizens suffer "physical insecurity" that lowers their enjoyment of income. But if the idea is that democrats can commit not to abuse human rights whereas dictators cannot, then the explanation for elections concerns accountability, not randomization or conflict resolution. If, by contrast, the model is interpreted literally, then it explains democracy by assuming that people prefer it, which is the opposite of the approach in Przeworski (1991).

<sup>8</sup>Boix (2003) and Benhabib and Przeworski (2005) take the same approach.

and the ruler has no incentive to hold them unless the public (or a large enough fraction of it) can pose a credible threat of rebellion. This introduces problems of coordination that I argue are at the heart of the problem of establishing and sustaining democracy.

### 3 Elections and rebellion

In this section I consider a series of simple models that together develop an explanation for the use of elections to allocate political power. All the models have the same players and basic structure. The players are always a ruler and a population of  $n \geq 1$  citizens, who interact in successive periods  $t = 0, 1, 2, \dots$ . In each period, the ruler moves first, deciding how to distribute resources worth  $v$ , with citizen  $i$  getting  $x_{it}$  and the remainder,  $v - \sum_i x_{it}$ , going to the ruler. The “citizens” may be thought of as individuals, or as groups or factions that are able to share information and coordinate on a common choice.

The “pie”  $v$  can be thought of as the national income and  $x_{it}$  what citizen  $i$  is left with after taxes in period  $t$ , or  $x_{it}$  could be considered an amount of some targetable “public good” that is produced by the state. What I have in mind, however, is that the government chooses from a rich and complicated set of policy instruments that ultimately determine the well-being ( $x_{it}$ ) of every citizen or group in each period.

In the first model, all citizens observe the complete allocation,  $\mathbf{x}^t = (x_{1t}, x_{2t}, \dots, x_{nt})$ . In the second model, citizens observe their own allocation  $x_{it}$  but not what anyone else got, and in the third model citizens privately observe a noisy signal of their own allocation. After observing whatever they observe, citizens individually choose whether to “rebel” or protest. The cost of rebelling is  $c > 0$ .

If a sufficient number protest ( $k$  is the threshold number), then the current ruler is deposed and a new ruler is drawn for the next period from a population of potential rulers. For notational

simplicity, I take the set of rulers to be large and distinct from the citizens. Provided  $n$  is large, we could modify the analysis below so that new rulers are drawn from the body of citizens at random without affecting any substantive results.

In the models with private observation of welfare and with noisy private signals, I show that the threat of rebellion alone may not be enough to allow the citizens to get as much as they could under perfect monitoring. For these cases I introduce variants in which the ruler has the option to hold elections and then step down depending on the electoral results.

### 3.1 Model 1: Public information about government actions

To review, the sequence of actions in each period  $t$  is:

1. The incumbent ruler  $R_t$  receives total tax revenues  $v > 0$ , and chooses how to distribute these. Individual  $i$  gets  $x_{it}$ , leaving  $v - \sum_i x_{it}$  for the ruler.
2. The citizens observe the complete allocation  $\mathbf{x}^t = (x_{1t}, \dots, x_{nt})$ , and then choose simultaneously whether to protest or not. Protesting costs an individual  $c > 0$  regardless of the outcome.
3. If  $k$  or more citizens protest (where  $2 \leq k \leq n$ ), then the current ruler is deposed, and a new ruler is drawn at random from the pool to start in the next period. Deposed rulers and rulers in the pool get zero each period.

All citizens, including the ruler, discount payoffs by a common factor  $\delta \in (0, 1)$ .

Not surprisingly, this game can have multiple equilibria, including ones in which the ruler consumes the whole “pie” and the citizens remain passive. If one expects that no one else will protest, then there is no reason to pay the cost oneself since one demonstrator or rebel will not change the regime. Call an equilibrium *dictatorial* if all citizens get zero on the equilibrium path. For later use, I spell out one such equilibrium.

**Proposition 1.** The above game has a dictatorial equilibrium in which players choose the following strategies. (1) The ruler always sets  $x_{it} = 0$  for all  $i$ , regardless of the history of play. (2) Citizens always choose not to rebel, regardless of the allocation  $\mathbf{x}^t$  in period  $t$  and the past history of allocations and protests.

**Proof.** We need only consider deviations in a single period. Clearly the ruler lowers her payoff by consuming less than  $v$  in period  $t$ , since doing so does not change the behavior of the citizens in any future period. And a one-stage deviation to protesting by any citizen merely costs this person  $c$ , with no future benefit, again since protesting does not change anyone else's behavior in the future.

□

Note that we can support the dictatorial equilibrium under two modifications of the model that will be considered below. First, we can change the model so that citizens observe only their own allocation  $x_{it}$  each period, and no one else's. Second, we could have citizens privately observe a noisy signal of what the ruler did, for example,  $z_{it} = x_{it} + \epsilon_{it}$  where  $\epsilon_{it}$  is a random variable with zero mean. As long as everyone expects that no one will rebel and that protesting by oneself will not spark a rebellion subsequently, it is individually rational to remain passive.

The more interesting question is whether and how we can support allocations in which the citizens get positive payoffs in equilibrium. To induce the ruler to distribute goods, she must face some punishment for failing to do so, which in the end must stem from the possibility of being kicked out and getting zero subsequently. And since protesting is individually costly, citizens will do it only if they expect that failure to act will switch them to a different (subgame) equilibrium in which they get a lower payoff. The next proposition gives conditions under which such a scheme can support an allocation  $\mathbf{x}$  that gives citizens positive payoffs in equilibrium.

**Proposition 2.** We will say that a *social consensus* exists in period  $t$  if whenever a ruler has in some prior period given less than  $x_i$  to any citizen, all citizens protested. The following strategies will then support  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  each period in a subgame perfect equilibrium, provided that

the two conditions on payoffs given below hold. (1) If a social consensus exists in period  $t$ , the ruler distributes  $\mathbf{x}$ ; otherwise, the ruler gives zero to all citizens. (2) If a social consensus exists in period  $t$ , citizen  $i$  protests if the ruler gave less than  $x_j$  to any citizen  $j$ , and not otherwise. If there is no social consensus, then citizen  $i$  does not protest regardless of what the ruler distributes.

The conditions for this strategy to form a subgame perfect equilibrium are

$$\sum_i x_i \leq \delta v \text{ and} \tag{1}$$

$$x_i \geq c(1 - \delta)/\delta \text{ for all } i. \tag{2}$$

**Proof.** When a social consensus exists, the ruler gets  $(v - \sum_i x_i)/(1 - \delta)$  by sticking with the proposed equilibrium strategy in period  $t$ , whereas she can get at most  $v$  by deviating (to keeping the whole pie in period  $t$  and then getting kicked out). So the ruler is happy to stick with the proposed strategy provided that  $\sum_i x_i \leq \delta v$ . Citizen  $i$  has no incentive to deviate to protesting on the equilibrium path, since this just costs  $c > 0$  and has no effect on what anyone does in future periods. Off the path, if some citizen  $j$  has received less than  $x_j$  in period  $t$  and a social consensus exists,  $i$  is willing to protest if  $-c + \delta x_i/(1 - \delta) \geq 0$ , or  $x_i \geq c(1 - \delta)/\delta$ . If a social consensus does not exist, the strategies are the same as for the dictatorial equilibrium in Proposition 1.  $\square$

Public welfare is maximized by setting  $\sum_i x_i = \delta v$ , so in an equal distribution each citizen gets  $\delta v/n$  and the ruler takes  $v(1 - \delta)$  each period. In this optimal allocation for the public, the ratio of the ruler's per-period payoff to the total received by the citizens is  $(1 - \delta)/\delta$ , and is thus determined completely by the discount factor.<sup>9</sup> The two conditions together imply that  $v \geq nc(1 - \delta)/\delta^2$  is necessary for such an equilibrium to obtain, which implies further that the players must be sufficiently patient.

---

<sup>9</sup>The discount factor here could represent the probability that the current ruler maintains power rather than being deposed in a coup led by a would-be ruler from the pool (coup attempts succeed with probability  $1 - \delta$ ). So exogenous political instability associates with lower maximum possible payoffs for the public.

In this equilibrium, everyone will protest and so depose the ruler if she gives less than what the “political contract” specifies to any citizen. The implicit threat of rebellion keeps the ruler honest. Obviously, no elections are needed here to get this outcome, and introducing elections would not allow the citizens to do any better. This is because condition (1) must obtain in *any* equilibrium of the game (and in all the versions considered below), since the ruler can always get  $v$  by grabbing the whole pie in any given period.

Since protesting is costly, there is a potential free rider problem among the citizens. If not everyone is needed to protest to depose the ruler ( $k \leq n$ ), why not let others do it if the circumstance arises? The free rider problem is solved in this equilibrium by means of a *collective reputation*. Citizens are willing to pay the costs of rebelling because if they don’t, they will lose their reputation, with all expecting to revert to the dictatorial equilibrium. For this to be credible, condition 2 must hold, which says that each citizen  $i$  has to be getting enough from the government to make it worthwhile to pay the costs of rebellion to restore  $x_i$ . (Put differently, the worst off would not fight simply to restore the old regime.)

A few points are worth making about Model 1 before proceeding to the next version. First, it can be viewed as an  $n$ -player, distributive politics version of the game analyzed by Weingast (1997). In Weingast’s model, two groups observe each period whether a sovereign “transgresses” against the rights of one, both, or neither group, and then decide whether to acquiesce or challenge the sovereign. Challenges are costly, and only a joint challenge can reverse the transgression. Weingast describes an efficient equilibrium in which both groups rebel if *either* one is transgressed against, because each expects that if it does not come to aid of the other, the “social consensus” on reciprocity will be broken and the sovereign will subsequently transgress against both. The same dynamic occurs here. The equilibrium above depends on citizens rebelling even if the ruler transferred benefits away from some other citizen to him or her. The logic is that the ruler cannot be trusted to continue to benefit you in the future, since your fellow citizens will no longer come

to your aid by rebelling if you are “cut out” in the next period. Hence it makes sense to support the rebellion even if you benefited from the ruler’s violation of the political contract.<sup>10</sup>

Second, it seems rather strong that to maintain the collective reputation, *everyone* must rebel if the ruler violates the contract. And in fact we can change the equilibrium strategies so that for each period  $t$  there is a subset of the citizens with at least  $k$  people who know that if a violation occurs, they are ones expected to rebel to depose the leader.<sup>11</sup> They are again induced to do so by the belief that if each one does not do his part, the social consensus will be broken and dictatorship will ensue. This is perhaps a bit more plausible than imagining that every citizen has to protest to maintain the consensus. But even so, in a very large population who would notice if I stayed home?

One possible response is to argue that it makes more sense to take the “citizens” in the model to be *groups* that have already overcome internal collective action problems. This would suggest an argument for the importance of civil society. Perhaps branches of government other than the executive, or non-state, societal groups and associations, help prevent dictatorship by lessening the free rider problem and so making a collective reputation for willingness to contest dictatorship easier to sustain.<sup>12</sup>

Another possible response is to note that the expected costs of protesting generally decrease with the number of protesters. Indeed, if so many protest that it becomes fairly safe, it may actually be *fun* to join in; see for a recent example the Orange Revolution in Ukraine. If we assume in model 1 that protesting is costless when at least  $k$  protest, then condition 2 disappears and there is no problem motivating people to rebel against violations of the political contract.

---

<sup>10</sup>As in Weingast’s model, note that the “social consensus” does not refer to a consensus on preferences or values, in that the citizens have sharply conflicting preferences over possible allocations  $x$ . Rather, “social consensus” here refers to a reciprocity agreement in opposing deviations from the political contract by the ruler.

<sup>11</sup>This allows for a slight weakening of condition 2, since it needs to be satisfied only for the  $k$  people who are on call to protest in each period  $t$ .

<sup>12</sup>On the role of civil society here, see Putnam (1993).

Third, we can also support equilibria in Model 1 in which a subset of the citizens with at least  $k$  people (or groups) get positive payoffs, whereas all other citizens get nothing. The former set receive payoffs because they have coordinated to maintain a collective reputation, so that they can credibly threaten to depose the ruler if she does not distribute as much as they expect. The rest get nothing because they are atomized, each for himself. This mirrors the situation in many dictatorships in which the regime requires the support of a small number of social groups, such as the military and large landowners, to stay in power. Interestingly, the smaller the number it takes to depose the ruler in the model ( $k$ ), the larger the maximum possible payoffs for the included group. Arguably, this resonates with the situation in many subSaharan states, where small groups can often succeed in winning control of the state through a coup. In these countries the mass of the population is typically rural, ethnically divided, and seemingly unable to coordinate on collective opposition, while leaderships are often highly kleptocratic.

### **3.2 Model 2: Private observations of government performance**

In Model 1, the ruler's choice of how to allocate taxes and benefits (or other policies that affect citizens' welfare) is publicly observed and common knowledge. This made it easy to construct an equilibrium in which the mass of citizens hold the ruler accountable, by making it possible for them to coordinate a collective response to the ruler's actions.

If we think of the allocation  $\mathbf{x}^t$  as the net effect on individuals of a complex set of public policies bearing on taxes and benefits, or if we think of the citizens as individuals in a large population, then it is implausible that everyone observes what everyone else gets from the ruler. So consider what happens if we change Model 1 by assuming that citizens observe their own allocation  $x_{it}$  but no one else's. To be complete, assume that potential rulers "in the pool" observe only the history of protests, and citizens do not observe anyone else's payoffs at the end of each round.

Call a strategy for a citizen *stationary* if he conditions rebellion in period  $t$  only on the

allocation  $x_{it}$  received in that period. Call an equilibrium *peaceful* if there is no protest on the equilibrium path, and note that if an equilibrium is not peaceful, it is not Pareto efficient. Then we have the following result:

**Proposition 3.** Suppose that citizens observe only their own allocation  $x_{it}$  in each period, and all citizens use stationary strategies. Then there is no peaceful Nash equilibrium in which any citizen receives a positive payoff. In other words, if citizens use stationary strategies, a peaceful equilibrium is necessarily dictatorial when citizens observe only their own allocation.

**Proof.** Suppose to the contrary that some citizen  $i$  gets a payoff  $x_{it} > 0$  on the equilibrium path. By deviating to  $x_{it} = 0$  (and changing no other allocations), the ruler increases her payoff in period  $t$ . Since in a peaceful equilibrium no  $j \neq i$  protests in period  $t$ , protest by  $i$  will not unseat the ruler, and since citizen strategies in subsequent periods do not condition on what happened in period  $t$ ,  $i$ 's protest in period  $t$  cannot affect the ruler's payoffs subsequently. Thus the ruler does better by deviating and  $x_{it} > 0$  cannot hold on the equilibrium path.  $\square$

The dilemma is that collective action is required to hold the ruler accountable, but information about violations of the political contract is held privately. For the citizens to pose a credible threat of rebellion they require a means of making their private signals public. Below I will argue that by threatening to rebel if the ruler fails to hold elections (a publicly observable signal), the citizens get the government to provide it with a cheap means of revealing their private information.

Arguably, though, elections are not necessary to expose depredations of the ruler. In the model as it stands, it is possible, though not particularly easy, to construct equilibria in which citizens play nonstationary strategies and get public goods. In these, citizens use individual protests as a way of signaling to other citizens that the ruler has violated the political contract.

To ease notation and make the set up more natural, I introduce a slight modification to the structure of Model 1. Suppose that after privately observing their allocations  $x_{it}$ , the citizens choose simultaneously whether to protest. They observe how many protested in total, and then

**Table 1.** Private observations of ruler policies, with and without elections

Model 2	Model 2 <sup>DEM</sup>
<ol style="list-style-type: none"> <li>1. <math>R_t</math> distributes <math>\mathbf{x}_t</math>; citizen <math>i</math> observes only <math>x_{it}</math>.</li> <li>2. Citizens simultaneously choose whether to protest, and observe the total number protesting.</li> <li>3. If ruler was not deposed in step 2, citizens again simultaneously choose whether to protest.</li> </ol>	<ol style="list-style-type: none"> <li>1. <math>R_t</math> distributes <math>\mathbf{x}_t</math>; citizen <math>i</math> observes only <math>x_{it}</math>.</li> <li>2. <math>R_t</math> chooses whether to hold elections.</li> <li>3. If no elections, citizens simultaneously choose whether to protest. If an election was held, citizens observe total voting for and against <math>R_t</math>.</li> <li>4. If an election was held, <math>R_t</math> chooses whether to step down.</li> <li>5. Citizens observe <math>R_t</math>'s decision and again simultaneously choose whether to protest.</li> </ol>

simultaneously choose *again* whether to protest (if the result of the first protest was not to depose the current leader). In other words, we allow two potential phases of protest within a period, thus admitting the possibility that if a citizen sees a protest or rebellion in the first phase, he may choose to join in the second. The sequence of actions for the resulting stage game, call it Model 2, is shown in Table 1.

The assumption that individuals never observe what other citizens get makes Model 2 more difficult to analyze than Model 1. The game now has only one proper subgame, the whole game itself, which means that subgame perfection has no ability to refine away Nash equilibria that rely on incredible threats. For this we need to use perfect Bayesian equilibrium, specifying citizens' beliefs about what other citizens got when the ruler does something unexpected.

**Proposition 4.** We will say that a social consensus exists in period  $t$  if in any prior period in which between one and  $k - 1$  citizens rebelled in phase 1, all citizens rebelled in phase 2, or all but one if there was only one protester in phase 1. Under conditions (1) and (2) above and assuming that  $k < n - 1$ , the following strategies can support both a subgame perfect and a perfect

Bayesian equilibrium in Model 2 in which rulers distribute  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  each period on the equilibrium path.

- If no social consensus exists, citizens never rebel and the ruler keeps everything for herself.
- If a social consensus exists, in period  $t$  citizen  $i$  rebels in phase 2 if and only if any other citizen  $j \neq i$  rebelled in phase 1.
- If a social consensus exists, each citizen  $i$  can be in one of two “states” with respect to the current ruler  $R_t$ . Whenever a new ruler comes to power, all  $i$  “reset” to state 1.

State 1:  $i$  has always received at least  $x_i$  from the current ruler, or if not,  $i$  rebelled. In state 1,  $R_t$  gives  $x_{it} = x_i$ , and  $i$  rebels in phase 1 if and only if  $x_{it} < x_i$ .

State 2:  $i$  failed to rebel after receiving  $x_{is} < x_i$  from the current ruler in some period  $s < t$ . In state 2,  $R_t$  gives  $x_{it} = \hat{x}_i \equiv x_i - c(1 - \delta)/\delta$ , and  $i$  rebels in phase 1 if and only if  $x_{it} < \hat{x}_i$ .

**Proof.** See Appendix.

So despite citizens’ observing only their own allocation from the ruler, we can support exactly the same set of payoffs for citizens in equilibrium as in Proposition 2, which considered the case where the ruler’s actions were common knowledge. The trick is to have citizens condition a general rebellion on seeing any one person protest. To make this work, the public as a whole must be motivated to act collectively in support of an individual (or small group) who protests, and individuals must be motivated to publicly reveal violations by the ruler. We accomplish the former by the same device as in the public information case – by having each member of the public fear losing a collective reputation for willingness to rebel if any one member fails to join in.<sup>13</sup>

We accomplish the latter by giving individuals an incentive to preserve a *private reputation* in their dealings with the ruler. Each citizen expects that if he fails to rebel when shortchanged by

---

<sup>13</sup>Again, note that a collective reputation is not necessary if we are willing to assume that more than  $k$  citizens are happy to protest if they are assured that it will result in the government falling.

the ruler, the ruler will switch to giving him a smaller payoff ( $\hat{x}_i < x_i$ ) that makes him indifferent between rebelling to get a new leader and keeping the current ruler (who will continue giving  $\hat{x}_i$ ). If this ruler gives him less than  $\hat{x}_i$ , the citizen is thus willing to rebel, so the ruler does not deviate to something even less than  $\hat{x}_i$ .

If we think of the players as individuals in a large population, then this equilibrium depends on the implausible possibility that a protest by any one person can spark a successful revolution. In a large population a protest is likely to be invisible in the sense of not being observable by and common knowledge among all citizens unless it passes some threshold size. Or there may be a certain level of background noise in form of a random number of people who protest regardless of the consequences. Or perhaps very small protests can be quietly quashed by the government well before they become common knowledge.

Returning to the model, suppose that to be “visible” to the other citizens, a protest or demonstration requires at least  $m$  people, where  $m$  is greater than one.<sup>14</sup> Then we have a much more general analogue to Proposition 2, which is proved by essentially the same argument.

**Proposition 5.** In Model 2, if a protest is not publicly visible to nonparticipants unless it has at least  $m > 1$  participants, then any peaceful Nash equilibrium is dictatorial.

One way to solve this problem (it is a problem from the perspective of the public) is to have a small number of groups in society that can share information internally and make collective decisions. In the model, if  $n$  is small and the players are groups, then it is far more plausible that any one group could mobilize in such a way as to make its protest common knowledge. Such a solution brings to mind the arguments some have made on behalf of the role of civil society in supporting democracy. But this solution may also characterize relatively institutionalized oligarchies where the executive depends on the support of a small number of organized groups, such as the military, industry, trade unions, and so on. Or we might think of clan or ethnic group leaders with an

---

<sup>14</sup> $m$  could be a random variable drawn at the start of each period and observed by neither citizens nor ruler.

informal agreement to police the actions of a paramount ruler, or branches of government in a separation-of-powers system.

Alternatively, the problem can be neatly solved by the institution of elections, even in the case of a large, atomized population. With elections, the public induces the ruler to provide a cheap means of aggregating private information about what the ruler did, by threatening to rebel if she fails to hold elections (a publicly observable signal).

Consider the following modification of model 2, called  $2^{DEM}$ . After distributing  $\mathbf{x}^t$ , the ruler in period  $t$  can opt to hold an election in which citizens vote for or against the current ruler. If no election is held the citizens simultaneously choose whether to rebel. If an election is held the numbers voting for and against are publicized, and the ruler then has the opportunity to step down and return to the pool of rulers. After this choice, citizens again have the opportunity to rebel. If the ruler steps down, his opponent in the election (a random draw from the pool) takes power.<sup>15</sup> The sequence of events for the stage game is given in Table 1. We do not need to specify any particular election rule (e.g., unanimity, majority rule) since the election rule can be endogenous to the equilibrium.

**Proposition 6.** We will say that a social consensus exists in period  $t$  if in any prior period when a ruler failed to hold elections, or failed to step down after receiving less than  $n$  “Yes” votes, all citizens rebelled. If conditions (1) and (2) are satisfied, then Model  $2^{DEM}$  has a perfect Bayesian equilibrium with the following strategies.

- If no social consensus exists, citizens never rebel and the ruler keeps everything for herself.
- If a social consensus exists, the ruler in period  $t$  distributes  $\mathbf{x}^t = (x_1, x_2, \dots, x_n)$ , holds an election, and steps down if there was not unanimous support.
- If a social consensus exists, citizen  $i$  rebels in the first phase in period  $t$  if the ruler failed to

---

<sup>15</sup>If the ruler steps down without an election, a new ruler is drawn at random from the pool.

hold elections, and in the second phase if the leader failed to step down after not receiving unanimous support. Otherwise  $i$  does not rebel. If elections are held in period  $t$ , citizen  $i$  votes in favor of the incumbent ruler if he received  $x_{it} \geq x_i$ , and against if he got less.

**Proof.** See Appendix.

The institution of elections thus shifts the burden of spreading news about the ruler's behavior from individual citizens – where it may be too great to bear, as Proposition 5 suggested – to the government itself, by threatening rebellion if fair elections are not held on schedule.

The equilibrium described in Proposition 6 uses a unanimity rule, but we can support the same maximum payoffs for citizens (on average) using simple majority rule. Modify Proposition 6 so that a social consensus exists if all citizens rebelled whenever a ruler failed to hold elections or failed to step down after receiving less than half the votes. Let the ruler randomly choose  $(n + 1)/2$  citizens each period (if  $n$  is odd), and give  $x^* = 2\delta v/(n + 1)$  to each of these citizens in each period on the equilibrium path. Citizen  $i$  votes Yes if he receives at least  $x^*$ , and No otherwise. Then we have an equilibrium in which citizens receive an expected payoff of  $\delta v/n$ , as in the best equal distribution equilibrium in the full information case. Note again that the election rule is a social construct in this model rather than an exogenously specified given.<sup>16</sup>

### 3.3 Model 3: Noisy government performance

The previous section showed that if citizens privately observe what they get from the government, then they require some means of publicizing their individual information if they are to pose a credible threat of collective action, which is in turn necessary to keep the government honest. Elections can solve this dilemma nicely. In principle, the possibility of individual protest could as

---

<sup>16</sup>Majority rule does have implications for the full set of  $\mathbf{x}$  that can be supported on the equilibrium path. For instance, with  $\mathbf{x}$  the same every period, the ruler's incentive is to distribute positive amounts only to the  $(n + 1)/2$  citizens with the smallest  $x_i$ 's.

well. But this requires the implausible assumption that protest by any one person can reliably start a mass movement.

This section explores a second type of problem with relying on protests alone rather than elections with protests as the implicit threat. If the citizens observe not the government's choice of policy but only a noisy signal of this choice – for example, their self-perceived welfare – then their only hope of keeping the government honest is to condition the threat of rebellion on the noisy signal. This means that mistakes will occur even if the ruler is sticking to the distributive bargain, so that citizens will sometimes have to pay the costs of rebelling. Periodic rebellions are of course inefficient. They also lower the maximum payoff possible for the public, both directly via the costs of rebelling and indirectly by lowering the ruler's value for office. (More likely to be deposed in the future, she has to be paid more to make corruption relatively less attractive than compliance.)

Elections greatly alleviate the problem by providing the citizens with a means of publicly committing to rebel if the ruler does not step down peacefully. Instead of having to exercise the costly threat, elections allow the public to put at risk a reputation for being willing to protest, and so to credibly threaten rebellion without having to rebel.

Because this mechanism has nothing to do with coordination issues in a large population, and to show how it works as clearly as possible, I consider a model with a single citizen who can unseat the current ruler by rebelling (at cost  $c > 0$ ). The “citizen” could also be thought of as a faction, group, or branch of government that can act coherently.

The stage game for Model 3 is as follows (as before, payoffs are discounted by the factor  $\delta$ ).

1. The ruler in period  $t$ ,  $R_t$ , chooses to give  $x_t \in [0, v]$  to the citizen and keeps the remainder for herself. The citizen does not observe the ruler's choice of  $x_t$ .
2. Both citizen and ruler observe the citizen's level of welfare,  $z_t$ , but potential rulers not in office in period  $t$  do not (they observe only the history of protests).  $z_t$  equals  $x_t$  with probability

$\alpha \in (0, 1)$ , and zero with probability  $1 - \alpha$ . Thus, there is a  $1 - \alpha$  chance that the citizen experiences “bad luck” or “bad times” even if government policies were favorable.

3. The citizen chooses whether to rebel or not. Let  $d_t \in \{0, 1\}$  represent this decision. Rebellion deposes the current ruler and installs a new one from the pool, but at cost  $c > 0$  for the citizen.

Arguably, it is more realistic to assume that the ruler does not observe the citizen’s welfare  $z_t$ . For one thing, welfare is subjective. Below I consider the model under this latter assumption, but for clarity I begin with the simpler case.

Proposition 7 in the Appendix describes a perfect Bayesian equilibrium of Model 3 in which, on the equilibrium path, the ruler distributes  $x^* = v\alpha\delta$  each period to the citizen, and the citizen deposes the ruler whenever he observes welfare  $z_t < x^*$ . Thus the ruler loses power with probability  $1 - \alpha$  each period. The citizen is motivated to pay the costs of rebellion by the expectation that if he does not rebel when  $z_t < x^*$ , the current ruler will subsequently switch to giving him  $\hat{x} = x^* - c(1 - \alpha\delta)/\alpha\delta$ . Off the path (after failing to rebel if he got less than  $x^*$  from the current ruler), the citizen rebels if he sees  $z_t < \hat{x}$ .  $\hat{x}$  is the amount that makes the citizen indifferent between rebelling to get a new ruler (who would give  $x^* > \hat{x}$ ) and continuing with the current ruler (who will give  $\hat{x}$  until the citizen throws him out in “bad times”).

The proof of Proposition 7 also demonstrates that the citizen’s (time-averaged) expected payoff in this equilibrium,

$$(1 - \delta)V^C = v\alpha^2\delta - c(1 - \alpha), \quad (3)$$

is the highest possible equilibrium payoff for the citizen in Model 3. Finally, this equilibrium exists provided that  $v/c \geq (1 - \alpha\delta)/(\alpha\delta)^2$ . (If this condition on the parameters does not hold, then the only equilibria are dictatorial.)

To understand the impact of imperfect monitoring on the citizen’s ability to control the government, compare expression (3) with the analogous one for Model 1, in which the citizens were

assumed to observe exactly what the government chooses. The relevant comparison is to Model 1 with one citizen who can depose the ruler ( $n = k = 1$ ), and with the citizen receiving  $x_t$  with probability  $\alpha$  if given  $x_t$ , and zero with probability  $1 - \alpha$  (but observing the ruler's choice of  $x_t$ ). In this case, Proposition 2 implies that the (time-averaged) maximum equilibrium payoff for the citizen is  $v\alpha\delta$ , which is attainable provided that  $v/c \geq (1 - \delta)/\alpha\delta^2$ .

With imperfect monitoring, the citizen's best possible payoff (equation 3) is reduced compared to the perfect monitoring case, both by an additional factor  $\alpha$  applied to the term  $v\alpha\delta$ , and by the expected per-period cost of rebellion  $c(1 - \alpha)$ . Both reductions result from the citizen's inability to distinguish whether bad times are the result of ruler malfeasance or bad luck. To keep the ruler honest the citizen can at best condition rebellion on whether times are bad rather than the ruler's actions, which means that the citizen pays the expected cost  $c(1 - \alpha)$  each period. Moreover, the  $1 - \alpha$  chance of rebellion reduces the ruler's value for office, which makes her more tempted to steal the whole pie. The citizen must, in effect, pay the ruler more to keep her honest, and this also reduces the maximum the citizen can get relative to perfect monitoring.<sup>17</sup>

Allowing for an "election" can improve matters for the citizen. To start simply, suppose we allow the citizen to make a public announcement of "Leave" or "Stay" after observing  $z_t$ . We also allow the ruler to choose whether to step down or stay in office after the citizen's announcement.<sup>18</sup>

Even though the announcement is pure cheap talk (it has no effect on any player's payoffs), the game now has an equilibrium in which the ruler gives  $x^* = v\alpha\delta$  every period on the equilibrium

---

<sup>17</sup>Notice that the constraint for equilibrium in Model 1,  $v/c \geq (1 - \delta)/\alpha\delta^2$ , is less strong than the analogous constraint in Model 3, due to additional  $\alpha$ 's in the numerator and the denominator. The constraint reflects the fact that the ruler needs to get enough in equilibrium to prefer distributing  $x^*$  to grabbing the whole pie, while at the same time the citizen must be getting enough to make rebelling worthwhile if the ruler deviates. With imperfect monitoring there is an additional  $\alpha$  in the denominator because the ruler's payoff is reduced by a higher likelihood of getting thrown out. The additional  $\alpha$  that increases the numerator results from the fact that rulers in the pool do not observe  $z_t$ , and so cannot continue the punishment of a "state 2" citizen after  $R_t$  has been deposed. This aspect of imperfect monitoring constrains the maximum punishment for a citizen, which makes for a larger set of parameters such that the citizen cannot be motivated to rebel after defection at the same time as the ruler is motivated not to steal.

<sup>18</sup>Potential rulers "in the pool" can observe the citizen's announcement and whether  $R_t$  complies.

path, and the citizen never rebels. Say that a social consensus exists if the citizen has always rebelled if a ruler did not step down after the citizen announced “Leave.” Consider the following strategies:

- If a social consensus exists, the ruler in period  $t$  distributes  $x^* = v\alpha\delta$ , and steps down if the citizen says “Leave.”
- If a social consensus exists, the citizen says “Stay” in period  $t$  if he observes  $z_t \geq x^*$ , and says “Leave” otherwise. If the ruler in period  $t$  fails to step down when the citizen said “Leave,” the citizen rebels.
- If no social consensus exists, the ruler gives  $x_t = 0$ , the citizen says anything (Leave or Stay), the citizen never rebels, and the ruler never leaves.

With these strategies, the citizen will rebel if the ruler fails to comply with an order to leave provided that  $-c + \delta\alpha x^*/(1 - \delta) \geq 0$ , or  $x^* \geq c(1 - \delta)/\alpha\delta$ . The ruler has no incentive to grab the whole pie and get kicked out for sure provided that

$$\frac{v - x^*}{1 - \alpha\delta} \geq v,$$

or  $x^* \leq v\alpha\delta$ . And the ruler is willing to step down if asked to leave, since she gets zero either way (she will be deposed in a rebellion if she stays).<sup>19</sup> Thus, as long as  $v/c \geq (1 - \delta)/(\alpha\delta)^2$ , the above strategies can form an equilibrium in Model-3-plus-an-election. Since the citizen never needs to rebel on the equilibrium path, his (time-averaged) expected payoff is  $\alpha x^* = v\alpha^2\delta$ , which is strictly greater than the maximum feasible payoff without elections,  $v\alpha^2\delta - c(1 - \alpha)$ .<sup>20</sup>

<sup>19</sup>To provide a positive incentive, we could have a punishment for rulers who fail to step down when asked.

<sup>20</sup>There is another, more subtle effect: With elections we can sustain an equilibrium in which the ruler gives out  $x^* = v\alpha\delta$  each period under parameter conditions such that this is impossible without elections (the constraint is weaker with elections). With elections, the citizen can in effect commit to get zero payoffs forever if he fails to rebel after getting shortchanged, under the assumption that electoral results are observed by potential rulers in the pool. Without elections, if the citizen’s failure to rebel when shortchanged is not observed by rulers in the pool, then new rulers cannot continue the punishment.

In this account, elections serve as a cheap talk signal that enables the public to commit to rebel if a losing incumbent does not step down. In equilibrium, everyone knows that the electoral results “draw a line in the sand,” engaging a reputation that is valuable for the citizens when they cannot directly observe the government’s policy choices or their impact on public welfare. With imperfect monitoring, motivating good choices by politicians requires that politicians sometimes lose office. It is socially much more efficient for them to exit under an implicit threat of rebellion than via actual rebellion.

In Model 3 there is only one citizen or social group, so all that is needed is an announcement rather than an elaborate system of ballots, voting booths, and electoral commissions. By contrast, a mass public cannot simply announce Leave or Stay. It can only “speak” by means of some technology, such as elections (or, in principle, polls). Organization is required to set this up. So, with a mass public, self-enforcing democracy must have the public threatening to rebel not only if the ruler fails to step down after losing, but also if the ruler fails to provide the technology for them to speak (that is, hold fair elections).

What happens if the ruler does not observe the citizen’s perceived welfare,  $z_t$ , in addition to the citizen not observing the ruler’s choice of policy,  $x_t$ ? This creates a moral hazard problem for *the citizen*. He would like to be able to commit to rebel if he observes low welfare, but because this is not observable by the government, he can have an incentive to “let it slide,” since the ruler will not know if he failed to rebel or if he perceived good times.

As a result the citizen cannot be motivated to pay the costs of rebellion by the fear that the government will shortchange him in the future if he fails to rebel given low welfare. Instead, to have a credible threat of rebelling if he sees low welfare, the citizen has to expect to get a higher payoff from a new ruler than from sticking with the current ruler. Indeed, the added benefit of a new ruler must exactly compensate for the cost of rebelling, so that the citizen is indifferent between rebelling and not rebelling. Otherwise he would want to rebel never or always, and in

neither case does the ruler have any incentive to distribute goods.

To get a sense for how this works formally, let  $V_1^C$  be the citizen's continuation value for getting a new ruler in the next period, and let  $V_s^C$  be the citizen's continuation value for the  $s$ -th period a ruler's reign.<sup>21</sup> Then to be willing to rebel in the  $s$ -th period of the current ruler's rule, the citizen must have

$$-c + \delta V_1^C \geq \delta V_s^C.$$

If this inequality is strict, then the ruler expects to be deposed in the  $s - 1$ -th period of her rule, in which case she has no incentive to distribute anything in that period. So to sustain an equilibrium in which the ruler always distributes positive amounts, we need an exact equality. This implies that  $V_s^C$  must be constant for all  $s > 1$ , and that in equilibrium  $V_1^C = V_s^C + c/\delta$  for all  $s > 1$ . As detailed in the proof of Proposition 8 in the Appendix, this implies further that rulers must distribute an amount  $x_1$  in their first period in office, and  $x_2 = x_1 - c/\delta$  in all subsequent periods.

Proposition 8 describes the best such equilibrium for the citizen, in which rulers distribute an amount  $x_1^* > 0$  in their first period in office, and then  $x_2^* < x_1^*$  subsequently. The proof of Proposition 8 also establishes that the citizen can do no better than in this equilibrium of the game without elections.

The citizen's time-averaged, ex ante expected payoff in this equilibrium is the same as in the case where the ruler observes  $z_t$ , *provided that* more restrictive conditions on the parameter  $v/c$  hold. Private or subjective observation of the citizen's welfare imposes a new constraint, namely that a ruler can offer no more than  $v$  in her first period in office.<sup>22</sup> For small enough  $v$  and  $\alpha\delta > 1/2$ , the citizen's best payoff is reduced below what it can be when  $z_t$  is observed by

---

<sup>21</sup>By not conditioning these values on the prior history perceived by the citizen (e.g.,  $V_1^C(h_t^C)$ ), I am considering an equilibrium in which all rulers are expected to act the same way once in office, regardless of the prior history of rebellions by the citizen. In the Appendix, I show that this must be the case in the best possible equilibrium for the citizen.

<sup>22</sup>Assuming that neither foreign loans nor foreign aid are available. Even if there were, there would be borrowing costs to spread over time.

the ruler, and for low enough  $v$  only dictatorship is now possible. Democracy in the form of a cheap talk announcement by the citizen enables the equilibrium described above under the same parameter conditions, so that the potential welfare gains for the citizen under democracy are even larger when  $z_t$  is privately observed.

## 4 Conclusion

A government capable of enforcing contracts, maintaining basic public security, and providing national defense is certainly capable of using its power against the interests of the citizens. Indeed, even a weak government that scarcely fulfills these core functions has the power to commit extreme abuses in the interests of those who control it. To prevent abuse and to ensure optimal provision of public goods, individuals (or groups in society or branches of government) must be able to credibly threaten to coordinate to oppose deprecations by the ruler (Weingast 1997). But such coordination is highly problematic given that different parties observe different aspects and effects of the ruler's policies.

The main argument here is that electoral democracy is a natural and indeed ingenious way of solving this coordination problem. In a functioning democracy, the public (or important social or political groupings) implicitly threaten to rebel if elections are not held according to a commonly understood electoral calendar. Whether elections are held according to the schedule is a publicly observable signal, which thus permits coordination and makes the implicit threat credible. The electoral results themselves then aggregate and publicize private information about the ruler's actions and performance. The public nature of the electoral results again allows the citizens to credibly threaten to coordinate on rebellion if the ruler does not obey the results, while at the same time giving the ruler an incentive to provide public goods.

The potential Achilles' Heel of this solution is subtle electoral fraud and other actions by the

ruler that chip away at the fairness of elections (such as controlling the media). If citizens receive different signals about whether the reported vote totals reflect the true distribution, then they again face a difficult coordination problem that may undermine the implicit threat to rebel.

As an empirical matter, a common path from democracy to dictatorship is for an elected ruler to undermine democracy gradually, in steps any one of which is too small to provide a clear signal that rebellion is called for. A particular opposition figure may be banned or accused of a crime, media outlets bought or subtly threatened, electoral commissions may be packed with supporters or gradually bought off. Votes may be padded in certain places, but not so much that it is glaringly obvious to all that the ruler would have lost a fair election. Putin has been gradually undermining democracy in Russia in this manner, for example.<sup>23</sup>

The theoretical analysis here thus supports the idea that international and domestic organizations that monitor whether elections are “free and fair” can significantly favor democratic consolidation if they send clearer public signals about whether elections were highly fraudulent. Such organizations spread rapidly in the 1980s, so much that Huntington (1991, 8) could remark that “By 1990 the point had been reached where the first election in a democratizing country would only be accepted as legitimate if it was observed by one or more reasonably competent and detached teams of international observers.” Carothers (1997) notes a number of instances where election monitors apparently played the role suggested by the analysis above. More recently, reports by domestic and foreign monitoring groups prompted the mass protests in Ukraine in the winter of 2004-2005. These ultimately unseated incumbents trying to undermine democracy through electoral fraud.

As Carothers argues, the proliferation of monitoring organizations often makes for confused signals. Carothers also points out that what constitutes a “free and fair” election involves judgement calls and fuzziness – it is not an easy coding decision. Nonetheless, in so far as considerable

---

<sup>23</sup>McFaul (2006). Some other recent examples include Thaksin in Thailand, Museveni in Uganda, Afwerki in Eritrea, and several autocrats in Central Asia who may have been elected in relatively fair votes the first time but much less so subsequently. Huntington (1991) noted that this form of undermining democracy has grown more common.

agreement may be attained on whether there was blatant fraud, domestic and international election monitoring groups may facilitate the popular coordination stressed by the analysis here, so deterring some moves toward dictatorship.<sup>24</sup>

More generally, the analysis directs our attention away from the slow-moving economic and cultural variables that have dominated empirical and theoretical analysis of conditions for democracy, and towards factors that favor or disfavor coordination of a body of citizens. Consider Figure 1, which shows the percentage of democracies in the world by year from 1946 to 2002.<sup>25</sup> The most striking aspect of these data is the extremely rapid increase in the proportion of democracies from 1989 to 1992, from about one third to almost one half in just four years. Since the level does not drop rapidly back in the late 1990s (but actually increases further), many of these new democracies have survived at least two elections. One might wonder if the rapid increase is entirely due to the fall of communism in Eastern Europe and the former Soviet Union. But Figure 2, which breaks the data down by regions, suggests that this is part but not all of the story. Rapid increases in the percentage democratic are observed in the same period of time in Latin America, Asia, and sub-Saharan Africa. Of the 31 democratic transitions from 1989 through 1992, 13 occurred in Eastern Europe and the former Soviet Union, and 18 outside it.

It is highly unlikely that this clustering in time resulted from all these countries happening to simultaneously cross some threshold level of per capita income, income equality, or “democratic capital” that caused them to become democratic.<sup>26</sup> It is far more plausible that the clustering

---

<sup>24</sup>Hyde (2006) exploits a natural experiment in the de facto random assignment of election monitors to polling stations in the 2003 Armenian presidential election. She finds that the incumbent (Kocharian) received almost 6% less on average at the monitored polling stations in the first round of voting, with an even larger difference in urban areas.

<sup>25</sup>The measure is from the Polity IV project. Countries are coded as democratic when the “polity2” variable, which ranges from -10 to 10, is greater than 5. This is a fairly conventional cutpoint, and little depends on it.

<sup>26</sup>This is true whether the threshold levels are assumed to be country specific or not. Lipset (1963) stressed high income as a determinant of democracy. Acemoglu and Robinson’s (2000, 2001) and Boix’s (2003) models tend to focus on income inequality as the key barrier to democracy. Democratic, or social capital, is emphasized by Putnam (1993), Persson and Tabellini (2005), Almond and Verba (1963) (“civic culture”), among others.

results from some kind of diffusion effect, or “snowballing” (Huntington 1991), associated with the collapse of communism in the Eastern bloc, from *Solidarność*, *perestroika* and the fall of the Berlin Wall to the break up of the Soviet Union at the end of 1991. Indeed, Huntington (1991, 100-108) traces out a series of consequential demonstration effects, including direct learning and instruction, from the Southern European transitions in the mid 1970s (Greece, Portugal, Spain), to Latin America in the late 1970s and 1980s, to Poland and Hungary and then Eastern Europe and the Soviet Union at the end of the 1980s.<sup>27</sup>

I would not claim that the models analyzed in section 3 explain diffusion effects. But I do think that the gist of argument – that establishing and maintaining democracy is a matter of citizens and institutions resolving a coordination problem concerning how to combine to oppose dictatorship – holds more empirical promise than accounts in terms of class or party conflict and their relation to economic fundamentals (Acemoglu and Robinson 2001, Przeworski 1991). As Schelling (1960) stressed, coordination problems are susceptible to resolution via focal points and focal principles. Observing coordination to demand democracy or oppose dictatorship in one country may trigger the same sort of coordination in another country.<sup>28</sup> And by facilitating social coordination, the spread of a global norm in the sense of a commonly understood democratic script or model may be a more powerful long-run cause of democracy than incremental changes in per capita income, inequality, or more deeply held individual-level norms and culture attributes.

Economic modernization may well have indirect effects that favor popular coordination to institute or maintain democracy. Some non-trivial level of urbanization, for example, is practically

---

<sup>27</sup>The succession of regional waves is evident in Figure 2.

<sup>28</sup>Dictators live in tremendous fear that popular coordination against authoritarian government in other countries will stimulate coordination and rebellion in their own country. For example, Mugabe’s brutal efforts to rid Harare of street vendors and urban squatters were probably influenced by fear of a Ukraine-style uprising (International Crisis Group 2005, 4). Putin’s recent attacks on western democracy promotion efforts and election monitoring groups suggest fear of a “color revolution” (Ukraine, Georgia, Kyrgyzstan) spreading to Russia. The extremely violent response of Uzbekistan’s government to demonstrations in Andijon province in May 2005 may be partly explained by fear of a repeat of the uprising that deposed the Kyrgyz dictator Akaev in March.

a precondition for a serious popular revolt. Literacy and the multiplication of interlinked corporate actors that come with economic development may also help. But if the central problem is how to shift from a bad equilibrium in which no one expects coordinated opposition to significant electoral fraud or suspension of elections, to a good equilibrium in which individual expectations and institutions support this expectation, then it would be surprising if economic modernization inevitably produced or maintained democracy in some deterministic way. Instead, what will matter most are things like historical accidents that happen to yield mass coordination; repeated experience with such events; demonstration effects; and the construction of institutions like independent electoral commissions, judiciaries, or branches of government distinct from the executive that have corporate integrity and so can coordinate opposition to piecemeal tyranny.

## 5 Appendix

**Proof of Proposition 4.** If no social consensus exists, the strategies are the same as for the dictatorial equilibrium in Proposition 1, and in a PBE  $i$  can hold any beliefs about what the other citizens will do since it is not rational to rebel if one expects to get zero from any subsequent ruler.

If a social consensus does exist, any deviation from the equilibrium path by the ruler that lowers any citizen's payoff below  $x_i$  provokes a successful rebellion given the citizens' strategy. So the ruler's best deviation would be to keep  $v$  and get kicked out, which leads to condition (1) as before. No citizen wants to deviate from the equilibrium path by rebelling when given  $x_i$  as this would just cost  $c$  and bring in a new ruler giving out the same payoffs as the old ruler. Hence the proposed strategies form a Nash equilibrium.

Off the equilibrium path, if a social consensus exists and if  $i$  observes  $x_{it} < x_i$ , let  $i$  believe that  $R$  kept the difference but did not change anyone else's allocation.<sup>29</sup> Given this belief  $i$  expects to be the only protester in phase 1 and so  $i$  expects to get  $-c + \delta x_i / (1 - \delta)$  by protesting. If  $i$  does not rebel, then  $i$  gets  $\delta \hat{x}_i / (1 - \delta)$ , where  $\hat{x}_i$  was chosen so that  $i$  is indifferent and thus willing to rebel.

In state 2, if  $i$  rebels after receiving  $x_{it} = \hat{x}_i$   $i$  gets  $-c + \delta x_i / (1 - \delta)$ , versus  $\delta \hat{x}_i / (1 - \delta)$  by not rebelling. By construction these are equal.  $i$  faces the same choice if he sees  $x_{it} < \hat{x}_i$ , and so is willing to rebel in this case as the strategy calls for (again, we need to assume that if  $i$  sees  $x_{it} < \hat{x}_i$  in this case,  $i$  believes that the ruler deviated only against him).

If, in any period  $t$ , one or more citizens protested in phase 1, then if any  $i$  who is obligated to protest in phase 2 fails to do so the social consensus is broken and every citizen's continuation payoff is zero. So it is better to protest provided that condition (2) above holds for all  $i$ . Condition (2) also guarantees that  $\hat{x}_i \geq 0$  for all  $i$ , so it is feasible that  $R_t$  can offer it  $\square$

**Proof of Proposition 6.** If a social consensus exists, the ruler's best deviation is to distribute nothing and then leave after losing the election (the ruler is indifferent between leaving and being forced out, so stepping down is a best reply). This yields  $v$ , which is less than the equilibrium path payoff by condition (1). If elections are not held, or if elections are held and the ruler lost but does not step down, then citizen  $i$  expects to get  $-c + \delta x_i / (1 - \delta)$  by rebelling and zero by not rebelling, which yields condition (2). If elections are held, citizen  $i$  is indifferent between voting for or against  $R_t$  since both courses will yield future payoffs of  $\delta x_i / (1 - \delta)$  (if the citizen votes against, the current ruler steps down and the new ruler also distributes  $x_i$ ). Thus the citizen is willing to vote Against if he observed  $x_{it} < x_i$  and For otherwise. If no social consensus exists,

---

<sup>29</sup>I will not exhaustively characterize the belief system for a PBE; in fact this is the only situation where the choice matters.

rebellion costs citizen  $i$   $c$  but will not change any ruler's strategy of giving zero, and of course no ruler wants to give any positive amount away if there is no threat of being deposed.  $\square$

**Proposition 7.** Provided that  $v/c \geq (1 - \alpha\delta)/(\alpha\delta)^2$ , Model 3 has a perfect Bayesian equilibrium with the following strategies. In period  $t$ , the citizen can be in one of two "states" with respect to the current ruler  $R_t$ . Whenever a new ruler comes to power, the state is reset to state 1.

State 1: The citizen has always observed  $z_s \geq x^*$  from the current ruler,  $s < t$ . In state 1,  $R_t$  gives  $x^* \equiv v\alpha\delta$ , and the citizen rebels in period  $t$  if and only if  $z_t < x^*$ .

State 2: The citizen failed to rebel after observing  $z_s < x^*$  in some period  $s < t$  but under the current ruler. In state 2,  $R_t$  gives  $\hat{x} \equiv x^* - c(1 - \alpha\delta)/\alpha\delta$ , and the citizen rebels if and only if  $z_t < \hat{x}$ .

Moreover, no perfect Bayesian equilibrium of Model 3 gives the citizen a higher time-averaged ex ante expected payoff (which here equals  $v\alpha^2\delta - c(1 - \alpha)$ ).

**Proof.** Under the proposed strategies, the citizen's payoffs in state 1 satisfies  $V_1^C = \alpha x^* - c(1 - \alpha) + \delta V_1^C$ , or  $V_1^C = (\alpha x^* - c(1 - \alpha))/(1 - \delta)$ . Let  $V_2^C$  be the citizen's expected payoff going into a period in state 2. For the citizen to be willing to rebel in state 1 if he observes  $z_t < x^*$ , it must be the case that  $-c + \delta V_1^C \geq \delta V_2^C$ . And for a citizen in state 2 to prefer *not* to rebel if he sees  $z_t \geq \hat{x}$  (so as to get a new ruler and thus  $x^*$ ), it must be the case that  $\delta V_2^C \geq -c + \delta V_1^C$ . Thus an equilibrium of the form proposed in the proposition requires that

$$-c + \delta V_1^C = \delta V_2^C. \quad (4)$$

In state 2, the citizen's expected payoff going into a period is

$$V_2^C = \alpha(\hat{x} + \delta V_2^C) + (1 - \alpha)(-c + \delta V_1^C).$$

Using (4), this become  $V_2^C = \alpha(\hat{x} + \delta V_2^C) + (1 - \alpha)\delta V_2^C$  or  $V_2^C = \alpha\hat{x}/(1 - \delta)$ . Using (4) again implies

$$\begin{aligned} \frac{\alpha x^* - c(1 - \alpha)}{1 - \delta} &= \frac{\alpha \hat{x}}{1 - \delta} + \frac{c}{\delta} \\ \alpha x^* &= \alpha \hat{x} + c(1 - \alpha) + c(1 - \delta)/\delta \\ \hat{x} &= x^* - c \left( \frac{1 - \alpha\delta}{\alpha\delta} \right). \end{aligned}$$

The ruler's expected payoff on the equilibrium path is  $(v - x^*)/(1 - \alpha\delta)$ , and the ruler's best possible deviation from the equilibrium path is to give nothing, which yields  $v$ . So the ruler prefers to continue giving  $x^*$  provided that  $x^* \leq v\alpha\delta$ . Since  $\hat{x} < x^*$ ,  $x^* \leq v\alpha\delta$  implies that  $R_t$  will not

want to deviate in state 2 either. Thus the best possible payoff for the citizen in an equilibrium of this form involves the ruler giving  $x^* = v\alpha\delta$  each period on the path.

The constraint that  $\hat{x} \geq 0$  implies the condition  $v/c \geq (1 - \alpha\delta)/(\alpha\delta)^2$ , which is always stronger than the condition for  $V_1^C \geq 0$ . This is sufficient to show that the strategies given in the proposition can form a perfect Bayesian equilibrium given the condition. (Rulers in the pool believe that equilibrium strategies have been followed at all times.)

It remains to be shown that the citizen can do no better in any other equilibrium of Model 3. [to be completed ...]

**Proposition 8.** Consider Model 3 altered so that rulers do not observe  $z_t$ . (a) Provided that the ratio of total benefits to the costs of rebellion,  $v/c$ , satisfies

$$\frac{v}{c} > \max \left\{ \frac{1}{1 - \alpha\delta}, \frac{1 - \alpha\delta}{(\alpha\delta)^2} \right\}, \quad (5)$$

the game has a perfect Bayesian equilibrium with the following strategies. Rulers offer  $x_1^* = \alpha\delta v + c$  in their first period in office, and  $x_2^* = x_1^* - c/\delta\alpha$  in all subsequent periods. The citizen rebels for sure if he observes  $z_t < x_1^*$  in the first period of a ruler's tenure, and if he observes  $z_t < x_2^*$  subsequently.

The citizen's (time-averaged) expected payoff in this case is  $(1 - \delta)V_1^C = v\alpha^2\delta - c(1 - \alpha)$ .

(b) If  $\alpha\delta > 1/2$  and  $v/c \in [1/\alpha\delta, 1/(1 - \alpha\delta)]$ , then the game has a perfect Bayesian equilibrium in which the ruler in period  $t$  offers  $x_1^* = v$  in her first period in office, and  $x_2^* = v - c/\delta\alpha$  in all subsequent periods. The citizen rebels for sure if he observes  $z_t < x_1^*$  in the first period of a ruler's tenure, and if he observes  $z_t < x_2^*$  subsequently.

The citizen's time-averaged expected payoff in this case is  $(1 - \delta)V_1^C = \alpha v - c$ .

(c) If  $v/c \leq \min\{1/\alpha\delta, (1 - \alpha\delta)/(\alpha\delta)^2\}$ , then the only equilibria are dictatorial.

(d) The parameter conditions for (a), (b), and (c) are exclusive and exhaustive, and the equilibrium in each case yields the citizen's highest attainable payoff in any perfect Bayesian equilibrium.

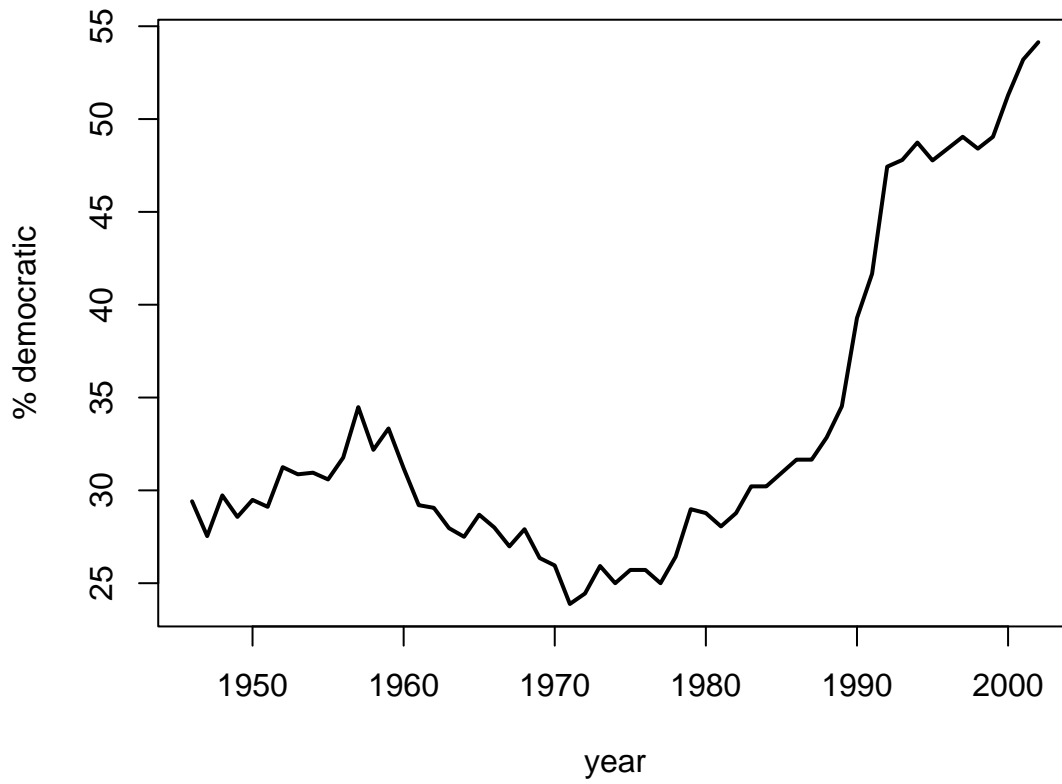
**Proof of Proposition 8.** The proof has two parts. The first develops the conditions for an equilibrium in which rulers give  $x_1 > 0$  in their first period of rule and  $0 < x_2 < x_1$  in any later period. The second part establishes that the optimal equilibrium for the citizen must have this form. [to be completed ...]

## References

- Acemoglu, Daron and James A. Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics* 115(4):1167–1199.
- Acemoglu, Daron and James A. Robinson. 2001. "A Theory of Political Transitions." *American Political Science Review* 91(4):938–63.
- Almond, Gabriel A. and Sidney Verba. 1963. *The Civic Culture*. Princeton, NJ: Princeton University Press.
- Baron, David and John Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83:1181–1206.
- Barrington Moore, Jr. 1966. *Social Origins of Dictatorship and Democracy*. Boston, MA: Beacon Press.
- Benhabib, Jess and Adam Przeworski. 2005. "The Political Economy of Redistribution under Democracy." *Economic Theory* .
- Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government*.
- Boix, Carles. 2003. *Democracy and Redistribution*. Cambridge: Cambridge University Press.
- Carothers, Thomas. 1997. "The Observers Observed." *Journal of Democracy* 8(3):17–31.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50:5–25.
- Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- Hyde, Susan. 2006. "Can Election Observers Deter Election Day Fraud? Evidence from a Natural Experiment." Chapter 7 of "Observing Norms: Explaining the Causes and Consequences of International Monitored Elections, Ph.D. Dissertation, UCLA.
- International Crisis Group. 2005. "Zimbabwe's Operation Murambatsvina: The Tipping Point?" Africa Report Number 97, 17 August.
- Lipset, Seymour Martin. 1963. *Political Man: The Social Bases of Politics*. Garden City, NY: Doubleday.
- McFaul, Michael. 2006. Russia. In *Countries at the Crossroads 2005*, ed. Freedom House. New York: Freedom House.
- Persson, Torsten and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

- Persson, Torsten and Guido Tabellini. 2005. "Democratic Capital: The Nexus of Political and Economic Change." Unpublished, Stockholm University and University of Bocconi.
- Przeworski, Adam. 1991. *Democracy and the Market*. New York: Cambridge University Press.
- Przeworski, Adam. 2005. "Democracy as an Equilibrium." *Public Choice* 123:253–273.
- Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. New Haven: Yale University Press.
- Schmitt, Carl. 1988[1923]. *The Crisis of Parliamentary Democracy*. Cambridge, MA: MIT Press.
- Schumpeter, Joseph A. 1975[1942]. *Capitalism, Socialism, and Democracy*. New York: Harper and Row.
- Weingast, Barry R. 1997. "Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91:245–63.

**Figure 1. % democracies, 1946–2002 (Polity > 5)**



**Figure 2. % democracies by region (Polity > 5)**

