The Watchful Eye:
Information Transmission and Political Failure

Branislav L. Slantchev[*]

Department of Political Science, University of California – San Diego

December 29, 2004

Abstract.　Domestic audience costs can help leaders establish credible commitments by tying their hands. Most studies assume these costs without explaining how they arise. I link domestic audience costs to the citizens' ability to sanction the leadership for pursuing policies they would not want if they had the same information about their quality. How can citizens learn about policy quality? I model two information transmission mechanisms: one potentially contaminated by politically-motivated strategic behavior (leader and opposition), and another that is noisy and possibly biased (media). In equilibrium, audience costs can arise from strategic sources only in mixed regimes under relatively restrictive conditions, and cannot arise in autocracies or democracies. However, in democratic polities the media can play a mitigating role and does enable leaders to generate audience costs. Still, their ability to do so depends on the institutional protections guaranteeing freedom of the media from political manipulation. Domestic audience costs are not necessarily linear in regime type, as often assumed in applied research.

Tying hands can be an effective way to communicate the credibility of one's commitment (Schelling 1966, Jervis 1970). Audience costs are one mechanism for doing so that has become fairly popular in recent studies of international behavior (Martin 1993, Fearon 1994*a*). Briefly, if leaders take actions that increase the costs of backing down from their position, then they can effectively commit to holding out for concessions. However, as Smith (1998) and Schultz (1999) note, this mechanism lacks microfoundations: the theoretical models that investigate the impact of audience costs on behavior have largely taken them for granted. This article clarifies what an interpretation of domestic audience costs would look like, and investigates the theoretical possibility for generating such costs endogenously. Under what conditions would a rational audience impose such costs on a leader? How do these conditions depend on the institutional structure of the polity?

The results suggest that while it is possible for these costs to arise, their generation is far from straightforward. In particular, if one relies solely on strategic sources of information (government, opposition parties), citizens of either democracies or autocracies are unlikely to learn enough to credibly threaten to sanction their leaders for bad behavior. Even though the reasons for such failure differ between the two regime types, the consequences are similar. This suggests that the widespread identification of regime type with audience costs may be seriously misleading, with attendant consequences for empirical studies. For example, it would not be correct to say that a democracy would necessarily signal better than an autocracy because of audience costs.[1]

On the other hand, I find a somewhat mitigating factor that might recover some of the democratic polity's ability to learn more than its autocratic counterpart. A non-political (exogenous) source of information, such as a free press, could have a serious impact even if the signal it sends is noisy and potentially biased. While citizens of a democracy can impose costs on their leaders only imperfectly, they could still do so provided the alternative sources of information are not highly politically manipulable. This implies that democracies themselves can be ranked with respect to their ability to generate audience costs: The ones with more media protections would enable their citizens to sanction leaders much better.

Taken together, these findings suggest that we should pay closer attention to the causal mechanism from which audience costs are supposed to arise and most certainly should not take them for granted as an assumption in our models. While much of the emphasis has been on the strategic sources of information, perhaps we should investigate in much more depth the interaction between government, opposition parties, and media in the public forum with domestic audiences. As Miller and Krosnick (2000) have shown, citizens infer the importance of an issue from the extent of its media coverage, making media accuracy and credibility significant factors in that process.

1   Credible Commitments, Audience Costs, and Citizen Control

The resolution of conflict in many international situations turns on the ability of actors to commit credibly to some course of action. In a crisis, the opponent would pay attention to

---

[1] For empirical applications and other models featuring audience costs as an assumption, see Eyerman and Hart, Jr. (1996), Regan (1998), Busch (2000), Partell (1997), Papayoanou (1997), Fearon (1994*b*), and Baum (2004).

a threat to resort to arms only if it is credible; an alliance would have a deterrent effect only if the defender's promise to come to the aid of his protege is credible; and so on.

Under asymmetric information, communicating commitments is fraught with difficulty. The actor with a genuine threat must somehow separate himself from the plethora of possible threateners who are mere bluffers. One possibility is to engage in an action that would not be in the interest of a bluffer to carry out: Doing so should provide convincing proof of one's intentions. But what would such an action be? In a crisis, the negotiated outcome turns on the difference between the expected payoffs from war and peace: an actor has to concede more to an opponent who believes he would do well fighting than to an opponent who believes himself too weak to fight. An actor with a high expected payoff from war relative to peace could demand a large concession because he would have to be compensated for not fighting.

The goal, then, is to persuade the opponent either that one's expected payoff from war is relatively high, or that one's expected payoff from peace is relatively low. With the exception of nuclear crises, one can envision circumstances where deliberate attack is a rational decision. If one succeeds in making war better than peace, one has effectively established a genuine threat to attack should the circumstances ever arise.

This implies that actors would engage in behavior that is designed to alter the strategic environment such that war becomes optimal if the opponent fails to concede enough. One straightforward way of doing that is through actions that decrease the value of peace obtained by one's own backing down: if peace is less palatable, the one is less likely to opt for it. Fearon (1994*a*) postulates just such a mechanism for tying one's hands: a leader who escalates a crisis also increases his *audience costs*, which he has to pay if he backs down. Since escalating further may get the opponent to concede, leaders would be tempted to do just that in order to avoid paying these costs. The more a leader escalates, the higher the costs of backing down, and the more credible the threat not to quit. If both leaders escalate too far, they can become locked in a position from which neither one would recede, thereby ensuring war. In a way, the cure (costly signaling through audience costs) can be worse than the disease (inability to signal resolve).

Through their tying hands effect, audience costs can influence crisis behavior in a fairly well-understood way, which probably accounts for the popularity of this commitment mechanism in recent studies. There is no question that *if* leaders can generate such costs, then they would be able to derive bargaining advantage from them under some conditions.[2] However, this is a big "if."

In the original article, Fearon (1994*a*) assumes that citizens punish leaders who bluff by escalating a crisis and then back down; hence escalation involves increasing audience costs. He assumes that audience costs exist and justifies this assumption with an appeal to "national honor"; that is, citizens punish the leader for failing to uphold it. It is unclear why citizens should punish their leaders for getting caught in a bluff when bluffing may be an optimal strategy (Schultz 1999), or when they may be happy that the leader avoided a costly foreign entanglement (Smith 1998). Without microfoundations, the domestic audience cost

---

[2]Baum (2004) studies when leaders might prefer to make their demands in private and forego public commitments if the latter generate audience costs. For summaries and empirical evaluations of citizen rationality, see Shapiro and Page (1988) and Knopf (1998).

story sounds a lot less persuasive.[3]

Schultz (1998) shows that an opposition could buttress the leader's signaling even in the absence of domestic audience costs. This transfers the signaling burden onto the political system and the presence of alternative sources of information. Schultz assumes that citizens are not entirely strategic but behave according to a simple retrospective evaluation framework that apportions blame or credit to the government and the opposition depending on policy outcomes. In other words, the crucial mechanism is again relegated to an assumption, even though it is a much more plausible one.

Fundamentally, audience costs are supposed to "arise from the action of domestic political audiences concerned with whether the leadership is successful or unsuccessful at foreign policy" (Fearon 1994*b*, p. 241). Why would citizens punish their leader? Presumably, the only circumstance where they would be willing to do it is when they find out that his behavior was different from what they would have done if they had the same information (Downs and Rocke 1995). That is, citizens do not punish for honest mistakes but would like to punish deliberate malfeasance. What matters is not whether citizens like the outcomes or not, but whether they would have wanted the policy if they knew all that the leader knows. In other words, leaders are judged on how faithful agents they are of the citizenry who act as the principal.

When we ask the question in this way, it becomes immediately clear that the idea of audience costs is closely related to the question of citizen control. Audience costs arguments are essentially about informational asymmetries and the ability of citizens to sanction their government for "inappropriate" behavior. Audience costs are the *direct reduction in the leader's reselection probability due to citizens inferring information unfavorable to the incumbent from the actions of the government, the opposition, and an independent source of information.* Since citizens care about policy quality, they may seek to replace a leader who has presided over bad policy with an yet-untried opposition candidate. If citizens are less likely to keep the incumbent because of his policies, then they are imposing audience costs.

The reduction in reselection chances should arise in equilibrium because of the way citizens react to the policies implemented by the government. That is, instead of assuming audience costs, we want to obtain them as consequence of equilibrium behavior; we want them to arise endogenously. We thus arrive at the connection between policy quality, government selection, information available to the citizens, and audience costs. The question becomes: If citizens are rational and fully strategic (that is, they use all the information

---

[3]Audience costs are imposed on the leader by the audience of his behavior. The audience can be domestic (citizens) or international (other states). I am aware of one attempt to provide microfoundations for domestic audience costs by Smith (1998), which I discuss later. An important recent research venue has been the investigation of the second source: Perhaps foreign audiences could generate audience costs? Sartori (2002) makes this argument in the context of repeated interaction with foreign rivals where reputational losses result in costs that may deter one from bluffing. The analysis relies more on the repeated framework (expected future punishment for bluffing) than it does on signaling through present actions. In a clever article, Ramsay (2004) demonstrates that the presence of a foreign actor could discipline, albeit in rather limited ways, the domestic opposition to credibly reveal information about the leader even if its statements are cheap talk. The opposition's behavior in our model is not cheap talk because of the possibility of repression, and thus the standard results do not apply (Crawford and Sobel 1982, Farrell and Gibbons 1989). Further, the focus of the present article is on purely domestic sources.

available and do not arbitrarily punish or reward either the incumbent or the opposition), how would they behave and what impact would their behavior have on the government's policies? In other words, would leaders be able to generate audience costs through their actions?

Framing the issue in this way immediately reveals the fundamental problem: How can citizens learn what they need to know to make the necessary inferences? To put it bluntly, how do they know that any given policy is bad? I consider two potential sources of information (in addition to the behavior of the leader): a political opposition and a politically-independent source, such as the media. The first source may be "contaminated" by the opposition's desire to gain office, and the second source may be "noisy" in that there is a chance that it would misreport a good policy as bad and vice versa.

The question then becomes: Under what circumstances would citizens be able to learn enough in order to punish the leader? These circumstances would be the only situations where the leader can generate audience costs, and hence the only situations in which the mechanism identified by Fearon and others would have a chance to operate. In all other situations, citizens cannot learn, so they cannot punish, and hence the leader faces no domestic audience costs for his behavior. In these situations, signaling due to such costs would be fairly circumscribed.

I do not assume that the government and the opposition can share credit or blame for existing policy (as Schultz (1998) does) or that the opposition can discipline itself through two-dimensional preferences (as Ramsay (2004) does). Instead, the opposition is just like the leader in that it is first and foremost office-seeking, and, given that, prefers good policies to bad ones, just like the rest of the citizens.[4] The only potentially disciplining device is at the leader's disposal: the possibility to repress the opposition if it dissents. Hence, the opposition is free to make any statement but since the leader can repress it, dissent may carry risks.

Citizens are fully strategic actors and attempt to make best possible use of the information available to them. They can freely replace an incumbent if they so choose or revolt if the leader has repressed the opposition, thereby removing the election option. To decide what to do, citizens use all the information provided by the strategic actions of the politicians and the reportage by the media. To focus on the domestic incentives for information transmission, the model follows Smith (1996) in simplifying the environment by ignoring the presence of a foreign actor. Insofar as a decision to continue a bad policy rests with the government which can always repeal it, this assumption should not be too distorting.

I define two types of political failure: In Type I failure, the leader repeals good policies; and in Type II failure, the leader continues bad policies.[5] The question becomes: under what conditions would political failure occur, and if it does, what type is it likely to be?

---

[4]There are many assumptions one could make about the opposition that would immediately lead to truth-telling equilibrium behavior. For example, if a statement created a commitment from which would be costly to reverse, or if citizens punished the opposition for incorrect predictions, the opposition would have less incentive to lie. However, this would beg the original puzzle: why would citizens adopt such strategies with respect to the opposition? Ascribing blame or credit for a policy that the opposition had no hand in implementing is a dubious assumption that would require a theoretical investigation of its own.

[5]Heuristically, these are intended to parallel the two errors in statistical hypothesis testing, where Type I error means erroneously rejecting a correct null hypothesis, while Type II error means erroneously failing to reject a wrong null.

The next natural question is to ask whether different regimes are more or less susceptible to political failure. Fearon (1994*a*) surmises that since democracies have elections as a low-cost way for citizens to express their disapproval, democratic regimes should be able to generate audience costs at higher rates than non-democratic ones. That is, the magnitude of audience costs is increasing with the openness of the political regime.

As Schultz (2001) notes, while it is easier to remove democratic leaders, their punishments will tend to be a lot less severe than those for removed authoritarian leaders. Therefore, it is unclear under which regime type the selection threat will be more credible. However, he argues that the magnitude of audience costs is not as important as the ability to convey that they exist to the foreign rival.

I conceptualize regimes along two dimensions. First, in terms of the efficacy and costliness of the repressive apparatus available to the government. While some regimes do make any sort of opposition illegal, most contemporary ones tend to put a facade of legitimacy by seemingly allowing it. Instead of assuming the effectiveness of opposition, I prefer to derive it from a more basic model, in which all opposition is *potentially effective*, but may turn out to be *actually ineffective* because of the strategies it pursues. As we shall see, it is the credibility of the threat of repression that influences opposition's behavior and its ultimate effectiveness. Anticipating some of the results, I note that even a potentially effective opposition becomes nothing but a blind supporter of government policy in repressive societies. One difference is that this is now equilibrium behavior, not an assumption in the model.

Second, regimes differ in terms of the bias of the alternative sources of information. While all such sources will be noisy, the signal to noise ratio will vary according to how protected from government interference these sources are. Citizens in a polity with constitutionally protected media freedoms, for example, will be more likely to receive both good and bad news from an exogenous source of information than citizens in a policy with tightly-controlled media. Thus, a democracy would be characterized by high costs of repression and high signal to noise ratio in the exogenous signal. Conversely, an authoritarian regime would be characterized by low costs of repression and an exogenous signal biased toward good news. Mixed-regimes, on the other hand, would have intermediate costs of repression with varying degrees of control of alternative sources. This setup allows us to examine variations among regimes types but also within types (e.g. democracies that differ in the amount of protection offered to media).

With these ideas in mind, I present a simple stylized formalization of a strategic interaction among three players: a leader (government), an opposition, and citizens. This model is not a faithful description of reality but an analytic tool to sharpen our intuitive understanding of these phenomena. By reducing the strategic context to a very stark and small (but certainly not minimal) set of actions, it can illuminate the conditions that are most conducive to political failure. I assume a homogenous electorate and non-rival, non-excludable (public good) policies. This abstracts away from distributive conflict and coordination problems. The goal is to give the informational theory the most permissive environment where it can operate, a strategic context that is entirely defined in terms of the informational asymmetries. Note that there are no audience costs in this model; instead, this is a model of how audience costs—defined as the reduction in the leader's reselection chances—may arise endogenously as citizens learn about the the quality of the leader's policies.

2 The Model

To characterize policy failure in a state with possibly restricted political competition, I present a simple two-period analytical framework, which builds on Dur (2001). The nation consists of a large number of identical citizens, each of whom derives utility from an existing public good policy depending on its quality. The preferences of the representative citizen are entirely policy-based, and the citizen receives $b \in (0, 1)$ if the policy is *good*, $-b$ if the policy is *bad*, and 0 if no policy is currently in place.[6]

Office-holders are drawn from the population of ordinary citizens but in addition to the policy benefits, they obtain ego rents, $v \in (0, 1)$, from holding the position of leadership. I assume that all else equal, ego rents are more important to office-holders than policy benefits. In other words, the leadership position is highly desirable.

At the beginning of each period, the current leader implements a policy, which can turn out to be either good or bad. The policies implemented by more competent leaders are more likely to be good. Denote the probability that the policy is good by p if it is chosen by a competent leader, and $q < p$ if it is chosen by an incompetent one. The common prior probability that the incumbent is competent, is μ, and thus, the probability that the existing policy is good is:

$$g = \mu p + (1 - \mu)q.$$

Leader's competence is unobservable and unknown to all players, including the leader.[7]

An opposition competes with this leader for office, which it may obtain either through elections or revolution, which we collectively call *the citizen's selection*. In both cases, only the citizens can put the opposition in office by replacing the leader. The probability that the opposition is competent is denoted by μ^o and is drawn randomly immediately prior to the citizen's selection from the uniform distribution function $F(\cdot)$, and so $F(\mu)$ is the probability that $\mu^o \leq \mu$.[8] Let μ_e^o denote the ex ante expected value of the opposition competence. Under the information structure specified below, the fundamental difference between the leader and the opposition is that while citizens may be able infer something about the incumbent's competence, they have no way of knowing anything about the opposition's.[9] Further, the opposition has no way of evaluating its own competence until it actually governs and observes the quality of the policy it implements.

---

[6]This eliminates distributive politics. While one can imagine many situations in which it will be reasonably satisfied (e.g. losing a war is a universal bad), the reason for having it in this model is to create an environment in which whatever distortion occurs in the leader's behavior would be due entirely to informational issues. It is not difficult to generate inefficient behavior in distributive settings where some fraction of the population likes one policy and others prefer another.

[7]This is a departure from existing models where the incumbent knows his own competence, but the flavor is quite similar. I prefer the policy quality formulation because it is more natural to think of the exogenous source report being conditional on the policy, and not on the intrinsic characteristics of the leader. Given the two possible pieces of private information a leader could possess (policy quality and own competence), it would complicate the model considerably to consider both, hence the assumption that the leader does not know, and therefore cannot condition his behavior on, his own competence.

[8]The consequence of this assumption is to exclude cases where the leader is certain to stay in office or be removed regardless of the policy decision. In these cases there are no incentives to distort information. The assumption of uniform distribution does not affect the results but does help in simplifying notation and math.

[9]This structure can be found in many other models. Rogoff (1990) provides a canonical example and also justifies it on the basis of empirical findings that show that "for U.S. presidential elections voters do not take

While both the leader and the opposition observe a perfectly informative signal about the policy quality before taking any actions, citizens can only observe a noisy signal and only if the policy does not get repealed. There are two possible signals from non-repealed policies, *success* and *failure*. A good policy produces success with probability $\alpha \in (\frac{1}{2}, 1)$; and a bad policy produces failure with probability $\beta \in (\frac{1}{2}, 1)$. That is, good policies are more likely to send the positive signal than bad policies. These signals come from an exogenous non-strategic source of information. For simplicity, I shall sometimes refer to them as "policy outcomes" with the understanding that the main concern is with the probability that the "outcomes" correctly reflect the quality of the policy in place. A government-controlled source would be biased toward reporting success in the sense that the probability of a good policy producing the success outcome is very high, while the probability of a bad policy producing failure is quite low. An extremely unbiased source would generate these outcomes with correspondingly high probabilities, allowing the citizens to infer policy quality with great precision.

The sequence of the game is as follows. In each period, chance determines the quality of the existing policy, and both the leader and the opposition learn it. In the first period, the leader may then repeal the policy or continue it. Continuing the policy enables the opposition to endorse it or dissent. Should the opposition dissent, the leader can repress it, which eliminates the possibility of elections and instead leaves the citizens with the option of a costly revolution. In all other cases, citizens may costlessly replace the leader with the opposition. In cases of non-repealed policies, citizens observe the noisy signal about their quality immediately prior to their selection, which is binding. Figure 1 illustrates the sequence of actions in the first period of the game. In the second period, there is no citizen's selection, and so the period ends with the incumbent's decision whether to repeal the policy or continue it.

Society is endowed with an institutional structure that determines the costliness and efficacy of the repressive instrument. Let $c \in (0, 1)$ denote the cost that the leader must pay for using repression, and let $(1 - c)$ denote the costs suffered by the opposition whenever it gets repressed.[10]

Citizens also have to pay for removing an incumbent through revolutionary means. Let $k \in (0, 1)$ denote the cost of rebeling. There is no obvious relationship between k and how painful the revolution will be to the leader. I assume that violent removal is catastrophic, with the leader losing at least the equivalent of the ego rents.

I make two simplifying assumptions when it comes to outcomes. First, if repression is followed by policy success, the leader stays in office with certainty. Repression raises the

Citizens (vote)

$1 - g_1$ $\quad$ g_1

repeal $\qquad$ repeal

Leader $\qquad$ $(1 - g)$ $\qquad$ **N** $\qquad$ (g) $\qquad$ **Leader**

bad $\qquad$ *good*

continue $\qquad$ continue

$1 - g_2$ **Citizens** g_2

(vote)

$(1 - \beta)$ $\qquad$ *success* $\qquad$ *success* $\qquad$ (α)

Opposition support **N** $\qquad$ **N** support **Opposition**

failure $\qquad$ *failure*

(β) $\qquad$ $(1 - \alpha)$

Citizens

$1 - g_3$ (vote) g_3

dissent $\qquad$ dissent

$1 - g_4$ **Citizens** g_4

(vote)

$(1 - \beta)$ $\qquad$ *success* $\qquad$ *success* $\qquad$ (α)

Leader allow **N** $\qquad$ **N** allow **Leader**

failure $\qquad$ *failure*

(β) $\qquad$ $(1 - \alpha)$

Citizens

$1 - g_5$ (vote) g_5

repress $\qquad$ repress

N (β) $1 - g_6$ **Citizens** g_6 $(1 - \alpha)$ **N**

failure (revolt) *failure*

$(1 - \beta)$ *success* $\qquad$ *success* (α)
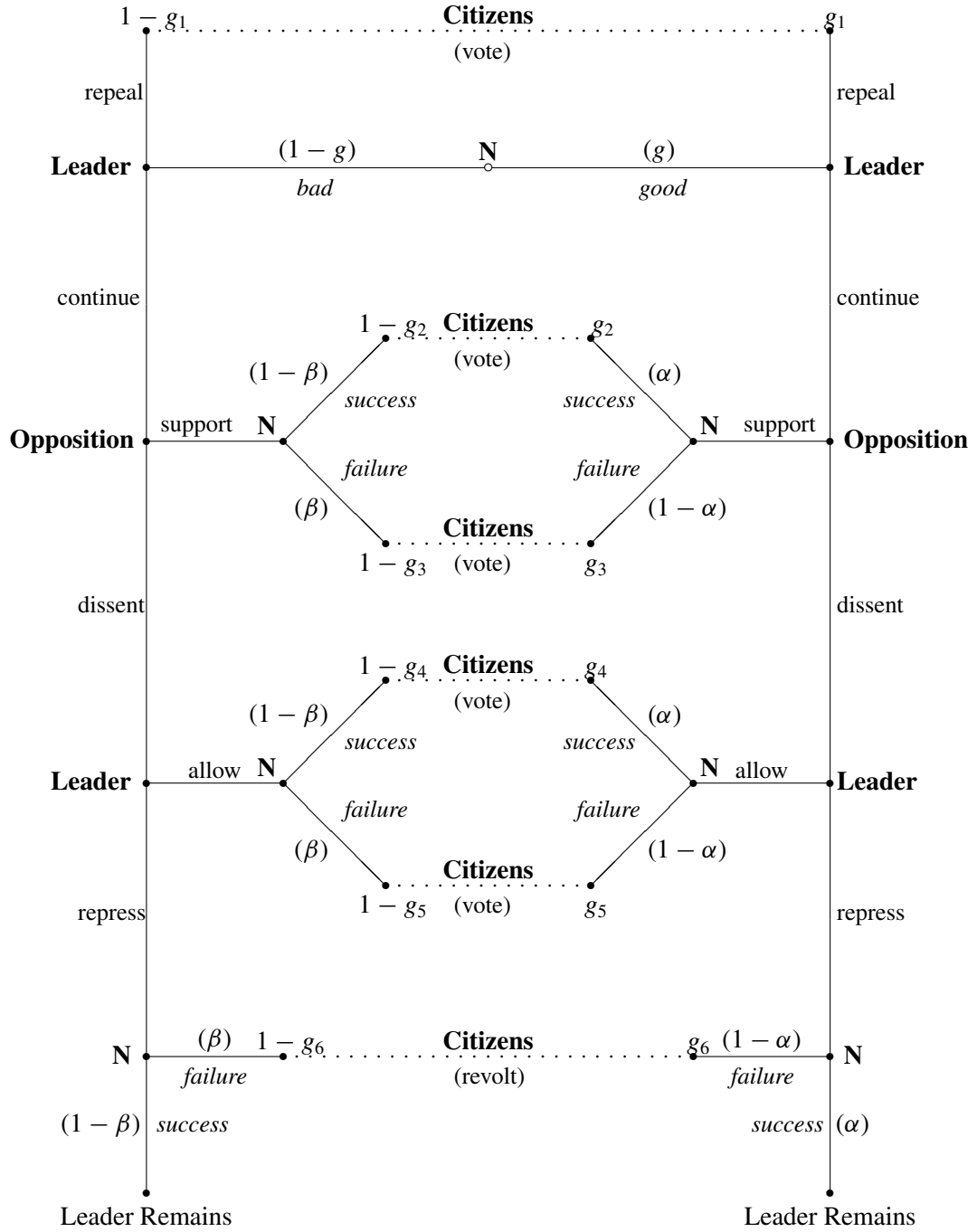
Leader Remains $\qquad$ Leader Remains

Figure 1: Schematic Representation of the First Period of the Game.

costs of replacing the leader, and policy success can only raise them even higher because the repressive apparatus is unlikely to have been damaged, which is what may happen following

policy failure. Second, if the citizens revolt, the revolution succeeds. The basic results do not change if we make revolutionary success a probabilistic event although the expression become quite a bit more cumbersome.

3 The Citizen Strategy

At the time of *selection*, the policy outcome in the first period is realized and the policy benefits are sunk. The citizens will only care about the expected payoff in the next period. In the second period, the (possibly new) incumbent has no reason to distort policy for electoral gain. Therefore, all leaders repeal bad policies and continue good ones. This strategy is optimal regardless of the competence of the incumbent.

The citizens' choice is therefore between keeping a leader about whose competence something can be inferred from strategies and policy outcome in the first period, or replacing that leader, possibly at some cost. How do citizens update their beliefs about the leader's competence?

Whenever the policy quality is known (either observed directly by the leader and the opposition, or inferred from the strategies and outcome by the citizens), it is possible to update beliefs about the competence of the leader by Bayes rule. The posterior belief is then:

$$\hat{\mu} = \begin{cases} \frac{\mu p}{\mu p + (1-\mu)q} \equiv \mu^G & \text{if the policy is good,} \\ \frac{\mu(1-p)}{\mu(1-p)+(1-\mu)(1-q)} \equiv \mu^B & \text{if the policy is bad.} \end{cases}$$

It is not difficult to show that because $p > q$, it follows that $\mu^G > \mu > \mu^B$.

Given the strategy of the incumbent in the second period, the expected payoff for the citizens is:

$$\pi(\hat{\mu}) = b[\hat{\mu}p + (1 - \hat{\mu})q],$$

where $\hat{\mu}$ denotes the probability that the incumbent is competent. Because $p > q$, it follows that $\pi(\cdot)$ is strictly increasing in $\hat{\mu}$. In words, the expected payoff to the citizen is strictly increasing in the expected competence of the incumbent.

To simplify notation, I shall use $\pi^G \equiv \pi(\mu^G)$, and $\pi^B \equiv \pi(\mu^B)$ to denote the expected payoffs based on beliefs about leader's competence when the policy is good and bad, respectively. Similarly, I shall use $\pi_e^o \equiv \pi(\mu_e^o)$ when beliefs are based on expected competence of the opposition.

Because the citizens observe the probability that the opposition is competent, μ^o, prior to making their choice, the expected payoff from replacing the incumbent is $\pi(\mu^o)$. Letting $\hat{g}$ denote the posterior probability that the first-period policy was good, the expected payoff from retaining the incumbent is $\hat{g}\pi^G + (1 - \hat{g})\pi^B$. Let $K = 0$ when the selection is done through elections, and $K = k$ when selection is done through revolution. The citizens will retain the incumbent whenever $\hat{g}\pi^G + (1 - \hat{g})\pi^B \geq \pi(\mu^o) - K$, or, expressed directly in terms of beliefs, whenever:

$$\hat{g}\mu^G + (1 - \hat{g})\mu^B \geq \mu^o - \frac{K}{b(p - q)}.$$

Let $e = 1$ denote a decision to retain the leader, and $e = 0$ denote a decision to replace the leader with the opposition. The citizens' decision rule is a function of the updated beliefs

and is given by:

$$e(\hat{g}, K) = \begin{cases} 1 & \text{if} \quad \hat{g} \geq \frac{\mu^o - \mu^B}{\mu^G - \mu^B} - \frac{K}{(p-q)(\mu^G - \mu^B)} \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

In other words, the citizens will keep the leader if they believe that first-period policy was good with sufficiently high probability. This, of course, gives the leader incentives to conceal information when the policy is bad, which may in turn lead to policy distortions for electoral gain. On the other hand, it also gives the opposition an incentive to reveal the information when the policy is bad to obtain electoral advantage.

Note that the belief required to retain the leader in elections is strictly higher ($K = 0$) compared to the belief necessary to retain the leader by not revolting. If citizens re-elect the leader given some belief about competence, they will never revolt if they have this belief. The converse, however, is not true. There is a range of (pessimistic) beliefs where citizens would not revolt but would replace the leader in elections if given a chance. When the only option is a costly revolution, the citizens must be convinced that the leader is truly incompetent to engage in violence. Thus, the leader will have incentives to confront the citizens with a choice between two unpalatable alternatives: retain someone they believe is not very competent, or overthrow the leader violently at great cost.

While the citizen's choice is deterministic (because at the time selection takes place μ^o is realized and observable), selection appears probabilistic to the other players. Rewriting the decision rule in terms of the opposition's competence yields:

$$\mu^o \leq \hat{g}\mu^G + (1 - \hat{g})\mu^B + \frac{K}{b(p - q)} \equiv \overline{\mu},$$

and thus the ex ante probability that the citizens will retain the leader is:

$$\Pr(\mu^o \leq \overline{\mu}) = F(\overline{\mu}).$$

I now make the following assumption to make the game substantively interesting:

ASSUMPTION 1 (Selection Incentives). $\quad \mu^o \sim U[\mu^B, \mu^G]$.

That is, $F(\cdot)$ is the uniform distribution with support $[\mu^B, \mu^G]$. This assumption implies that if citizens know that the policy is good, they always re-elect the leader: $F(\mu^G) = 1$. If they know that the policy is bad, they always replace the leader: $F(\mu^B = 0)$. This assumption further implies that the citizens will never revolt when they know that the policy is good.

The results do not depend on this distribution being uniform. However, the assumption considerably simplifies notation because the probability that citizens retain the leader in elections is:

$$F(\hat{g}\mu^G + (1 - \hat{g})\mu^B) = \frac{\hat{g}\mu^G + (1 - \hat{g})\mu^B - \mu^B}{\mu^G - \mu^B} = \hat{g}.$$

That is, calculations can be done directly in terms of beliefs at the information sets, which also allows for simple closed form expressions in the solutions. The citizens have six information sets in this game. Figure 1 lists the notational shortcuts, g_i for all $i \in 1, 2, \ldots, 6$,

10

that denote the citizens' beliefs that the policy is good at these sets. The probability that citizens retain the leader following repression is:

$$\gamma_6 \equiv g_6 + m, \quad \text{where} \quad m \equiv \frac{k}{b(p-q)(\mu^G - \mu^B)} > 0.$$

Let $V^j = v + \pi^j$ with $j = G, B$ denote the leader's policy quality-dependent expected payoff from retaining office, let $V_e^o = v + \pi_e^o$ denote the opposition's expected payoff from gaining office, and assume that being in office is rewarding:

ASSUMPTION 2 (Office-Seeking). $v > b + b(p-q)(\mu^G - \mu^B)$.

We now have $V_e^o - \pi^G = v + \pi_e^o - \pi^G = v - b(p-q)(\mu^G - \mu_e^o) > 0$, where the inequality follows from the assumption and $\mu_e^o > \mu^B$. The assumption therefore implies that $V_e^o > \pi^G$. That is, the opposition's expected payoff from holding office is strictly greater than what it would get if a leader whose policy happens to be good remains in office. In other words, the opposition has incentives to seek office even when the policy implemented by the leader is good.[11]

4   The Unique Fully Revealing Efficient Equilibrium

An equilibrium is *fully revealing* if in it citizens are able to infer the policy quality with certainty from the strategies of the other players. It is *partially revealing* if they can do so probabilistically. It is *uninformative* if the only new information comes from the noisy signal. Audience costs are only meaningful in fully or partially revealing equilibria because these are the only equilibria in which the probability of political failure depends on the citizens' selection decision.

The following proposition, whose proof is in Appendix A, demonstrates that if the costs of repression are intermediate, then there exists a unique fully revealing equilibrium.

PROPOSITION 1. *Let $\underline{c} = V^B - \pi_e^o < V^G - \pi_e^o = \overline{c}$. If the costs of repression are intermediate, that is if $c \in [\underline{c}, \overline{c}]$, then there exists a unique fully informative equilibrium in which the leader continues good policies and repeals bad ones; the opposition supports only good policies, and the leader represses dissent only when the policy is good. The citizens always retain leaders who continue policies, and remove leaders who repeal policies.*

Intuitively, if the costs are too high ($c > \overline{c}$), then repression will not be optimal for the leader when the policy is good even if repressing would convince the citizens the policy is good while not repressing would convince them it is bad. Repression is simply too expensive. On the other hand, if the costs are too low ($c < \underline{c}$), then repression is too cheap and it is worth repressing dissent unless allowing it would convince the citizens that the policy is good.

Because $\underline{c} < \overline{c}$, the interval $[\underline{c}, \overline{c}]$ exists for any value of the priors. Whenever repression costs are in this range, the leader strictly prefers to allow dissent and be removed instead of repressing and keeping office. The reason is that when the policy is bad, the leader's own

---

[11]This assumption can be replaced by the more intuitive, but a bit more restrictive, requirement that $v \geq 2b$, which is another way of saying that ego-rents are at least twice the policy benefit from a good policy.

estimate of the expected payoff of remaining in office is low because of the updated belief of competence relative to the expected competence of the opposition. Then, if the costs are not too low, the double whammy of a policy expected to be worse in the next period and the costs or repression necessary to ensure remaining in office outweighs the ego-rents, and the leader prefers to "go quietly into the night."

This equilibrium is efficient because the probability of political failure is zero. Leaders continue only good policies and repeal only bad ones. How stringent are the conditions for its existence? The width of the cost range is:

$$\bar{c} - \underline{c} = \pi^G - \pi^B = b(p-q)(\mu^G - \mu^B) = \frac{b\mu(1-\mu)(p-q)^2}{[\mu p + (1-\mu)q] - [\mu p + (1-\mu)q]^2} > 0.$$

The maximum width, at $\mu = 1/2$, is $\frac{b(p-q)^2}{(p+q)(2-p-q)} < b$. With reasonable values for p and q (that is, values such that it is not simultaneously the case that p is close to 1 and q is close to 0), the width is much smaller. This means that the range of costs that can sustain this equilibrium may be quite small indeed. In the numeric example below, the costs would have to be in the interval [.6625, .7325], anything smaller or higher would not work, and this is the best-case scenario with $\mu = 1/2$.
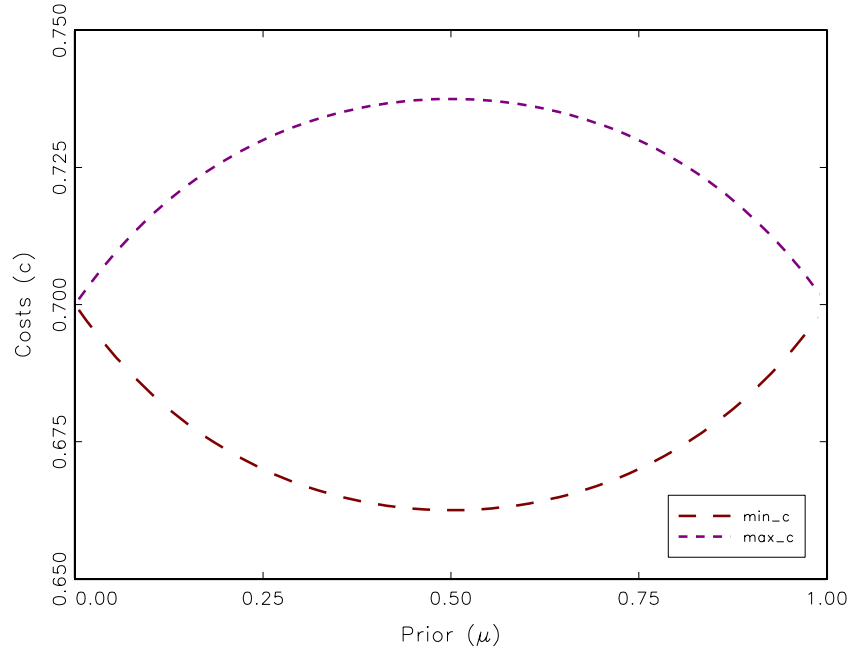

Figure 2: Cost Range for the Fully Informative Equilibrium ($p = .75$, $q = .25$, $b = .3$).

The cost range dramatically shrinks as the prior beliefs are biased either for or against the incumbent (that is, as μ moves away from $1/2$), as shown in Figure 2. For each value of the priors, the costs that can support the fully revealing equilibrium are inside the lens shape. As the figure demonstrates, this range strictly decreases with any bias for or against the incumbent. This implies that the conditions for this equilibrium may be quite stringent if there is any significant bias in the priors regardless of the parameters. Thus, although

12

this efficient equilibrium always exists if the repression costs are in the intermediate range, the condition may be restrictive. (I investigate other implications of comparative statics in Section 7.) What may one expect if the costs are outside that range? One important immediate consequence of Proposition 1 is:

COROLLARY 1. *All other sequential equilibria are either partially revealing or uninformative. In all of them political failure occurs with positive probability.*

There are no other equilibria in which the citizens can infer the policy quality with certainty from the strategies of the leader and the opposition. This means that there are no other equilibria in which the leader repeals only bad policies and continues only good ones.

5   Uninformative Equilibria: Failure Rate Independent of Exogenous Signals

For the remainder of this paper, I analyze the properties of equilibria when the costs of repression are outside the range where the efficient equilibrium can be supported. This is made on the charitable assumption that in the presence of multiple equilibria, the most informative one will be selected (recall that we are investigating the best-case scenario for information transmission).[12]

5.1   Repressive Society: Universal Endorsement

Suppose that the repressive instrument is not too costly; that is, $c < \underline{c}$. In such a society the leader can credibly threaten to repress all forms of dissent regardless of policy quality. I first show that given these low costs of repression, it is optimal for the leader to repress all dissent. The necessary condition for repression to be optimal when the policy is good is:

$$c \leq V^G - \pi_e^o - (1 - \alpha)(V^G + v)(1 - \gamma_6), \qquad (2)$$

and the analogous condition when the policy is bad is:

$$c \leq V^B - \pi_e^o - \beta(V^B + v)(1 - \gamma_6). \qquad (3)$$

Because $0 < m \leq \gamma_6 \leq 1$, it follows that whenever $c \leq \underline{c}$, the condition in (3) will be satisfied as well. I now show that (3) is, in fact, the binding condition that ensures that repression is optimal regardless of policy quality. Subtracting (3) from (2), and using $\beta > \frac{1}{2} > 1 - \alpha$, yields:

$$V^G - V^B + (1 - \gamma_6)\big[\beta(V^B + v) - (1 - \alpha)(V^G + v)\big] > \left[\frac{1 + \gamma_6}{2}\right](V^G - V^B) > 0.$$

---

[12]Characterizing all sequential equilibria is rather tedious, and many of them can be eliminated as unintuitive (Cho and Kreps 1987). The complete characterization of the game's equilibria and other proofs are available from the author. All other pure-strategy equilibria with the leader playing a separating repressive strategy are unintuitive. This leaves only equilibria where the pools on repression or no repression. Further, equilibria where the leader allows dissent always but the opposition plays a separating strategy inducing the leader to repeal all policies are also unintuitive. Thus, we only have two sets of uninformative equilibria, where both the leader and the opposition play pooling strategies.

This means that if condition (3) is satisfied, then (2) will be satisfied as well. That is, if it is optimal to repress when the policy is bad, it is also optimal to repress when it is good. As we have seen, optimality of repression when the policy is bad is ensured whenever $c \leq \underline{c}$. We conclude that in this case, repression is optimal regardless of policy quality.

Given that the leader represses dissent, in any equilibrium the opposition must be supporting both the good (by Lemma 6) and the bad (by Lemma 3) policies. If the leader always represses, then the opposition always endorses. By Corollary 1, in all these equilibria the leader either always continues the policies or always repeals them. Therefore, there are two pure-strategy equilibria when $c < \underline{c}$: In both, the leader always represses and the opposition always dissents; in one of them, the leader repeals all policies, and in the other the leader continues them.

Let g_A solve the equation $g = \frac{b}{V^B - \pi_e^o}$, and let g_B solve the equation $g = \frac{b}{V^G - \pi_e^o}$. Note that $V^B < V^G \Rightarrow g_A < g_B$. The following proposition, whose proof is in Appendix A, establishes the existence of these uninformative equilibria.

PROPOSITION 2.   *If $c < \underline{c}$, then there exist sequential equilibria in which leader always represses dissent, and the opposition endorses all policies regardless of quality. If the prior belief that the policy is good, $g \in [g_A, g_B]$, then only Type I failure can occur (leader repeals all policies). If $g \geq g_B$, then both failures can occur but with g sufficiently high, only Type II failure (leader continues all policies) is intuitive.*

If citizens attach a relatively low initial probability to the policy being good, then the leader will not continue policies, and so the only failure possible is that good policies get repealed. The reason for this is intuitive: Because in these uninformative equilibria citizens can only rely on the noisy signal to infer information, their updated beliefs will generally not be sufficiently favorable to the leader (because the prior is so low) to induce continuation of bad policies. On the other hand, if they are quite optimistic, then the leader may keep the policies.

5.2   Non-repressive Society: Endorsement Babble

Suppose now that the repression instrument is quite costly: $c > \bar{c}$. The leader can never credibly threaten to repress dissent. All uninformative equilibria in this case involve the leader allowing dissent regardless of policy quality. To see this, note that allowing dissent when the policy is good is optimal whenever:

$$c \geq (V^G - \pi_e^o)\big[1 - \alpha g_4 - (1 - \alpha)g_5\big] - (1 - \alpha)(V^G + v)(1 - \gamma_6). \qquad (4)$$

The necessary condition (at $g_4 = g_5 = 1$) is always satisfied because $c > 0$. Further note that the sufficient condition (at $g_4 = g_5 = 0$) is the converse of the necessary condition for repression in (2). Recalling that $\bar{c} = V^G - \pi_e^o$ from (10), we conclude that whenever $c > \bar{c}$, condition (4) will be satisfied.

Optimality of allowing dissent when the policy is bad requires:

$$c \geq (V^B - \pi_e^o)\big[1 - (1 - \beta)g_4 - \beta g_5\big] - \beta(V^B + v)(1 - \gamma_6). \qquad (5)$$

As before, the necessary condition is always satisfied. The sufficient condition is the converse of (3). We have already seen that this bound is strictly smaller than (2), and conclude

that (4) is the binding condition, That is, if it is optimal to allow dissent when the policy is good, it is also optimal to allow it when the policy is bad. Thus, whenever $c \geq \overline{c}$, allowing dissent regardless of policy quality is optimal.

When the leader allows dissent, the opposition is free to choose any strategy available. However, if it plays any separating strategy, then, by Corollary 1, in equilibrium the leader always repeals policies and does not allow it to signal quality to the citizens. These Type I failure equilibria, however, are all unintuitive. To see why, note that the leader could instead continue the good policy if that would convince citizens of its quality, but because the opposition is playing a separating strategy, continuing the bad policy cannot benefit from the updated belief because it reaches a different information set.

This leaves four equilibria in pure strategies in this range: the opposition either always dissents or always endorses, and the leader either always repeals or always continues the policy. The following proposition, whose proof is in Appendix A, establishes the existence of these uninformative equilibria.

PROPOSITION 3. *If $c > \overline{c}$, then there exist sequential equilibria in which the leader always allows dissent, and the opposition either always endorses or always dissents. If the prior belief that the policy is good, $g \in [g_A, g_B]$, then only Type I failure can occur. If $g \geq g_B$, then both failures can occur but with g sufficiently high, only Type II failure is intuitive.*

This parallels the results for repressive societies: citizens are unable to infer policy quality from the strategies of the leader and the opposition. The reason is slightly different, however. In a repressive society, the threat to repress is always credible, and so the opposition never dissents from any policies, which means citizens cannot learn anything from the strategies. In a non-repressive society, the threat to repress is never credible, and so the opposition can say whatever it wants, and given that it wants citizens to believe the policy is bad, its pronouncements about quality are never believable, and so the citizens cannot learn anything either. Only when the partially credible repressive threat induces the opposition into truth-telling can full information disclosure occur.

Note further that the non-strategic signal plays no role in the probability of political failure occurring in either the universal support or endorsement babble equilibria. Although citizens do infer information from the signal, their behavior does not affect the leader's strategy: leaders either always repeal or always continue all policies. The rate of failure only depends on the citizens' priors, that is, their bias toward or against the leader. If they are positively disposed, then leaders continue all policies. If they are not, then leaders repeal all policies.

In the next section I investigate the properties of partially revealing equilibria, in which the probability of failure does depend on the quality of the non-strategic signal, and which do exist even for pessimistic priors (recall that if citizens are quite pessimistic, $g < g_A$, then the pure-strategy equilibria do not exist).

6   Partially Revealing Equilibria: Endogenous Rate of Failure

While the results of the preceding section may appear discouraging, they do not tell the entire story. We have yet to investigate the role of noisy signals in depth. In the fully

revealing equilibrium, these signals play no role because citizens are able to infer policy quality from the strategies of the informed players. In the uninformative equilibria, citizens do update based on these signals but this is not sufficient to induce the informed players to adopt even semi-separating strategies. What role, then, can these exogenous signals have?

The model allows for exogenous signals of varied quality. If both α and β are close to 1, then the signals are of very high quality because they reveal the policy type with near certainty. Conversely, low values of these parameters imply very noisy signals. It is also possible to examine the bias of these signals. For example, keeping α close to 1 but β low describes the situation where the signals are biased toward good news in the sense that if the policy is good, they would report that with very high probability but if it is bad, the corresponding probability is low.

If the probability of failure depends on the quality of the exogenous signals, then the leader must be playing a mixed strategy at the repeal stage. By Lemma 2, these equilibria cannot involve continuing the bad policy with certainty while repealing the good one with positive probability. Thus, we shall look for equilibria in which the leader continues good policies always, and continues bad ones with probability r. The equilibrium probability r^* is then the endogenously determined rate of (Type II) failure.

Suppose then that the leader plays this semi-separating continuation strategy. By lemmas 4 and 5 it follows that the opposition and the leader must be pooling at the remaining information sets. The discussion and the result in this section is stated for the case where the opposition always dissents and never gets repressed, but the cases of the opposition always supporting and the leader pooling on repression or non-repression can be easily established analogously.

Thus, suppose $c > \bar{c}$ and suppose the leader continues good policies always, and continues bad ones with probability r. The optimality of the leader allowing dissent and the opposition always dissenting is established in the proof of Proposition 3. By Bayes rule, $g_1 = 0$ because conditional on observing repeal, citizens would conclude the policy must have been bad because good policies are never repealed. Further by this rule, the posteriors at the two information sets along the equilibrium path are:

$$g_4 = \frac{g\alpha}{g\alpha + (1-g)r(1-\beta)}; \qquad g_5 = \frac{g(1-\alpha)}{g(1-\alpha) + (1-g)r\beta}. \qquad (6)$$

Continuing the good policy is always optimal because repealing it yields (at $g_1 = 0$) exactly $\pi_e^o < b + \pi_e^o$, which is the least the leader could get by continuing it.

Since the leader is willing to randomize when the policy is bad, it follows that the payoff of repeal and continuation are the same. Repeal gets π_e^o and continuing gets $-b + \pi_e^o + (V^B - \pi_e^o)[(1-\beta)g_4 + \beta g_5]$. Setting these payoffs equal to one another and solving yields:

$$(1-\beta)g_4 + \beta g_5 = \frac{b}{V^B - \pi_e^o}. \qquad (7)$$

At $r = 0$, we have $g_4 = g_5 = 1$ (that is complete separation). This cannot be an equilibrium because the leader has an incentive to continue the bad policy if winning the election is guaranteed (as it would be with these beliefs). At $r = 1$, we have a pooling equilibrium only when $(1-\beta)g_4 + \beta g_5 \geq \frac{b}{V^B - \pi_e^o}$; that is, only when the prior g is sufficiently high.

16

Otherwise, we can solve (7) for any prior $g \in (0, 1)$, and find the optimal r^* that satisfies it. To find this probability, we use (6) in (7), which yields the quadratic $Ar^2 + Br + C = 0$ with coefficients:

$$A = -(1 - g)^2 \beta(1 - \beta)b,$$
$$B = g(1 - g)\big[(V^B - \pi_e^o)\beta(1 - \beta) - b(1 - \alpha - \beta + 2\alpha\beta)\big],$$
$$C = g^2\alpha(1 - \alpha)(V^B - \pi_e^o - b).$$

The positive root (the only valid one because we need it to be a probability) is then:

$$r^* = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \tag{8}$$

Since $A < 0$ and $C > 0$, the discriminant is always strictly positive, so the square root exists. If $r^* \leq 1$, the specified strategies constitute a sequential equilibrium, otherwise the Type II equilibrium of Proposition 3 exists. The above reasoning then yields the following result:

PROPOSITION 4. *If $c > \bar{c}$, then there exists a sequential equilibrium in which the leader always allows dissent and the opposition always dissents. The leader continues good policies always and continues bad ones with probability r^* from (8). The probability of (Type II) political failure, r^*, is strictly decreasing in the quality of the exogenous signal.*

It is clear that there exists an analogous equilibrium with the opposition always endorsing the policy regardless of quality, and in fact the probability of continuing the bad policy is exactly the same. The comparative statics on the quality of the exogenous signal and their implications are in the section that follows. The intuition for the above result is straightforward. The opposition cannot reveal the policy quality credibly if the leader has no repressive threat. Given the opposition's babbling, the leader seems to have no incentive to reveal any information either. However, this turns out not to be the case in the presence of exogenous signals.

With these signals, the citizens can infer something about the policy quality anyway, reducing the incentives to persist in bad policies. If the signal to noise ratio of this exogenous source of information is high, then these incentives diminish significantly. That is, the threat of independent revelation of policy quality may help keep the leader in check. This is an imperfect mechanism, unlike the strategic one in the fully revealing equilibrium from Proposition 1, but a mechanism nevertheless. I now turn to the analysis of substantive implications of these findings.

7 Discussion

That citizens cannot rely on the government coming clean about what it knows is clear. Citizen control through the threat of the re-selection sanction alters the strategic incentives of the government even in our reduced context where their preferences over policies are identical. The government is usually better informed about the effects of its policies than the

citizens. Unfavorable information would adversely affect its chances of retaining power because citizens would become pessimistic about its competence. Therefore, the government may seek to strategically misrepresent what it knows for electoral purposes.

We found that rational retrospective forward-looking citizens will attempt to infer the likelihood of getting good policies in the future by evaluating the past performance of the current leader. Should they become quite pessimistic about these prospects, they would replace the leader with the untried opposition unless doing so is prohibitively costly. That is, they can impose audience costs on the leader for pursuing policies contrary to the wishes of the citizenry.

The fate of the leader turns on the probability of this sanction, which depends on the citizen's beliefs and costs of replacement. Thus, the leader has a two-pronged strategy: obfuscate unfavorable information or raise these costs. However, its effectiveness may be blunted by the presence of an informed opposition with incentives to thwart the leader's attempts to hide such information. These incentives are themselves subject to distortion through the leader's repressive action.

7.1 Repressive Threat Credibility and Citizen Learning

Can the opposition credibly reveal to citizens the policy quality? In general, the answer seems to be negative. Only when the leader's repressive threat is partially credible can the opposition commit to a strategy that would reveal this information (Proposition 1). If repression is cheap, then the credible threat to stifle any dissent causes the opposition to endorse everything the leader does (Proposition 2). If repression is costly, then the opposition can do anything it wants with impunity because the leader cannot credibly threaten it with repression. This behavior cannot reveal information precisely because the opposition would always want the citizens to believe that the leader's policies are bad. Given this endorsement babble (mostly dissenting, but not necessarily so), the leader would appear to have no incentive whatsoever to condition the repeal strategy on anything but the citizen's prior estimate of its quality (Proposition 3). This polity can experience lots of dissent, but it will not be informative. Thus, *from the strategic informational perspective*, citizens under a democracy are no more able to learn the qualities of the policies implemented by their government than citizens under an autocracy.

What happens in mixed regimes where the leader can credibly threaten to repress some, but not all, dissent? We examine the cost range that supports the fully revealing equilibrium. As already noted, it shrinks when citizens are biased for or against the incumbent relative to the opposition: *Audience costs are most likely to matter in closely contested polities, when citizens are favoring neither the incumbent nor the opposition too much.*

This effect of bias in priors in the fully informative equilibrium is closely related to the one in Smith (1998, p. 629), who also finds that as bias increases in either direction, the probability of intervention (that reveals information) decreases. The logic is the same: when bias exists, the impact that new information will have on citizen behavior is comparatively weak, so there is less incentive to reveal such information and risking associated costs.[13]

---

[13]Smith (1998) presents the sole existing model of endogenous audience costs. In that model, war reveals competence fully and in the separating equilibrium, only competent leaders threaten and intervene; incompetent ones do not threaten and do not intervene. Why does an incompetent leader never threaten? The argument for

The cost range increases in b : as policies become more important (good ones deliver more and bad ones cost more), full revelation becomes easier because the range of costs that would support it widens. This implies that *audience costs are most likely to matter only for salient policies of great national importance*. This finding is intuitive and seems corroborated by empirical evidence (Aldrich, Sullivan and Borgida 1989).

It is worth noting that, perhaps counter-intuitively, the cost range width is unrelated to the private benefits from holding office providing Appendix 2 is satisfied (they are sufficiently larger than public benefits). On the other hand, the interval is sensitive to the expected benefit from having a competent leader versus an incompetent one: the larger the difference in producing a good policy between the two types, the wider the range that can support full efficiency in equilibrium. The cost range increases in p and decreases in q for all $p > q$. Very high p means that the competent leader is very likely to produce good policies, and very low q means that the incompetent leader is very unlikely to do so. When p is close to q, the expected difference competence will make is too small to matter. This suggests that domestic audience costs are likely to matter more for complex matters where competence is deemed crucial than for issues where leader's competence is not expected to make much of a difference.

7.2 The Watchful Eye in Non-Repressive Societies

Given that strategic information transmission will be non-existent except in mixed-regimes, the exogenous source of information becomes quite important, even if imperfect, disciplining device (Proposition 4). This result, along with Proposition 1 demonstrate how audience costs can arise endogenously through two different mechanisms for information transmission: strategic behavior of the informed players (perfect audience costs) and non-strategic revelation by a non-political source (weak audience costs).

An important claim related to audience costs is that they are linear in regime type. That is, democracies are able to generate them at higher rates than other regimes. In this model, neither democracies nor autocracies can generate perfect costs. However, democracies may be able to generate weak costs through their independent signaling sources. There is a *nonlinear relationship between regime type and audience costs, with audience costs being nearly non-existent in autocracies, then sharply increasing for mixed regimes, then declining for democracies, where their level depends on media protections*.

Turning to the effect of the non-political source, Figure 3 demonstrates the probability of political failure across the entire range of citizen prior beliefs and for varying quality of the noisy signal.

---

separation works only if the loss from a called bluff outweighs the gain from a bluff that succeeds. Since in this equilibrium the threat is credible, the likelihood that the bluff is called cannot be too high, and so the expected payoff from bluffing should be relatively high unless one assumes catastrophic consequences to having one's bluff called. It is such consequences that the proof assumes: if the leader fails to intervene, citizens are assumed to conclude that the leader is the least competent type. Since Bayes rule is undefined after zero-probability events, one must postulate the beliefs in this crucial contingency. Off-the-path beliefs can be extremely important in generating equilibrium behavior, and so the informative equilibrium essentially relies on an assumption that allows citizens to threaten with beliefs instead of actions. This is a shortcoming of game theory that has not been adequately resolved even though many refinements exist that specify what beliefs should be "reasonable" after zero-probability events.
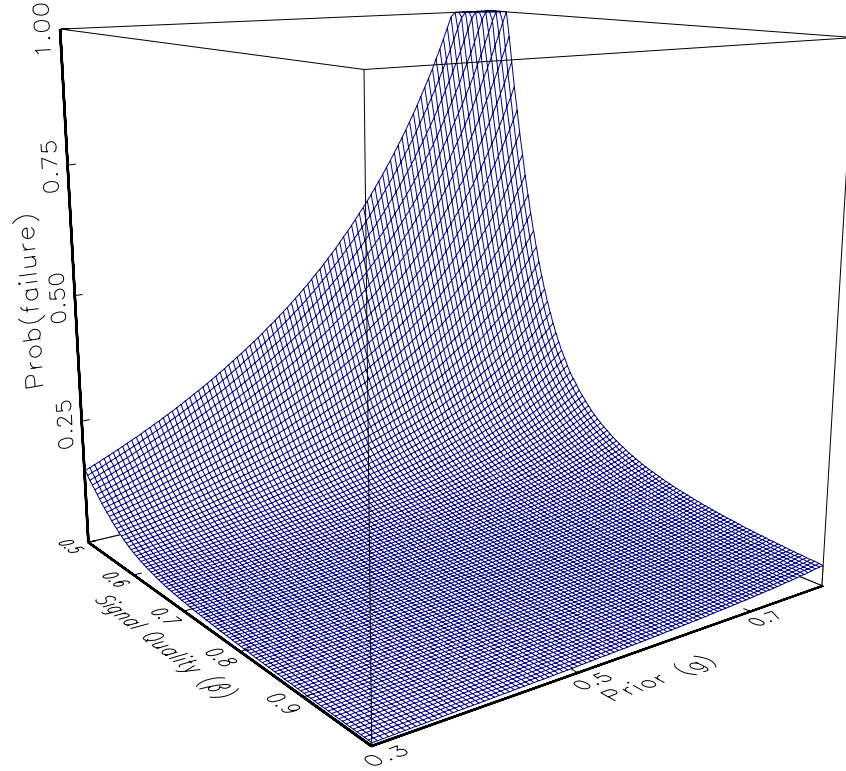
Figure 3: Impact of Signal Quality and Prior Bias on Policy Failure.

The x-axis represents the citizen's prior belief that the policy is good. This can be interpreted as their incumbent bias. Low values represent bias against the leader, and high values represent bias in leader's favor. The y-axis represents the probability that a bad policy generates a correct signal.[14] The z-axis represents the probability that the bad policy is continued in equilibrium, and hence it represents the complement of the probability of audience costs being imposed.

The probability of political failure is increasing in the bias toward the leader. The more optimistic citizens are about the incumbent, the more likely is the latter to continue bad policies, and the less likely are citizens to impose audience costs on him for doing so. The magnitude of this effect is conditioned by the quality of the exogenous signal. If the signal is of very high quality (meaning that a bad policy is likely to be revealed as such), the probability of failure is low and its increase with bias very limited. In the example, it starts from close to 0 percent and goes up to about 5 percent even at the most extreme pro-leader bias.

The situation differs dramatically when the signal's quality is bad. In this case, the probability of failure is higher from the start and quickly reaches unity. In the example, it begins at about 12 percent when citizens have anti-leader bias, and rapidly increases to certainty

---

[14]The probability that a good policy generates a correct signal is constant and high because I assume that this would be the case regardless of regime type: all governments have incentives to make sure citizens learn about their successes, but governments who can control the media can suppress its ability to reveal their failures.

20

with the bias shifting in favor of the leader. In other words, under these conditions, citizens are quite unlikely to be able to impose any audience costs on the leader.

The impact of signal quality is much more pronounced than that of prior bias. To see this, note the abrupt drop-off in probability of failure even at the high pro-leader bias of about 80 percent: it begins from certainty when the signal is bad and then plunges to 25 percent with a moderate increase in signal quality of less than 20 percent. To effect the corresponding drop holding the signal at a constant low quality, the bias has to shift against the leader by over 45 percent.

While prior bias is doubtless important, the "watchful eye" that produces the exogenous signal is crucial for the rate at which leaders risk continuing bad policies against the wishes of their citizenry. Thus, from the *non-strategic informational perspective*, citizens under a democracy are far more likely to learn the qualities of the policies implemented by their government. This ability varies with the quality of the watchful eye, and can provide an imperfect sanctioning mechanism.[15]

7.3   Information and the Fate of Leaders

One consequence of the nonlinear audience costs is that *leaders in mixed regimes are more vulnerable to repealing policy than either autocrats or democrats.* This is because repeal signals so unequivocally that policy is bad. While both democrats and autocrats can repeal policies and remain in office, it is generally not possible for leaders in mixed regimes to do so. Further, even in the watchful eye equilibrium in which repeal also implies losing office, the probability of repeal is strictly smaller than the probability of the policy being bad.

This implies that in mixed regimes leaders will have the shortest predicted term in office following policy repeal. On the average, we expect that leaders in mixed regimes are removed at higher rates than democrats or autocrats following repeal of some bad policy. For example, losing a war can be seen as repealing of a bad policy, and so our model would predict that leaders of mixed regimes would tend to lose office much more often than their counterparts in democracies or autocracies.

The model permits extreme flexibility at the policy repeal stage. However, in reality leaders may find it very difficult to simply stop policies, especially if this is going to signal their quality to the electorate. Because repealing a policy has such grave consequences for leaders of mixed regimes, one would expect the strongest resistance to such perfect signaling there. In other words, leaders in mixed regimes have great incentives to stick with a bad policy as long as the opposition and the citizens cannot infer for sure that it is bad. Once the leadership admits to its quality by repealing it, the game is up. Thus, if the leaders can either convince the opposition to toe the line and somehow ensure that the noisy signal received by the citizens is not too damaging, mixed-regime leaders would tend to prolong bad policies. However, if either the opposition can no longer be subdued or the truth hidden from the citizens, repealing the policy would immediately lead to the leader being removed from power. For example, such considerations explain why Germany prolonged its participation in the First World War for as long as it did and why its leaders got what they did once they surrendered (Goemans 2000).

---

[15]Empirically, there may be limits to how much citizens may be able to learn even from intensive coverage (Bennett 1994).

The results further imply that in non-repressive regimes, exogenous signals would be much more important than strategic behavior of the leadership and the opposition. For democracies, this is in keeping with the finding that the president's "decisions on policy do not affect his popularity so much as their results" (Kernell 1978). In particular, the fate of democratic leaders would crucially depend on the chance revelation of some particularly favorable or damaging piece of information that is independently verifiable and therefore credible. Thus, democrats may be unwilling to repeal bad policies until forced by sudden and unexpected revelation of such information (e.g. the Pentagon Papers). Brody and Shapiro (1989) show how the news of the Iran-Contra Affair led to a sharp decline in President Reagan's popularity, contrary to the expectations of the "rally 'round the flag" theory (Mueller 1973).

8   Conclusion

Domestic political audiences could help leaders commit credibly to a course of action by tying their hands. To do that, these audiences must be able to sanction the leader, and rational audiences would only sanction a leader who behaves contrary to their interests. It is not difficult to generate such "bad" behavior if the leader has policy preferences that are very different from the citizens. Consequently, I constructed a model where leaders would have least incentive to deviate from their responsibilities as agents of the people. In such an environment, office-holding motivations may induce distorting behavior by the incumbents, and this is precisely what citizens would wish to sanction. Ironically, the very existence of a sanctioning device to control leaders can produce perverse incentives for behavior it is designed to eliminate. Domestic audience costs constitute the reduction in probability of retaining office for pursuing policies contrary to the interests of one's constituency. Hence, generating such costs crucially depends on the citizens' ability to infer policy quality from information available to them.

I considered two potential sources of such information: politically-motivated strategic behavior of better-informed actors like the government and the opposition, and a non-political but potentially biased source, like the media. The results suggest that perfect audience costs can arise endogenously only in mixed regimes where the costs of repressing dissent are neither too high nor too low. The opposition's ability to credibly reveal private information is severely curtailed by its incentives to seek office. Only under limited circumstances can they commit to a fully revealing strategy, where the credibility of their commitment is induced by the partial repressive threat of the leader. More generally, however, the opposition's behavior degenerates either in universal endorsement of policies when the leader represses all dissent (autocracies), or endorsement babble when the leader never represses (democracies).

Thus, the exogenous noisy signaling mechanism plays a more important role in disciplining leaders, who will condition their behavior on the chance that the citizens would learn quite a bit about policy quality anyway. This may help explain the consequences of *glasnost* for the Soviet decision to withdraw from the war in Afghanistan: as the quagmire became common knowledge, the pressure to repeal this bad policy increased substantially.

These findings show that while audience costs can arise endogenously given a strategic citizenry, their relationship with regime type is not linear. Thus, the model provides a micro-

foundation for this assumption but also sounds a cautionary note about its use in applied research.

Numerous empirical hypotheses can be derived from this model. For example, audience costs are most likely in closely contested polities, and when policies involve complex matters of great salience. Further, consider leaders who are perceived to have pursued bad policies. Because repeal is such a strong signal in mixed regimes, one would expect that the tenure of leaders in mixed regimes is most adversely affected by failure. The fate of democrats is heavily dependent on an independent signal received by the citizens about the policy. These signals are not systematic, which may make them quite hard to incorporate into statistical models. It is not the opposition that disciplines democratic governments but the watchful eye of sources of information free of government control. However, this also suggests that the behavior of democracies should be systematically related to the institutional provisions protecting the freedom of the press: not all democracies are equal.

A Proofs

We now round up a herd of lemmas that significantly simplify equilibrium analysis by eliminating a large number of candidate strategy profiles.

LEMMA 1. *In any sequential equilibrium, either $g_4 = g_5 \in \{0, 1\}$ or else $0 < g_5 < g_4 < 1$; and either $g_2 = g_3 \in \{0, 1\}$ or else $0 < g_3 < g_2 < 1$.*

Proof. The only difference between g_4 and g_5 is induced by the different non-strategic probabilities associated with the two policy types. Let $\hat{g}$ denote the citizens' updated belief *before* observing the noisy signal. Given the information obtained from the signal, Bayes rule yields $g_4 = \frac{\hat{g}\alpha}{\hat{g}\alpha+(1-\hat{g})(1-\beta)}$ and $g_5 = \frac{\hat{g}(1-\alpha)}{\hat{g}(1-\alpha)+(1-\hat{g})\beta}$. For $\hat{g} \in \{0, 1\}$, it is readily seen that $g_4 = g_5 = \hat{g}$. For any $\hat{g} \in (0, 1)$, it is easily verified that $0 < g_5 < \hat{g} < g_4 < 1$. Because this holds for any sequence of completely mixed strategies, it must hold in any sequential equilibrium. The proof for g_2, g_3 is analogous. □

LEMMA 2. *There is no sequential equilibrium in which the leader repeals the good policy and continues the bad one with positive probability.*

Proof. Suppose the leader repeals the good policy and continues the bad one with positive probability in some equilibrium. Bayes rule pins down the posterior $g_1 = 1$. Given this posterior, repealing the bad policy would yield V^B. We now show that under no circumstances would the leader continue the bad policy. There are three potential strategy profile types to consider for expected payoffs when the policy is bad.

Case 1: the opposition dissents and the leader allows it. In this case the leader would get *at most* $V^B - b < V^B$, so continuation is not rational. Case 2: the opposition dissents and the leader represses it. In this case the leader would get *at most* $V^B - b - c < V^B$, so continuation is not rational. Case 3: the opposition endorses the policy. In this case the leader would get *at most* $V^B - b < V^B$, so continuation is not rational. □

LEMMA 3. *There is no sequential equilibrium in which the opposition dissents when the policy is bad and gets repressed.*

23

Proof. Suppose it is optimal for the leader to repress when the policy is bad, and thus $-c + V^B - \beta(V^B + v)(1 - \gamma_6) \geq \pi_e^o + (V^B - \pi_e^o)\big[(1 - \beta)g_4 + \beta g_5\big]$. Because the RHS is minimized at $g_4 = g_5 = 0$, it follows that a necessary condition for the existence of this equilibrium is that $c \leq V^B - \pi_e^o - \beta(V^B + v)(1 - \gamma_6) \equiv \bar{c}$. Suppose now that it is also optimal for the opposition to dissent, and thus $\pi^B - (1 - c) + \beta(V_e^o - \pi^B)(1 - \gamma_6) \geq V_e^o - (V_e^o - \pi^B)\big[(1 - \beta)g_2 + \beta g_3\big]$. Because the RHS is minimized at $g_2 = g_3 = 1$, it follows that another necessary condition for the existence of this equilibrium is that $c \geq 1 - \beta(V_e^o - \pi^B)(1 - \gamma_6) \equiv \underline{c}$. Thus, such an equilibrium requires $c \in [\underline{c}, \bar{c}]$. However, this interval does not exist. To see that $\underline{c} > \bar{c}$, note that $\underline{c} - \bar{c} = 1 - V^B + \pi_e^o + \beta(1 - \gamma_6)(v + 2\pi^B - \pi_e^o) > 0$, where the inequality follows from $1 - V^B + \pi_e^o = 1 - v + \pi_e^o - \pi^B > 0$ because $1 > v$ and $\pi_e^o > \pi^B$, and from $v + \pi^B - \pi_e^o = v - b(p - q)(\mu_e^o - \mu^B) > 0$ from Assumption 2. Thus, there can exist no value for c that would satisfy both necessary conditions for this equilibrium. Consequently, such an equilibrium cannot exist. □

LEMMA 4. *Suppose that the leader continues all policies. Then there is no sequential equilibrium in which the opposition plays a separating strategy.*

Proof. Suppose first that the opposition dissents only when the policy is good, which implies $g_2 = g_3 = 0$. Given these beliefs, the payoff from deviating to endorsing the good policy is V_e^o. The payoff from dissenting from the good policy and getting repressed is strictly smaller than $V_e^o - (1 - c) < V_e^o$, so dissent cannot be optimal. If the leader does not repress this dissent, then Bayes rule pins down $g_4 = g_5 = 1$, and the expected payoff from dissenting is $\pi^G < V_e^o$, and so dissent cannot be optimal in this case either.

Suppose now that the opposition dissents only when the policy is bad, which implies $g_2 = g_3 = 1$. If the leader allows this dissent, Bayes rule pins down $g_4 = g_5 = 0$. The leader could strictly improve the payoff by repealing the bad policy, which would yield *at least* $\pi_e^o > \pi_e^o - b$, which is what continuing it gets. Thus, in any such equilibrium the leader must be repressing dissent when the policy is bad. But Lemma 3 shows that no such equilibria exist. □

LEMMA 5. *Suppose that the leader continues all policies and the opposition always dissents. Then there is no sequential equilibrium in which the leader's plays a separating repression strategy.*

Proof. If the leader represses only when the policy is good, Bayes rule pins down $g_6 = 1$. Endorsing the good policy yields *at least* $\pi^G > \pi^G - (1 - c)$, which is what the opposition gets by dissenting. By Lemma 3, the profile where the leader represses dissent when the policy is bad cannot be an equilibrium either. □

Proof of Proposition 1. First, I derive the conditions that the following strategies as equilibrium. Next, I show that there are no other fully revealing equilibria. The claim is that if $c \in [\underline{c}, \bar{c}]$, then the following strategies and beliefs constitute a unique (up to specification of off-the-path beliefs) sequential equilibrium:

- Leader: when the policy is good, continue and repress dissent; when the policy is bad, repeal and do not repress dissent; in the second period, continue good policy and repeal bad policy.

- Opposition: support good policy and dissent from bad policy;
- Citizens: select according to (1) and the following updated beliefs:
    - along the equilibrium path, by Bayes rule, $g_1 = 0$, and $g_2 = g_3 = 1$;
    - off the equilibrium path, $g_4 = g_5 = 0$, and $g_6 = \underline{g_6}$.

In this equilibrium, Bayes rule pins down beliefs $g_1 = 0$, and $g_2 = g_3 = 1$. Optimality of repressing when the policy is *good* implies that:

$$-c + \alpha V^G + (1-\alpha)(-v) + \gamma_6(1-\alpha)(V^G + v) \geq \pi_e^o + (V^G - \pi_e^o)[\alpha g_4 + (1-\alpha)g_5]. \quad (9)$$

The necessary condition for this to hold, at $g_4 = g_5 = 0$, is then $c \leq V^G - \pi_e^o - (1 - \gamma_6)(1-\alpha)(V^G + v)$. This yields the upper bound on the costs at $\gamma_6 = 1$:

$$c \leq V^G - \pi_e^o \equiv \overline{c}. \quad (10)$$

To obtain an expression that satisfies $\gamma_6 = 1$, we solve for g_6, which gives $1 - m$. Thus, we have a bound for this belief: $\underline{g_6} = \max\{0, 1 - m\}$. Any $g_6 \geq \underline{g_6}$ guarantees that $\gamma_6 = 1$. I now show that (9) also establishes an upper bound on the admissible off-the-path beliefs g_4 and g_5. Rewriting the expression for the necessary condition with $\gamma_6 = 1$ yields:

$$\alpha g_4 + (1-\alpha)g_5 \leq 1 - \frac{c}{V^G - \pi_e^o} \equiv \overline{g} < 1. \quad (11)$$

By Lemma 1, this implies that $g_4, g_5 < 1$. Let $\overline{g}$ denote the largest belief that is necessary to render repression optimal. Optimality of allowing dissent when the policy is *bad* implies: $\pi_e^o + (V^B - \pi_e^o)[(1-\beta)g_4 + \beta g_5] \geq -c + (1-\beta)V^B + \beta(-v) + \gamma_6\beta(V^B + v)$. We now have $g_5 \leq g_4 \Rightarrow (1-\beta)g_4 + \beta g_5 \leq \alpha g_4 + (1-\alpha)g_5 \leq \overline{g}$, where the first inequality is established by Lemma 1, and the second follows from $\alpha, \beta > \frac{1}{2}$, and the last is from (11). Thus, given the maximal belief allowable in equilibrium by (11), optimality of dissent yields the necessary condition: $c \geq (1 - \overline{g})(V^B - \pi_e^o) - (1 - \gamma_6)\beta(V^B + v)$. We already know that $\overline{g} < 1$, and that γ_6 must be sufficiently high, so we can rewrite this as a sufficient condition with $g_4 = g_5 = 0$ and $\gamma_6 = 1$, or: $c \geq V^B - \pi_e^o \equiv \underline{c}$. When the policy is *good*, the opposition supports it and if it dissents, it will be repressed, and thus: $\pi^G \geq \pi^G + (1 - \gamma_6)(1-\alpha)(V_e^o - \pi^G) - (1 - c)$. The necessary condition, at $\gamma_6 = 1$, is clearly satisfied. When the policy is *bad*, the opposition dissents and does not get repressed, and so: $V_e^o - (V_e^o - \pi^B)[(1-\beta)g_4 + \beta g_5] \geq \pi^B$. This inequality holds strictly for all admissible g_4, g_5.

Continuing the *good* policy given that the opposition would support it yields an expected payoff of $V^G + b > \pi_e^o$, which is what the leader would obtain by deviating and repealing it. Therefore, this decision is optimal. Repealing the *bad* policy yields an expected payoff of π_e^o, while continuing it, given that the opposition would dissent without getting repressed, yields $-b + \pi_e^o + (V^B - \pi_e^o)[(1-\beta)g_4 + \beta g_5] = \pi_e^o - b < \pi_e^o$, where we use the fact that since in equilibrium some beliefs g_4, g_5 prevent this deviation, then preventing it under $g_4 = g_5 = 0$ will be sufficient. This establishes that these strategies do form a sequential equilibrium.

The next step is to demonstrate uniqueness. The above equilibrium is unique in its class in the sense that all sequential equilibria in this class have the same equilibrium path of

play specified by the strategies.[16] Showing that this equilibrium is the only informative one is more involved and requires the methodical elimination of all other possible strategy profiles.

I first show that there are no other equilibria in which the leader repeals the bad policy and continues the good one. Suppose the leader continues the good policy and repeals the bad one. In all such separating equilibria, Bayes rule pins down $g_1 = 0$.

Case 1: the opposition endorses all policies regardless of quality. These strategies imply $g_2 = g_3 = 1$ by Bayes rule. By continuing the bad policy, the leader would obtain $V^B - b > \pi_e^o$, which is what repealing it yields. Therefore, none of these profiles can be equilibria.

Case 2: the opposition dissents from good policy. Suppose first the leader represses this dissent. These strategies imply $g_6 = 1$ by Bayes rule. The opposition's expected payoff then is $\pi^G - (1 - c) < \pi^G$, which is *the least* it would obtain by supporting the good policy instead. Thus, none of of these profiles can be equilibria.

Suppose now the leader allows this dissent. These strategies imply $g_4 = g_5 = 1$ by Bayes rule. Allowing dissent when the policy is bad yields $V^B > V^B - c$, which is *the most* the leader could expect by repressing it. Therefore, in any such sequential equilibrium, the leader must allow dissent when the policy is bad. Suppose now that the opposition dissents from the bad policy. The leader could strictly improve the payoff by continuing the bad policy and getting $V^B - b > \pi_e^o$, which is what repeal yields. Therefore, in any such equilibrium, the opposition must endorse the bad policy. Suppose now the opposition endorses the bad policy. Because it dissents from the good policy, it follows that $\alpha g_4 + (1 - \alpha) g_5 = 1 \leq \alpha g_2 + (1 - \alpha) g_3$, which implies that $g_2 = g_3 = 1$. Continuing the bad policy then yields $V^B - b > \pi_e^o$, which is what the leader gets from repealing it. Therefore, this cannot be a sequential equilibrium either. Thus, none of the profiles with the leader allowing dissent can be equilibria. This exhausts all profiles with the opposition dissenting only from good policies.

Case 3: the opposition endorses only the good policy. These strategies imply $g_2 = g_3 = 1$ by Bayes rule. Suppose the leader allows dissent when the policy is good. Deviation to dissent is profitable as long as it is not the case that $g_4 = g_5 = 1$. Therefore, in any such equilibrium, it must be the case that $g_4 = g_5 = 1$, which implies that the leader would always allow dissent when the policy is bad because this would yield $V^B > V^B - c$, which is *the most* what repression would yield. Given that dissent from bad policy would be allowed, the leader could improve the expected payoff by continuing the bad policy, which would yield $V^B - b > \pi_e^o$, which is what repeal gets. Thus, there are no equilibria where the leader allows dissent when the policy is good.

Suppose now the leader represses regardless of policy quality. To prove that this cannot be an equilibrium, we show that the optimality of dissent when when the policy is bad is incompatible with the optimality of repressing it. Suppose this profile is an equilibrium. Since it is optimal to dissent from a bad policy despite being repressed, it follows that

---

[16] That is, because of the latitude in specifying the beliefs g_4, g_5, and g_6 within the limits established by the argument, we can support an infinite number of these sequential equilibria. For example, any pair g_4, g_5 that satisfies (11) would work, including beliefs strictly greater than zero. Similarly, we do not need to use $g_6 = \underline{g_6}$, and any $g_6 \geq \underline{g_6}$ would work just as fine in giving $\gamma_6 = 1$. However, these equilibria only differ in these off-the-path beliefs and the equilibrium path of play in all of them is the same, so our substantive predictions would remain the same. It is worth noting that these different beliefs imply shifts in the cost range as well.

$-(1-c) + \pi^B + (1-\gamma_6)\beta(V_e^o - \pi^B) \geq \pi^B$, which implies that there exists an upper bound on permissible beliefs: $\gamma_6 \leq 1 - \frac{1-c}{\beta(V_e^o - \pi^B)} \equiv \overline{\gamma}$. Turning now to the leader's strategy under bad policy, optimality of repression implies: $-c + (1-\beta)V^B + \beta(-v) + \gamma_6\beta(V^B + v) \geq \pi_e^o + (V^B - \pi_e^o)\big[(1-\beta)g_4 + \beta g_5\big]$. In equilibrium, this decision is optimal for some beliefs g_4, g_5, and so it must be optimal under beliefs $g_4 = g_5 = 0$, which simplifies the expression and establishes a lower bound on permissible beliefs: $\gamma_6 \geq 1 - \frac{V^B - \pi_e^o - c}{\beta(V^B + v)} \equiv \underline{\gamma}$. I now show that $\overline{\gamma} < \underline{\gamma}$, which implies that there exist no beliefs that can satisfy both optimality requirements. Note that $\overline{\gamma} < \underline{\gamma}$ reduces to $\frac{1-c}{V_e^o - \pi^B} > \frac{V^B - \pi_e^o - c}{V^B + v}$. From $v < 1$, it follows that $V^B - \pi_e^o = v - b(p-q)(\mu_e^o - \mu^B) < 1$, and therefore $1 - c > V^B - \pi_e^o - c$. But $V^B + v - V_e^o + \pi^B = \pi^B + v - b(p-q)(\mu_e^o - \mu^B) > 0$, where the inequality follows from Assumption 2. Thus, the numerator on the LHS is strictly greater than the numerator on the RHS, and the denominator on the LHS is strictly smaller than the denominator on the RHS, which establishes the inequality. Thus, the profile cannot be an equilibrium.

Therefore, all remaining sequential equilibria must involve policy failure. I now show that no remaining equilibria are fully revealing. Lemma 2 exhausts the possibilities for signaling at the repeal stage. Therefore, in any remaining fully revealing equilibrium, the leader must be continuing regardless of policy quality. Lemma 4 shows that there are no such equilibria with the opposition playing a separating strategy. Therefore, in any remaining fully revealing equilibrium, the opposition must be pooling on dissent. Lemma 5 shows that there are no such equilibria with the leader playing a separating repression strategy. $\qquad\square$

LEMMA 6. *There is no sequential equilibrium, in which the opposition dissents when the policy is good if the leader represses regardless of quality.*

Proof. Suppose in some equilibrium it is optimal for the leader to repress when the policy is bad. This implies that $-c + V^B - \beta(V^B + v)(1-\gamma_6) \geq \pi_e^o + (V^B - \pi_e^o)\big[(1-\beta)g_4 + \beta g_5\big]$. Because the RHS is minimized at $g_4 = g_5 = 0$, it follows that a necessary condition for the existence of this equilibrium is $c \leq V^B - \pi_e^o - \beta(V^B + v)(1-\gamma_6) \equiv \hat{c}$. Suppose that it is optimal for the opposition to dissent despite getting repressed when the policy is good. This implies $\pi^G + (1-\alpha)(V_e^o - \pi^G)(1-\gamma_6) - (1-c) \geq V_e^o - (V_e^o - \pi^G)\big[\alpha g_2 + (1-\alpha)g_3\big]$. Because the RHS is minimized at $g_2 = g_3 = 1$, it follows that a necessary condition for the optimality of this strategy is that $c \geq 1 - (1-\alpha)(V_e^o - \pi^G)(1-\gamma_6) \equiv \hat{\hat{c}}$. Thus, a necessary condition for the existence of this equilibrium is that $c \in [\hat{\hat{c}}, \hat{c}]$. However, this interval does not exist. That is, $\hat{c} < \hat{\hat{c}}$ because $V^B - \pi_e^o - \beta(V^B + v)(1-\gamma_6) < 1 - (1-\alpha)(V_e^o - \pi^G)(1-\gamma_6)$. This inequality follows from $V^B - \pi_e^o = v + \pi^B - \pi_e^o < v < 1$, and $\beta > 1 - \alpha$ with $V^B + v > V_e^o - \pi^G$, which imply that $\beta(V^B + v)(1-\gamma_6) \geq (1-\alpha)(V_e^o - \pi^G)(1-\gamma_6)$. Thus there can be no c that would simultaneously satisfy both necessary conditions for the existence of this equilibrium. $\qquad\square$

Proof of Proposition 2. The argument in the text establishes the optimality of the pooling repressive strategy, while Lemmas 3 and 6 establish the optimality of the opposition's pooling endorsement strategy. By Corollary 1, the leader's strategy must also be pooling at the initial stage. I now establish the conditions necessary to support these two sequential equilibria given that $c \leq \underline{c}$.

In the continuation (Type II) equilibrium, Bayes rule pins down the posterior beliefs $g_2 = \frac{g\alpha}{g\alpha + (1-g)(1-\beta)}$ and $g_3 = \frac{g(1-\alpha)}{g(1-\alpha)+(1-g)\beta}$. Assign the following off-the-path beliefs: $g_1 = g_4 = g_5 = 0$, and $g_6 = 1$. Optimality of continuing the good policy requires $b + \pi_e^o + (V^G - \pi_e^o)[\alpha g_2 + (1-\alpha)g_3] \geq \pi_e^o + g_1(V^G - \pi_e^o)$, which, at $g_1 = 0$, is always satisfied regardless of beliefs g_2, g_3. Optimality of continuing the bad policy requires $-b + \pi_e^o + (V^B - \pi_e^o)[(1-\beta)g_2 + \beta g_3] \geq \pi_e^o + g_1(V^B - \pi_e^o)$, which, at $g_1 = 0$, yields the necessary condition:

$$g > (1-\beta)g_2 + \beta g_3 \geq \frac{b}{V^B - \pi_e^o}. \tag{12}$$

The first inequality follows from $g_3 < g < g_2$, and $\beta > \frac{1}{2}$. Both g_2 and g_3 are strictly increasing in g. This implicitly defines a critical value for g. Let g_B be that value. For all $g \geq g_B$, the condition is satisfied, and the Type II equilibrium exists.

In the repeal (Type I) equilibrium, Bayes rule pins down the posterior belief $g_1 = g$. As before, we assign off the path beliefs as follows, $g_2 = g_3 = g_4 = g_5 = 0$, and $g_6 = 1$. Optimality of repealing the bad policy requires the necessary condition, at $g_2 = g_3 = 0$, of $-b \leq g(V^B - \pi_e^o)$, which is always satisfied. Optimality of repealing the good policy requires the necessary condition:

$$g \geq \frac{b}{V^G - \pi_e^o}, \tag{13}$$

which implicitly defines a critical value for g. Let g_A be that value. For any $g \geq g_A$, the conditions are satisfied and the Type I equilibrium exists. This establishes the first claim in the proposition.

Because $V^G - \pi_e^o > V^B - \pi_e^o$, it follows that the RHS in (12) is strictly larger than the RHS in (13). That is, the Type II equilibrium bound is larger than the Type I equilibrium bound. Whenever g satisfies (12), it will necessarily satisfy (13), although the converse is not true. Thus, $g \geq g_B \Rightarrow g > g_A$. Thus, if $g \geq g_B$, then both Type I and Type II equilibria can be supported. If, however $g \in [g_A, g_B]$, then only Type I equilibrium can exist. (If $g < g_A$, then only partially revealing equilibria exist.) Consider now the Type I equilibrium and suppose the leader could deviate and continue the good policy if that would convince the citizens of its quality, and so, given the opposition's strategy, citizens would update $g_2 = g_3 = 1$. This would give the leader a payoff of $V^G + b > \pi_e^o + g(V^G - \pi_e^o)$, which is what the equilibrium payoff is. If the leader continues the bad policy given these beliefs, the expected payoff is $V^B - b < \pi_e^o + g(V^B - \pi_e^o)$ for sufficiently high g. In other words, when g is sufficiently high, the leader cannot benefit from continuing the bad policy even if re-election is certain. Thus, repealing the good policy is equilibrium-dominated, and the Intuitive Criterion (Cho and Kreps 1987) rules out the Type I equilibrium in this region, leaving only the Type II equilibrium. $\square$

Proof of Proposition 3. By Corollary 1, if the opposition plays a separating strategy, the leader must be pooling on repeal. Two such equilibria exist, but they are unintuitive. The leader can always improve the payoff by continuing the good policy if that would convince the citizens of its quality. Because the opposition plays a separating strategy, the information sets that are reached by continuing the good policy are different from the

ones that are reached by continuing the bad policy, and so the leader cannot benefit from continuing the bad policy even under the new beliefs. This leaves four sequential equilibria, two for each failure type, distinguished by whether the opposition pools on endorsement or dissent. Deriving the conditions is analogous to the proof of Proposition 2. □

References

Aldrich, John H., John L. Sullivan, and Eugene Borgida. 1989. "Foreign Affairs and Issue Voting: Do Presidential Candidates 'Waltz Before a Blind Audience?'." *American Political Science Review* 83 (March): 123–141.

Baum, Matthew A. 2004. "Going Private: Presidential Rhetoric and the Domestic Politics of Audience Costs in U.S. Foreign Policy Crises." *Journal of Conflict Resolution* .

Bennett, Stephen Earl. 1994. "The Persian Gulf War's Impact on Americans' Political Information." *Political Behavior* 16 (June): 179–201.

Brody, Richard A., and Catherine R. Shapiro. 1989. "Policy Failure and Public Support: The Iran-Contra Affair and Public Assessment of President Reagan." *Political Behavior* 11 (December): 353–369.

Busch, Marc L. 2000. "Democracy, Consultation, and the Paneling of Disputes under GATT." *Journal of Conflict Resolution* 44 (August): 425–446.

Cho, In-Koo, and David M. Kreps. 1987. "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics* 102 (May): 179–221.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (November): 1431–1451.

Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.

Dur, Robert A.J. 2001. "Why do policy makers stick to inefficient decisions?" *Public Choice* 107 (June): 221–234.

Eisinga, Rob, Philip Hans Franses, and Dick J.C. van Dijk. 1998. Timing of Vote Decision in First and Second Order Dutch Elections, 1978-1995. In *Political Analysis*, ed. Walter R. Mebane Jr. Ann Arbor: University of Michigan Press pp. 117–142.

Eyerman, Joe, and Robert A. Hart, Jr. 1996. "An Empirical Test of the Audience Cost Proposition: Democracy Speaks Louder than Words." *Journal of Conflict Resolution* 40 (December): 597–616.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap Talk Can Matter in Bargaining." *Journal of Economic Theory* 48 (1): 221–237.

Fearon, James D. 1994*a*. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (September): 577–92.

Fearon, James D. 1994*b*. "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38 (June): 236–69.

Goemans, Hein E. 2000. *War and Punishment: The Causes of War Termination and the First World War*. Princeton: Princeton University Press.

Jervis, Robert. 1970. *The Logic of Images in International Relations*. New York: Columbia University Press.

Kernell, Samuel. 1978. "Explaining Presidential Popularity." *American Political Science Review* 72 (June): 506–522.

Knopf, Jeffrey W. 1998. "How Rational is "The Rational Public"? Evidence from U.S. Public Opinion on Military Spending." *Journal of Conflict Resolution* 42 (October): 544–571.

Martin, Lisa L. 1993. "Credibility, Costs, and Institutions: Cooperation on Economic Sanctions." *World Politics* 45 (April): 406–432.

Miller, Joanne M., and Jon A. Krosnick. 2000. "News Media Impact on the Ingredients of Presidential Evaluations: Politically Knowledgeable Citizens are Guided by a Trusted Source." *American Journal of Political Science* 44 (April): 301–315.

Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: John Wiley.

Papayoanou, Paul A. 1997. "Intra-Alliance Bargaining and U.S. Bosnia Policy." *Journal of Conflict Resolution* 41 (February): 91–116.

Partell, Peter J. 1997. "Executive Constraints and Success in International Crises." *Political Research Quarterly* 50 (September): 503–528.

Ramsay, Kristopher W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48 (August): 459–486.

Regan, Patrick M. 1998. "Choosing to Intervene: Outside Interventions in Internal Conflicts." *The Journal of Politics* 60 (August): 754–779.

Rogoff, Kenneth. 1990. "Equilibrium Political Budget Cycles." *The American Economic Review* 80 (March): 21–36.

Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (February): 121–149.

Schelling, Thomas C. 1966. *Arms and Influence*. New Haven: Yale University Press.

Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92 (December): 829–44.

Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53 (Spring): 233–266.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.

Shapiro, Robert Y., and Benjamin I. Page. 1988. "Foreign Policy and the Rational Public." *Journal of Conflict Resolution* 32 (June): 211–247.

Smith, Alastair. 1996. "Diversionary Foreign Policy in Democratic Systems." *International Studies Quarterly* 40 (March): 133–153.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.