Information Manipulation, Coordination, and Regime Change*

Chris Edmond[†]

This draft: July 2012

Abstract

This paper presents a model of information and political regime change. If enough citizens act against a regime, it is overthrown. Citizens are imperfectly informed about how hard this will be and the regime can, at a cost, engage in propaganda so that at face-value it seems hard. This coordination game with endogenous information manipulation has a unique equilibrium and the paper gives a complete analytic characterization of its comparative statics. For a fixed number of signals, if the signal precision is sufficiently high, then the regime is harder to overthrow. But if there are increasingly many high precision sources of information, the regime becomes easier to overthrow unless there are strong economies of scale in controlling multiple sources of information.

Keywords: global games, hidden actions, signal-jamming, propaganda, bias, media.

JEL classifications: C7, D7, D8.

*This is an extensively revised version of a chapter from my UCLA dissertation. Previous versions circulated under the title "Information and the limits to autocracy". Andrew Atkeson and Christian Hellwig provided much encouragement and many helpful discussions. I particularly thank Marco Ottaviani and several anonymous referees for their comments which greatly improved the paper. I also thank Marios Angeletos, Costas Azariadis, Heski Bar-Isaac, Adam Brandenburger, Ethan Bueno de Mesquita, Ignacio Esponda, Catherine de Fontenay, Matias Iaryczower, Simon Loertscher, Stephen Morris, Alessandro Pavan, Andrea Prat, Hyun Song Shin, Adam Szeidl, Laura Veldkamp, Iván Werning, seminar participants at the ANU, Boston University, University of Chicago (GSB and Harris), Duke University, FRB Minneapolis, La Trobe University, LSE, University of Melbourne, MIT, NYU, Northwestern University, Oxford University, University of Rochester, UC Irvine, University College London and UCLA and participants at the NBER political economy meetings for their comments. Vivianne Vilar provided excellent research assistance.

[†]University of Melbourne. Email: ⟨cedmond@unimelb.edu.au⟩.

1 Introduction

Will improvements in information technologies help in overthrowing autocratic regimes? Optimists on this issue stress the role of new technologies in facilitating coordination and in improving information about a regime's intentions and vulnerabilities. The "Arab Spring" of uprisings against regimes in Tunisia, Egypt, Libya and elsewhere that began in December 2010 has led to widespread discussion of the role of modern social media such as Facebook, Twitter, Skype and YouTube in facilitating regime change. Similar discussion followed the use of such technologies during the demonstrations against the Iranian regime in June 2009.[1]

But optimism about the use of new technologies in putting autocratic regimes under sustained pressure is hardly new; social media is only the latest technology to be viewed as a catalyst for regime change. Simple internet access, cell phones, satellite television, radio and newspaper have all been viewed as potential catalysts too. And while information optimism has a long and somewhat mixed history, it is also worth bearing in mind that the relationship between information technologies and autocratic regimes has a prominent dark side. Perhaps the most well known examples are the use of mass media propaganda by Nazi Germany and the Soviet Union. Moreover it has become increasingly clear that recent breakthroughs in information technology also provide opportunities for autocracies. During the Iranian demonstrations, technologies like Twitter allowed the regime to spread rumors and disinformation (Esfandiari, 2010). Similarly, the Chinese regime's efforts to counter online organization make use of the exact same technologies that optimists hope will help in bringing regime change (Kalathil and Boas, 2003; Fallows, 2008; Morozov, 2011).

So, should we be optimistic that recent breakthroughs in information technology will lead to the collapse of present-day autocratic regimes? To help address this question, I develop a simple model of information and regime change. While stylized, this model provides a number of insights into ways in which a regime's chances of survival are affected by changes in information technology. The model predicts that (i) an increase in the *precision* of individuals' information can increase the regime's chances of survival, but (ii) an increase in the *number* of sources of information can reduce the regime's chances unless the regime benefits from strong *economies of scale* in its ability to exert control over those sources of information.

The model clearly identifies situations where pessimism about the ability of new information technologies to threaten autocratic regimes is born out. In particular, if the number of sources of information is held fixed while the precision of information increases, then the regime is likely to be harder to overthrow. More generally, if we consider an *information revolution* that consists of both a dramatic increase in the number *and* precision of information sources, then the model predicts that whether we should be more pessimistic or more optimistic about regime change largely depends on the extent of economies of scale, if any, in the

[1]See, for instance, Kirkpatrick (2011) for an account of the use of social media during the Egyptian protests and Musgrove (2009) for the role of Twitter in the Iranian demonstrations.

regime's ability to control these sources of information. For example, suppose an information revolution consists of new technologies such as newsprint, radio, and cinema and that the regime is able to exert control over the new sources of information by deploying a large fixed propaganda apparatus that can then readily influence additional newspapers, radio stations or cinemas at low marginal cost. Then the regime may have sufficient economies of scale in its control over these new sources of information that its chances of survival will increase. By contrast, if the regime has no economies of scale in its control, then the model predicts the regime will become easier to overthrow as the number of sources of information increases.

Section 2 outlines the model. There is a single regime and a large number of citizens with heterogeneous information. Citizens can participate in an attack on the regime or not. If enough do, the regime is overthrown. The regime's ability to withstand an attack is given by a single parameter, the regime's type. Citizens are imperfectly informed about the regime's type and may coordinate either on overturning the regime or not. The regime is informed about its type and seeks to induce coordination on the status quo. It does this by taking a costly *hidden action* which influences the distribution of signals so that citizens receive *biased* information that, taken at face-value, suggests the regime is hard to overthrow. Citizens are rational and internalize the regime's incentives when forming their beliefs.

Section 3 gives the first main result of the paper: this coordination game with endogenous information manipulation has a unique perfect Bayesian equilibrium. Section 4 then turns to the effects of changes in the information precision, holding the number of sources of information fixed. The second main result of this paper is that the regime's information manipulation is *effective*, in the sense of increasing the regime's ex ante survival probability, when the intrinsic signal precision is sufficiently high. If the signal precision is sufficiently high, then *the regime survives in all states where it is possible for the regime to survive.*

The intuition for this result depends on two effects. First, when the signal precision is high, so that many citizens have signals near the mean, it takes only a small amount of bias to deliver a large fall in the size of the attack on the regime. In the limit where information is arbitrarily precise, even an infinitesimal amount of bias serves to drastically reduce the size of the attack on the regime. This gives the regime a *potentially* powerful tool for shaping the equilibrium outcome in its favor. Second, the regime is actually able to *use this tool in equilibrium.* Citizens face a difficult joint coordination and information filtering problem that inhibits their ability to infer the amount of bias in their signals. Citizens face a difficult filtering problem because the regime's policy is state-contingent, and, since citizens are imperfectly informed about the state, they are also imperfectly informed about the amount of manipulation that occurs, even in equilibrium. Moreover, citizens face a difficult coordination problem because they are playing a game of strategic complementarities where it is common knowledge that *any* equilibrium bias only serves to reduce the size of the aggregate attack and so reduces every individual's incentives to attack. By contrast, if either (i) the regime's policy was not state-contingent or (ii) the citizens were perfectly coordinated, then there would be no difficulty in inferring and discarding any bias.

Section 5 gives the third main result of the paper, that the regime's chances of surviving depend crucially on the extent of economies of scale in control over information outlets. Section 6 considers two extensions. Section 7 concludes. All proofs are in the appendix. A separate supplementary appendix contains further extensions and discussion.

Coordination games and imperfect information. Coordination games with *multiple equilibria* have been applied to a vast array of topics, including bank runs (Diamond and Dybvig, 1983), currency attacks (Obstfeld, 1986, 1996), debt crises (Calvo, 1988; Cole and Kehoe, 2000) and technology adoption (Katz and Shapiro, 1986), amongst many others. But as emphasized by Carlsson and van Damme (1993), when individuals have imperfect information about the structure of the coordination game, it is often possible to select a *unique* equilibrium. This is an attractive possibility, not least because it makes for sharper conclusions about how equilibrium outcomes vary with parameters of interest. By now this *global games* approach has likewise been applied to bank runs (Rochet and Vives, 2004; Goldstein and Pauzner, 2005), currency attacks (Morris and Shin, 1998), investment complementarities (Chamley, 1999; Dasgupta, 2007) and others.[2] In addition to this paper, applications of global games ideas to political economy include Boix and Svolik (2009), Bueno de Mesquita (2010), Chassang and Padro-i-Miquel (2010) and Shadmehr and Bernhardt (2011).

More generally, this paper concerns a policy maker who seeks to induce individual information receivers to coordinate on the policy maker's preferred outcome. Similar issues are likely to be at work in many scenarios. For example, the managers of a bank may try to influence information so as to prevent a run on deposits. I present this analysis in the context of a game of political regime change because that seems, to me, a particularly "pure" setting in which to first understand how information manipulation to induce a particular coordination outcome could work. By contrast, more market-oriented scenarios, like bank runs or currency attacks, are also likely to involve the further complication of market prices that endogenously aggregate information.[3] I view the results in this paper as (i) being directly applicable to the political regime change scenario and (ii) as suggestive of a key element that would be operative in a more complete, but also more complex, account of information manipulation and coordination in market-oriented coordination games.

The equilibrium uniqueness results in this paper stand in contrast to Angeletos, Hellwig and Pavan (2006), who were the first to emphasize the importance of endogenous information in global games. They consider a similar game where the policy maker's actions provide an endogenous signal to individual information receivers. This signaling reintroduces the possibility of multiple equilibria. The key reason for the uniqueness in my model relative

[2] Morris and Shin (2000, 2003) and Vives (2005) provide accessible introductions to these tools.

[3] See Angeletos and Werning (2006) and Hellwig, Mukherji and Tsyvinski (2006) fo global games with endogenous information aggregation by market prices. Other examples of global games with endogenous information include Angeletos, Hellwig and Pavan (2007), Dasgupta (2007) and Heidhues and Melissas (2006).

to theirs is that in my setting individuals' endogenous signals are a *monotone* function of the state whereas in Angeletos, Hellwig and Pavan (2006) the endogenous signals are a non-monotone function of the state that gives individual information receivers more common belief as to whether the status quo is likely to prevail or not. The greater amount of common belief in their setting makes it easier to sustain multiple equilibria.

In short, this paper provides a framework with a form of costly *noisy signaling* that, by contrast with the indeterminacies of many signaling models, features instead determinate outcomes of the kind familiar from standard global games.

2 Model of information manipulation and regime change

There is a unit mass of ex ante identical citizens, indexed by $i \in [0, 1]$. The citizens face a regime that seeks to preserve the status quo. Each citizen decides whether to participate in an attack on the regime, $s_i = 1$, or not, $s_i = 0$. The aggregate attack is $S := \int_0^1 s_i di$. The type of a regime θ is its private information. The regime is overthrown if and only if $\theta < S$.

Citizen payoffs. The payoff to a citizen is given by

$$u(s_i, S, \theta) = (\mathbb{1}\{\theta < S\} - p)s_i \tag{1}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function and where $p > 0$ is the opportunity cost of participating in an attack. A citizen will participate, $s_i = 1$, whenever they expect regime change to occur with more than probability p. To focus on scenarios where individual decisions are not trivial, I also assume that $p < 1$. Otherwise, if $p \geq 1$, it is optimal for an individual to not participate independent of θ and S. With $p < 1$ the individual s_i and aggregate S are *strategic complements*: the more citizens participate in an attack, the more likely it is that the regime is overthrown and so the more likely it is that any individual also participates.

Regime's benefit from surviving. If $\theta < S$ the regime is overthrown and obtains an outside option normalized to zero. Otherwise, if $\theta \geq S$, the regime obtains a benefit $B(\theta, S) = \theta - S$ from remaining in power. The key assumption here is that the benefit is *separable* in θ and S. Given separability, the *linearity* in θ and S is without further loss of generality; it simply represents a normalization of the type. The benefit is strictly decreasing in S, i.e., the regime wants to minimize the costs of unrest and so wants S small even when it survives.

Hidden actions and regime payoffs. Given θ, a regime may take a *hidden* action $a \geq 0$ in an attempt to convey that its type is $\theta + a$. These actions incur cost $C(a)$ where $C(0) = 0$, $C'(a) > 0$ for $a > 0$ and $C''(a) \geq 0$ with $C''(0) = 0$. The net payoff to a regime is therefore

$$\theta - S - C(a) \tag{2}$$

whenever $\theta \geq S$, and zero otherwise. Similar separable payoffs are used throughout the global games literature (e.g., Angeletos, Hellwig and Pavan (2006), amongst many others).

Citizen information. Citizens begin with common priors for θ, specifically the improper uniform on $\mathbb{R}$.[4] Following the regime's hidden action a, each citizen draws an idiosyncratic signal $x_i := y + \varepsilon_i$ centered on the regime's *apparent strength* $y = \theta + a$. The noise ε_i is IID normal[5] across citizens, independent of θ, with mean zero and precision $\alpha > 0$. Given the regime's apparent strength y, the density of signals is

$$f(x_i \,|\, y) := \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - y)), \qquad y = \theta + a \tag{3}$$

where $\phi(\cdot)$ denotes the standard normal PDF.

2.1 Equilibrium

A symmetric *perfect Bayesian equilibrium* of this model consists of individual beliefs $\pi(\theta \,|\, x_i)$ and participation decisions $s(x_i)$, an aggregate attack $S(y)$, and regime actions $a(\theta)$ such that: (i) a citizen with information x_i rationally takes into account the manipulation $y(\theta) = \theta + a(\theta)$ when forming their posterior beliefs, (ii) given these beliefs, $s(x_i)$ maximizes individual expected utility, (iii) the aggregate attack is consistent with the individual decisions, and (iv) the actions $a(\theta)$ maximize the regime's payoff given the aggregate attack. In equilibrium, the regime is overthrown if $\theta < S(y(\theta))$ and otherwise survives.

2.2 Exogenous information benchmarks

Two important special cases of the model are when: (i) the regime's θ is *common knowledge*, or (ii) there are no hidden actions and so the analysis reduces to a *standard global game*.

Common knowledge. If θ is common knowledge, costly hidden actions are pointless and $a(\theta) = 0$ for all θ. The model reduces to a standard coordination game. If $\theta < 0$, any attack $S \geq 0$ can overthrow the regime. It is optimal for any individual to participate, all do so, and the regime is overthrown. If $\theta \geq 1$, no attack can overthrow the regime. It is optimal for any individual not to participate, none do, and the regime survives. If $\theta \in [0, 1)$, the regime is "fragile" and multiple self-fulfilling equilibria can be sustained. For example, if each individual believes that everyone else will attack, it will be optimal for each citizen to do so and $S = 1 > \theta$ leads to regime change and the vindication of the initial expectations.

[4]The supplementary appendix discusses the case of informative priors/public information at length.

[5]The normal distribution is the natural assumption if we think of a central limit theorem applied to many bits of information (as in Section 5 below). Moreover, the normal distribution allows for a clean distinction between changes in the mean and changes in the precision, which simplifies the discussion. The main results require only that the noise has continuous, unbounded support and an increasing hazard function.

Standard global game. If there are no hidden actions, $a(\theta) = 0$ for all θ, then each citizen has signal $x_i = \theta + \varepsilon_i$ and the analysis reduces to a standard global game. As in Carlsson and van Damme (1993), Morris and Shin (1998) and much subsequent literature, this introduces the possibility of pinning down a unique equilibrium. In this equilibrium, strategies are threshold rules: there is a unique θ^* such that the regime is overthrown for $\theta < \theta^*$ and a unique signal x^* such that a citizen participates for $x_i < x^*$. These thresholds are given by:

Morris-Shin Benchmark. The equilibrium thresholds $x^*_{\text{MS}}, \theta^*_{\text{MS}}$ simultaneously solve

$$\text{Prob}[\theta < \theta^*_{\text{MS}} \,|\, x^*_{\text{MS}}] = \Phi(\sqrt{\alpha}(\theta^*_{\text{MS}} - x^*_{\text{MS}})) = p \tag{4}$$

$$\text{Prob}[x_i < x^*_{\text{MS}} \,|\, \theta^*_{\text{MS}}] = \Phi(\sqrt{\alpha}(x^*_{\text{MS}} - \theta^*_{\text{MS}})) = \theta^*_{\text{MS}} \tag{5}$$

where $\Phi(\cdot)$ denotes the standard normal CDF and p is the individual opportunity cost of attacking. In particular, $\theta^*_{\text{MS}} = 1 - p$ independent of α and $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha}$.

The first condition says that if the regime's threshold is θ^*_{MS}, the marginal citizen with signal $x_i = x^*_{\text{MS}}$ expects regime change with exactly probability p. The second says that if the signal threshold is x^*_{MS}, a regime with $\theta = \theta^*_{\text{MS}}$ will be indifferent to abandoning its position. I say that a regime's hidden action technology is *effective* if in equilibrium $\theta^* < \theta^*_{\text{MS}} = 1 - p$.

As the signal precision α becomes high, some regimes are faced with a powerful incentive to shift the signal mean in their favor. To see this, observe that the equilibrium aggregate attack consists of all citizens with $x_i < x^*_{\text{MS}}$, namely

$$S^*_{\text{MS}}(\theta) := \text{Prob}[x_i < x^*_{\text{MS}} \,|\, \theta] = \Phi(\sqrt{\alpha}(\theta^*_{\text{MS}} - \theta) - \Phi^{-1}(p)) \tag{6}$$

so that as the precision $\alpha \to \infty$, the attack $S^*_{\text{MS}}(\theta) \to \mathbb{1}\{\theta^*_{\text{MS}} > \theta\}$. If the regime's type is $\theta < \theta^*_{\text{MS}}$, it faces a unit mass of citizens and is overthrown. If the regime has $\theta > \theta^*_{\text{MS}}$ it faces zero unrest and survives. A *small* amount of bias would enable a regime with θ just below θ^*_{MS} to achieve a *large* reduction in the attack and to switch from being overthrown to surviving. As information becomes precise, there is a large incentive to engage in manipulation.

3 Unique equilibrium with information manipulation

The first main result of this paper is that there is a unique equilibrium:

Proposition 1. There is a unique perfect Bayesian equilibrium. The equilibrium is *monotone* in the sense that there exist thresholds x^* and θ^* such that $s(x_i) = 1$ for $x_i < x^*$ and zero otherwise, while the regime is overthrown for $\theta < \theta^*$ and not otherwise.

A detailed proof is given in Appendix A. Briefly, the proof involves first showing (i) that there is a unique equilibrium in monotone strategies, and (ii) that the unique monotone equilibrium is the only equilibrium which survives the iterative elimination of interim strictly dominated strategies. Here in the main text I briefly characterize the equilibrium.

3.1 Equilibrium characterization

Let $\hat{x}$ denote a candidate for the citizens' threshold and let $\Theta(\hat{x})$ and $a(\theta, \hat{x})$ denote candidates for the regime's threshold and hidden actions given $\hat{x}$.

Regime problem. Since citizens participate $s(x_i) = 1$ for $x_i < \hat{x}$, for any given $\hat{x}$ the aggregate attack facing a regime of apparent strength y is simply $\text{Prob}[x_i < \hat{x} \,|\, y]$, that is

$$S(y) = \int_{-\infty}^{\infty} s(x_i) f(x_i \,|\, y)\, dx_i \tag{7}$$

$$= \int_{-\infty}^{\hat{x}} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - y))\, dx_i = \Phi(\sqrt{\alpha}(\hat{x} - y)), \qquad y = \theta + a \tag{8}$$

(using the expression for the signal density given in (3) above). Since the regime is overthrown for $\theta < \Theta(\hat{x})$, hidden actions are $a(\theta, \hat{x}) = 0$ for all $\theta < \Theta(\hat{x})$, otherwise the regime would be incurring a cost but receiving no benefit. For all $\theta \geq \Theta(\hat{x})$, the regime chooses

$$a(\theta, \hat{x}) \in \underset{a \geq 0}{\text{argmin}} \left[\Phi(\sqrt{\alpha}(\hat{x} - \theta - a)) + C(a)\right] \tag{9}$$

A key step in proving equilibrium uniqueness is to recognize that hidden actions are given by $a(\theta, \hat{x}) = A(\theta - \hat{x})$, where the auxiliary function $A : \mathbb{R} \to \mathbb{R}_+$ is *exogenous* and in particular does *not* depend on the citizen threshold $\hat{x}$. Using (8), this function is defined by

$$A(t) := \underset{a \geq 0}{\text{argmin}} \left[\Phi(\sqrt{\alpha}(-t - a)) + C(a)\right] \tag{10}$$

The regime threshold $\Theta(\hat{x})$ is then found from the indifference condition

$$\Theta(\hat{x}) = \Phi(\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x}))) + C(A(\Theta(\hat{x}) - \hat{x})) \tag{11}$$

This condition requires that total costs equal total benefits at the extensive margin. For any given candidate citizen threshold $\hat{x}$, equations (10)-(11) determine the regime threshold $\Theta(\hat{x})$ and hidden actions $a(\theta, \hat{x}) = A(\theta - \hat{x})$ solving the regime's problem.

Citizen problem. An individual participates only if they believe the regime will be overthrown with probability at least p. Given $\hat{x}$ and the solution to the regime's problem, the posterior probability of regime change for an individual with arbitrary signal x_i is

$$\text{Prob}[\theta < \Theta(\hat{x}) \mid x_i,\, a(\cdot, \hat{x})] := \frac{\int_{-\infty}^{\Theta(\hat{x})} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta))\, d\theta}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta' - a(\theta', \hat{x}))\, d\theta'}$$

(using $a(\theta, \hat{x}) = 0$ for all $\theta < \Theta(\hat{x})$ in the numerator). Writing the hidden actions in terms of the auxiliary function $a(\theta, \hat{x}) = A(\theta - \hat{x})$, evaluating at $x_i = \hat{x}$, and then equating the result to the opportunity cost p gives the indifference condition characterizing the citizen threshold

$$\frac{\int_{-\infty}^{\Theta(\hat{x})} \sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \theta))\, d\theta}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \theta' - A(\theta' - \hat{x})))\, d\theta'} = p \tag{12}$$

Monotone equilibrium. As shown in Appendix A, there is a unique monotone equilibrium with thresholds x^* and θ^* simultaneously solving conditions (11) and (12). The regime's equilibrium hidden actions are then given by $a(\theta) = A(\theta - x^*)$ using the auxiliary function from (10). A key step in the proof is showing that the posterior probability on the left hand side of equation (12) depends only on the difference $\Theta(\hat{x}) - \hat{x}$ and is monotone increasing in this argument so that (12) can be solved for a unique difference $\theta^* - x^*$. Similarly, the right hand side of (11) only depends on the difference $\Theta(\hat{x}) - \hat{x}$ so we can take the unique solution $\theta^* - x^*$ from (12) and plug it into the right hand side of (11) to determine θ^* separately.

The appendix goes on to show that this monotone equilibrium is the *only* equilibrium.

3.2 Hidden actions

To give an intuitive characterization of the hidden actions taken by a regime, it is helpful to recast the problem in slightly more general terms. To this end, let $S(y)$ denote the aggregate attack facing a regime that chooses apparent strength $y \geq \theta$ and let $B(\theta, S)$ denote the benefit the regime obtains from surviving. Thus the regime's net payoff can be written

$$y(\theta) \in \underset{y \geq \theta}{\operatorname{argmax}} \left[B(\theta, S(y)) - C(y - \theta) \right] \tag{13}$$

with $B(\theta, S) = 0$ whenever the regime is overthrown. We then have:

PROPOSITION 2. If $B(\theta, S)$ satisfies the Spence-Mirrlees *sorting condition* $B_{\theta S} \leq 0$, then for any aggregate attack $S(y)$ the solution $y(\theta)$ to (13) is increasing in θ.

The proof is given in Appendix A. The key steps are to observe (i) that for given θ, at any interior solution higher values of y cost more and so will be chosen only if they reduce $S(y)$, and (ii) since $B(\theta, S)$ satisfies a sorting condition in θ and S, stronger regimes are at least weakly better off from a smaller attack than are weaker regimes,[6] hence stronger regimes are more willing, at the margin, to pay the cost of choosing a larger y.

Thus we can establish the monotonicity of $y(\theta)$ for a general benefit function $B(\theta, S)$ satisfying the sorting condition without restricting attention to the separable case used above.

3.3 Discussion

Can we transform this to a standard global game? This monotonicity result suggests an alternative approach to establishing uniqueness. Namely, given monotonicity of $y(\theta)$, perhaps we can transform this game by taking the relevant notion of the "state" to be the outcome y (rather than θ) so that the regime is overthrown if its y is too low, with standard global games arguments then being invoked to obtain uniqueness. Unfortunately, there is

[6] I thank an anonymous referee for suggesting this line of argument. Observe also that since the regime's payoff $B(\theta, S) - C(a)$ is separable in a, the condition $B_{\theta S} \leq 0$ implies that the regime's indifference curves in (S, a) satisfy a standard *single-crossing* property.

a key difficulty with this approach. To ensure we are playing the same game, the citizens' prior over y in the transformed game has to be consistent with their beliefs about $y(\theta)$ in the original game. We cannot simply take the prior over y to be the improper uniform, for instance, since that would be completely inconsistent with the regime's manipulation. In fact, to be consistent with the regime's manipulation, the prior over y will have (at least one) *gap*. Because $y(\theta)$ typically jumps from $y(\theta) = \theta$ for $\theta < \theta^*$ to some $y^* = y(\theta^*) > \theta^*$ at the critical threshold, the prior will have to give exactly zero probability density to the interval (θ^*, y^*). Standard global games arguments crucially assume, however, that the prior has strictly positive density over its support (e.g., Carlsson and van Damme (1993), Assumption A2) and that property would not hold in the transformed game. Worse yet, if, to take what will be a leading example, the function $y(\theta)$ is *constant* on an interval of θ, then this will put an *atom* in the prior over y. Standard global games arguments generally require that the prior be sufficiently *diffuse*,[7] and that property is harder to obtain if the prior has an atom.

Proposition 1 shows that this game *does* have a unique equilibrium and that this equilibrium has many of the familiar global games properties. Rather than attempting to transform this game to a purportedly equivalent standard global game, the proof uses the usual global games *methodology* of establishing dominance regions and iteratively eliminating (interim) strictly dominated strategies. In employing this methodology, I make extensive use of the assumption that the benefit $B(\theta, S)$ is separable in θ and S, a standard payoff assumption in the global games literature. In my model separability also allows the actions to be written $a(\theta, \hat{x}) = A(\theta - \hat{x})$ in terms of the exogenous function defined in (10). This greatly simplifies the equilibrium fixed point problem.

Comparison with Angeletos, Hellwig and Pavan. The uniqueness result in Proposition 1 contrasts with the equilibrium multiplicity results in the closely related model of Angeletos, Hellwig and Pavan (2006). In their benchmark model, individuals get one noisy observation of θ plus one observation of a signal a chosen at cost $C(a)$ by the regime which may also be informative for θ. In this *signaling* game, there is typically an uninformative pooling equilibrium and many separating equilibria. For example, if each individual expects no manipulation, individual strategies and hence the aggregate mass S will be independent of a. The regime then has no incentive to manipulate and so validates the original expectation.

The more interesting kind of multiplicity, however, are the multiple separating equilibria where the regime is *active*. In the version of their model closest to this paper, Angeletos, Hellwig and Pavan let individuals receive two noisy signals, (i) a pure signal of the regime's type θ, and (ii) a separate signal on the endogenous action $a(\theta)$. In my model, by contrast, individuals get one noisy observation of the sum $y(\theta) = \theta + a(\theta)$, not separate signals for the two constituent parts. This makes an important difference.

In particular, in their setting individuals have two qualitatively different kinds of infor-

[7] See Hellwig (2002) and Morris and Shin (2000, 2003) for further discussion.

mation; the pure signal on θ is a monotone function of the regime's type, but, crucially, the separate signal on the endogenous action $a(\theta)$ is a non-monotone function of the regime's type. In their setting the regime chooses $a(\theta) = a^*$ for some critical interval $\theta \in [\theta^*, \theta^{**})$ and $a(\theta) = \underline{a} < a^*$ otherwise. To see the importance of this, consider the benchmark version of their model where individuals observe this endogenous action without noise. In this case, the action creates *common certainty* that the regime's type is either in the critical interval $[\theta^*, \theta^{**})$ or not. This common certainty then serves as the basis of multiple possible coordinated responses. In the variant where $a(\theta)$ is observed with unbounded idiosyncratic noise there is not common certainty but, so long as the noise on $a(\theta)$ is not too imprecise, there is enough *common p-belief* (Monderer and Samet, 1989) that individuals can likewise coordinate on multiple responses to their observations of the endogenous action.

In my setting, by contrast, while the *actions* $a(\theta)$ may be non-monotone in θ, what matters for individual *signals* is the regime's apparent strength $y(\theta) = \theta + a(\theta)$ which, from Proposition 2, *is* monotone in θ. Put differently, in their setting the actions $a(\theta)$ simply *are* the signal mean, whereas in my setting the actions are only part of the signal mean. The monotonicity of the signal is the key force driving equilibrium uniqueness in my setting; with monotonicity, there is not enough common p-belief to coordinate on multiple responses.

4 Information manipulation and signal precision

The most interesting implication of this model is that the regime's information manipulation — or *signal-jamming* — is more effective if the signal precision α, is sufficiently high. Section 4.1 provides a detailed analysis of the equilibrium in the case of a linear cost function and Section 4.2 then gives a characterization of the conditions under which signal-jamming is effective for this special case. Section 4.3 provides further intuition. Section 4.4 shows that the effects of changes in the signal precision are inversely proportional to the effects of changes in the costs of manipulation. Section 4.5 gives results for general cost functions.

Terminology. I measure the effectiveness of signal-jamming by its ability to reduce the regime's threshold θ^* below the Morris-Shin level $\theta^*_{\text{MS}} = 1 - p$. A lower θ^* increases the regime's ex ante survival probability by making it more likely that nature draws a $\theta \geq \theta^*$. In principle, it might be the case that lower θ^* is achieved through large, costly, actions that give the regime a lower net payoff than they would achieve in the Morris-Shin world. But it turns out that as $\alpha \to \infty$ and θ^* falls, hidden actions also become small so that the fall in θ^* represents a genuine increase in payoffs, at least in the limit.

4.1 Equilibrium with linear costs

Let the cost function be $C(a) := ca$ for some constant marginal cost $c > 0$.

Overview of the regime's problem. Now let $S(y)$ denote the size of the aggregate attack if the regime chooses apparent strength $y = \theta + a$. With linear costs, if the choice of y is to induce survival it must minimize $S(y) + cy$ subject to the constraint $y \geq \theta$. Since the first order condition for interior solutions is $-S'(y) = c$, all regimes with interior solutions choose *the same* apparent strength y^*. Corner solutions can arise for two reasons: (i) for all θ such that $\theta < S(y^*) + c(y^* - \theta)$ the regime is too weak to sustain the cost of taking y^* and hence is overthrown, and (ii) for all $\theta \geq y^*$ the innate strength of the regime is sufficiently high that they do not need to undertake any costly manipulation (i.e., the non-negativity constraint $y \geq \theta$ is binding). All regimes at a corner simply have apparent strength equal to actual strength, $y(\theta) = \theta$. All regimes that are at an interior solution have $y(\theta) = y^*$. In short,

$$y(\theta) = \begin{cases} \theta & \text{if } \theta \notin [\theta^*, \theta^{**}) \\ y^* & \text{if } \theta \in [\theta^*, \theta^{**}) \end{cases} \tag{14}$$

The critical point θ^* is just the threshold above which the regime survives and is implicitly determined by the indifference condition $\theta^* = S(y^*) + c(y^* - \theta^*)$. The higher critical point θ^{**} is simply equal to the apparent strength y^* since it is the smallest regime type for which the non-negativity constraint $y^* \geq \theta$ is binding. Figure 1 illustrates.

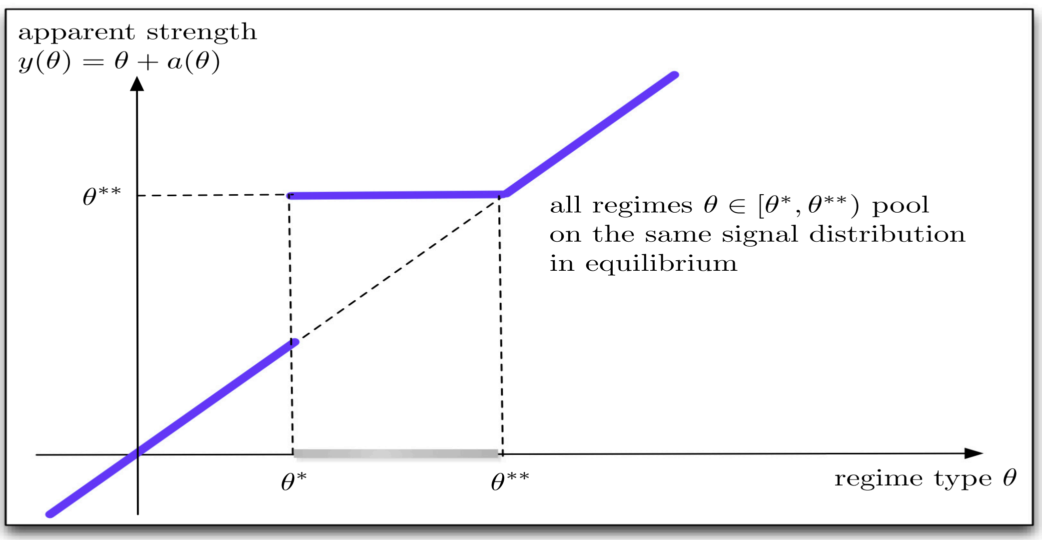


Figure 1: Signal-jamming with linear costs.

Apparent strength $y(\theta) = \theta + a(\theta)$ when the regime has linear costs of manipulation. All regimes with $\theta < \theta^*$ are overthrown. All regimes with $\theta \in [\theta^*, \theta^{**})$ choose the same apparent strength $y^* = \theta^{**}$ and thus generate the same signal distribution in equilibrium. They mimic a stronger regime θ^{**} and generate signals for the citizens that are (locally) uninformative about θ.

To summarise, regimes that are sufficiently weak and that will be overthrown, $\theta < \theta^*$, do not engage in any manipulation. Regimes that are sufficiently strong, $\theta > \theta^{**}$, have enough innate ability to resist an attack that they do not engage in any costly manipulation to shore up their position. Only intermediate "middling" regimes $\theta \in [\theta^*, \theta^{**})$ find it worthwhile to pay the costs of manipulation in equilibrium and all these middling regimes mimic the strength of a strong regime, θ^{**}.

No manipulation, Morris-Shin outcome. Intuitively, when the cost of manipulation is sufficiently high, no regime will find it worthwhile to manipulate and the unique equilibrium will coincide with the Morris-Shin outcome. Since the aggregate attack is $S(y) = \Phi(\sqrt{\alpha}(x^* - y))$, the first order necessary condition $-S'(y) = c$ can be written

$$\sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - y^*)) = c \tag{15}$$

Observe that $c/\sqrt{\alpha}$ must be sufficiently small for this condition to have a solution.[8] In particular, a necessary condition for the existence of an interior solution is that $c/\sqrt{\alpha} \leq \phi(0)$ where $\phi(0) = 1/\sqrt{2\pi} \approx 0.399$ is the maximum value of the standard normal density. If

$$\frac{c}{\sqrt{\alpha}} \geq \phi(0) \tag{16}$$

then the signal precision α is too low relative to the cost of manipulation and *all regimes* are at a corner with $y(\theta) = \theta$. In this case, $\theta^* = \theta^*_{\text{MS}} = 1 - p$ and $\theta^{**} = \theta^*$. By contrast, if the signal precision is high, *some* regimes will engage in information manipulation.

Active manipulation. In particular, from (16), let $\underline{\alpha}(c) := (c/\phi(0))^2$ denote the smallest signal precision such that (15) has a solution. Then all regimes that manipulate choose apparent strength

$$y^* = x^* + \gamma \tag{17}$$

where the coefficient γ solves $\sqrt{\alpha}\phi(\sqrt{\alpha}\gamma) = c$, that is

$$\gamma = \sqrt{\frac{1}{\alpha}\log\left(\frac{\alpha}{\underline{\alpha}(c)}\right)} > 0, \qquad \alpha > \underline{\alpha}(c) \tag{18}$$

The signal-jamming is *acute*. All regimes that manipulate information pool on the same distribution of signals, i.e., all regimes $\theta \in [\theta^*, \theta^{**})$ generate a mean of $y^* = \theta^{**}$ with signals $x_i = y^* + \varepsilon_i$ that are *locally completely uninformative* about θ. As a consequence, the equilibrium precision of a citizen's information is generally less than its intrinsic "fundamental" precision α.

Size of the attack. Now let $S(y(\theta))$ denote the size of the aggregate attack facing a regime of type θ when the signal-jamming has this form. Using (14) and (17), the attack is

$$S(y(\theta)) = \begin{cases} \Phi(\sqrt{\alpha}(x^* - \theta)) & \text{if } \theta \notin [\theta^*, \theta^{**}) \\ \Phi(-\sqrt{\alpha}\gamma) & \text{if } \theta \in [\theta^*, \theta^{**}) \end{cases} \tag{19}$$

[8] This first order condition may have zero or two solutions. If there are two solutions, it is straightforward to show that only the larger of these satisfies the second order condition.

Manipulation to convey a higher apparent strength causes the size of the attack to drop discontinuously at the regime threshold θ^*. As shown in Figure 2, at θ^*, the size of the attack jumps discretely from $\Phi(\sqrt{\alpha}(x^* - \theta^*))$ down to the lower value $\Phi(\sqrt{\alpha}(x^* - y^*)) = \Phi(-\sqrt{\alpha}\gamma)$. All regimes that manipulate, all $\theta \in [\theta^*, \theta^{**})$, face *the same sized attack* $\Phi(-\sqrt{\alpha}\gamma)$. Beyond that, for regimes with enough innate strength that no manipulation is required, the size of the attack continuously declines according to $\Phi(\sqrt{\alpha}(x^* - \theta))$.

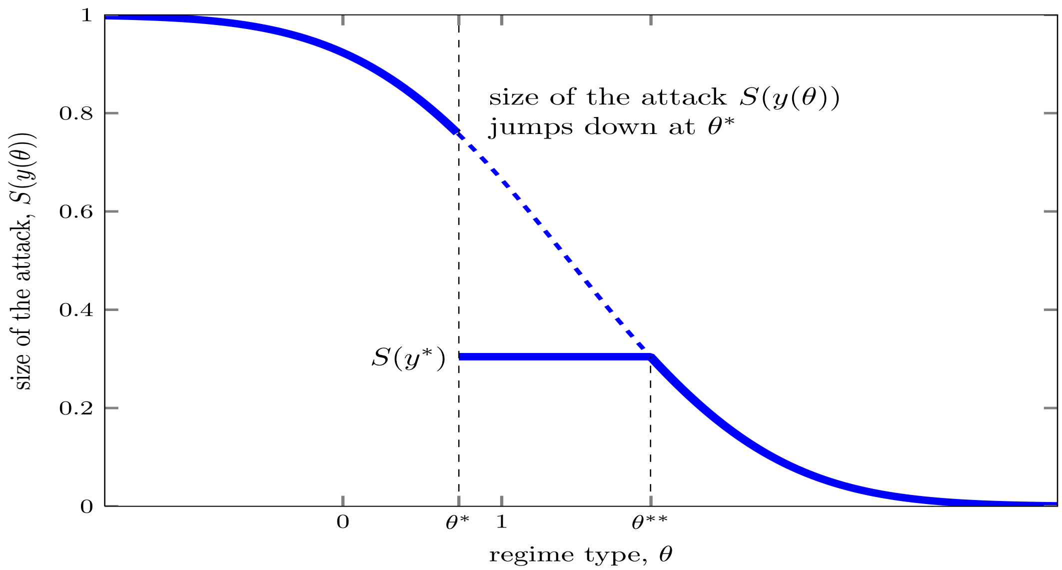


Figure 2: Manipulation leads to discrete fall in size of the attack.

All regimes with $\theta \in [\theta^*, \theta^{**})$ choose the same apparent strength $y^* = \theta^{**} = x^* + \gamma$ and consequently face the same sized attack $S(y^*) = \Phi(-\sqrt{\alpha}\gamma)$. All regimes that do not manipulate have $y(\theta) = \theta$ and face the attack $S(y(\theta)) = \Phi(\sqrt{\alpha}(x^* - \theta))$. Observe that the attack is continuous at the upper boundary θ^{**}

But is manipulation effective in equilibrium? Thus, as in classic signaling games, the regime is able to send a (noisy) signal in equilibrium and this enables some weaker regime types to pool with stronger regime types. But this pooling tells us nothing about whether the signal-jamming is *effective* in equilibrium. It may be that the only regimes that are able to imitate stronger regime types are those regimes that would have survived in the absence of the signal-jamming technology. Moreover, it may be that some weak regimes that would survive if they could commit to not use manipulation are overthrown because they cannot make that commitment. Put differently, the discussion so far has taken the thresholds θ^* and θ^{**} as given. But these thresholds are endogenous and, in principle, may shift strongly agains the regime so that in equilibrium information manipulation is ultimately ineffective.

It turns out that manipulation *is* effective when the signal precision α is sufficiently high. But to see this we have to solve the rest of the model.

Solving the model with linear costs. To solve the model, re-write the indifference condition of the marginal citizen (12) in terms of the equilibrium thresholds x^*, θ^* and the apparent strength $y(\theta)$, as follows

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p}\int_{\theta^*}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - y(\theta)))\, d\theta \tag{20}$$

Now use the first order condition (15) to simplify the right hand side integral

$$\begin{aligned}\int_{\theta^*}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - y(\theta)))\, d\theta &= \int_{\theta^*}^{\theta^{**}} c\, d\theta + \int_{\theta^{**}}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta))\, d\theta \\ &= (x^* - \theta^* + \gamma)c + \Phi(-\sqrt{\alpha}\gamma)\end{aligned}$$

where the first equality uses $y(\theta) = \theta$ for $\theta \geq \theta^{**}$ and the second equality uses $\theta^{**} = x^* + \gamma$. Plugging this back into (20) gives

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p}\left[(x^* - \theta^* + \gamma)c + \Phi(-\sqrt{\alpha}\gamma)\right] \tag{21}$$

It is straightforward to show that there is a unique threshold difference $\theta^* - x^*$ that solves this equation. The regime threshold θ^* is then determined using the indifference condition (11) which, with linear costs, can be written

$$\theta^* = (x^* - \theta^* + \gamma)c + \Phi(-\sqrt{\alpha}\gamma) \tag{22}$$

where the difference $\theta^* - x^*$ on the right hand side is uniquely determined by (21) above.

4.2 Changes in signal precision

With the equilibrium thresholds x^*, θ^* determined by (21)-(22), we can now solve for the conditions under which information manipulation is effective for the regime.

Effective manipulation when signal precision is high. For large enough α the regime threshold is strictly less than the Morris-Shin benchmark $\theta^*_{\text{MS}} = 1 - p$. In particular:

PROPOSITION 3. For each c there is a $\underline{\alpha}(c)$ such that for $\alpha \leq \underline{\alpha}(c)$ all regimes are at a corner with $y(\theta) = \theta$ for all θ and $\theta^* = \theta^*_{\text{MS}}$. Otherwise, for $\alpha > \underline{\alpha}(c)$, regimes $\theta \in [\theta^*, \theta^{**})$ are at an interior solution and there is a critical precision $\alpha^*(c,p) \geq \underline{\alpha}(c)$ given by

$$\alpha^*(c,p) := \underline{\alpha}(c) \exp\left(\max\left[0, \Phi^{-1}(p)\right]^2\right) \tag{23}$$

such that

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \qquad \text{for all} \qquad \alpha > \alpha^*(c,p) \tag{24}$$

and $\lim_{\alpha\to\infty} \theta^* = 0$. For α sufficiently high, θ^* is strictly less than the Morris-Shin benchmark.

There are two cases to consider. First, when the opportunity cost of attacking the regime is *low*, $p < 1/2$, then all that matters is whether the signal precision α is high enough to induce *any* regime to manipulate. If so, the regime threshold is monotonically declining in

α and so for all $\alpha > \underline{\alpha}$ the regime threshold is less than the Morris-Shin benchmark. This is shown in the left panel of Figure 3. Second, if the opportunity cost of attacking the regime is *high*, $p > 1/2$, then the regime threshold is "hump-shaped" in α, reaching a maximum at $\alpha^* > \underline{\alpha}$ before declining thereafter. This is shown in the right panel of Figure 3. In either case, in the limit as $\alpha \to \infty$, the regime threshold $\theta^* \to 0$ for any fixed c and p.

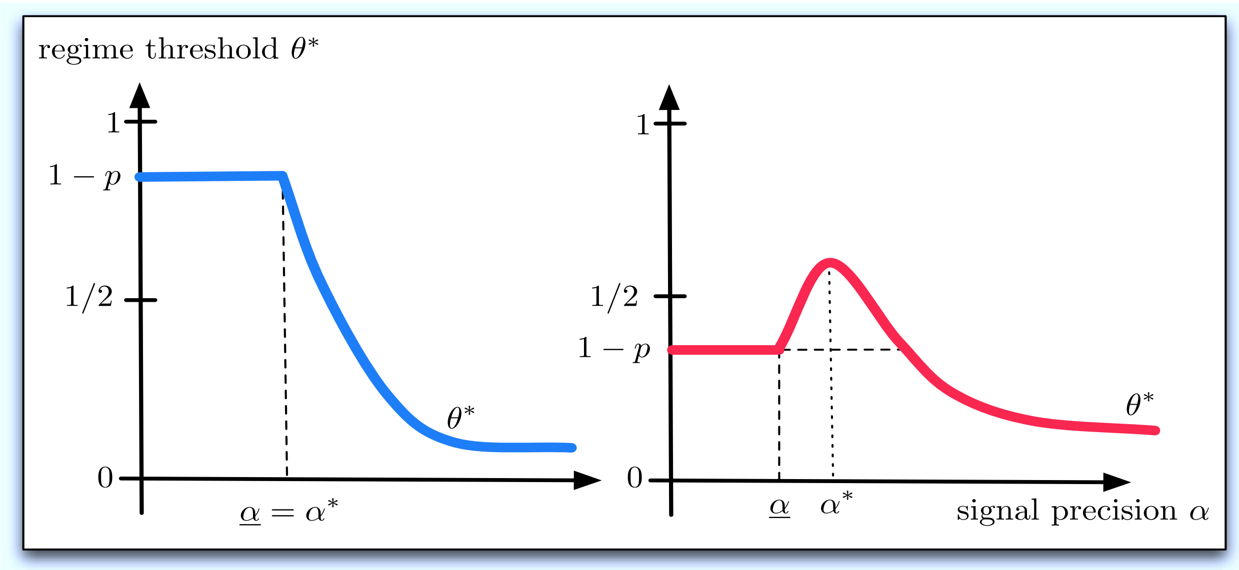


Figure 3: Information manipulation is effective when signal precision α is sufficiently high.

The left panel shows the case when $p < 1/2$ and the regime threshold θ^* is *monotone* decreasing in the precision α, the right panel shows the case when $p > 1/2$ so that θ^* is non-monotone in α. In both cases, for low enough precision ($\alpha < \underline{\alpha}$) the regime threshold coincides with the Morris-Shin benchmark $\theta^*_{\text{MS}} = 1 - p$ while for high enough α the threshold is less than $1 - p$.

Even the most fragile regimes can survive. This result is striking. As the precision becomes sufficiently high, *all* the regimes that can survive, do survive. To see an extreme example of this, consider an economy with opportunity cost $p \to 0$ so that it requires almost no individual sacrifice to participate in an attack on the regime. In the Morris-Shin benchmark we would have $\theta^*_{\text{MS}} \to 1$ and only the strongest regimes, those with $\theta \geq 1$, can survive. But with information manipulation, we have $\theta^* \to 0$ so all regimes $\theta \geq 0$ survive even though p is very low. If information can be manipulated and signals are sufficiently precise, then even the very most fragile regimes can survive.

Signal precision and the size of the marginal attack. To understand this result further, observe that the size of the attack facing the marginal regime, $S(y^*)$ is always monotonically declining in the signal precision α. This follows immediately on plugging the expression for γ from (18) into (19). But there is another, perhaps more insightful, way to see this too. Recall the problem of a regime minimizing $S(y) + cy$ with $S(y) = \Phi(\sqrt{\alpha}(x^* - y))$.

Since $S(y)$ is strictly monotone in y we can equivalently choose S to minimize

$$S - \frac{c}{\sqrt{\alpha}}\Phi^{-1}(S) \tag{25}$$

with first order condition $\phi(\Phi^{-1}(S)) = c/\sqrt{\alpha}$. Any solution is independent of the citizen threshold x^* and depends only on the ratio $c/\sqrt{\alpha}$ (rather than each parameter separately). The ratio $c/\sqrt{\alpha}$ determines the effective cost of obtaining a given-sized attack. And, from the second order condition, the size of the attack that solves this cost minimization problem must be at least locally increasing in $c/\sqrt{\alpha}$. From this point of view, we see clearly that an increase in the signal precision α reduces the cost of obtaining a smaller attack; the more probability density is located near the mean, the easier it is for the regime to achieve a relatively large change in the size of the attack from a relatively small perturbation to the signal mean. Other things equal, a regime can achieve a smaller attack when α is large.

Whenever $\alpha > \underline{\alpha}(c)$, we can solve (25) to obtain the marginal attack

$$S(y^*) = \Phi(-\sqrt{\alpha}\gamma) = \Phi\left(-\sqrt{\log\left(\frac{\alpha}{\underline{\alpha}(c)}\right)}\right) = \Phi\left(-\sqrt{2\log\left(\frac{\sqrt{\alpha}}{c}\phi(0)\right)}\right) \tag{26}$$

Otherwise, for $\alpha < \underline{\alpha}(c)$, the marginal attack is simply the Morris-Shin benchmark, $1-p$.

Critical signal precision. If the regime did not have to pay to obtain y^*, then we would be done; the reduction in the size of the marginal attack would directly translate into a reduction in θ^*. But since the manipulation is costly, the regime threshold satisfies $\theta^* = S(y^*)+c(y^*-\theta^*)$ so a fall in $S(y^*)$ is not enough to guarantee that θ^* falls as the signal precision α increases. In particular, if the cost of y^* increases sufficiently faster than the size of the attack $S(y^*)$ falls, then the threshold θ^* increases. When this happens, there is a non-monotone relationship, as in the right panel of Figure 3. Proposition 3 establishes that this non-monotonicity in α can only occur if the individual opportunity cost of attacking the regime is large ($p > 1/2$). The critical precision α^* is exactly that such that the marginal regime faces an attack equal to the Morris-Shin benchmark $1-p$. Using (26), the critical precision α^* solves

$$\Phi\left(-\sqrt{2\log\left(\frac{\sqrt{\alpha^*}}{c}\phi(0)\right)}\right) = 1-p \tag{27}$$

which can be inverted to obtain the expression in (23) above. When the signal precision is larger than α^* the regime threshold θ^* is decreasing in α and is eventually less than $1-p$.

4.3 Intuition for threshold reduction

The result that information manipulation is effective when the signal precision is sufficiently high depends on two conceptually distinct effects. First, manipulation of the signal mean is a *potentially* powerful tool for the regime when the signal precision is high, because then most citizens have signals near the mean and it takes only a small amount of bias to deliver a large change in the size of the attack. Second, the regime is actually able to *make use* of this powerful tool in equilibrium. The regime is able to use its tool because of the joint coordination and information-filtering problems facing individual citizens.

4.3.1 A powerful tool ...

To see why signal manipulation is potentially powerful, recall the Morris-Shin benchmark (6) where, as a function of the regime type θ, the size of the aggregate attack is $S^*_{\text{MS}}(\theta) = \Phi(\sqrt{\alpha}(x^*_{\text{MS}} - \theta))$ which approaches a step function $\mathbb{1}\{1 - p > \theta\}$ as $\alpha \to \infty$. Relative to this benchmark, a small "shock" that increased the signal mean from θ to $\theta + \tilde{a}$, say, would cause a shift of the step function so that the attack facing the marginal regime would fall from $1 - p$ to 0 and all regimes $\theta \geq 0$ would be able to survive. But this only delivers a reduction in the attack if $\tilde{a}$ is *unanticipated*. If the citizens could correctly anticipate such an increase in the signal mean, then they would discount their signals appropriately.

This discussion illustrates two points: (i) the ability to shift the mean is a potentially powerful technology when the precision is high, for then it takes only a small amount of bias to achieve a large reduction in the size of the attack and hence a large fall in θ^*, but (ii) for this technology to be operationally useful to the regime, it must be the case that the citizens are, somehow, impeded in their efforts to infer the bias.

4.3.2 ... and the regime can make use of it

In turn, two features of the model account for the citizen's inability to infer the bias. First, different regimes have different incentives so that given imperfect information about θ, there is also imperfect information about $a(\theta)$. If all regimes took the same action, $\tilde{a}$ say, it would be easy to undo. Second, citizens are imperfectly coordinated.

Regime actions are state-contingent. In this model, citizens know the regime's incentives, so why can they not use that information to back out the regime's manipulation? The short answer is that the regime's actions $a(\theta) = y(\theta) - \theta$ are *state-contingent* and while, in equilibrium, citizens know the function $a(\cdot)$, they are imperfectly informed about θ and hence cannot extract the exact value $a(\theta)$ needed to undo the manipulation. Thus, by contrast with traditional *career concerns* models, such as Holmström (1999) and Dewatripont, Jewitt and Tirole (1999), the signal receivers cannot perfectly decompose their signal into its true underlying state component and its endogenous bias component.

Imperfectly coordinated signal receivers. In traditional models of strategic information transmission such as Crawford and Sobel (1982) and Holmström (1999) there is one sender and one signal receiver. But in this paper, as in the similar model of Angeletos, Hellwig and Pavan (2006), there is instead a cross-section of imperfectly coordinated receivers. This makes a crucial difference to the sender's ability to shape outcomes in its favor.

In particular, citizens differ in their assessment of θ and hence differ in their assessment of $a(\theta)$. And from Proposition 2 it is common knowledge that *any* manipulation serves to increase the regime's apparent strength $y(\theta) = \theta + a(\theta)$ and hence to reduce the aggregate attack. Since the individual s_i and aggregate S are strategic complements, this makes each individual less likely to participate. Consequently, the size of the attack facing the marginal regime falls. Moreover, when α is high, it takes only a small amount of bias to achieve a large reduction in S. Knowing this makes individuals even more reluctant to participate. In the limit as $\alpha \to \infty$, the regime threshold is driven to zero because in that limit it takes only an infinitesimal amount of bias to deter *all* citizens from attacking.

This argument relies on the citizens being imperfectly coordinated, i.e., on individuals taking the aggregate S as given. To see the importance of this, suppose to the contrary that citizens were perfectly coordinated and able to act as a single large agent who could force regime change for all $\theta < 1$. This agent receives a signal $x = \theta + a + \varepsilon$ with precision $\alpha \to \infty$. For simplicity, suppose also that costs are strictly convex. This implies $y(\theta)$ is strictly increasing[9] and hence can be inverted to recover $\theta = y^{-1}(x)$. But knowing θ, the single agent can deduce any manipulation $a(\theta)$ and discard it — so the regime has no incentive to undertake costly manipulation. So if citizens are perfectly coordinated, they know $x \to \theta$ and attack if and only if $x = \theta < 1$ and all regimes $\theta \in [0, 1)$ are wiped out. By contrast, if citizens are imperfectly coordinated, all regimes $\theta \in [0, 1)$ survive. In this sense, the imperfect coordination drives the equilibrium selection in the regime's favor (see Appendix B for more details).

Manipulation without bias. In this model the regime can benefit from information manipulation because it is able to exploit the joint coordination and information-filtering problems facing individual citizens so as to introduce at least a small amount of bias into signals. But it turns out that the regime can benefit from information manipulation even if there is *no* bias.[10] Section 6.2 below provides one example of this, showing that the regime can benefit from manipulation that directly alters information precision while leaving the signal mean free of distortion. Edmond (2012) provides another example, developing a related model where citizens and the regime have quadratic preferences of the kind used in the cheap-talk literature and showing that while citizens in this setting can infer the bias in their information (and so there is no bias in equilibrium), their signals are endogenously noisier in a way that helps the regime prevent coordination against its interests.

[9] It is straightforward to show that if the cost function $C(\cdot)$ is strictly convex, the regime's problem is strictly supermodular in (θ, y) and hence $y(\theta)$ is strictly increasing in θ.

[10] While the existence of equilibrium bias is not necessary for manipulation to be effective, such manipulation is generally much *more* effective when it is able to affect mean beliefs.

4.4 Changes in the cost of manipulation

As we have seen, an increase in the signal precision α reduces a regime's cost of obtaining a given-sized reduction in the aggregate attack, and, at least when α is high enough, this reduces the regime threshold θ^*. This suggests there is a close link between changes in the signal precision and changes in the cost of manipulation. Indeed, from (26), the aggregate attack facing the marginal regime $S(y^*)$ depends only on the ratio $c/\sqrt{\alpha}$, not each of these parameters separately. The same is true for the regime threshold θ^*:

Proposition 4. The parameters α and c only affect the regime threshold θ^* through the ratio $c/\sqrt{\alpha}$. Consequently, an increase in the cost c of manipulation always has the opposite sign to the effect of of an increase in the signal precision

$$\frac{\partial \log \theta^*}{\partial \log c} = -2\frac{\partial \log \theta^*}{\partial \log \alpha} \tag{28}$$

In particular, for high enough α we have $\alpha > \alpha^*$ and an increase in the cost of manipulation c would increase the regime's threshold θ^* and reduce the regime's ex ante chances of surviving while at the same time an increase in the signal precision α would reduce the threshold and so increase its chances of surviving. In this example the *level* of the regime threshold may well be $\theta^* < \theta^*_{\text{MS}} = 1 - p$ so that the regime is benefitting from the ability to manipulate information — it is just that a marginal increase in c makes the regime marginally worse off.

4.5 Asymptotic results with general convex costs

The result that regimes benefit from manipulation when the signal precision is high enough is not special to the case of linear costs. In particular:

Proposition 5. For general convex cost functions $\lim_{\alpha\to\infty} \theta^* = 0^+$. Moreover, if the cost function is *strictly convex*, $C''(a) > 0$, then $\lim_{\alpha\to 0} \theta^* = 1^-$.

So, for high enough α, the regime is indeed able to reduce the threshold θ^* below the Morris-Shin benchmark $1 - p$. In addition, if the cost function is strictly convex, then as $\alpha \to 0$, the regime threshold rises *above* the Morris-Shin benchmark (as in Figure 4). With linear costs, as the signal precision becomes small enough ($\alpha < \underline{\alpha}$) it is common knowledge that no regime will find it worthwhile to manipulate and so the regime threshold approaches the Morris-Shin outcome. But with strictly convex costs the regime *always* has an incentive to manipulate even for very low levels of α, since it can always take a very small action with very small marginal cost. For low values of α, the regime would want to be able to credibly commit to refrain from all information manipulation. But because such commitments cannot be made, the regime is "trapped" into taking costly actions even as the signal precision falls. Indeed, the regime threshold is driven all the way to 1 as α falls to zero. In this sense, the manipulation completely backfires on the regime when the signal precision is low.

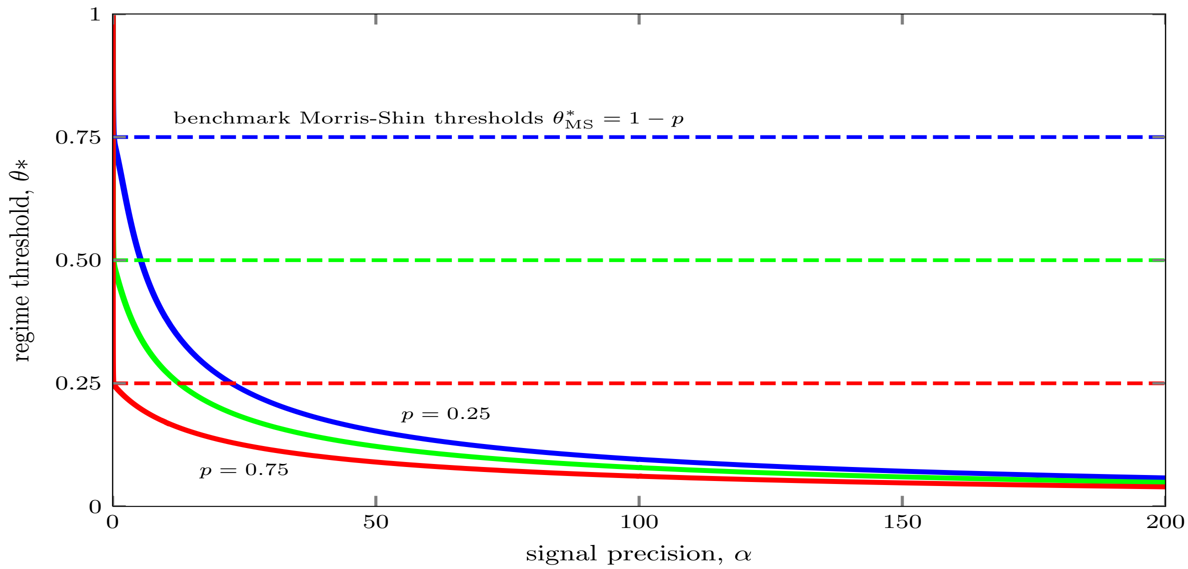


Figure 4: Information manipulation with strictly convex costs.

With information manipulation, θ^* is below $1 - p$ when α is sufficiently high. As $\alpha \to \infty$ the threshold $\theta^* \to 0$ so that all regimes $\theta \in [0, 1)$ survive. By contrast as $\alpha \to 0$ the threshold $\theta^* \to 1$. Information manipulation *backfires* when the signal precision is too low. These examples use strictly convex costs $C(a) = a^2/2$.

5 An information revolution

I now consider a situation where citizens have a large number of signals and where the regime's costs of information manipulation depend on both the amount of manipulation they seek to achieve and on the number of signals they need to control. In this world, an increase in the number of signals tends to make regimes easier to overthrow, i.e., tends to increase θ^*, unless there are strong *economies of scale* in the regime's ability to control sources of information (e.g., if the control of one dramatically reduces the cost of controlling another).

Many signals. Suppose citizens obtain information about the regime from $n \geq 1$ identical[11] media outlets (sources of information). Each outlet $j = 1, ..., n$ produces a signal mean $y = \theta + a$ and each citizen observes n signals of the form $x_{ij} = y + \varepsilon_{ij}$ where the noise ε_{ij} is IID across citizens and media outlets with mean zero and precision $\hat{\alpha} > 0$. This noise need not be normally distributed. But by a classical *Lindeberg-Lévy* central limit theorem, for large n the *average signals* $x_i := \frac{1}{n} \sum_{j=1}^{n} x_{ij}$ are approximately normally distributed with mean $y = \theta + a$ and precision $\alpha := n\hat{\alpha}$. Importantly, the precision of the average signal is proportional to n.

Regime's cost of controlling many signals. Let the regime's cost of injecting manipulation a into n signals be $C(a, n) := c(n)a$ with $c(n) > 0$ and increasing in n. The more

[11]The supplementary appendix considers an alternative setup with *heterogeneous* media outlets, some of which pass on the manipulation and some of which do not.

sources of information, the more costly it is for the regime to achieve a given level of manipulation a. For example, the marginal cost of a may be $\hat{c}$ *per signal*, i.e., $c(n) = \hat{c}n$. More generally, the curvature of $c(n)$ reflects the regime's ability to influence a changing informational environment. If it is harder to exert control over an ever-expanding array of media outlets, that suggests the average cost $c(n)/n$ is increasing in n, i.e., diseconomies of scale in controlling media outlets. Alternatively, if there are complementarities in media control, with influence over one media outlet facilitating influence over others, then that suggests average cost $c(n)/n$ is decreasing in n, i.e., economies of scale in controlling media outlets.

Equilibrium. With n symmetric media outlets the analysis proceeds as before except that citizens' types are represented by their average signals $x_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ which have mean $y = \theta + a$ and precision $\alpha = \hat{\alpha} n$ while the regime has costs of manipulation $c(n)a$. The cost for a regime to obtain a given-sized attack S is determined by the ratio

$$\frac{c(n)}{\sqrt{n\,\hat{\alpha}}} \tag{29}$$

namely, the generalization of (16) to this setting with n signals. If this ratio is sufficiently large, i.e., $\hat{\alpha} < (c(n)/\phi(0))^2/n =: \underline{\alpha}(n)$, then it is too expensive to manipulate and the equilibrium coincides with the Morris-Shin benchmark. Otherwise, there is an interval of regimes $[\theta^*, \theta^{**})$ that do manipulate and that choose apparent strength $y^* = x^* + \gamma$. These thresholds are determined as in (21)-(22) but with $c(n)$ replacing c and $n\hat{\alpha}$ replacing α, etc.

Effect of an increase in the number of signals. The results on the effects of signal precision and the costs of manipulation from Section 4 continue to apply so long as the number of signals is held fixed. But what happens if there is an increase in n?

Proposition 6. The effect of an increase in the number of media outlets n on the regime threshold θ^* is given by

$$\frac{\partial \log \theta^*}{\partial \log n} = -2\left(\frac{c'(n)n}{c(n)} - \frac{1}{2}\right)\frac{\partial \log \theta^*}{\partial \log \hat{\alpha}} \tag{30}$$

As before, if the signal precision is too low, $\hat{\alpha} < \underline{\alpha}(n)$, then all regimes are at a corner and $\theta^* = 1 - p$ independent of $\hat{\alpha}$ and n. In this case, both effects in (30) are zero. Otherwise, if $\hat{\alpha} > \underline{\alpha}(n)$, the effect of the signal precision depends on whether $\hat{\alpha}$ is larger than a critical signal precision $\alpha^*(n, p)$ analogous to (23) above. For $\hat{\alpha} > \alpha^*(n, p)$ an increase in signal precision reduces the regime threshold θ^*. For fixed n this would help regimes survive, but now with n increasing the overall effect also depends on the elasticity of $c(n)$.

Role of economies of scale in media control. To be specific, suppose we are in a sufficiently *high precision* environment so that an increase in signal precision reduces the regime threshold θ^*. Then from (30) an increase in n makes the regime threshold θ^* increase unless there are strong economies of scale in media control. For example, if the average cost function $c(n)/n = \hat{c}$ for all n, then in a high precision environment the regime is worse off. Observe from Proposition 5 that the magnitude of the cost effect on θ^* is exactly *twice as large* as the signal precision effect. So if an increase in n causes both the cost of manipulation and the signal precision to increase by the same magnitude, then the former effect dominates. In a high precision environment, that means θ^* rises and the regime is easier to overthrow.

More generally, let $\epsilon(n)$ denote the elasticity of the average cost function $c(n)/n$, namely

$$\epsilon(n) := \frac{c'(n)n}{c(n)} - 1$$

Then in a high precision environment, an increase in n increases the regime threshold θ^* and makes the regime easier to overthrow unless the average cost $c(n)/n$ declines sufficiently fast — specifically, unless $\epsilon(n) < -1/2$. This can be seen as a restriction on the number of media outlets n. For example, consider the cost function $c(n) = c_0 + c_1 n$ so that there is a "fixed cost" $c_0 > 0$ and a constant "marginal cost" of manipulating more signals $c_1 > 0$. Then the elasticity of average cost is $\epsilon(n) = -c_0/(c_0 + c_1 n)$ and the condition $\epsilon(n) < -1/2$ is satisfied only when the number of media outlets is relatively small, namely $n < (c_0/c_1)$. Thus in addition to influencing the information *content*, as measured by a, the regime may also have an incentive to directly seek control over the *number* of outlets n.

This suggests there is an important difference between a rapid increase in information originating from technologies that are relatively easy for the regime to centrally control as opposed to technologies that are not. Media like newspapers, radio, cinema and similar forms of public broadcasting are technologies for which a regime plausibly pays a high fixed cost to control but relatively small marginal cost (e.g., high fixed costs of a central propaganda office producing a stream of officially approved content, but low marginal cost of exerting control over additional newspapers, radio stations or cinemas). If so, these technologies are ones we would expect to see prove complementary to the regime's ability to survive. By contrast, diffuse modern technologies like social media may be harder to control in this fashion.

6 Extensions

I briefly consider two alternative setups. Section 6.1 allows for a *struggle* over information as an opposition group attempts to shift information against the regime. Section 6.2 considers a model where the regime's actions directly affect signal precision rather than the mean. Thus in this setting the regime engages in manipulation but cannot, by assumption, introduce any bias. The supplementary appendix contains more detail on these and other extensions.

6.1 Struggles over information

Suppose there is an *opposition* and that if a regime is of type θ and takes action a while the opposition takes action e, then citizens draw signals $x_i = \theta + a - e + \varepsilon_i$ where ε_i is again IID normal with mean zero and precision α. The regime's action a increases the signal mean while the opposition's action e decreases it. Both actions are unobserved by citizens.

To highlight the struggle over manipulating information, I assume that the regime and the opposition *both* know the regime's type θ. Along the equilibrium path citizens receive signals with mean $\theta + a(\theta) - e(\theta)$. If $a(\theta) = e(\theta)$, then the opposition simply undoes the efforts of the regime of type θ. I further assume that the opposition pays $C(\kappa e)/\kappa$ to take action e where $C(\cdot)$ is the same cost function as for the regime and where $\kappa > 0$. If $\kappa = 1$, the costs of the regime and opposition are the same, if $\kappa > 1$ the regime has a cost advantage.

The payoff to the opposition is of the form $S - C(\kappa e)/\kappa$ so that the opposition prefers the attack to be as large as possible (subject to the cost of taking action e), similar to the dissidents in Bueno de Mesquita (2010). Now let $S(\theta, a, e)$ denote the aggregate attack. Taking this as given, an equilibrium in the subgame between the regime and the opposition consists of hidden actions $a(\theta), e(\theta)$ that are mutual best responses

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \{B(\theta, S(\theta, a, e(\theta))) - C(a)\} \tag{31}$$

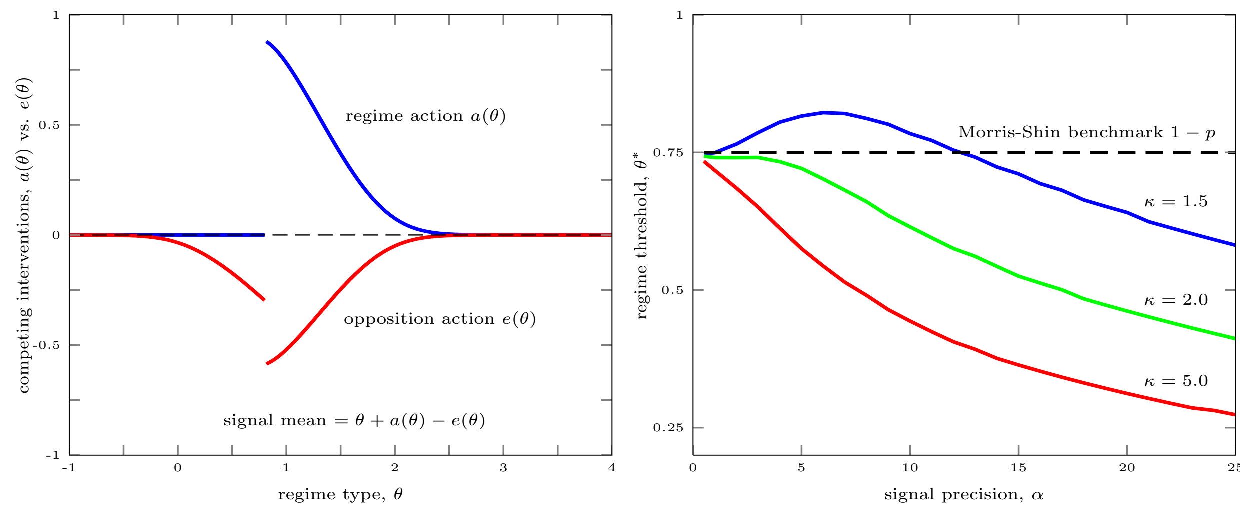
$$e(\theta) \in \underset{e \geq 0}{\operatorname{argmax}} \{S(\theta, a(\theta), e) - C(\kappa e)/\kappa\} \tag{32}$$

The regime's outside option introduces a key *asymmetry*. The regime does not care about the size of S in those states where it is overthrown. By contrast, the opposition cares about S both when the regime is overthrown and when it is not.

I consider a monotone equilibrium where the regime is overthrown if $\theta < \theta^*$ and citizens participate $s(x_i) = 1$ if their signal is $x_i < x^*$ for thresholds x^*, θ^* to be determined. In this equilibrium it is straightforward to show that $a(\theta) = 0$ when $\theta < \theta^*$ and $a(\theta) = \kappa e(\theta)$ when $\theta \geq \theta^*$. Thus, when the regime survives, its actions are larger than those of the opposition if and only if the regime has a cost advantage, $\kappa > 1$. Just as the regime's hidden actions jump discretely to $a(\theta^*) > 0$ at the threshold $\theta = \theta^*$, so too do the opposition's actions typically jump at the threshold (though their jump may be up *or* down, depending on parameters). The left panel of Figure 5 illustrates these action profiles with $\kappa > 1$ so that the regime's actions $a(\theta)$ are larger than the opposition's actions $e(\theta)$ on $\theta \geq \theta^*$.

Does the opposition's action undo the regime's efforts? On the one hand, it is true that the presence of the opposition generally moves the threshold θ^* against the regime (it is higher than it would be in the model where there is no opposition, $\kappa = \infty$). On the other hand, the regime still manipulates information and for high enough signal precision α is still better off than it would be in the Morris-Shin benchmark (as shown in Figure 5).

These results suggest that while the presence of organized opposition is important for understanding how much manipulation takes place and for the equilibrium level of the regime threshold θ^*, it is less important for the result that the regime threshold can be decreasing in the signal precision so that more precise signals move the threshold in the regime's favor.



(a) regime and opposition actions $a(\theta)$ and $e(\theta)$ as function of regime type θ

(b) regime threshold θ^* as function of signal precision α and opposition cost disadvantage κ

Figure 5: Hidden actions and regime threshold when there is an opposition.

Panel (a) shows the regime's and opposition's actions $a(\theta)$ and $e(\theta)$ when there is a struggle over information. For clarity the opposition's $e(\theta)$ is plotted on a negative scale. For $\theta < \theta^*$, only the opposition takes an action. For $\theta \geq \theta^*$ the actions satisfy $a(\theta) = \kappa e(\theta)$, where κ measures the costliness of the opposition's action. In this example, $\kappa = 1.5$ and the opposition's costs are greater than the regime's. Panel (b) shows θ^* as a function of the signal precision α for various κ. In these examples, the regime still benefits from information manipulation in that $\theta^* < 1-p$ when α is high enough. In these calculations, the opportunity cost is $p = .25$ and the regime's cost functions is $C(a) = a^2/2$ so that the opposition's cost function is $C(\kappa e)/\kappa = \kappa e^2/2$.

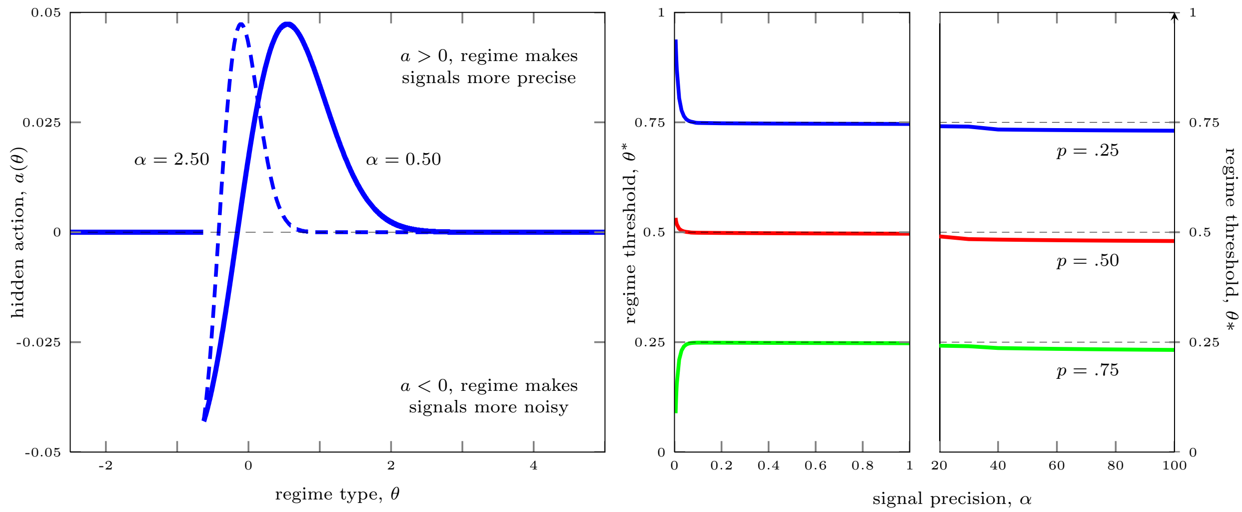
6.2 Manipulating signal precision

Until now, signal manipulation entered in an additive way, $x_i = \theta + a + \varepsilon_i$. With this specification the action shifts the signal mean and only indirectly influences the signal precision. I now consider an alternative approach where the regime can *directly* set the signal precision but *cannot introduce bias*. In particular, let signals be $x_i = \theta + \varepsilon_i$ where the ε_i is IID normal with mean zero and precision $\beta(a) > 0$ that depends on the regime's hidden action a. I adopt the specification

$$\beta(a) := \alpha\left(\frac{1}{2} + \Phi(a)\right), \qquad \alpha > 0$$

Thus when the regime takes no action, the precision is just α, which in this context should be thought of as the *intrinsic* signal precision.

Again, I consider a monotone equilibrium where the regime is overthrown for $\theta < \theta^*$ and citizens participate $s(x_i) = 1$ for $x_i < x^*$ for thresholds x^*, θ^* to be determined. It is straightforward to show that $a(\theta) = 0$ for $\theta < \theta^*$ before jumping discretely at θ^* and that the sign of $a(\theta)$ is the same as the sign of $\theta - x^*$. If $\theta > x^*$, the hidden actions are positive. If $\theta < x^*$, the hidden actions are negative.



(a) hidden actions $a(\theta)$ as function of regime type θ

(b) regime threshold θ^* as function of signal precision α

Figure 6: Hidden actions and regime threshold when regime can manipulate signal precision. Panel (a) shows the regime's actions to directly manipulate signal precision. For intermediate θ it may be optimal for $a(\theta) < 0$ so that the regime makes the signal less noisy than α. For high θ it is optimal for $a(\theta) > 0$ so that the regime clarifies its strength by making the signal more precise than α. In this example the opportunity cost is $p = .25$. Panel (b) shows θ^* as a function of the *intrinsic* signal precision α for various p. In these examples, the regime benefits from information manipulation in that $\theta^* < 1 - p$ when α is high enough. In all these calculations, the cost function is $C(a) = a^2/2$.

Intuitively, if the regime has intermediate type $\theta \in [\theta^*, x^*)$ then it makes the signal noisier than α (it *muddies* the signal) while if $\theta > x^*$, the regime makes the signal less noisy than α so as to *clarify* its position of strength. The left panel of Figure 6 illustrates with parameters chosen so that $x^* > \theta^*$, implying that the hidden actions jump down at θ^*. Diminishing returns to information manipulation set in faster when α is high. An increase in α from .5 to 2.5 hardly shifts the state threshold θ^* at all.

Despite the lack of bias in signals, the regime can still benefit from information manipulation in this setting. The right panel of Figure 6 shows θ^* as α varies from 0 to 100. The left panel shows the results for very low α. There is a brief interval of α where θ^* can initially rise. These effects play out very quickly and for high enough α we have $\theta^* < 1 - p$ so that the regime is again benefitting from information manipulation.

7 Conclusions

In this paper I develop a simple signal-jamming model of information and political regime change. Perhaps the most surprising result is that a regime's chances of surviving *increase* as the precision of the signals available to individuals becomes sufficiently high. In contrast with familiar signal-jamming games, the regime's manipulation presents citizens with a difficult signal-extraction problem and the manipulation is often payoff improving for the regime. Perhaps breakthroughs in information technologies may not be as threatening to autocratic regimes as is often supposed?

Offsetting this pessimistic result, the model also predicts that in a world with increasingly *many* sources of high precision information, a regime will be easier to overthrow unless there are strong *economies of scale* in the regime's control over multiple sources of information. If there are diseconomies or only mild economies of scale, then more sources of information will make the regime easier to overthrow.

The model thus allows for two kinds of information revolutions. In the first kind, perhaps best associated with the role of radio and mass newspapers under the totalitarian regimes of the early twentieth century, an information revolution consists of many increasingly high precision sources of information but these technologies are subject to strong economies of scale in control, e.g., a large fixed cost of a propaganda machinery with low marginal cost of controlling additional newspapers, radio stations or cinemas. This kind of information revolution is favorable to a regime's survival prospects. In the second kind, perhaps best associated with the rise of diffuse forms of social media and other decentralized sources of information, however, there are less likely to be strong economies of scale in information control. This kind of information revolution is unfavorable to a regime's survival prospects.

The coordination game studied in this paper is deliberately stylized so as focus attention on the effectiveness of the regime's manipulation and its sensitivity to changes in the information environment. The results suggest several directions for future research. For example, this model takes as given the degree of influence the regime has over citizens' sources of information. But there are clear incentives for the regime to attempt to exert influence both over the number (or type) of information outlets as well as information content. It would clearly be interesting to develop a richer model where the degree of influence over the media is itself an equilibrium outcome, in the spirit of Besley and Prat (2006) or Gehlbach and Sonin (2008), that needs to be determined simultaneously with the regime's manipulation and survival probability. Similarly, this paper has abstracted from all issues of individual and social learning and dynamics. Recent papers including Angeletos, Hellwig and Pavan (2007), Dasgupta (2007) and Heidhues and Melissas (2006), have made important advances in incorporating these features into global games models. It would be interesting to consider a dynamic version of the information manipulation model where both individual and social information can accumulate over time so that unrest against the regime builds (or dissipates). An extension along these lines would likely bring the analysis closer to the information cascade models of regime change developed by Kuran (1991), Lohmann (1994) and others, but would feature strategic interactions between the regime and citizens that are absent from their work.

Appendix

A Proofs and omitted derivations

A.1 Morris-Shin Benchmark

Let $\hat{x}, \hat{\theta}$ denote candidates for the critical thresholds. The posterior beliefs of a citizen with x_i facing $\hat{\theta}$ are given by $\text{Prob}[\theta < \hat{\theta} \,|\, x_i] = \Phi(\sqrt{\alpha}(\hat{\theta} - x_i))$. A citizen with x_i will subvert if and only if $\Phi(\sqrt{\alpha}(\hat{\theta} - x_i)) \geq p$. This probability is continuous and strictly decreasing in x_i, so for each $\hat{\theta}$ there is a unique signal for which a citizen is indifferent. Similarly, if the regime faces threshold $\hat{x}$ the mass of subversives is $\text{Prob}[x_i < \hat{x} \,|\, \theta] = \Phi(\sqrt{\alpha}(\hat{x} - \theta))$. A regime θ will not be overthrown if and only if $\theta \geq \Phi(\sqrt{\alpha}(\hat{x} - \theta))$. The probability on the right hand side is continuous and strictly decreasing in θ, so for each $\hat{x}$ there is a unique state for which a regime is indifferent. The Morris-Shin thresholds $x^*_{\text{MS}}, \theta^*_{\text{MS}}$ simultaneously solve these best response conditions as equalities, as stated in equations (4)-(5) in the main text. It is then straightforward to verify that there is only one solution to these equations and that $\theta^*_{\text{MS}} = 1 - p$ independent of α and $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha}$.

A.2 Proof of Proposition 1

The proof shows first that (i) there is a unique equilibrium in monotone strategies, and (ii) that the unique monotone equilibrium is the only equilibrium which survives the iterative elimination of interim strictly dominated strategies. For ease of exposition, the proof is broken down into separate lemmas.

(i) *Unique equilibrium in monotone strategies*

Regime problem. Let $\hat{x} \in \mathbb{R}$ denote a candidate for the citizens' threshold.

LEMMA 1. For each $\hat{x} \in \mathbb{R}$, the unique solution to the regime's decision problem is characterized by a pair of functions, $\Theta : \mathbb{R} \to [0, 1)$ and $A : \mathbb{R} \to \mathbb{R}_+$ such that if citizens subvert for all $x_i < \hat{x}$ then the best-response of the regime is to abandon if and only if its type is $\theta < \Theta(\hat{x})$ and to choose an action $a(\theta) = 0$ for $\theta < \Theta(\hat{x})$ and $a(\theta) = A(\theta - \hat{x})$ for $\theta \geq \Theta(\hat{x})$.

Proof of Lemma 1. To begin, let

$$S(w) := \Phi(-\sqrt{\alpha} w) \tag{33}$$

The auxiliary function $S(w)$ is exogenous and does not depend on $\hat{x}$. In terms of this function, the mass of subversives facing the regime is

$$\int_{-\infty}^{\hat{x}} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a))\, dx_i = \Phi(\sqrt{\alpha}(\hat{x} - \theta - a)) = S(\theta + a - \hat{x}) \tag{34}$$

Since the regime has access to an outside option normalized to zero, its problem can be written

$$V(\theta, \hat{x}) := \max[0, W(\theta, \hat{x})] \tag{35}$$

where $W(\theta, \hat{x})$ is the best payoff regime θ can get if it is not overthrown

$$W(\theta, \hat{x}) := \max_{a \geq 0} \left[\theta - S(\theta + a - \hat{x}) - C(a)\right] \tag{36}$$

From the envelope theorem, the partial derivative $W_\theta(\theta, \hat{x}) = 1 - S'(\theta - \hat{x} + a) > 1$ since $S'(w) < 0$ for all $w \in \mathbb{R}$. Since $S(w) \geq 0$ and $C(a) \geq 0$ we know $W(\theta, \hat{x}) < 0$ for all $\theta < 0$ and all $\hat{x}$. Similarly, $W(1, \hat{x}) > 0$ for all $\hat{x}$. So by the intermediate value theorem there is a unique $\Theta(\hat{x}) \in [0, 1)$ such that $W(\Theta(\hat{x}), \hat{x}) = 0$. And since $W_\theta(\theta, \hat{x}) > 1$ the regime is overthrown if and only if $\theta < \Theta(\hat{x})$. Since positive actions are costly, the regime takes no action for $\theta < \Theta(\hat{x})$. Otherwise, for $\theta \geq \Theta(\hat{x})$, the actions of the regime are given by

$$a(\theta) = A(\theta - \hat{x}) \tag{37}$$

where the function $A(t)$ is exogenous and does not depend on $\hat{x}$. This auxiliary function is defined by:

$$A(t) := \underset{a\geq 0}{\operatorname{argmin}}\,[S(t+a)+C(a)] \tag{38}$$

The first order necessary condition for interior solutions can be written $C'(a) = -S'(t+a)$, and, on using the formula for $S(\cdot)$ in equation (33) above, $C'(a) = \sqrt{\alpha}\phi(\sqrt{\alpha}(t+a))$ where $\phi(w) := \exp(-w^2/2)/\sqrt{2\pi}$ for all $w \in \mathbb{R}$. This first order condition may have zero, one or two solutions for each t. If for a given t there are zero (interior) solutions, then $A(t) = 0$. If for given t there are two solutions, one of them can be ruled out by the second order sufficient condition $\alpha\phi'(\sqrt{\alpha}(t+a)) + C''(a) > 0$. Using the property $\phi'(w) = -w\phi(w)$ for all $w \in \mathbb{R}$ shows that if there are two solutions to the first order condition, only the "higher" of them satisfies the second order condition. Therefore for each t there is a single $A(t)$ that solves the regime's problem. Making the substitution $t = \theta - \hat{x}$, the regime's threshold $\Theta(\hat{x})$ is then found from the indifference condition $W(\Theta(\hat{x}), \hat{x}) = 0$, or more explicitly

$$\Theta(\hat{x}) = S(\Theta(\hat{x}) - \hat{x} + A(\Theta(\hat{x}) - \hat{x})) + C(A(\Theta(\hat{x}) - \hat{x})) \tag{39}$$

Taking $\hat{x}$ as given, equations (38) and (39) give the regime threshold $\Theta(\hat{x})$ and the hidden actions $a(\theta) = A(\theta - \hat{x})$ that solve the regime's problem. □

Citizen problem. Let $\hat{\theta} \in [0,1)$ and $a : \mathbb{R} \to \mathbb{R}_+$ denote, respectively, a candidate for the regime's threshold and a candidate for the regime's hidden actions with $a(\theta) = 0$ for $\theta < \hat{\theta}$.

Lemma 2. For each $\hat{\theta} \in [0,1)$ and $a : \mathbb{R} \to \mathbb{R}_+$

(a) The unique solution to the problem of a citizen with signal x_i is given by a mapping $P(\cdot\,|\,a(\cdot)) : \mathbb{R}\times\mathbb{R} \to [0,1]$ such that the citizen subverts if and only if

$$P(x_i, \hat{\theta}\,|\,a(\cdot)) := \text{Prob}[\theta < \hat{\theta} \mid x_i, a(\cdot)] \geq p \tag{40}$$

where P is continuous and strictly decreasing in x_i with limits $P(-\infty, \hat{\theta}\,|\,a(\cdot)) = 1$ and $P(+\infty, \hat{\theta}\,|\,a(\cdot))) = 0$ for any $\hat{\theta}$ and function $a(\cdot)$ satisfying $a(\theta) = 0$ for $\theta < \hat{\theta}$.

(b) For any candidate citizen threshold $\hat{x}$, with implied regime threshold $\Theta(\hat{x})$ and hidden actions $A(\theta-\hat{x})$, an individual citizen with signal x_i subverts if and only if its signal is such that

$$K(x_i, \hat{x}) := \text{Prob}[\theta < \Theta(\hat{x}) \mid x_i, A(\cdot)] \geq p \tag{41}$$

where $K : \mathbb{R}\times\mathbb{R} \to [0,1]$ is continuous, strictly increasing in x_i with limits $K(-\infty, \hat{x}) = 0$ and $K(+\infty, \hat{x}) = 1$ for any $\hat{x}$. Moreover, $K(x_i, \hat{x}) = \text{Prob}[\theta < \Theta(\hat{x}) - \hat{x} \mid x_i - \hat{x}, A(\cdot)]$ for any $\hat{x}$.

Proof of Lemma 2. (a) For notational simplicity, write x for an individual's signal, θ for the state threshold, and $P(x,\theta)$ for the probability an individual with x assigns to the regime's type being less than θ when the actions are $a : \mathbb{R} \to \mathbb{R}_+$. That is,

$$P(x,\theta) = \frac{\int_{-\infty}^{\theta} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - t))\,dt}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - t - a(t))]\,dt} \tag{42}$$

where the numerator uses $a(t) = 0$ for $t < \theta$. Hence $P : \mathbb{R}\times\mathbb{R} \to [0,1]$ is continuous in x, θ. This probability can be written

$$P(x,\theta) = \frac{N(\theta - x)}{N(\theta - x) + D(x,\theta)} \tag{43}$$

where

$$N(\theta - x) := \Phi(\sqrt{\alpha}(\theta - x)), \quad \text{and} \quad D(x,\theta) := \int_{\theta}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x - \xi - a(\xi)))\,d\xi \tag{44}$$

Differentiating (43) shows $P_x < 0$ if and only if $N'/N > -D_x/D$. Calculating the derivatives shows that this is equivalent to

$$H(\sqrt{\alpha}(x-\theta)) > -\frac{\int_{\theta}^{\infty} \phi'(\sqrt{\alpha}(x - y(\xi)))\,d\xi}{\int_{\theta}^{\infty} \phi(\sqrt{\alpha}(x - y(\xi)))\,d\xi} = \frac{\int_{\theta}^{\infty} \sqrt{\alpha}(x - y(\xi))\phi(\sqrt{\alpha}(x - y(\xi)))\,d\xi}{\int_{\theta}^{\infty} \phi(\sqrt{\alpha}(x - y(\xi)))\,d\xi} \tag{45}$$

where $H(w) := \phi(w)/(1-\Phi(w)) > 0$ denotes the standard normal *hazard function* for $w \in \mathbb{R}$, where $y(\xi) := \xi + a(\xi)$ is the mean of the signal distribution if $\xi \geq \theta$, and where the equality follows from $\phi'(w) = -w\phi(w)$ for all w. Now define a density $\varphi(\xi \,|\, x) > 0$ by

$$\varphi(\xi \,|\, x) := \frac{\phi(\sqrt{\alpha}(x - y(\xi)))}{\int_\theta^\infty \phi(\sqrt{\alpha}(x - y(\xi'))) \, d\xi'}, \qquad \xi \in [\theta, \infty) \tag{46}$$

Then after a slight rearrangement of terms in (45), $P_x < 0$ if and only if

$$H(\sqrt{\alpha}(x-\theta)) - \sqrt{\alpha}(x-\theta) > \sqrt{\alpha}\left[\theta - \int_\theta^\infty y(\xi)\varphi(\xi \,|\, x)\, d\xi\right] \tag{47}$$

Since the hazard function satisfies $H(w) > w$ for all $w \in \mathbb{R}$ and $\alpha > 0$, it is sufficient that

$$\int_\theta^\infty y(\xi)\varphi(\xi \,|\, x)\, d\xi \geq \theta \tag{48}$$

But since $y(\xi) := \xi + a(\xi)$, $\xi \geq \theta$, and $a(\xi) \geq 0$, condition (48) is always satisfied. Therefore $P_x < 0$. Since $N' > 0$ and $D_\theta < 0$, $P_\theta > 0$ for all x, θ. Moreover, since $N(-\infty) = 0$ and $D > 0$ we have $P(x, -\infty) = 0$ for all x. Similarly, since $a(\xi) = 0$ for all $\xi < \theta$ as $\theta \to \infty$ we have $D(x,\theta) \to 1 - N(\theta - x)$ and since $N(+\infty) = 1$ this means $D(x, +\infty) = 0$ for all x. Therefore $P(x, +\infty) = 1$ for all x. The limit properties in x are established in parallel fashion.

(b) Fix a $\hat{x} \in \mathbb{R}$ and let $A(\theta - \hat{x})$ denote the associated hidden actions. Analogous to (43), write $P(x, \theta, \hat{x}) = N(\theta - x)/[N(\theta - x) + D(x, \theta, \hat{x})$ where $N : \mathbb{R} \to [0,1]$ is defined as in (44) above and where

$$D(x, \theta, \hat{x}) := \int_\theta^\infty \sqrt{\alpha}\phi(\sqrt{\alpha}(x - t - A(t - \hat{x})))\, dt \tag{49}$$

Now define $K(x, \hat{x}) := P(x, \Theta(\hat{x}), \hat{x})$. That $K(x, \hat{x})$ is continuous and decreasing in x is immediate from part (a) above. Finally, for $K(x, \hat{x}) = P(x - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0)$ it is sufficient that $D(x, \Theta(\hat{x}), \hat{x}) = D(x - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0)$. From (49) and using the change of variables $\xi := \theta - \hat{x}$ we have

$$D(x, \Theta(\hat{x}), \hat{x}) = \int_{\Theta(\hat{x}) - \hat{x}}^\infty \sqrt{\alpha}\phi(\sqrt{\alpha}(x - \hat{x} - \xi - a(\xi)))\, d\xi = D(x - \hat{x}, \Theta(\hat{x}) - \hat{x}, , 0) \tag{50}$$

Therefore $K(x, \hat{x}) = P(x - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0) = \text{Prob}[\theta < \Theta(\hat{x}) - \hat{x} \mid x - \hat{x}, A(\cdot)]$ as claimed. □

Fixed point. A citizen with signal x_i will subvert the regime if and only if $K(x_i, \hat{x}) \geq p$. Since $K(x_i, \hat{x})$ is strictly increasing in x_i with $K(-\infty, \hat{x}) < p$ and $K(+\infty, \hat{x}) > p$ for any $\hat{x} \in \mathbb{R}$, there is a unique signal $\psi(\hat{x})$ solving

$$K(\psi(\hat{x}), \hat{x}) = p \tag{51}$$

such that a citizen with signal x_i subverts if and only if $x_i < \psi(\hat{x})$.

Lemma 3. The function $\psi : \mathbb{R} \to \mathbb{R}$ is continuous and has a unique fixed point $x^* = \psi(x^*)$ with derivative $\psi'(x^*) \in (0,1)$ at the fixed point. Moreover $\psi(x) \leq x^*$ for all $x < x^*$ and $\psi(x) \geq x^*$ for all $x > x^*$.

Proof of Lemma 3. Since $K(x, \hat{x})$ is continuously differentiable in x, an application of the implicit function theorem to (51) shows that $\psi(\cdot)$ is continuous. Fixed points of $\psi(\cdot)$ satisfy $x^* = \psi(x^*)$. Equivalently, by part (b) of Lemma 2, they satisfy $K(x^*, x^*) = P(0, \Theta(x^*) - x^*, 0) = p$, where $\Theta(\hat{x})$ is the critical state in the regime's problem (38)-(39). By Lemma 2 and the intermediate value theorem there is a unique $z^* \in \mathbb{R}$ such that $P(0, z^*, 0) = p$. Then applying the implicit function theorem to (38)-(39) gives

$$\Theta'(\hat{x}) = \frac{\sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x})))}{1 + \sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x})))} \in (0,1) \tag{52}$$

Since $\Theta(-\infty) = 0$ and $\Theta(+\infty) = 1$, there is a unique $x^* \in \mathbb{R}$ such that $\Theta(x^*) - x^* = z^*$, hence $\psi(\cdot)$ has a unique fixed point, the same x^*. Now using part (b) of Lemma 2 and implicitly differentiating (51) we have

$$\psi'(\hat{x}) = 1 + \frac{P_\theta(\psi(\hat{x}) - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0)}{P_x(\psi(\hat{x}) - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0)}(1 - \Theta'(\hat{x})) \tag{53}$$

By Lemma 2, $P_\theta > 0$ and $P_x < 0$ and $\Theta'(\hat{x}) \in (0,1)$ from (52). Therefore $\psi'(\hat{x}) < 1$ for all $\hat{x}$. To see that $\psi'(x^*) > 0$, first notice that it is sufficient that $P_\theta/P_x \geq -1$ when evaluated at $\hat{x} = x^*$. Calculating the derivatives shows that this is true if and only if

$$\phi(\sqrt{\alpha}(y(\theta^*) - x^*)) + \int_{\theta^*}^{\infty} \sqrt{\alpha}\phi'(\sqrt{\alpha}(y(\theta) - x^*))\, d\theta \leq 0 \tag{54}$$

where $\theta^* := \Theta(x^*)$ and where $y(\theta) = \theta + a(\theta)$ is the mean of the signal distribution from which a citizen is sampling if the regime has type $\theta \geq \theta^*$. To show that this condition always holds, we need to consider the cases of linear costs and strictly convex costs separately. If costs are linear, $C(a) = ca$, then if $c \geq \bar{c} := \sqrt{\alpha}\phi(0)$ the result is trivial because $a(\theta) = 0$ for all $\theta \in \mathbb{R}$. So suppose $c < \bar{c}$. Then $a(\theta) = \max[0, x^* + \gamma - \theta]$ where $\gamma := \sqrt{2\log\left(\sqrt{\alpha}\phi(0)/c\right)/\alpha} > 0$. Calculating the integral and then simplifying shows that (54) holds if and only if $-\alpha\gamma\phi(\sqrt{\alpha}\gamma)a(\theta^*) \leq 0$ which is true because $a(\theta^*) \geq 0$. If costs are strictly convex, then from the optimality conditions for the regime's choice of action we have that $a(\theta) > 0$ for all $\theta \geq \theta^*$ and

$$\sqrt{\alpha}\phi(\sqrt{\alpha}(y(\theta) - x^*)) = C'(a(\theta)), \quad \theta \geq \theta^* \tag{55}$$

Differentiating with respect to θ gives

$$\alpha\phi'(\sqrt{\alpha}(y(\theta) - x^*))y'(\theta) = C''(a(\theta))a'(\theta), \quad \theta \geq \theta^* \tag{56}$$

Using the associated second order condition shows that $y'(\theta) > 0$ for $\theta \geq \theta^*$. Since $y(\cdot)$ is invertible, a change of variables shows that (54) holds if and only if

$$\int_{\theta^*}^{\infty} \phi'(\sqrt{\alpha}(y(\theta) - x^*))\frac{a'(\theta)}{y'(\theta)}\, d\theta \geq 0 \tag{57}$$

Using (56) we equivalently have the condition

$$\int_{\theta^*}^{\infty} \frac{\phi'(\sqrt{\alpha}(y(\theta) - x^*))^2}{C''(a(\theta))}\, d\theta \geq 0 \tag{58}$$

which is true since the integrand is non-negative. Therefore, $P_\theta/P_x \geq -1$ at $\hat{x} = x^*$ and $\psi'(x^*) > 0$.

Finally, $\psi(\hat{x}) \leq x^*$ for every $\hat{x} < x^*$ is proven by contradiction. Suppose not. Then by continuity of ψ there exists $\tilde{x} < x^*$ such that $\psi(\tilde{x}) = x^*$. Moreover, since $\psi'(x^*) > 0$, we must have $\psi'(\tilde{x}) < 0$ for at least one such $\tilde{x}$. Since $\psi(\tilde{x}) = x^*$ and $K(x^*, x^*) = p$, under this hypothesis we can write $K(\psi(\tilde{x}), \psi(\tilde{x})) = p$ so by the implicit function theorem $\psi(\tilde{x})$ must satisfy

$$\psi'(\tilde{x})[K_1(x^*, x^*) + K_2(x^*, x^*)] = 0 \tag{59}$$

where the hypothesis $\psi(\tilde{x}) = x^*$ is used to evaluate the partial derivatives K_1 and K_2. Since $\psi'(\tilde{x}) < 0$, this can only be satisfied if $K_1(x^*, x^*) + K_2(x^*, x^*) = 0$. But for any $\hat{x} \in \mathbb{R}$, the value $\psi(\hat{x})$ is implicitly defined by $K(\psi(\hat{x}), \hat{x}) = p$ so that by the implicit function theorem $\psi'(\hat{x}) = -K_2(\psi(\hat{x}), \hat{x})/K_1(\psi(\hat{x}), \hat{x})$. From (53) we know $\psi'(\hat{x}) < 1$ for any $\hat{x}$ and since $K_1 < 0$ from Lemma 2 we conclude $K_1(\psi(\hat{x}), \hat{x}) + K_2(\psi(\hat{x}), \hat{x}) < 0$ for any $\hat{x}$. For $\hat{x} = x^*$ in particular, $K_1(x^*, x^*) + K_2(x^*, x^*) < 0$ so we have the needed contradiction. Therefore $\psi(\hat{x}) \leq x^*$ for every $\hat{x} < x^*$. A symmetric argument shows $\psi(\hat{x}) \geq x^*$ for every $\hat{x} > x^*$. □

Concluding that there is a unique equilibrium in monotone strategies. To conclude part (i) of the proof, we take an arbitrary $\hat{x} \in \mathbb{R}$ and solve the regime's problem to get $\Theta(\hat{x})$ and $a(\theta, \hat{x}) = A(\theta - \hat{x})$ using the auxiliary function from Lemma 1. We use these functions to construct $K(x_i, \hat{x})$ from (41) for each signal $x_i \in \mathbb{R}$ and use Lemma 2 to conclude that in particular $K(\hat{x}, \hat{x}) = P(0, \Theta(\hat{x}) - \hat{x}, 0)$. We then use the intermediate value theorem to deduce that there is a unique $z^* \in \mathbb{R}$ such that $P(0, z^*, 0) = p$. This gives a unique difference $z^* = \theta^* - x^*$ that can be plugged into the regime's indifference condition (39) to get the unique $\theta^* = \Theta(x^*) \in [0, 1)$ such that the regime is overthrown if and only if $\theta < \theta^*$. The unique signal threshold is then $x^* = \theta^* - z^*$ and the unique hidden action function is given by $a(\theta) := A(\theta - x^*)$.

(ii) *Iterative elimination of interim strictly dominated strategies*

We can now go on to show that there is no other equilibrium. The argument begins by showing that for sufficiently low signals it is a dominant strategy to subvert the regime and for sufficiently high signals it is a dominant strategy to not subvert the regime.

Dominance regions. If the regime has $\theta < 0$, any mass $S \geq 0$ can overthrow the regime. Similarly, if the regime has $\theta \geq 1$ it can never be overthrown. Any regime that is overthrown takes no action, since to do so would incur a cost for no gain. Similarly, any regime θ that is not overthrown takes an action no larger than the a such that $\theta = C(a)$. Any larger action must result in a negative payoff which can be improved upon by taking the outside option. Given this:

LEMMA 4. There exists a pair of signals $\underline{x} < \overline{x}$, both finite, such that $s(x_i) = 1$ is strictly dominant for $x_i < \underline{x}$ and $s(x_i) = 0$ is strictly dominant for $x_i > \overline{x}$.

Proof of Lemma 4. The most *pessimistic* scenario for any citizen is that regimes are overthrown only if $\theta < 0$ and that regimes take the largest hidden actions that could be rational $\underline{a}(\theta) := C^{-1}(\theta)$ for $\theta \geq 0$ and zero otherwise. Let $\underline{P}(x_i) := \text{Prob}[\theta < 0 \mid x_i,\ \underline{a}(\cdot)]$ denote the probability the regime is overthrown in this most pessimistic scenario. Part (a) of Lemma 2 holds for hidden actions of the form $\underline{a}(\theta)$ and implies $\underline{P}'(x_i) < 0$ for all x_i, and since $\underline{P}(-\infty) = 1$ and $\underline{P}(+\infty) = 0$ by the intermediate value theorem there is a unique value, $\underline{x}$, finite, such that $\underline{P}(\underline{x}) = p$. For $x_i < \underline{x}$ it is (iteratively) strictly dominant for $s(x_i) = 1$. Similarly, the most *optimistic* scenario for any citizen is that regimes are overthrown if $\theta < 1$ and that regimes take the smallest hidden actions that could be rational $\overline{a}(\theta) := 0$. Let $\overline{P}(x_i) := \text{Prob}[\theta < 1 \mid x_i,\ \overline{a}(\cdot)]$ denote the probability the regime is overthrown in this most optimistic scenario. A parallel argument establishes the existence of a unique value, $\overline{x}$, finite, such that $\overline{P}(\overline{x}) = p$. For $x_i > \overline{x}$ it is (iteratively) strictly dominant for $s(x_i) = 0$. ☐

Iterative elimination. Starting from the dominance regions implied by $\underline{x}$ and $\overline{x}$ it is then possible to iteratively eliminate (interim) strictly dominated strategies. Recall that $S(w) := \Phi(-\sqrt{\alpha}w)$ and $A(t) := \text{argmin}_{a\geq 0}\,[S(t+a) + C(a)]$. Again, these auxiliary functions do not depend on any endogenous variable and in particular do not depend on citizen thresholds.

LEMMA 5. Let $x_{n+1} = \psi(x_n)$ for $n = 0, 1, 2, ...$ where

$$K(\psi(x_n), x_n) = p$$

(a) If it is strictly dominant for $s(x_i) = 1$ for all $x_i < \underline{x}_n$, then the regime is overthrown for at least all $\theta < \underline{\theta}_n := \Theta(\underline{x}_n)$ where the function $\Theta : \mathbb{R} \to [0, 1)$ solves

$$\Theta(x) = S(\Theta(x) - x + A(\Theta(x) - x)) + C((\Theta(x) - x)) \tag{60}$$

Similarly, if it is strictly dominant for $s(x_i) = 0$ for all $x_i > \overline{x}_n$, then regime is not overthrown for at least all $\theta > \overline{\theta}_n := \Theta(\overline{x}_n)$.

(b) Moreover, if it is strictly dominant for $s(x_i) = 1$ for all $x_i < \underline{x}_n$, then it is strictly dominant for $s(x_i) = 1$ for all $x_i < \underline{x}_{n+1} = \psi(\underline{x}_n)$. Similarly, if it is strictly dominant for $s(x_i) = 0$ for all $x_i > \overline{x}_n$, then it is strictly dominant for $s(x_i) = 0$ for all $x_i > \overline{x}_{n+1} = \psi(\overline{x}_n)$.

Proof of Lemma 5. (a) Fix an $\underline{x}_n$ and $\overline{x}_n$ such that citizens with signals $x_i < \underline{x}_n$ have $s(x_i) = 1$ and likewise citizens with signals $x_i > \overline{x}_n$ have $s(x_i) = 0$. From Lemma 4 this can be done at least for the signals $\underline{x}, \overline{x}$ that determine the bounds of the dominance regions. All citizens with signals $x_i < \underline{x}_n$ have $s(x_i) = 1$ so the mass of subversives is at least $\Phi(\sqrt{\alpha}(\underline{x}_n - \theta - a))$. To acknowledge this, write the total mass of subversives as

$$\Phi(\sqrt{\alpha}(\underline{x}_n - \theta - a)) + \Delta(\theta + a) \tag{61}$$

for some function $\Delta : \mathbb{R} \to [0, 1]$. First consider the case $\Delta(\cdot) = 0$ where *only* citizens with $x_i < \underline{x}_n$ subvert the regime. From Lemma 1 there is a unique threshold $\underline{\theta}_n := \Theta(\underline{x}_n) \in [0, 1)$ sustained by hidden actions $a(\theta) = A(\theta - \underline{x}_n)$ solving (38)-(39) such that the regime is overthrown if $\theta < \underline{\theta}_n = \Theta(\underline{x}_n)$. Now consider the case $\Delta(\cdot) > 0$ where *some* citizens with signals $x_i \geq \underline{x}_n$ also subvert the regime. The proof that the regime is overthrown for at least all $\theta < \Theta(\underline{x}_n)$ is by contradiction. Suppose that when $\Delta(\cdot) > 0$ regime change occurs for all $\theta < \tilde{\theta}_n$ for some $\tilde{\theta}_n \leq \Theta(\underline{x}_n)$. A marginal regime $\tilde{\theta}_n$ must be indifferent between being overthrown and taking the outside option, so this threshold satisfies $\tilde{\theta}_n = S(\tilde{\theta}_n + \tilde{a}_n - \underline{x}_n) + C(\tilde{a}_n)$ where $\tilde{a}_n \geq 0$ is the optimal action for the marginal regime $\tilde{\theta}_n$. Then observe

$$\begin{aligned}
\Theta(\underline{x}_n) &= \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - A(\Theta(\underline{x}_n) - \underline{x}_n))] + C[A(\Theta(\underline{x}_n) - \underline{x}_n))] \\
&\leq \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - a)] + C(a), \\
&< \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - a)] + \Delta(\tilde{\theta}_n + \tilde{a}_n) + C(a), \qquad \text{for any } a \geq 0
\end{aligned}$$

where the first inequality follows because $A(\cdot)$ minimizes $\Phi[\sqrt{\alpha}(\underline{x}_n - \theta - a)] + C(a)$ and where the second inequality follows from $\Delta(\cdot) > 0$. Taking $a = \tilde{a}_n \geq 0$ we then have

$$\begin{aligned} \Theta(\underline{x}_n) &< \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - \tilde{a}_n)] + \Delta(\tilde{\theta}_n + \tilde{a}_n) + C(\tilde{a}_n) \\ &= \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - \tilde{a}_n)] + \Delta(\tilde{\theta}_n + \tilde{a}_n) + C(\tilde{a}_n) \\ &\quad + \Phi[\sqrt{\alpha}(\underline{x}_n - \tilde{\theta}_n - \tilde{a}_n)] - \Phi[\sqrt{\alpha}(\underline{x}_n - \tilde{\theta}_n - \tilde{a}_n)] \\ &= \tilde{\theta}_n + \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - \tilde{a}_n)] - \Phi[\sqrt{\alpha}(\underline{x}_n - \tilde{\theta}_n - \tilde{a}_n)] \\ &\leq \tilde{\theta}_n \end{aligned}$$

where the last inequality follows because the hypothesis $\tilde{\theta}_n \leq \Theta(\underline{x}_n)$ implies $\Phi[\sqrt{\alpha}(\underline{x}_n - \tilde{\theta}_n - \tilde{a}_n)] \geq \Phi[\sqrt{\alpha}(\underline{x}_n - \Theta(\underline{x}_n) - \tilde{a}_n)]$. This is a contradiction, and so $\tilde{\theta}_n > \Theta(\underline{x}_n)$. Therefore, the regime is overthrown for at least all $\theta < \Theta(\underline{x}_n)$. A parallel argument shows that if it is strictly dominant for $s(x) = 0$ for all $x_i > \overline{x}_n$, then the regime is not overthrown for at least all $\theta > \overline{\theta}_n := \Theta(\overline{x}_n)$.

(b) Since cumulative distribution functions are non-decreasing, for any beliefs of the citizens, the posterior probability assigned by a citizen with signal x_i to the regime's overthrow is at least as much as the probability they assign to $\theta < \Theta(\underline{x}_n)$. Equivalently, $K(x_i, \underline{x}_n) - p$ is the most conservative estimate of the expected gain to subverting. From Lemma 2 and the intermediate value theorem, there is a unique $\underline{x}_{n+1} = \psi(\underline{x}_n)$ solving $K(\psi(\underline{x}_n), \underline{x}_n) = p$ such that if it is strictly dominant for $s(x_i) = 1$ for all $x_i < \underline{x}_n$, then it is strictly dominant for $s(x_i) = 1$ for all $x_i < \underline{x}_n$. Similarly, there is a unique $\overline{x}_{n+1} = \psi(\overline{x}_n)$ solving $K(\psi(\overline{x}_n), \overline{x}_n) = p$ such that if it is strictly dominant for $s(x_i) = 0$ for all $x_i > \overline{x}_n$, then it is strictly dominant for $s(x_i) = 0$ for all $x_i > \overline{x}_{n+1}$. Applying the proof of part (a) at each step then completes the argument. □

Concluding that there is no other equilibrium. Let $\underline{x}_0 := \underline{x}$ and $\overline{x}_0 := \overline{x}$ and generate sequences $\{\underline{x}_n\}_{n=0}^{\infty}$ from $\underline{x}_{n+1} = \psi(\underline{x}_n)$ and $\{\overline{x}_n\}_{n=0}^{\infty}$ from $\overline{x}_{n+1} = \psi(\overline{x}_n)$ where

$$K(\psi(\overline{x}_k), \overline{x}_k) = p \tag{62}$$

and

$$K(\psi(\underline{x}_k), \underline{x}_k) = p \tag{63}$$

Part (a) of Lemma 5 maps the sequences of citizen thresholds $\{\underline{x}_n\}_{n=0}^{\infty}$ and $\{\overline{x}_n\}_{n=0}^{\infty}$ into *monotone* sequences of regime thresholds, $\{\underline{\theta}_n\}_{n=0}^{\infty}$ from $\underline{\theta}_n := \Theta(\underline{x}_n)$ and $\{\overline{\theta}_n\}_{n=0}^{\infty}$ from $\overline{\theta}_n := \Theta(\overline{x}_n)$. Moreover, by Lemma 3 the function $\psi(\cdot)$ generating the sequences $x_{n+1} = \psi(x_n)$ is continuous, has a unique fixed point $x^* = \psi(x^*)$ with derivative $\psi'(x^*) \in (0,1)$ at this fixed point and upper bound $\psi(\underline{x}_n) \leq x^*$ for all $\underline{x}_n < x^*$. From below, the sequence $\{\underline{x}_n\}_{n=0}^{\infty}$ is bounded above, strictly monotone increasing and so converges $\underline{x}_n \nearrow x^*$ as $n \to \infty$. Similarly the sequence $\{\underline{\theta}_n\}_{n=0}^{\infty}$ is bounded above, strictly monotone increasing and so converges $\underline{\theta}_n \nearrow \theta^* =: \Theta(x^*)$ as $n \to \infty$. From above, symmetrically, the sequence $\{\overline{x}_n\}_{n=0}^{\infty}$ is bounded below, strictly monotone decreasing and so converges $\overline{x}_n \searrow x^*$ as $n \to \infty$. Similarly the sequence $\{\overline{\theta}_n\}_{n=0}^{\infty}$ is bounded below, strictly monotone decreasing and so converges $\overline{\theta}_n \searrow \theta^* =: \Theta(x^*)$ as $n \to \infty$. After a finite n iterations, the only candidates for a citizen's equilibrium strategy all have $s(x_i) = 1$ for $x_i < \underline{x}_n$ and $s(x_i) = 0$ for $x_i > \overline{x}_n$ with $s(x_i)$ arbitrary for $x_i \in [\underline{x}_n, \overline{x}_n]$. Similarly, the only candidate for the regime's strategy has the regime abandoning for all $\theta < \underline{\theta}_n$, not abandoning for $\theta \geq \underline{\theta}_n$ with arbitrary choices for $\theta \in [\underline{\theta}_n, \overline{\theta}_n]$. At each iteration, these regime thresholds are implicitly determined by hidden actions $\underline{a}_n(\theta) := A(\theta - \underline{x}_n)$ and $\overline{a}_n(\theta) := A(\theta - \overline{x}_n)$ respectively. In the limit as $n \to \infty$, the only strategy that survives the elimination of strictly dominated strategies is the one with $s(x_i) = 1$ for $x_i < x^*$ and $s(x_i) = 0$ otherwise for citizens, with the regime abandoning for $\theta < \theta^* = \Theta(x^*)$ and hidden actions given by $a(\theta) = A(\theta - x^*)$. Therefore the only equilibrium is the unique monotone equilibrium. ■

A.3 Proof of Proposition 2

Since the regime benefit function $B(\theta, S)$ satisfies the Spence-Mirrlees sorting condition we have, for any $\theta \geq \theta'$ and any $S \geq S'$, that

$$B(\theta, S') - B(\theta, S) \geq B(\theta', S') - B(\theta', S) \tag{64}$$

That is, stronger regimes benefit at least weakly more from a smaller aggregate attack. The proof that stronger regimes choose higher apparent strengths is by contradiction.[12] Suppose that regime $\theta_H \geq \theta_L$

[12] I thank an anonymous referee for suggesting this proof.

chooses apparent strength $y_H = y(\theta_H) < y_L = y(\theta_L)$. Then the weaker regime must be paying a higher cost

$$C(y_L - \theta_L) > C(y_H - \theta_L) \geq C(y_H - \theta_H) \tag{65}$$

Moreover, since y_L is optimal for θ_L

$$B(\theta_L, S(y_L)) - B(\theta_L, S(y_H)) \geq C(y_L - \theta_L) - C(y_H - \theta L) > 0 \tag{66}$$

That is, since the cost of choosing y_L is greater, the benefit must be greater too. But then since $B(\theta, S)$ is strictly decreasing in S it must be that $S_L = S(y_L) < S_H = S(y_H)$. In short, a higher apparent strength is chosen only if it induces a smaller aggregate attack. But since $\theta_H \geq \theta_L$ and $S_H \geq S_L$, from the sorting condition (64), we then have

$$B(\theta_H, S_L) - B(\theta_H, S_H) \geq B(\theta_L, S_L) - B(\theta_L, S_H) > 0 \tag{67}$$

Since higher types can achieve a desired apparent strength at lower cost, i.e., $C(y_L - \theta_H) < C(y_H - \theta_H)$, (67) can only be true if

$$B(\theta_H, S_L) - C(y_H - \theta_H) > B(\theta_H, S_H) - C(y_H - \theta_H) \tag{68}$$

But this contradicts the optimality of y_H for θ_H. Hence $y_H \geq y_L$, and hence $y(\theta)$ is increasing in θ. ∎

A.4 Proofs of Proposition 3, Proposition 5, and Proposition 6

Preliminaries

The proofs of Propositions 3, 5 and 6 are similar. To begin, substitute the regime indifference condition (22) into the citizen indifference condition (21) to obtain

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p}\theta^* \qquad \Leftrightarrow \qquad \theta^* - x^* = \frac{1}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) \tag{69}$$

And now substitute this expression for the difference $\theta^* - x^*$ back into the regime indifference condition (22) to get a single equation characterizing the critical regime threshold θ^*, namely

$$\theta^* + \frac{c}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = \sqrt{\alpha}\gamma\phi(\sqrt{\alpha}\gamma) + \Phi(-\sqrt{\alpha}\gamma) \tag{70}$$

(using the fact that γ is implicitly defined by $c = \sqrt{\alpha}\phi(\sqrt{\alpha}\gamma)$). Now define the composite parameter $z := c/\sqrt{\alpha} > 0$ and observe that in terms of this parameter

$$\sqrt{\alpha}\gamma = \sqrt{2\log\left(\frac{\phi(0)}{z}\right)} =: \delta(z)$$

so that (70) can be written in terms of the composite parameter z alone, namely

$$T(\theta^*) := \theta^* + z\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = \delta(z)\phi(\delta(z)) + \Phi(-\delta(z)) \tag{71}$$

All of the results in Propositions 3, 5 and 6 follow straightforwardly from the comparative statics of θ^* with respect to z. Implicitly differentiating with respect to z gives

$$T'(\theta^*)\frac{\partial\theta^*}{\partial z} + \Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = \delta(z)\phi'(\delta(z))\delta'(z) + \delta'(z)\phi(z) - \phi(-\delta(z))\delta'(z) \tag{72}$$

The right hand side can be simplified using symmetry, $\phi(-\delta(z)) = \phi(\delta(z))$, and the property $\phi(\delta(z)) = z$ so that $\phi'(\delta(z))\delta'(z) = 1$. Thus

$$T'(\theta^*)\frac{\partial\theta^*}{\partial z} + \Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = \delta(z) \tag{73}$$

Then because $T'(\theta) > 0$ for all θ we have that

$$\frac{\partial}{\partial z}\theta^* > 0 \quad \Leftrightarrow \quad \theta^* < \theta_{\text{crit}} := \frac{1-p}{p}\Phi(\delta(z)) \tag{74}$$

And because $T'(\theta) > 0$ for all θ we have $\theta^* < \theta_{\text{crit}}$ if and only if $T(\theta^*) < T(\theta_{\text{crit}})$. Applying $T(\cdot)$ to both sides of equation (74) and simplifying we have that the regime threshold θ^* is increasing in z if and only if

$$p < \Phi(\delta(z)) \tag{75}$$

Since $\delta(z) > 0$ and $\Phi^{-1}(p) < 0$ for any $p < 1/2$, this condition is necessarily satisfied if $p < 1/2$. Using the definition of $\delta(z)$ and rearranging then gives

$$\frac{\partial}{\partial z}\theta^* > 0 \quad \Leftrightarrow \quad z < z^* := \phi(0)\exp\left(-\frac{1}{2}\max\left[0, \Phi^{-1}(p)\right]^2\right) \tag{76}$$

This condition is the key to the proofs of each of Propositions 3, 5 and 6 below.

Proof of Proposition 3

If $\alpha \leq \underline{\alpha}(c) := (c/\phi(0))^2$, any regime is at a corner solution and has hidden actions $a(\theta) = 0$. In this case the regime threshold is the same as in the Morris-Shin benchmark economy, $\theta^* = 1 - p$ for all $\alpha \leq \underline{\alpha}$. Otherwise, for interior solutions, the comparative statics are found by using the definition of the composite parameter $z = c/\sqrt{\alpha}$ and the chain rule

$$\frac{\partial}{\partial \alpha}\theta^* = -\frac{1}{2}\frac{c}{\alpha\sqrt{\alpha}}\frac{\partial}{\partial z}\theta^* \tag{77}$$

Hence

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \quad \Leftrightarrow \quad \frac{\partial}{\partial z}\theta^* > 0 \quad \Leftrightarrow \quad z < z^* \tag{78}$$

And then on using the definition of z and rearranging

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \quad \Leftrightarrow \quad \alpha > \alpha^*(c,p) := \underline{\alpha}(c)\exp\left(\max\left[0, \Phi^{-1}(p)\right]^2\right) \tag{79}$$

Now to establish that $\lim_{\alpha\to\infty}\theta^* = 0$, observe that for any $w \in \mathbb{R}$ the cumulative density $\Phi(\sqrt{\alpha}w) \to \mathbb{1}\{w > 0\}$ as $\alpha \to \infty$, i.e., to the indicator function that equals one if $w > 0$ and zero otherwise. Moreover, as $\alpha \to \infty$ the coefficient $\gamma = \sqrt{\log(\alpha/\underline{\alpha})/\alpha} \to 0$ and $\Phi(-\sqrt{\alpha}\gamma) \to 0$. Applying these to (21) we see that for large α solutions to the citizen's indifference condition are approximately the same as solutions to

$$\mathbb{1}\{\theta^* - x^* > 0\} = -\frac{p}{1-p}(\theta^* - x^*)c \tag{80}$$

The only solution to (80) is $\theta^* - x^* = 0$. So as $\alpha \to \infty$, solutions to (21) approach zero too. Then from the regime's indifference condition (22), if $\theta^* - x^* \to 0$ it must also be the case that $\theta^* \to 0$ as claimed. ■

Proof of Proposition 5

Similarly from the chain rule

$$\frac{\partial}{\partial c}\theta^* = \frac{1}{\sqrt{\alpha}}\frac{\partial}{\partial z}\theta^* \tag{81}$$

Hence eliminating the derivative with respect z between (77) and (81) we have

$$\frac{\partial}{\partial c}\theta^* = -2\frac{\alpha}{c}\frac{\partial}{\partial z}\theta^* \tag{82}$$

Multiplying both sides by $c/\theta^* > 0$ then gives the stated relationship between elasticities (28). Thus the cost effect is twice as large in magnitude as the signal precision effect. Observe that for corner solutions, $\alpha \leq \underline{\alpha}(c) := (c/\phi(0))^2$, both these effects are zero. Otherwise, if $\alpha > \underline{\alpha}(c)$ the signs of the two effects are determined by whether α is larger or smaller than $\alpha^*(c,p)$ but are always opposite to each other. ■

Proof of Proposition 6

Now let the composite parameter be

$$z = \frac{c(n)}{\sqrt{n\hat{\alpha}}}$$

Following calculations identical to (69)-(76) above, the equilibrium threshold θ^* depends on n and $\hat{\alpha}$ only through this composite parameter with the comparative statics with respect to $\hat{\alpha}$ being obtained exactly as in Proposition 3 above. Then using the chain rule

$$\frac{\partial \theta^*}{\partial n} = \frac{\partial \theta^*}{\partial z}\frac{\partial z}{\partial n}, \qquad \text{and} \qquad \frac{\partial \theta^*}{\partial \hat{\alpha}} = \frac{\partial \theta^*}{\partial z}\frac{\partial z}{\partial \hat{\alpha}} \tag{83}$$

Eliminating the derivative with respect to z then gives

$$\frac{\partial \theta^*}{\partial n} = \frac{\partial z}{\partial n}\left(\frac{\partial z}{\partial \hat{\alpha}}\right)^{-1}\frac{\partial \theta^*}{\partial \hat{\alpha}} \tag{84}$$

Calculating the derivative of z with respect to $\hat{\alpha}$ and then multiplying both sides by $n/\theta^* > 0$ then gives the relationship between elasticities

$$\frac{\partial \log \theta^*}{\partial \log n} = -2\frac{\partial \log z}{\partial \log n}\frac{\partial \log \theta^*}{\partial \log \hat{\alpha}} \tag{85}$$

And then using

$$\frac{\partial \log z}{\partial \log n} = \frac{c'(n)n}{c(n)} - \frac{1}{2}$$

gives the stated relationship between elasticities, equation (30). ■

A.5 Proof of Proposition 4

For each precision α, there is a unique equilibrium characterized by x^*, θ^* and $a(\theta)$. I calculate the asymptotic behavior of the equilibrium by a guess-and-verify method. In particular, I consider a constrained problem consisting of the original system of equations plus a set of auxiliary constraints. I show that this constrained problem has a unique solution for each α. Then because the equilibrium conditions have a unique solution for each α, the solution to the original problem and to the constrained problem coincide.

The equilibrium conditions can be written

$$(1-p)\Phi(\sqrt{\alpha}\,(\theta^* - x^*)) = p\int_{\theta^*}^{\infty}\sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - A(\theta - x^*)))\,d\theta \tag{86}$$

and

$$\theta^* = \Phi(\sqrt{\alpha}(x^* - \theta^* - A(\theta^* - x^*))) + C(A(\theta^* - x^*)) \tag{87}$$

with hidden actions solving

$$\sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - A(\theta - x^*))) = C'(A(\theta - x^*)), \qquad \theta \geq \theta^* \tag{88}$$

where $A(\cdot)$ is the function given in (38) above. The auxiliary constraints that govern the asymptotic behavior of the equilibrium are

$$\lim_{\alpha\to\infty}\sqrt{\alpha}(x^* - \theta^* - A(\theta^* - x^*)) = \lim_{\alpha\to\infty}\sqrt{\alpha}(\theta^* - x^*) = -\infty \tag{89}$$

If (89) holds, from (87) we have $\theta^* = C(A(\theta^* - x^*))$. Similarly, if (89) holds, then $\Phi(\sqrt{\alpha}(\theta^* - x^*)) \to 0$ and the value of the integral on the right hand side of (86) converges to zero. From (86) and (88), this requires

$$\lim_{\alpha\to\infty}\int_{\theta^*}^{\infty} C'(A(\theta - x^*))\,d\theta = 0$$

Since $\theta^* \in [0, 1)$ for any α and $C'(A(\theta - x^*)) \geq 0$ and is uniformly continuous in α for any θ, this can only be true if $A(\theta - x^*) \to 0^+$ for all $\theta \geq \theta^*$. But then if $A(\theta^* - x^*) \to 0^+$, $C(A(\theta^* - x^*)) \to 0^+$ and so $\theta^* \to 0^+$ too. Hence we have the result that for high enough signal precision, even the most fragile regime can survive.

For completeness, observe that if both constraints are to hold simultaneously for large α, $x^* - \theta^*$ is positive and $x^* - \theta^* - A(\theta^* - x^*)$ is negative. For both constraints to have the same sign, x^* can neither diverge nor converge to either a strictly positive or a strictly negative number. So $x^* \to 0^+$ too.

Now for the opposite limit as $\alpha \to 0^+$. Recall that for this part we assume strictly convex costs. Let $\alpha \to 0^+$ and guess that $\sqrt{\alpha}x^* \to \infty$ holds. Then $x^* \to \infty$. Since $\theta^* \in [0, 1)$, we have $\sqrt{\alpha}(x^* - \theta^*) \to \infty$ and the integral on the right hand side of (86) must converge to zero. Hence, by (88), $A(\theta - x^*) \to 0^+$ for all $\theta \geq \theta^*$ (the strict convexity of $C(\cdot)$ is assumed here so that (88) holds for all θ even as $\alpha \to 0^+$; with constant marginal costs, this would not be true). But if $A(\theta^* - x^*) \to 0^+$, $\theta^* \in [0, 1)$, and $\sqrt{\alpha}x^* \to \infty$, then (87) requires that $\theta^* \to 1^-$. ■

B Role of coordination

This appendix highlights the role of imperfect coordination in enabling the regime to survive even when signals are precise. Suppose to the contrary that citizens are perfectly coordinated and receive one $x = \theta + a + \varepsilon$. Collectively, they can overthrow the regime if $\theta < 1$. In a monotone equilibrium the mass attacks the regime, $S(x) = 1$, if and only if $x < x^*$ where x^* solves $\text{Prob}[\theta < 1 \,|\, x^*] = p$.

The regime now faces aggregate uncertainty. It does not know what value of x will be realized. The regime chooses its hidden action to maximize its expected payoff

$$a(\theta) \in \underset{a \geq 0}{\text{argmax}} \left[-C(a) + \int_{-\infty}^{\infty} \max[0, \theta - S(x)] \sqrt{\alpha}\phi(\sqrt{\alpha}(x - \theta - a))\, dx \right] \tag{90}$$

In a monotone equilibrium, the regime's objective simplifies to

$$-C(a) - \min[\theta, 1]\Phi(\sqrt{\alpha}(x^* - \theta - a)) \tag{91}$$

Regimes with $\theta < 0$ are overthrown and so never engage in costly manipulation.

Example: strictly convex costs. Suppose that costs are *strictly convex*, $C''(a) > 0$. This implies all regimes $\theta > 0$ will choose some positive manipulation $a(\theta) > 0$ even regimes that are overthrown ex post. The key first order necessary condition for the regime's choice of action $a(\theta)$ is

$$\min[\theta, 1]\sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - a)) = C'(a), \qquad \theta \geq 0 \tag{92}$$

As usual, there may be two solutions to this first order condition; if so, the smaller is eliminated by the second order condition. An equilibrium of this game is constructed by simultaneously determining $a(\theta)$ and the x^* that solves $\text{Prob}[\theta < 1 \,|\, x^*] = p$.

The first order condition implies that taking as given x^* the regime's $a(\theta) \to 0$ as $\alpha \to \infty$. Given this, the probability of overthrowing the regime $\text{Prob}[\theta < 1 \,|\, x] \to \mathbb{1}\{x < 1\}$ as $\alpha \to \infty$. This implies $x^* \to 1$. With arbitrarily precise information, the regime takes no action and so x is very close to θ. The mass attacks only if it believes $\theta < 1$ and since x is close to θ attacks only if $x < 1$. So if citizens are perfectly coordinated then for precise information regime change occurs for all $\theta < 1$. By contrast, if citizens are imperfectly coordinated then for precise information all regimes $\theta \geq 0$ survive.

Angeletos, Hellwig and Pavan (2006) provide a related analysis. In their model, if agents are imperfectly coordinated then for precise information θ^* can be any $\theta \in (0, \theta^*_{\text{MS}}]$ where $\theta^*_{\text{MS}} = 1 - p < 1$. But if agents are perfectly coordinated then for precise information regime change occurs for all $\theta < 1$. Thus when information is precise the two models agree about the regime change outcome when agents are perfectly coordinated but come to different conclusions when agents are imperfectly coordinated.

References

Angeletos, George-Marios and Iván Werning, "Crises and Prices: Information Aggregation, Multiplicity and Volatility," *American Economic Review*, 2006, *96* (5), 1720–1736.

___ , **Christian Hellwig, and Alessandro Pavan**, "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, 2006, *114* (3), 452–484.

___ , ___ , **and** ___ , "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks," *Econometrica*, 2007, *75* (3), 711–756.

Besley, Timothy and Andrea Prat, "Handcuffs for the Grabbing Hand? The Role of the Media in Political Accountability," *American Economic Review*, 2006, *96* (3), 720–736.

Boix, Carles and Milan Svolik, "The Foundations of Limited Authoritarian Government: Institutions and Power-Sharing in Dictatorships," 2009. Princeton University working paper.

Bueno de Mesquita, Ethan, "Regime Change and Revolutionary Entrepreneurs," *American Political Science Review*, 2010, *104* (3), 446–466.

Calvo, Guillermo A., "Servicing the Public Debt: The Role of Expectations," *American Economic Review*, September 1988, *78* (4), 647–671.

Carlsson, Hans and Eric van Damme, "Global Games and Equilibrium Selection," *Econometrica*, 1993, *61* (5), 989–1018.

Chamley, Christophe P., "Coordinating Regime Switches," *Quarterly Journal of Economics*, August 1999, *114* (3), 869–905.

Chassang, Sylvain and Gerard Padro-i-Miquel, "Conflict and Deterrence under Strategic Risk," *Quarterly Journal of Economics*, 2010, *125* (4), 1821–1858.

Cole, Harold L. and Timothy J. Kehoe, "Self-Fulfilling Debt Crises," *Review of Economic Studies*, January 2000, *67* (1), 91–116.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Dasgupta, Amil, "Coordination and Delay in Global Games," *Journal of Economic Theory*, 2007, *134*, 195–225.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 1999, *66* (1), 183–198.

Diamond, Douglas W. and Philip H. Dybvig, "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy*, June 1983, *91* (3), 401–419.

Edmond, Chris, "Information Manipulation and Social Coordination," July 2012. University of Melbourne working paper.

Esfandiari, Golnaz, "The Twitter Devolution," *Foreign Policy*, June 2010.

Fallows, James, "The Connection Has Been Reset," *Atlantic Monthly*, 2008, *301* (2).

Gehlbach, Scott and Konstantin Sonin, "Government Control of the Media," 2008. University of Wisconsin working paper.

Goldstein, Itay and Ady Pauzner, "Demand Deposit Contracts and the Probability of Bank Runs," *Journal of Finance*, June 2005, *60* (3), 1293–1327.

Heidhues, Paul and Nicholas Melissas, "Equilibria in a Dynamic Global Game: The Role of Cohort Effects," *Economic Theory*, 2006, *28* (3), 531–557.

Hellwig, Christian, "Public Information, Private Information, and the Multiplicity of Equilibria in Coordination Games," *Journal of Economic Theory*, 2002, *107*, 191–222.

_ , Arijit Mukherji, and Aleh Tsyvinski, "Self-Fulfilling Currency Crises: The Role of Interest Rates," *American Economic Review*, December 2006, *96* (5), 1769–1787.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 1999, *66* (1), 169–182.

Kalathil, Shanthi and Taylor C. Boas, *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule*, Carnegie Endowment for International Peace, 2003.

Katz, Michael L. and Carl Shapiro, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, August 1986, *94* (4), 822–841.

Kirkpatrick, David D., "Wired and Shrewd, Young Egyptians Guide Revolt," *New York Times*, February 2011.

Kuran, Timur, "Now Out of Never: The Element of Surprise in the European Revolution of 1989," *World Politics*, 1991, *44* (1), 7–48.

Lohmann, Susanne, "The Dynamics of Information Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91," *World Politics*, 1994, *47* (1), 42–101.

Monderer, Dov and Dov Samet, "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1989, *1* (2), 170–190.

Morozov, Evgeny, *The Net Delusion*, Allen Lane, 2011.

Morris, Stephen and Hyun Song Shin, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 1998, *88* (3), 587–597.

_ and _ , "Rethinking Multiple Equilibria in Macroeconomic Modeling," in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER Macroeconomics Annual*, MIT Press 2000, pp. 139–161.

_ and _ , "Global Games: Theory and Applications," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press 2003.

Musgrove, Mike, "Twitter Is a Player In Iran's Drama," *Washington Post*, June 2009.

Obstfeld, Maurice, "Rational and Self-Fulfilling Balance-of-Payments Crises," *American Economic Review*, 1986, *76* (1), 72–81.

— , "Models of Currency Crises with Self-Fulfilling Features," *European Economic Review*, 1996, *40* (3–5), 1037–1047.

Rochet, Jean-Charles and Xavier Vives, "Coordination Failures and the Lender of Last Resort: Was Bagehot Right after All?," *Journal of the European Economic Association*, December 2004, *2* (6), 1116–1147.

Shadmehr, Mehdi and Dan Bernhardt, "Collective Action with Uncertain Payoffs: Coordination, Public Signals and Punishment Dilemmas," *American Political Science Review*, 2011, *105* (4), 829–851.

Vives, Xavier, "Complementarities and Games: New Developments," *Journal of Economic Literature*, June 2005, *43* (2), 437–479.