Information Manipulation, Coordination and Regime Change[*]

Chris Edmond[†]

First draft: October 2003. This draft: August 2007

Abstract

This paper studies endogenous information manipulation in games where a population can overthrow a regime if individuals coordinate. The benchmark game has a unique equilibrium and in this equilibrium propaganda is effective if signals are sufficiently precise. Despite playing against perfectly rational individuals, a regime is able to manipulate information in a way that exploits heterogeneity in individual beliefs so that at equilibrium its chances of surviving are higher than they otherwise would be. This result is robust to alternative payoffs where the regime cares only for survival and to a number of alternative information structures, including situations where individuals have access to high-quality private information that is entirely uncontaminated by the regime.

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

JEL classifications: C72, D82, D84.

In coordination games, the information available to agents is a crucial determinant of equilibrium outcomes. Typically, the truthfulness of that information is never in question. But when some players choose what information to release, that information is likely to be biased.[1] The optimal amount of bias used by senders of information will depend on the way that receivers typically filter their information. In many settings, receivers find it easy to infer bias from their signals and information manipulation has no effect on equilibrium beliefs. In a coordination setting, however, the way that an individual receiver filters information will also depend on the filtering behavior of others. In equilibrium, senders may be able to manipulate information so as to exploit heterogeneity in receivers' beliefs and thereby affect outcomes.

---

[1]Important recent contributions to the study of endogenous information in coordination games include Angeletos, Hellwig, and Pavan (2006, 2007) and Angeletos and Werning (2006).

This paper studies the use of biased information to manipulate outcomes in coordination games of regime change.[2] In these games, there are two outcomes of interest — either the preservation of a status quo or its overthrow.[3] I show that a regime with an interest in ensuring society coordinates on the status quo may be able to manipulate information to achieve that outcome in equilibrium.

Section 1 outlines the model. Individuals with an interest in preserving the status quo are treated as a single agent playing against a heterogeneous population of small agents. Each small player can take either an action that adversely affects the regime or an action that supports it. The actions of the small players are *strategic complements*. If enough people take the action that undermines the regime, then it is overturned. As in Morris and Shin (1998) and similar models, the ability of the regime to prevail is determined by a single parameter. Individuals are imperfectly informed about this parameter and in principle may coordinate either on overturning the regime or on living with the status quo. In this paper, however, the regime is endowed with the ability to take a costly *hidden action* that influences the distribution of information. In the benchmark model, this *signal-jamming* technology allows the regime to shift the mean of the distribution from which individuals sample so that individuals receive information that at face-value suggests the regime is difficult to overthrow. Rational individuals understand that their information is contaminated by the regime's propaganda and evaluate their signals accordingly.

This paper makes two contributions. I first show that this coordination game with endogenous information manipulation has a unique equilibrium. I then show that signal-jamming is effective in equilibrium when the intrinsic *precision* of the signal distribution is high. As signal precision becomes increasingly high, so does the ex ante probability of the regime surviving. This suggests that a regime's propaganda apparatus will be more useful to it when individuals are receiving, from a technological standpoint, intrinsically high quality signals. Perhaps the information technology revolution may not be as threatening to autocratic regimes as is sometimes supposed.[4]

Equilibrium uniqueness is proved in Section 2. A key intermediate result is that, despite the regime's manipulation, in any candidate equilibrium individuals' assessments of the probability of the regime being overthrown are monotonic in their signals. This makes it possible to use 'global games' arguments along the lines of Carlsson and van Damme (1993) and Morris and Shin (1998).[5] In particular, it is possible to first show that there exists a unique equilibrium when strategies are restricted to be monotone and then to show that this is the only equilibrium which survives the iterative elimination of strictly dominated strategies.

---

[2]Cheli and Della-Posta (2002) study a coordination game where individuals' signals are exogenously biased. By contrast, this paper studies an equilibrium problem where the amount of bias introduced by the regime has to be compatible with rational beliefs on the part of information receivers.

[3]The underlying coordination game employed in this paper has been used for many applications, including currency and debt crises, bank runs, and political unrest. See, for example, Obstfeld (1986, 1996), Morris and Shin (1998, 2003, 2004), Chamley (1999) and Atkeson (2000). See also Cooper (1999) and Chamley (2004, 237-267) for textbook overviews with many additional applications.

[4]See Kalathil and Boas (2003) for a discussion of the conventional wisdom on the role of information in undermining authoritarian regimes. They argue that the revolution in communication technologies like the internet is not likely to help in overthrowing such regimes.

[5]Carlsson and van Damme (1993) introduced the term 'global game' to describe coordination games of incomplete information.

The effectiveness of signal-jamming when signals are precise is proved in Section 3. In equilibrium regimes are overthrown if their type is below an endogenous threshold. If a regime manipulates it generates a signal distribution with an artificially high mean that is strictly greater than the threshold. So if signals are precise, in this situation many individuals have signals suggesting the regime will survive. And consequently it is rational for any individual, when contemplating the beliefs of others, to assign relatively high probability to the event that they mostly have signals near this artificially high mean. At the margin this makes any individual less likely to attack and so the aggregate mass who do is relatively low. This in turn makes it more likely that the regime does in fact manipulate and create an artificially high signal mean thereby validating the original beliefs. Two things are critical to this argument. First, different types of regimes must take different actions so that there is uncertainty about the amount by which individuals should discount their signals: if all regimes took the same action, it would be easy to undo. Second, individuals must be imperfectly coordinated. If individuals were perfectly coordinated and information precise, all regimes that could be overthrown would be overthrown.

The equilibrium uniqueness result contrasts with Angeletos, Hellwig, and Pavan (2006), who were the first to emphasize endogenous information in global coordination games. Angeletos, Hellwig, and Pavan (2006) have in mind situations where a regime's policy choice (e.g., a central bank's interest rate in anticipation of a speculative attack) provides an additional signal about the regime's type. Since individual strategies may or may not condition on the signal, this gives rise to both pooling and separating equilibria. In this paper, however, individuals' information is contaminated by a behind-the-scenes choice of the regime (e.g., pressure on a recalcitrant editor or general) that is commonly known to be possible but cannot be observed. Individuals have to disentangle the truth from the effects of the hidden action. Signal-jamming games like this typically have fewer equilibria than their signaling counterparts. For example, in Angeletos, Hellwig, and Pavan (2006) a pooling equilibrium occurs if all individuals believe that the regime will not intervene, for then the regime has no incentive to make a costly intervention. The regime is 'trapped' in an inactive policy equilibrium where the regime merely validates the belief that it will not intervene. But in this paper if all individuals believe that the regime will not take any action, the regime has a powerful incentive to manipulate information so there can be no pooling equilibrium.

Traditional theories of strategic information transmission, such as Crawford and Sobel (1982) and Matthews and Mirman (1983), focus on a sole information receiver. In this paper there is instead a large cross-section of signal receivers. Since the incentives of the sender to manipulate depend on the aggregate behavior of the receivers and since individual actions are strategic complements, implicitly, each individual solves a filtering problem that depends simultaneously on the solutions to the filtering problems of all other receivers.

Section 4 shows that results are robust to alternative payoffs for the regime that give it no direct incentive to manipulate. Section 5 extends the model to cover alternative information structures, including situations where some information is *uncontaminated* by the regime's manipulation and an alternate technology for the regime where its actions directly affect signal precision.

1 Model of information manipulation and regime change

There is a unit mass of citizens, indexed by $i \in [0, 1]$. Citizens are ex ante identical. After drawing a signal (discussed below) each citizen decides whether to subvert the regime, $s_i = 1$, or not, $s_i = 0$. The population mass of subversives is $S := \int_0^1 s_i di$. If a citizen subverts, she pays a fixed opportunity cost $p > 0$.

The citizens face a regime indexed by a hidden state variable, denoted θ, that is the regime's private information. The type θ is normalized such that the regime is overthrown if and only if $\theta < S$. The payoff to a citizen is

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

where $\mathbb{1}$ denotes the indicator function. Individual actions s_i and the population aggregate S are *strategic complements*: the more citizens subvert the regime, the more likely it is that the regime is overthrown and so the more likely it is that any individual citizen's best action is to also subvert.

After learning θ, a regime may take a hidden action $a \geq 0$ that incurs a convex cost $C(a)$ where $C(0) = 0$, $C'(a) > 0$ for $a > 0$ and $C''(a) \geq 0$ for all a. I begin by assuming that the regime obtains a benefit $\theta - S$ from remaining in power. The regime is not just concerned with remaining in power but also wants to keep S small when it does survive. If $\theta < S$, the regime is overthrown and obtains an outside option with value normalized to zero. The payoff to a regime is therefore

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

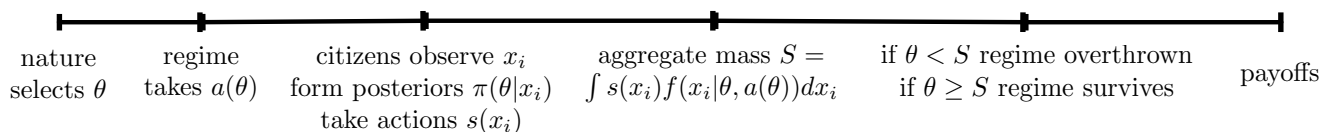where $B(S, \theta) := \mathbb{1}\{\theta \geq S\}(\theta - S)$.

| nature selects θ | regime takes $a(\theta)$ | citizens observe x_i form posteriors $\pi(\theta\|x_i)$ take actions $s(x_i)$ | aggregate mass $S = \int s(x_i) f(x_i\|\theta, a(\theta)) dx_i$ | if $\theta < S$ regime overthrown if $\theta \geq S$ regime survives | payoffs |

Figure 1: The timing of the model.

Following a regime's hidden action a, each citizen simultaneously draws an idiosyncratic signal $x_i := \theta + a + \varepsilon_i$ where the noise ε_i is independent of θ and is IID normally[6] distributed with mean zero and *precision* α (that is, variance α^{-1}). So the density of signals is $f(x_i|\theta, a) := \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a))$ where ϕ denotes the standard normal density. I begin by assuming that citizens have common priors for θ and that this prior is the (improper) uniform distribution over the whole real line.[7] The realization of the signal x_i is informative for both the type of the regime θ and the hidden action a. This action is itself informative about the regime's type and rational citizens take this account when forming their beliefs. In equilibrium, the action taken by a regime and the beliefs of citizens will need to be mutually consistent. The timing of the model is shown in Figure 1.

---

[6]Normality is assumed for the sake of exposition. The critical assumption is that the support of ε_i is unbounded. In previous drafts of this paper I dealt with more general continuous distributions. Details are available on request.
[7]I discuss informative priors and heterogeneous priors in Section 5 below.

1.1 Equilibrium

A symmetric perfect Bayesian equilibrium is an individual's posterior density $\pi(\theta|x_i)$, individual subversion decision $s(x_i)$, mass of subversives $S(\theta, a)$ and hidden actions $a(\theta)$ such that

$$\pi(\theta|x_i) = \frac{f(x_i|\theta, a(\theta))}{\int_{-\infty}^{\infty} f(x_i|\theta, a(\theta))d\theta} \tag{3}$$

$$s(x_i) \in \underset{s_i \in \{0,1\}}{\operatorname{argmax}} \left\{ \int_{-\infty}^{\infty} u(s_i, S(\theta, a(\theta)), \theta)\pi(\theta|x_i)d\theta \right\} \tag{4}$$

$$S(\theta, a) = \int_{-\infty}^{\infty} s(x_i)f(x_i|\theta, a)dx_i \tag{5}$$

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \left\{ B(S(\theta, a), \theta) - C(a) \right\} \tag{6}$$

The first condition says that a citizen with information x_i takes into account the regime's manipulation $a(\theta)$. The second says that given these beliefs, $s(x_i)$ is chosen to maximize its expected payoff. The third condition aggregates individual decisions to give the mass of subversives. The final condition says that the actions $a(\theta)$ maximize the regime's payoff. In equilibrium, the regime is overthrown if $R(\theta) := \mathbb{1}\{\theta < S(\theta, a(\theta))\} = 1$ while the regime survives if $R(\theta) = 0$.

1.2 Discussion of model

Regime's payoffs. Conditional on surviving a regime obtains benefit $\theta - S$ and so has a *direct* aversion to S. Regimes prefer to avoid a Prague Spring or a Tiananmen Square. There are two distinct motivations for this. First, suppressing a riot is likely to be resource costly to the regime and it seems natural for this cost to be increasing in the mass of rioters. Second, loosely speaking, a benefit $\theta - S$ provides a simple way of modeling aversion to information generated by observations of large S. For example, in a dynamic version of this model citizens would observe (possibly with noise) the mass of people who have in the past attempted to overthrow the regime. The size of previous attacks contains information about θ — in particular, low S is associated with high θ — and so by keeping S low even when it will survive this period, the regime may have another tool for convincing individuals not to attack.[8] Section 4 below shows that the main results of this paper go through when payoffs are instead determined by $B(S, \theta) = \mathbb{1}\{\theta \geq S\}\theta$ so that the regime cares only about survival.

Absence of free-rider problems. If a citizen is almost sure the regime will be overthrown she will subvert and will not *free-ride* on others. Although this seems unappealing, a bad free-rider problem is simply one of many factors that might lead to a high opportunity cost p. To see this formally, suppose citizens get a larger expected payoff if they were actively involved in the downfall of the

---

[8]More crudely, observations of large S might be able to convince foreign powers that it would be easy to assist the regime's opponents it in bringing the regime down. The $\theta - S$ specification also facilitates comparison with Angeletos, Hellwig, and Pavan (2006), ensuring that the information structure is the only difference.

regime.[9] Specifically, if the regime is overthrown a citizen gets random payoff $w \in \{\underline{w}, \overline{w}\}$ with $\underline{w} < \overline{w}$ and $\Pr(w = \underline{w}) := \mu(s_i)$ with $0 \leq \mu(1) < \mu(0) \leq 1$. That is, a citizen may free-ride but if they do so they face a higher probability of being caught and penalized by getting only $\underline{w}$. Then if the utility cost of subverting the regime is normalized to 1 and a citizen believes the regime will be overthrown with probability $P(x_i)$, she will participate if and only if

$$P(x_i)[\overline{w} - \mu(1)(\overline{w} - \underline{w})] - 1 \geq P(x_i)[\overline{w} - \mu(0)(\overline{w} - \underline{w})]$$

equivalently, if and only if

$$P(x_i) \geq \frac{1}{(\overline{w} - \underline{w})(\mu(0) - \mu(1))} =: p \tag{7}$$

If $p \geq 1$ it can never be rational for a citizen to participate in subversion. To make the model interesting, then, we need:

ASSUMPTION 1. The opportunity cost of subversion is not too high: $p < 1$.

Intuitively, free-riding is less of a problem if the probability of being caught out is sufficiently high or if the penalty from being caught is sufficiently severe. With this assumption, free-riding is possible in principle but does not occur in equilibrium.

Interpretation of hidden actions. The hidden action $a \geq 0$ of the regime gives it the ability to choose the common component of individual beliefs and the potential to bias the information that citizens receive. If $a > 0$ citizens draw from a signal distribution that at face value suggests the regime will be more difficult to overthrow. This represents a situation where it is common knowledge that the regime is able to exert pressure on editors, force recalcitrant generals to stand on parade, etc — so as to depict itself as difficult to overthrow — but where it is not possible to observe that pressure or manipulation directly and it instead must be inferred. In Section 5 below I endow citizens with informative priors for θ, thereby giving individuals extra information that is independent of a.

1.3 Exogenous information benchmarks

Two important special cases of the model are when: (i) the regime's type is common knowledge, or (ii) hidden actions are prohibitively expensive. In each case, citizens have *exogenous information.*

If θ is common knowledge, costly hidden actions are pointless and $a(\theta) = 0$ all θ. The model reduces to the coordination game used by Obstfeld (1986, 1996) to discuss self-fulfilling currency crises. If $\theta < 0$, any crowd $S \geq 0$ can overthrow the regime. It is optimal for any individual to subvert, all do so, and the regime is overthrown. If $\theta \geq 1$, no crowd can overthrow the regime. It is optimal for any individual to not riot, none do, and the regime survives. If $\theta \in [0, 1)$, the regime

---

[9]Perhaps it is more likely that a citizen will secure an influential position in the new regime if she participated in the overthrow of the old regime. Or perhaps retribution is exacted on those who are thought to have let others take the risks in overthrowing the regime. See, respectively, Jackson (2001) and Frommer (2005) for discussion of the retribution exacted on collaborators after the liberation of France and Czechoslovakia from Nazi rule.

is 'fragile' and multiple self-fulfilling equilibria can be sustained. For example, if each individual believes that everyone else will riot, it will be optimal for each citizen to do so and $S = 1 > \theta$ leads to the regime's overthrow and the vindication of the initial expectations.

If hidden actions are prohibitively expensive, $a(\theta) = 0$ all θ and each citizen has private signal $x_i = \theta + \varepsilon_i$. Because each citizen has a signal of the regime's type, expectations are no longer arbitrary. As discussed by Carlsson and van Damme (1993), Morris and Shin (1998) and subsequent literature, this introduces the possibility of pinning down a unique equilibrium outcome.[10] In this equilibrium, strategies are threshold rules: there is a unique type θ^* such that the regime is overthrown for $\theta < \theta^*$ and a unique signal x^* such that a citizen subverts for $x < x^*$.

PROPOSITION 1. (Morris-Shin): The unique equilibrium thresholds $x_{\mathrm{MS}}^*, \theta_{\mathrm{MS}}^*$ simultaneously solve

$$\Phi\left[\sqrt{\alpha}\left(\theta_{\mathrm{MS}}^* - x_{\mathrm{MS}}^*\right)\right] = p \tag{8}$$
$$\Phi\left[\sqrt{\alpha}\left(x_{\mathrm{MS}}^* - \theta_{\mathrm{MS}}^*\right)\right] = \theta_{\mathrm{MS}}^* \tag{9}$$

where Φ denotes the standard normal cumulative distribution. In particular, $\theta_{\mathrm{MS}}^* = 1 - p$ independent of α.

The first condition says that if the regime's threshold is θ_{MS}^*, a citizen with signal $x_i = x_{\mathrm{MS}}^*$ will be indifferent between subverting or not. The second condition says that if the signal threshold is x_{MS}^*, a regime with type $\theta = \theta_{\mathrm{MS}}^*$ will be indifferent. In the analysis below, I will say that a regime's hidden action technology is *effective* if in equilibrium $\theta^* < \theta_{\mathrm{MS}}^* = 1 - p$.

To understand the consequences of information sets that are a function of the regime's manipulation, we need to study a more difficult equilibrium problem where the regime's manipulation is not trivial and citizens internalize a regime's incentives.

2   Unique equilibrium with hidden actions

The first main result of this paper is:

THEOREM 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 $R(\theta) = 1$ for $\theta < \theta^*$ and zero otherwise.

Details of the proof are given in Appendix A. The proof shows that (i) there is a unique equilibrium when agents' strategies are restricted to be monotone, and (ii) that the unique monotone equilibrium is the only equilibrium which survives the iterative elimination of (interim) strictly dominated strategies.[11]

---

[10]This result depends on a relatively diffuse common prior. See Hellwig (2002) and Morris and Shin (2000, 2003, 2004) for discussion of the possibility of multiple equilibria in coordination games when public information is sufficiently informative. Public information is introduced into the model with hidden actions in Section 5 below.

[11]The proof strategy is similar to Carlsson and van Damme (1993) and Morris and Shin (1998) but details differ because of the endogenous information.

The unique monotone equilibrium. Let $\hat{x}, \hat{\theta}$ denote candidates for the thresholds in a monotone equilibrium and let $a(\theta|\hat{x})$ denote a candidate for the regime's hidden actions taking $\hat{x}$ as given. Because of the additive signals, $x = \theta + a + \varepsilon$, a unit increase in θ and a unit increase in x perfectly offset each other in terms of their effect on the regime's desired action. Given this, equilibrium actions will depend only on the difference[12] $\theta - \hat{x}$ and can be written $a(\theta|\hat{x}) = a(\theta - \hat{x})$.

Taking $\hat{x}$ as given the mass of citizens facing the regime is

$$\hat{S}(\theta + a) := \int_{-\infty}^{\hat{x}} f(x_i|\theta, a)dx_i = \Phi[\sqrt{\alpha}(\hat{x} - \theta - a)] \tag{10}$$

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

$$V(\theta) := \max[0, W(\theta)] \tag{11}$$

where $W(\theta)$ is the best payoff regime θ can get if it is not overthrown

$$W(\theta) := \max_{a \geq 0} \left[\theta - \hat{S}(\theta + a) - C(a)\right] \tag{12}$$

Using the envelope theorem and the definition of $\hat{S}(\theta + a)$ in equation (10) shows that $W'(\theta) > 1$ all θ. And since $W(\theta) < 0$ for $\theta < 0$ and $W(1) > 0$, by the intermediate value theorem there is a unique $\hat{\theta} \in [0, 1)$ such that $W(\hat{\theta}) = 0$. Using (11), the regime is overthrown if and only if $\theta < \hat{\theta}$. For $\theta \geq \hat{\theta}$, the actions of the regime solve

$$a(\theta - \hat{x}) \in \operatorname*{argmax}_{a \geq 0} \left[\theta - \hat{S}(\theta + a) - C(a)\right], \quad \theta \geq \hat{\theta} \tag{13}$$

Threshold $\hat{\theta}$ is found from

$$W(\hat{\theta}) = \hat{\theta} - \hat{S}[\hat{\theta} + a(\hat{\theta} - \hat{x})] - C[a(\hat{\theta} - \hat{x})] = 0 \tag{14}$$

Taking $\hat{x}$ as given, equations (13)-(14) determine the threshold $\hat{\theta}$ and the hidden actions $a(\theta - \hat{x})$ that characterize the solution to the regime's problem.

Given the solution to the regime's problem, a citizen with signal x_i assigns probability $P(\hat{\theta}, x_i)$ to the regime being overthrown

$$P(\hat{\theta}, x_i) := \Pr(\theta < \hat{\theta}|x_i) = \frac{\int_{-\infty}^{\hat{\theta}} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_i - \theta)]d\theta}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_i - \theta - a(\theta - \hat{x}))]d\theta} \tag{15}$$

where the numerator uses the fact that the regime takes no action for $\theta < \hat{\theta}$. The most important properties of $P : \mathbb{R} \times \mathbb{R} \to [0, 1]$ are:

LEMMA 1. The function P is continuous, strictly increasing in $\hat{\theta}$ and strictly decreasing in x_i, and for any $\hat{x} \in \mathbb{R}$ satisfies

$$P(\hat{\theta}, x_i) = P(\hat{\theta} - \hat{x}, x_i - \hat{x}) \tag{16}$$

---

[12]Technically, the candidate action profile $a(\theta - \hat{x})$ also depends on $\hat{\theta}$, but the only reason for keeping track of this dependence is to demarcate the region where $a = 0$ from the region where $a > 0$. Since this is easy to remember, I suppress the dependence on $\hat{\theta}$ wherever this ought to cause no confusion.

A citizen with signal x_i will subvert the regime if and only if $P(\hat{\theta}, x_i) \geq p$. Therefore given the solution to the regime's problem as implied by (13)-(14), the signal threshold $\hat{x}$ solves

$$P(\hat{\theta}, \hat{x}) = P(\hat{\theta} - \hat{x}, 0) = p \qquad (17)$$

To calculate a monotone equilibrium, we first solve the decision problem in (13) for arbitrary $\hat{x}, \hat{\theta}$. This gives a family of hidden actions $a(\theta - \hat{x})$ which can be used to construct the P function in (15). Since P is continuous and strictly increasing in $\hat{\theta}$ with $P(-\infty, 0) = 0$ and $P(\infty, 0) = 1$, there is a unique difference $\theta^* - x^* \in \mathbb{R}$ such that $P(\theta^* - x^*, 0) = p$. This solution can be plugged into (14) to obtain a unique threshold $\theta^* \in [0, 1)$ so that we know both θ^* and $x^* \in \mathbb{R}$ separately. The equilibrium hidden actions are then $a(\theta) := a(\theta - x^*)$.

Iterative elimination of dominated strategies. We can now go on to show that there is no other equilibrium. From Lemma 1, there exist unique signals $\underline{x} < \overline{x}$ such that

$$P(0, \underline{x}) = p = P(1, \overline{x}) \qquad (18)$$

Since $p \in (0, 1)$ and P is continuous with $P(0, \infty) = 0$ and $P(1, -\infty) = 1$ both $\underline{x}$ and $\overline{x}$ are finite. Any citizen with a sufficiently low signal $x_i < \underline{x}$ has posterior $\Pr(\theta < 0 | x_i) > p$ and finds it strictly dominant to subvert the regime. Similarly, any citizen with a sufficiently high signal $x_i > \overline{x}$ has posterior $\Pr(\theta < 1 | x_i) < p$ and finds it strictly dominant to not subvert the regime.

Now suppose that almost all citizens subvert the regime if and only if they have signals $x < \hat{x}$. Then (13) and (14) determine a threshold $\hat{\theta}$ and hidden actions $a(\theta - \hat{x})$. In this situation, the expected gain from subverting to a citizen with signal x_i is $P(\hat{\theta} - \hat{x}, x_i - \hat{x}) - p$. Since P is continuous and strictly decreasing in x_i with $P(\hat{\theta} - \hat{x}, -\infty) = 1$ and $P(\hat{\theta} - \hat{x}, \infty) = 0$ for any $\hat{\theta} - \hat{x}$, for each $\hat{x} \in \mathbb{R}$ there is a unique signal $\psi(\hat{x})$ such that

$$P(\hat{\theta} - \hat{x}, \psi(\hat{x}) - \hat{x}) = p \qquad (19)$$

where $\hat{\theta}$ implicitly depends on $\hat{x}$ from (13) and (14). When other citizens subvert the regime if and only if they have signals $x < \hat{x}$, a citizen with x_i finds it optimal to subvert the regime if and only if they have $x_i < \psi(\hat{x})$. The most important properties of $\psi : \mathbb{R} \to \mathbb{R}$ are:

LEMMA 2. The function ψ is continuous and has a single fixed point $x^* = \psi(x^*)$ with $\psi'(x^*) \in (0, 1)$. Moreover $\psi(\hat{x}) < x^*$ for all $\hat{x} < x^*$ and $\psi(\hat{x}) > x^*$ for all $\hat{x} > x^*$.

We can now use ψ to iteratively eliminate strictly dominated strategies.[13] We build up a sequence $\{\underline{x}_n\}_{n=0}^{\infty}$ by iterating on ψ starting from the initial condition $\underline{x}_0 := \underline{x}$. Since ψ is continuous, has a unique fixed point $x^* = \psi(x^*)$ with $\psi'(x^*) \in (0, 1)$ and $\psi(\underline{x}_n) < x^*$ for all $\underline{x}_n < x^*$, this sequence is strictly monotone increasing, bounded above by x^* and converges to x^* in the limit. Similarly, the sequence $\{\overline{x}_n\}_{n=0}^{\infty}$ obtained by iterating on ψ starting from $\overline{x}_0 := \overline{x}$ is strictly monotone decreasing, bounded below by x^* and converges to x^* in the limit.

---

[13]Given the properties of ψ as listed in Lemma 2, the remaining arguments are standard in the literature [Morris and Shin (1998, 2003) or for a textbook discussion Chamley (2004)]. They are outlined here for completeness.

Intuitively, these sequences represent the elimination of strictly dominated strategies 'from below' and 'from above'. For example, even if no-one else subverts the regime, a citizen with signal $x_i < \underline{x}_1 = \psi(\underline{x}_0)$ will find it optimal to subvert the regime. Therefore *all* citizens with signals $x < \underline{x}_1$ have $s(x) = 1$. But if so, all citizens with signals less than $\underline{x}_2 = \psi(\underline{x}_1)$ also find it optimal to subvert the regime. And so on. After n iterations, the only candidates for a citizen's equilibrium strategy all have the property that $s(x_i) = 1$ for $x_i < \underline{x}_n$ and similarly $s(x_i) = 0$ for $x_i > \overline{x}_n$ with $s(x_i)$ arbitrary for $x_i \in [\underline{x}_n, \overline{x}_n]$. 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. In short, the unique monotone equilibrium is the only equilibrium.

2.1 Discussion

Signaling vs. signal-jamming. Theorem 1 contrasts with Angeletos, Hellwig, and Pavan (2006), who were the first to emphasize endogenous information in a global games context. In their 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. Individual strategies $s(x_i, a)$ may condition on a. 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. Given this, the regime has no incentive to manipulate and so validates the original expectation.

In this paper, individuals get one noisy observation of the sum $\theta + a$ and so cannot condition their strategies on a. Loosely speaking, *signal-jamming* games like this typically have fewer equilibria than the corresponding signaling game because there is less freedom to specify the signal receiver's beliefs about θ. For example, in this paper the impossibility of conditioning on a almost immediately rules out the existence of a perfectly pooled equilibrium: to see why, suppose that all individuals believe that no manipulation takes place. The best payoff a regime can get if it is not overthrown $W(\theta)$ is strictly increasing in θ and there is a unique $\hat{\theta} \in [0, 1)$ such that the regime is overthrown for all $\theta < \hat{\theta}$ [as in equation (12) above]. Each regime $\theta \geq \hat{\theta}$ will have an incentive to take an action $a(\theta) \geq 0$ (strictly positive if $C'(0)$ is not too large). If so, rational individuals will not believe that no manipulation takes place and there cannot be a pooling equilibrium.

Apart from questions of equilibrium uniqueness, the presence of a signal-jamming technology also raises the issue of whether the signal sender, the regime, is *effective* in its goal of deterring citizens from participating. This parallels the debate in the industrial organization literature on limit pricing and entry deterrence. In Milgrom and Roberts (1982) an incumbent firm may price below its static monopoly price to signal that it has low costs in the hope of deterring entry. This game has a pooling equilibrium and many separating equilibria where the incumbent's price is indeed below its monopoly price but where limit pricing is ineffective in that a rational entrant correctly infers the incumbent's costs from its price and so enters exactly when it would if the incumbent's costs were observable. By contrast, Matthews and Mirman (1983) study a signal-jamming problem where the entrant observes the incumbent's price with noise. They show there is

often a unique equilibrium and that it exhibits both limit pricing and successful entry deterrence.[14]

Imperfectly coordinated signal receivers. In traditional models of strategic information transmission like Crawford and Sobel (1982) and Matthews and Mirman (1983) there is one sender and one receiver. But in this paper and in Angeletos, Hellwig, and Pavan (2006) there is a large cross-section of signal receivers. Since the incentives of the sender to manipulate depend on the aggregate behavior of the receivers and since individual actions are strategic complements, implicitly, each individual receiver's filtering problem depends simultaneously on the solution to the filtering problem of all other receivers. If some citizens believe the regime is manipulating, they will conclude that the regime will survive (since only regimes with $\theta \geq \theta^*$ actively manipulate) and will not subvert. Because of the strategic complementarity, this will make other individuals less likely to subvert.

That receivers are imperfectly coordinated gives rise to effects absent from the traditional one sender/one receiver game. As discussed in Section 3.3 below, when precision is very high and citizens are imperfectly coordinated, signal-jamming is maximally effective: as the precision $\alpha \to \infty$, all the fragile regimes with $\theta \in [0, 1)$ survive. But if instead citizens are perfectly coordinated then all the regimes with $\theta \in [0, 1)$ are overthrown.[15]

3   Equilibrium information manipulation

The second main result of this paper is that the regime's signal-jamming technology becomes more effective as the precision α increases. Section 3.1 prepares the ground by characterizing equilibrium hidden actions. Section 3.2 and Section 3.3 show that signal-jamming can be effective in equilibrium even when the regime is playing against rational citizens who internalize the regime's incentives. To give intuition for this, I first use a simple example with constant marginal costs before turning to more general convex costs. Section 3.4 characterizes the equilibrium bias a regime injects into individual beliefs.

Terminology. Throughout this section I draw a distinction between whether signal-jamming *occurs* in equilibrium (when $a(\theta) > 0$ for some θ) and whether it is *effective*. I measure the effectiveness of signal-jamming by its ability to reduce the regime's threshold θ^* relative to the Morris-Shin level of $\theta^*_{\mathrm{MS}} = 1 - p$ (as in Proposition 1). A lower θ^* increases the regime's ex ante survival probability by making it more likely that nature draws a $\theta \geq \theta^*$. For high enough α the regime has $\theta^* < 1 - p$ so that endowing a regime with a signal-jamming technology allows it to survive in more states than it would in the Morris-Shin world.

I say that the regime *benefits* from lower θ^* even though this does not necessarily increase the regime's payoff. In principle, it might be the case that lower θ^* is achieved through large, costly

---

[14]See Wilson (1992), Riley (2001, 455-459) and Tirole (1988, 364-374) for further discussion.

[15]Similarly in Angeletos, Hellwig, and Pavan (2006) if agents cannot coordinate θ^* can be any $\theta \in (0, \theta^*_{\mathrm{MS}}]$ where $\theta^*_{\mathrm{MS}} = 1 - p < 1$, but if agents can coordinate then as $\alpha \to \infty$ regime change occurs for all $\theta < 1$.

actions that give the regime a lower payoff than they would achieve in the Morris-Shin world. But it turns out that as precision $\alpha \to \infty$ 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.

3.1 Regime's hidden actions

In equilibrium, hidden actions $a(\theta)$ are characterized by the first order necessary condition[16]

$$C'(a) = \sqrt{\alpha}\phi[\sqrt{\alpha}(x^* - \theta - a)], \qquad \theta \geq \theta^* \tag{20}$$

The marginal benefit of an action is the associated reduction in the mass of subversives and at an interior solution this is equated to $C'(a)$. For signal-jamming to occur (meaning $a(\theta) > 0$ for at least some θ), the cost function either has to be either (i) strictly convex, or (ii) if marginal costs are constant, $C'(a) = c$ all a, then the level of c cannot be 'too high': $c < \sqrt{\alpha}\phi(0) =: \bar{c}$. If either of these conditions is satisfied, then actions are zero for all $\theta < \theta^*$ before jumping up discontinuously to a positive value at the threshold θ^*. As the fundamentals of the regime become strong, costly actions taken to generate a favorable signal distribution begin to encounter diminishing returns and the action profile dies to zero. Figure 2 illustrates.
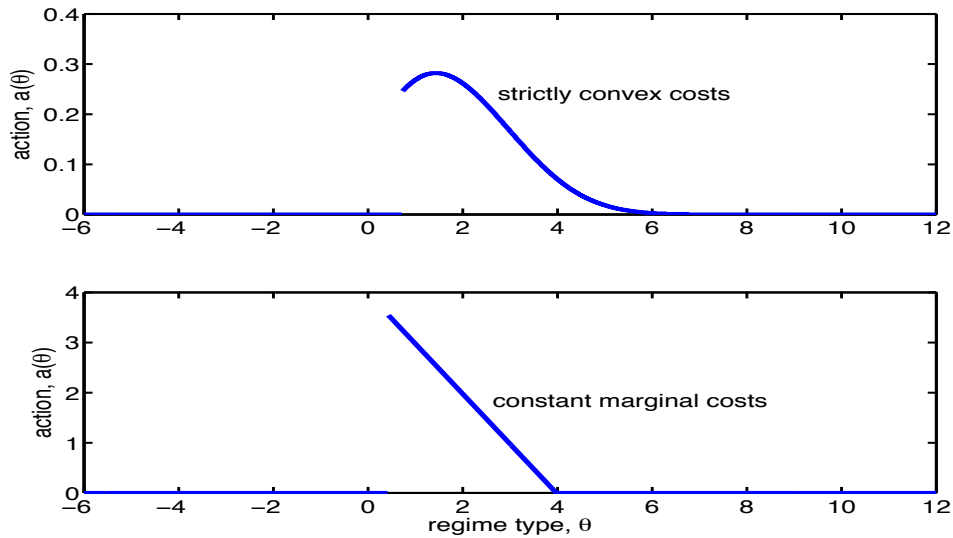

Figure 2: Equilibrium hidden actions $a(\theta)$.

Even if some regimes manipulate information in equilibrium, it is not necessarily the case that the existence of the signal-jamming technology is to the regime's benefit. To understand when a regime does benefit from the signal-jamming technology, I first specialize to the case of constant marginal costs. This example allows for explicit calculations showing that signal-jamming is effective when the signal precision α is sufficiently high.

---

[16]The first order condition (20) for $a(\theta) > 0$ may have zero, one or two solutions. In the event of two solutions, only the higher solution satisfies the second order condition.

3.2 Signal-jamming example: constant marginal costs

Let $C(a) := ca$ for some constant $c \in (0, \bar{c})$ where $\bar{c} := \sqrt{\alpha}\phi(0)$ so that $a(\theta) > 0$ for some θ. Then manipulating equation (20) shows that interior solutions to the regime's problem are given by

$$a(\theta) = x^* + \gamma - \theta, \qquad \theta \in [\theta^*, \theta^{**}) \tag{21}$$

where $\theta^{**} := x^* + \gamma$ and where

$$\gamma := \sqrt{\frac{2}{\alpha} \log\left(\frac{\sqrt{\alpha}\phi(0)}{c}\right)} > 0 \tag{22}$$

In this case, the signal-jamming is particularly acute. All regimes that manipulate information pool on the same distribution of signals. Citizens receive signals $x_i = x^* + \gamma + \varepsilon_i$ that are *locally completely uninformative* about θ.

More precise information makes signal-jamming effective. In the absence of manipulation, higher precision α makes signals more tightly clustered around the true θ? Does this kind help citizens overthrow fragile regimes $\theta \in [0, 1)$ that survive only because of imperfect coordination? In the Morris-Shin case, the answer is 'no': the equilibrium threshold $\theta^*_{\mathrm{MS}} = 1 - p$ is invariant to the precision of the private information.[17] With endogenous information manipulation matters are even worse for the citizens: a regime's signal-jamming is more effective for high values of α than for low values.

To see this formally, use (21) and rearrange (14) and (17) to get

$$\Phi[\sqrt{\alpha}(\theta^*_\alpha - x^*_\alpha)] = \frac{p}{1-p}\left[c(x^*_\alpha - \theta^*_\alpha + \gamma_\alpha) + \Phi(-\sqrt{\alpha}\gamma_\alpha)\right] \tag{23}$$

and

$$\theta^*_\alpha = c(x^*_\alpha - \theta^*_\alpha + \gamma_\alpha) + \Phi(-\sqrt{\alpha}\gamma_\alpha) \tag{24}$$

where subscript α acknowledges dependence on the signal precision. For each $\alpha > 0$, these two equations uniquely determine $x^*_\alpha, \theta^*_\alpha$ [solve (23) for the unique difference $\theta^*_\alpha - x^*_\alpha$ and then plug into (24) to get θ^*_α]. The equilibrium mass of subversives that makes the regime indifferent is

$$S^*_\alpha := \Phi(-\sqrt{\alpha}\gamma_\alpha) = \Phi\left[-\sqrt{2\log\left(\frac{\sqrt{\alpha}\phi(0)}{c}\right)}\right] \tag{25}$$

The mass S^*_α is strictly decreasing in α and $S^*_\alpha \to 0^+$ as $\alpha \to \infty$. High α helps the regime engineer a small mass of subversives. In turn, this means that as $\alpha \to \infty$ all regimes with $\theta \geq 0$ will survive. To see this note for large α solutions to equation (23) are approximately the same as solutions to

$$\mathbb{1}\{\theta^*_\alpha - x^*_\alpha \geq 0\} = -\frac{p}{1-p}c(\theta^*_\alpha - x^*_\alpha) \tag{26}$$

---

[17]See Proposition 1 for more discussion. The invariance of the state threshold to the signal precision is an artifact of the assumption of a diffuse prior for θ. I test the robustness of my comparative statics results to informative priors in Section 5 below.

The only solution to equation (26) is $\theta_\alpha^* - x_\alpha^* = 0$. So as $\alpha \to \infty$, solutions to equation (23) approach zero too. From equation (24) we now know that $\theta_\alpha^* \to 0^+$. Therefore, the signal-jamming technology is effective when the precision α is large enough. For large α the regime's threshold θ_α^* is less than the Morris-Shin level of $\theta_{\text{MS}}^* = 1 - p$ and the regime's survival probability is correspondingly higher.

Intuition for the result. Why does a higher α help signal-jamming regimes? If a regime manipulates, then from (21), it generates a signal distribution with an artificially high mean that is strictly greater than the threshold, $x^* + \gamma = \theta^{**} > \theta^*$. So if signals are precise, in this situation many individuals have signals suggesting the regime will survive. And consequently it is rational for any individual, when contemplating the beliefs of others, to assign relatively high probability to the event that they mostly have signals near this artificially high mean $\theta^{**} > \theta^*$. At the margin this makes any individual less likely to attack and so the aggregate mass S who do is relatively low. This in turn makes it more likely that the regime does in fact manipulate and create an artificially high signal mean thereby validating the original beliefs. Two things are critical to this argument. First, different types of regimes must take different actions so that there is uncertainty about the amount by which individuals should discount signals: if all regimes took the same action, it would be easy to undo. And second, as discussed below, individuals must be imperfectly coordinated.

From a technological standpoint, signals may be intrinsically precise (of high quality). But this does not necessarily translate into reduced posterior uncertainty for individuals. The direct effect of higher α is to reduce posterior uncertainty, but there is also an indirect effect through the regime's hidden actions. Suppose the regime's policy was linear, $a(\theta) = a_0(\alpha) + a_1(\alpha)\theta$ for some coefficients $a_0(\alpha), a_1(\alpha)$ (this can't be true in equilibrium, but it's instructive nonetheless). If so, citizens would have normal posteriors with precision $\alpha[1 + a_1(\alpha)]^2$. Then if $\alpha \to \infty$ but $a_1(\alpha) \to -1$ sufficiently fast, the signals x_i have no local information about θ even when α is large. The example with constant marginal cost has slope coefficient exactly -1 whenever a regime manipulates and so in this case signals are locally uninformative about θ.

3.3   Signal-jamming: the general case

The effectiveness of signal-jamming when α is large extends to the case of general convex costs:

THEOREM 2. As the signal precision $\alpha \to \infty$ the limiting equilibrium thresholds and hidden action profile are

$$\lim_{\alpha \to \infty} \theta_\alpha^* = 0^+, \qquad \lim_{\alpha \to \infty} x_\alpha^* = 0^+, \quad \text{and} \quad \lim_{\alpha \to \infty} a_\alpha(\theta) = 0^+ \quad \text{for all } \theta$$

If in addition the cost of information manipulation is strictly convex, $C''(a) > 0$ all a, then as $\alpha \to 0^+$ the limiting equilibrium thresholds and hidden action profile are

$$\lim_{\alpha \to 0^+} \theta_\alpha^* = 1^-, \qquad \lim_{\alpha \to 0^+} x_\alpha^* = +\infty, \quad \text{and} \quad \lim_{\alpha \to 0^+} a_\alpha(\theta) = 0^+ \quad \text{for all } \theta$$

14

So for high enough α the signal-jamming technology is maximally effective. Moreover, there is a partial converse.[18] If costs are strictly convex then for low enough α the signal-jamming technology is ineffective in that $\theta^* > \theta^*_{\mathrm{MS}} = 1 - p$. If so, regimes would want to be able to credibly commit to not use the signal-jamming technology.

To characterize signal-jamming in the general case, rearrange (21) to get an alternative implicit characterization of the hidden actions

$$a(\theta) = x^* - \theta + \sqrt{\frac{2}{\alpha} \log \left(\frac{\sqrt{\alpha}\phi(0)}{C'[a(\theta)]} \right)}, \qquad \theta \geq \theta^* \tag{27}$$

which is (21) generalized to arbitrary convex costs but at the expense of losing the closed-form solution. By implicitly differentiating with respect to θ and rearranging, it's possible to show that $a'(\theta) \geq -1$ with strict equality if $C''(a) > 0$. But as $\alpha \to \infty$, regimes that manipulate have $a'(\theta) \to -1$ so that signals are locally uninformative. Moreover, since generally $a(\theta^*) > 0$, the signal mean for a regime that intervenes is strictly larger than θ^* so that signal-jamming is effective as α becomes sufficiently large because each individual worries about a large number of others drawing signals that suggest the regime is not going to be overthrown.

Numerical examples. With general cost functions the model cannot be solved analytically. Figure 3 shows θ^*_α as a function of precision α under the assumption that $C(a) := 0.5a^2$ for three levels of p. The higher the individual opportunity cost p, the lower the threshold and the thresholds are decreasing in the signal precision. In these examples, the speed of convergence to the limit is faster if p is high and slower if p is low. Regimes that inhabit a world where the individual cost of subversion p is high may benefit most from a given increase in α.

Interpretation. These results suggest that a regime's less overt propaganda apparatus (pressure exerted on editors, generals forced to stand on parade, etc) will be more useful when individuals are receiving, from a technological standpoint, intrinsically high quality signals. In equilibrium signals may be very uninformative, but that is precisely because the regime is co-opting the technology to its own ends.[19] A regime will want to exert a strong influence over the media when the signal precision is high enough.

Role of imperfect coordination. Disaggregated information is essential to the regime's ability to benefit from higher signal precision. To see this, suppose that citizens were perfectly coordinated and able to act as a large agent of mass one who could force regime change for all $\theta < 1$ (as detailed in Appendix B). Suppose also that this mass receives one signal x with precision α that

---

[18]As discussed following equation (21) if the marginal cost at zero is too large $C'(0) > \bar{c} := \sqrt{\alpha}\phi(0)$, then the cost of information manipulation is so high that the model reduces to the standard Morris-Shin game. When we take $\alpha \to \infty$ this bound does not matter. When we take $\alpha \to 0^+$ this bound will be violated. Consequently, the second part of Theorem 2 deals only with the case of strictly convex costs.

[19]A clear example of a regime co-opting new technologies for propaganda purposes is the sponsored diffusion of the cheap *Volksempfänger* radio set in 1930s Germany. By 1939, 70% of households owned a set — the highest proportion in the world at the time (Zeman, 1973, 34-62).
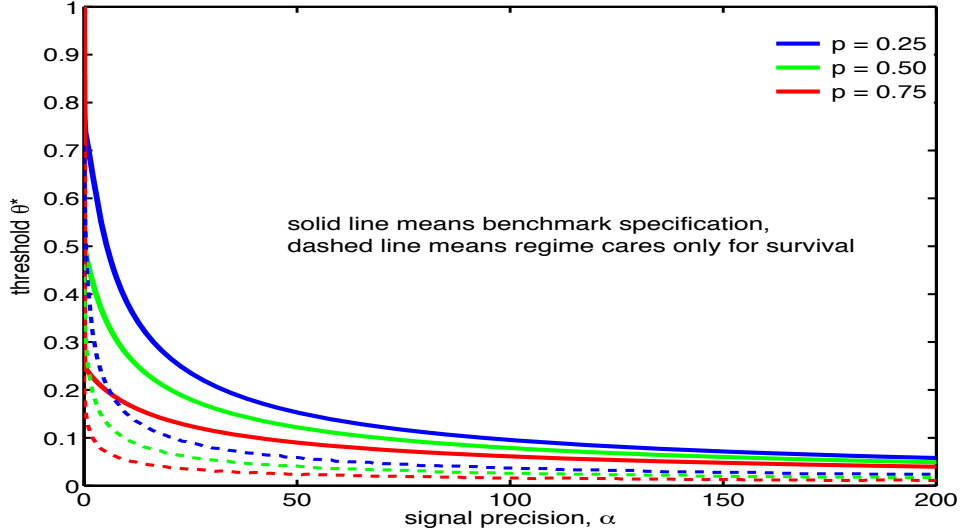
Figure 3: In the Morris-Shin game, thresholds are $\theta^*_{MS} = 1 - p$ all α. With information manipulation, large α implies $\theta^*_\alpha \to 0^+$, so in the limit all 'fragile' regimes with $\theta \in [0,1)$ survive. Solid lines show the benchmark specification while dashed lines show the alternative specification where the regime cares only about survival (see Section 4 below). In either case, better information increases a regime's likelihood of survival. All calculations use $C(a) := 0.5a^2$.

is contaminated by the regime's action. As in the benchmark model, for high precision the regime receives almost no benefit from taking a costly action and $a(\theta) \to 0$ for all θ as $\alpha \to \infty$. But if citizens are coordinated, they then know that $x \to \theta$ and so attack if and only if $x = \theta < 1$. If citizens are perfectly coordinated, all fragile regimes with $\theta \in [0,1)$ are wiped out. But if citizens are imperfectly coordinated all fragile regimes with $\theta \in [0,1)$ survive.

3.4   Equilibrium bias

Suppose that a citizen with signal x_i was 'naive' and believed she lived in the Morris-Shin world where no manipulation takes place. Her posterior expectation for θ would equal x_i. As a simple measure of any bias in equilibrium beliefs, let $b(x_i)$ denote the difference between the rational expectation of θ and the naive expectation

$$b(x_i) := \int_{-\infty}^{\infty} \theta \pi(\theta|x_i) - x_i \tag{28}$$

If $b(x_i) > 0$, the regime's propaganda has successfully induced this citizen to believe that θ is relatively high. Otherwise, if $b(x_i) < 0$, this citizen believes θ is relatively low. The average $\int b(x_i) f(x_i|\theta, a(\theta)) dx_i = -a(\theta) \leq 0$, so on average the population generally discounts signals.

Depending on α either (i) no-one is induced to believe that θ is relatively high, or (ii) there is disagreement about whether θ is relatively high or not. If α is sufficiently high, the distribution of x_i is almost degenerate for any θ so the average b can only be non-positive if each $b(x_i) \leq 0$ for all x_i. Alternatively, if α is sufficiently low, then the conditional expectation of θ is almost

16

independent of x_i and so there exist some x_i who are fooled in equilibrium, $b(x_i) > 0$. The top panel of Figure 4 illustrates.
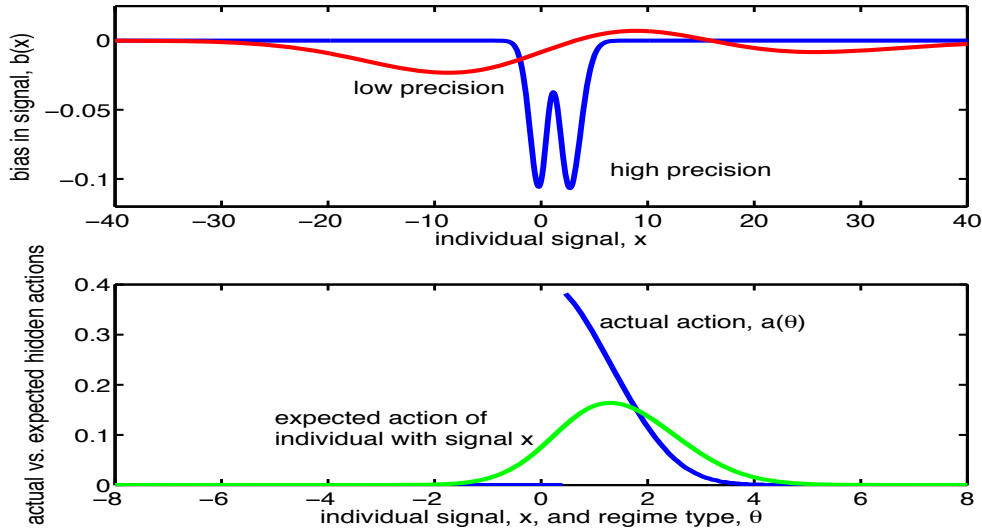

Figure 4: Equilibrium bias: if signal precision is low, some citizens believe the regime's propaganda and have bias $b(x_i) > 0$, but this does not happen if signal precision is high (top panel). While citizens on average discount their signals, they may not discount their signals enough. There are many θ such that $a(\theta) \geq \int a(\theta)\pi(\theta|x_i)d\theta$ for every x_i (bottom panel).

Citizens on average discount their information in an attempt to filter out the regime's manipulation. But this does not mean they discount their signals *enough*. The bottom panel of Figure 4 shows the relationship between $a(\theta)$, which is the amount by which citizens should discount their information, and $\int a(\theta)\pi(\theta|x_i)d\theta$, which is the amount that someone with x_i does discount their signal. Critically, there are regimes with $\theta \geq \theta^*$ that engage in more manipulation in equilibrium than is believed by *any* citizen. Citizens 'under-discount' the manipulation done by fragile regimes just above the threshold θ^*. Similarly, since $a(\theta) = 0$ for $\theta < \theta^*$ and $a(\theta) \to 0$ for large θ, most citizens 'over-discount' the manipulation done by either very weak or very strong regimes.

The model that delivers these results is stylized and it's natural to question its robustness. To investigate this, Section 4 extends the model to consider regimes that only care about survival. Regimes still benefit from better information. Alternative information structures are covered in Section 5: for some but not all of these structures, regimes benefit from better information.

4   Regime cares only about survival

Let the payoff to a regime be $B(S, \theta) - C(a)$ where now

$$B(S, \theta) = \mathbb{1}\{\theta \geq S\}\theta \tag{29}$$

so that a regime cares about S only to the extent that a high S makes it harder to survive.

Characterization of equilibrium. With these payoffs only fragile regimes with $\theta \in [0,1)$ could possibly have an incentive to manipulate. The mass of subversives is $S(\theta, a) = \Phi[\sqrt{\alpha}(x^* - \theta - a)]$ and if an action $a(\theta) > 0$ is chosen, it is set at just the right level to ensure that $\theta = S(\theta, a)$. Specifically

$$a(\theta) = x^* - \theta - \frac{1}{\sqrt{\alpha}}\Phi^{-1}(\theta), \qquad \theta \in [\theta^*, \theta^{**}) \tag{30}$$

and zero otherwise. Any higher an action would incur a cost for no benefit. The upper boundary $\theta^{**} \in [\theta^*, 1)$ is the unique solution to $\theta^{**} = \Phi[\sqrt{\alpha}(x^* - \theta^{**})]$.

Given a signal threshold x^*, equation (30) determines the action policy and the threshold below which the regime is overthrown is determined by $\theta^* = C[a(\theta^*)]$. This implicitly determines θ^* as a function of x^*. The action profile in (30) can then be used to calculate posterior beliefs for a citizen with signal x_i as in equation (15) and this implicitly determines a threshold x^* as a function of θ^*. An equilibrium is calculated by simultaneously solving these two conditions for x^*, θ^*.

It turns out that as, in the benchmark specification, signal-jamming is effective when α is high: $\theta^*_\alpha \to 0^+$ as $\alpha \to \infty$.

Constant marginal costs revisited. To provide intuition for this, it's helpful to again specialize to the case where $C(a) := ca$ and to consider the case when α is large. Let $x^*_\alpha, \theta^*_\alpha$ and $a_\alpha(\theta)$ denote the equilibrium objects in the benchmark economy and let $\hat{x}^*_\alpha, \hat{\theta}^*_\alpha$ and $\hat{a}_\alpha(\theta)$ denote the same objects in the economy where the regime cares only about survival.

From (21) we have approximately $a_\alpha(\theta) = \max[0, x^*_\alpha - \theta]$ when α is large. But from (30) we also have approximately $\hat{a}_\alpha(\theta) = \max[0, \hat{x}^*_\alpha - \theta]$ when α is large. So if $x^*_\alpha = \hat{x}^*_\alpha$, then $a_\alpha(\theta) = \hat{a}_\alpha(\theta)$ for each θ. The thresholds $\theta^*_\alpha, \hat{\theta}^*_\alpha$ are determined similarly: θ^*_α is determined by $\theta^*_\alpha = ca_\alpha(\theta^*_\alpha)$ from (24) when $\alpha \to \infty$ and if the regime cares only about survival $\hat{\theta}^*_\alpha = c\hat{a}_\alpha(\hat{\theta}^*_\alpha)$ for any α. So again if $x^*_\alpha = \hat{x}^*_\alpha$, then $a_\alpha(\theta^*_\alpha) = \hat{a}_\alpha(\hat{\theta}^*_\alpha)$ so that $\theta^*_\alpha = \hat{\theta}^*_\alpha$. But for either payoff specification for the regime the signal threshold is uniquely determined by the citizen indifference condition (17) once the hidden actions and state threshold are known. So for large α indeed $x^*_\alpha = \hat{x}^*_\alpha$ and the entire equilibrium will be approximately the same under either specification.

To further check robustness, I have solved the model numerically with strictly convex costs and when α is small. Figure 3 shows θ^*_α as a function of α for various levels of p. The qualitative similarity between the model with the benchmark payoffs and the alternative payoffs is clear even for relatively low α.

5   Alternative information structures

Until now, improvements in the quality of information were modeled as increases in the precision α. But in the benchmark model there is only one kind of information and it is affected by the regime's action. Section 5.1 allows citizens to have clean sources of information *uncontaminated* by the regime's actions. Section 5.2 changes the technology available to the regime, giving it the ability to manipulate the precision of signal information rather than the signal mean.

5.1 Uncontaminated information

I first consider the case where citizens are endowed with a second private signal independent of the regime's a and then turn to the case of a public signal. This second exercise connects the model to recent papers in the global games literature regarding the differential effects of private and public information on equilibrium uniqueness.

Heterogeneous informative priors.   Let citizens have two idiosyncratically noisy signals

$$\begin{aligned} x_i &= \theta + a &+& \varepsilon_{x,i} \\ z_i &= \theta &+& \varepsilon_{z,i} \end{aligned} \tag{31}$$

where $\varepsilon_{z,i}$ is normal noise with precision α_z independent of $\varepsilon_{x,i}$ which has precision α_x. The z_i signals can be treated as heterogeneous priors and unlike x_i are not contaminated by the regime's choice of a. This is equivalent to idiosyncratically noisy information on the hidden action, say $y_i = a + \xi_i$ where $y_i := x_i - z_i$ and $\xi_i := \varepsilon_{x,i} - \varepsilon_{z,i}$.

A monotone equilibrium consists of hidden actions $a(\theta)$, state threshold θ^* and signal thresholds $x^*(z_i)$ such that a citizen with x_i, z_i subverts if $x_i < x^*(z_i)$. Unfortunately, it is difficult to provide an analytic characterization of the equilibrium with both contaminated and uncontaminated private information, so the discussion below merely provides some numerical examples. The key finding is that an exogenous increase in the precision of *uncontaminated* information α_z may still reduce θ^*. See Appendix C for details on the computations.
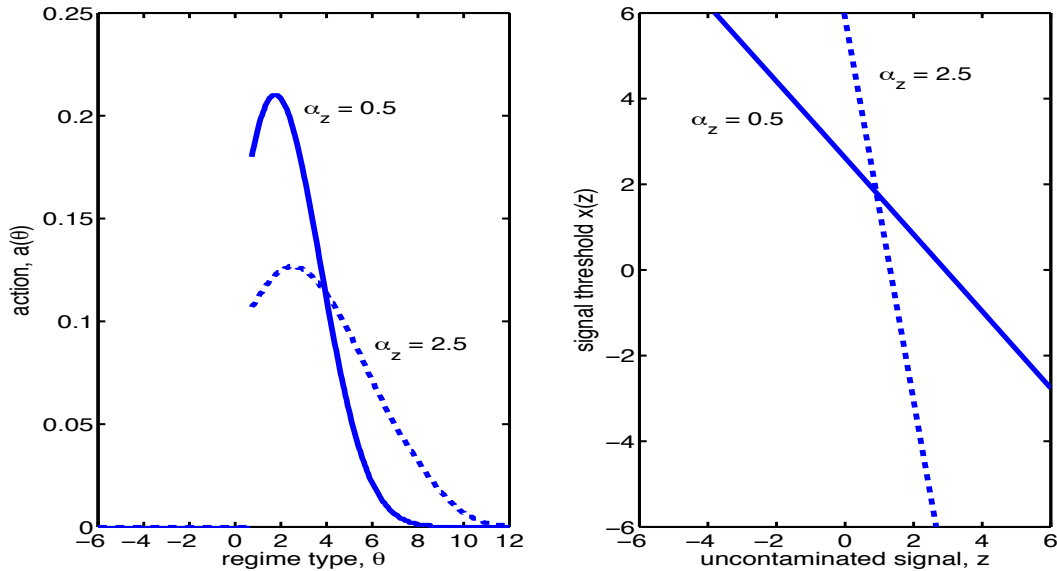

Figure 5: As precision α_z of uncontaminated private information increases, regimes near θ^* take smaller actions $a(\theta)$ but diminishing returns do not set in at such low values of θ. Citizens give more weight to their uncontaminated information so $x^*(z_i)$ is steeper: for low z_i it takes a higher contaminated x_i to induce subversion.

Figure 5 shows the actions $a(\theta)$ and the thresholds $x^*(z_i)$ for $\alpha_z = \alpha_x = 0.5$ and for $\alpha_z = 2.5$. As the uncontaminated information increases in quality, regimes with types larger than θ^* take
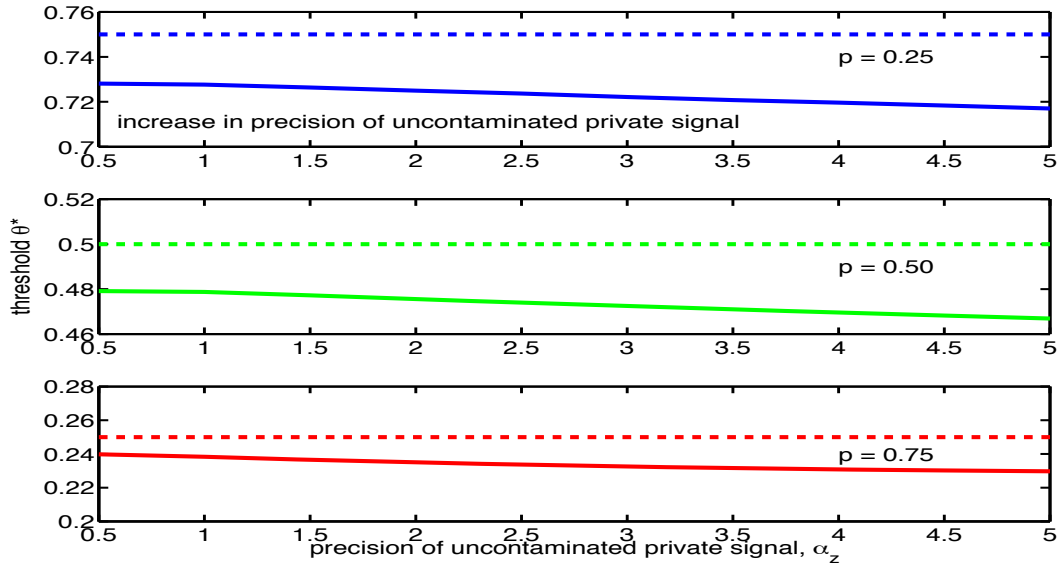
Figure 6: Thresholds θ^* as function of precision of uncontaminated private information α_z. In these examples a regime may benefit even from an increase in the quality of 'clean' information. Equilibrium thresholds are relatively insensitive to α_z.

smaller actions. But when α_z is high, diminishing returns do not set in so quickly: $a(\theta)$ is noticeably positive for more θ. The right hand panel shows the corresponding thresholds $x^*(z_i)$; these are decreasing because if an individual gets a low z_i it takes a high x_i to induce subversion. As α_z increases, $x^*(z_i)$ becomes steeper so that the z_i are weighed more heavily and it takes, for example, an even bigger x_i to compensate for a low z_i.

Figure 6 shows θ^* as a function of the uncontaminated precision α_z for various opportunity costs p. The θ^* are lower than the Morris-Shin levels $1 - p$ so that signal-jamming is effective and are decreasing in the precision α_z of the uncontaminated private information. But compared to the model with only contaminated information, the effects of an increase in α_z on θ^* are small.

These examples are only suggestive of what can happen in equilibrium. But it is clear that introducing uncontaminated private information does not necessarily overturn the results of Section 3. A *clean* information revolution may still increase the survival probability of a regime.

Common informative priors/public signal. Let citizens have a public signal

$$
\begin{aligned}
x_i &= \theta + a &+ \varepsilon_{x,i} \\
z &= \theta &+ \varepsilon_z
\end{aligned}
\tag{32}
$$

The realization of z is common knowledge and again is an uncontaminated signal of θ.

If no hidden actions are possible, the model reduces to a setup studied by Angeletos and Werning (2006), Hellwig (2002), Metz (2002), and Morris and Shin (2000, 2003) and others. For each z, a monotone equilibrium is a threshold $x^*(z)$ such that individuals subvert if $x_i < x^*(z)$ and a $\theta^*(z)$ such that regimes are overthrown if $\theta < \theta^*(z)$. It is well known that if public information is too precise relative to private information, there may be multiple monotone equilibria [see, e.g.,

Hellwig (2002) and Morris and Shin (2003, 2004)]. If public information is too precise, there is 'approximate' common knowledge of θ. Hellwig (2002) derived a sufficient condition for a unique monotone equilibrium in a game of this kind, namely $\alpha_z/\sqrt{\alpha_x} < \sqrt{2\pi}$. In the discussion that follows, I assume that this is satisfied.

What happens if the precision of private information increases? As $\alpha_x \to \infty$ for given α_z (or as $\alpha_z \to 0$ for given α_x) the public signal becomes uninformative and we revert to the Morris-Shin game with $x^*(z) \to \theta^*(z)$ and $\theta^*(z) \to 1 - p$ independent of the realization of the public signal z. If the quality of private information is sufficiently good, the public information z is irrelevant.

Metz (2002) characterized the direction from which $\theta^*(z)$ converges to $1 - p$ as $\alpha_x \to \infty$. If the parameters p or z are favorable to the regime, then $\theta^*(z) \nearrow 1 - p$ but if the parameters p or z are unfavorable to the regime, $\theta^*(z) \searrow 1 - p$. Why? Both x_i and z are informative about θ, but only x_i is informative about the role of coordination. Also, it is common knowledge that signals are bunched around θ and that every citizen gives weight to their idiosyncratic signal in proportion to its quality. Now consider an economy with high z (which suggests the regime is going to be difficult to beat, since high z is correlated with high θ). For moderate α_x, citizens will give some weight to this public signal and will be less inclined to engage in subversion. So for moderate precision, $\theta^*(z)$ is low and the ex ante survival probability of the regime is high. But as α_x increases, the influence of the high realized z diminishes because everybody knows that everybody gives less weight to z when α_x increases. In the limit, only the opportunity cost p matters and $\theta^*(z) \nearrow 1 - p$. I refer to this as the *coordination effect* from increasing idiosyncratic signal precision.

With hidden actions the coordination effect is dominant for low α_x and $\theta^*(z)$ is increasing in α_x if p or z is high but decreasing in α_x if p or z is low. But the coordination effect is limited: when α_x is large, people ignore z and the effects of further increase in α_x are almost nil. Now recall the basic hidden action model with no public signal: the state threshold approached zero as the idiosyncratic signal precision became large. So we expect for high α_x, the existence of a public signal is almost immaterial and the *hidden action effect* of information manipulation is dominant.

Figure 7 suggests that this intuition is correct. When primitives are favorable to the regime, the coordination effect and the hidden action effect pull in opposing directions. The coordination effect tends to drive the equilibrium threshold up and to reduce the survival probability of the regime. But for high α_x the coordination effect is irrelevant while the hidden action effect still matters. In this case, the thresholds have a 'hump-shaped' look and asymptote to zero. But if primitives are unfavorable to the regime, both the coordination and hidden action effects are in the regime's favor and reinforce each other. In this case, the threshold is monotone decreasing and asymptotes to zero. If primitives are unfavorable to the regime, there is a large benefit from information manipulation (this contrasts with the benchmark model where regimes with favorable primitives benefitted more from a given increase in α_x).

Uncontaminated public information does not undo the central message of the hidden actions model. For high α_x signal-jamming is effective, $\theta^* < \theta^*_{\text{MS}} = 1 - p$. Better private information may increase the regime's ex ante survival probability. But perhaps this is not the most interesting
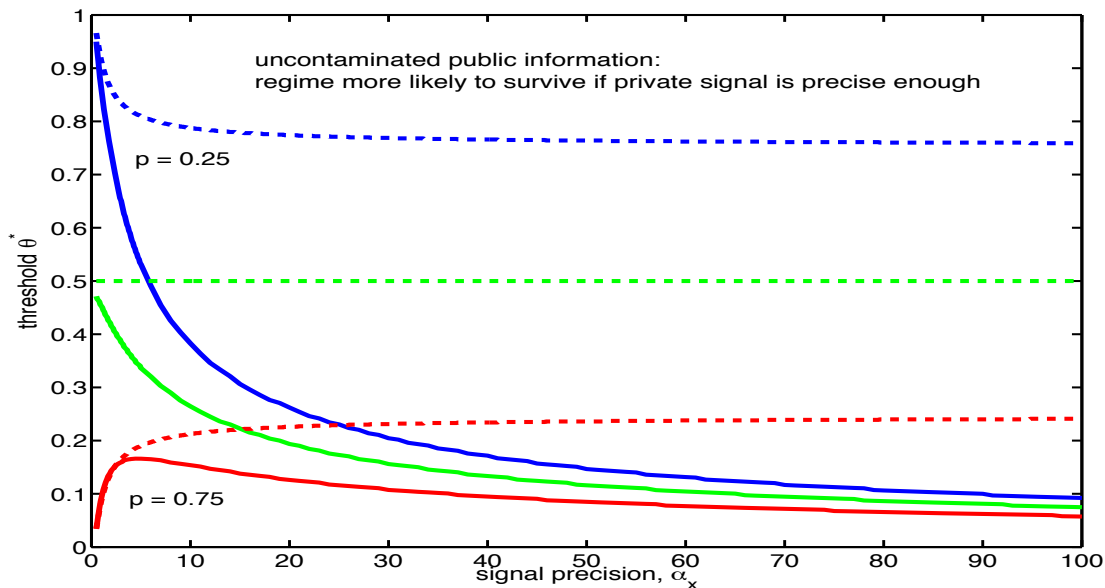
Figure 7: Thresholds θ^* as functions of α_x. Dashed lines show the thresholds when there are no hidden actions. Solid lines show θ^* when hidden actions contaminate the private information. If primitives are favorable (say $p = 0.75$), the coordination effect identified by Metz (2002) and the hidden action effect pull in opposite directions, giving rise to a 'hump-shaped' function. If primitives are unfavorable to the regime (say $p = 0.25$), both forces drive θ^* down.

comparative static. What if the precision of the *public* signal increases? Then we run into the problem of multiplicity. For given α_x, the inequality $\alpha_z/\sqrt{\alpha_x} < \sqrt{2\pi}$ will eventually be violated and there are multiple monotone equilibria. In this case, we lose the ability to draw sharp conclusions. What if both the precisions of the public and private information increase together? Then if the bound on the relative precision of the public information is satisfied the analysis goes through essentially as above. At a given level of the relative precision $\alpha_z/\sqrt{\alpha_x}$, the threshold may be increasing or decreasing (depending on the relative strengths of the coordination effect and the hidden action effect), but as the ratio $\alpha_z/\sqrt{\alpha_x}$ becomes small the hidden action effect dominates.

5.2   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 *mean* of the signal distribution without (directly) influencing the precision. Now let signals be $x_i := \theta + \varepsilon_i$ with the precision of ε_i given by

$$\beta(a) := \alpha \left(\frac{1}{2} + \Phi(a) \right), \qquad \alpha > 0 \tag{33}$$

With this specification, negative actions $a < 0$ are possible and we need to reinterpret the cost function C. When $a > 0$, the regime sets a precision $\beta(a) > \alpha$ and when $a < 0$, the regime sets a precision $\beta(a) < \alpha$. Accordingly, I refer to α as the *intrinsic precision* of private information.

In equilibrium, hidden actions $a(\theta)$ are characterized by the first order necessary condition

$$C'(a) = (\theta - x^*)\sqrt{\beta(a)}\phi\left[\sqrt{\beta(a)}(\theta - x^*)\right]\frac{\alpha}{2}, \qquad \theta \geq \theta^* \tag{34}$$
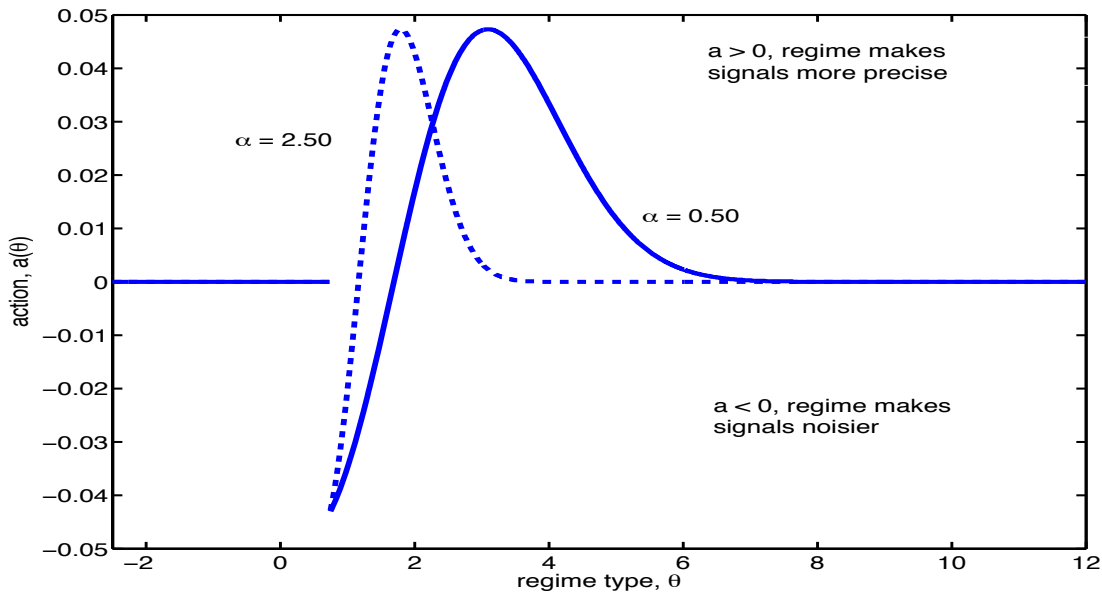
22

Figure 8: Equilibrium hidden actions $a(\theta)$ when regime can manipulate signal precision. For intermediate θ it may be optimal for the regime to 'muddy' the signal, $a(\theta) < 0$, relative to the benchmark α. For high θ it is optimal for the regime to clarify its strength, $a(\theta) > 0$.

Actions are zero for $\theta < \theta^*$ before jumping discontinuously at θ^*. If $x^* > \theta^*$, the jump is down but if $x^* < \theta^*$ the jump is up. Intuitively, if the regime has intermediate type $\theta \in [\theta^*, x^*)$ then it worsens the signal noise relative to α (it *muddies* the signal). Otherwise, if $\theta > x^*$ then the regime finds it worthwhile to *clarify* its position of strength and so increase the signal precision relative to α. Figure 8 illustrates with parameters chosen so $x^* > \theta^*$ and the profile jumps down at θ^*.

Transparency.   In related work Bannier and Heinemann (2005) and Heinemann and Illing (2002) study higher signal precision (more 'transparency') in the Morris-Shin game by allowing the regime to choose the individuals' signal precision to minimize the probability of a successful attack. But in their formulation the regime's choice of precision does not depend on its type θ and so when forming beliefs individuals can take this precision as exogenous. In Figure 8, a regime reduces transparency in this sense if $\theta \in [\theta^*, x^*)$ and increases transparency if $\theta > x^*$.

6   Final remarks

The coordination games studied in this paper are deliberately stylized so as to focus attention on the use of information manipulation as a tool for influencing equilibrium outcomes. In keeping things simple, I have abstracted from issues that might play a role in the development of a more complete model of information, coordination and regime change. For example, information accumulation over time, non-simultaneous moves by groups of citizens and strategic communication between individuals might all be important components of a richer analysis. Moreover, for some applications it might be desirable to have several 'large' players with competing interests over equilibrium

outcomes (for example, if one large player has an interest in the status quo while a rival large player seeks its overthrow). In such situations, individuals beliefs and equilibrium outcomes would depend on the interactions between several hidden actions each of which is trying to push beliefs in a different direction.

A   Proofs

PROOF OF PROPOSITION 1. (Morris-Shin): Let $\hat{x}, \hat{\theta}$ denote candidate thresholds. Posterior beliefs of a citizen with x_i facing threshold $\hat{\theta}$ are $\Pr(\theta < \hat{\theta}|x_i) = \Phi[\sqrt{\alpha}(\hat{\theta} - x_i)]$. 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 monotonically 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 a threshold rule about $\hat{x}$ it faces mass $\hat{S}(\theta) = \Phi[\sqrt{\alpha}(\hat{x} - \theta)]$. 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 monotonically 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. It is straightforward to verify that there is only one solution to these equations and that $\theta_{\text{MS}}^* = 1 - p$ independent of α. □

PROOF OF LEMMA 1. Fix a $\hat{x} \in \mathbb{R}$ and, for notational simplicity, write θ for the state threshold and x for an individual's signal. From (15), the probability $P(\theta, x)$ that an individual with x assigns to the regime's type being less that θ can be written

$$P(\theta, x) = \frac{A(\theta - x)}{A(\theta - x) + B(\theta, x)} \tag{35}$$

where

$$A(\theta - x) := \Phi[\sqrt{\alpha}(\theta - x)], \quad \text{and} \quad B(\theta, x) := \int_{\theta}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x - \xi - a(\xi - \hat{x}))]d\xi \tag{36}$$

Since $A' > 0$ and $B_{\theta} < 0$, $P_{\theta} > 0$ for all θ, x. But $P_x < 0$ if and only if $A'/A > -B_x/B$. Calculating the derivatives shows that this is equivalent to

$$H[\sqrt{\alpha}(x - \theta)] > -\frac{\int_{\theta}^{\infty} \phi'[\sqrt{\alpha}(x - y(\xi))]d\xi}{\int_{\theta}^{\infty} \phi[\sqrt{\alpha}(x - y(\xi))]d\xi} = \frac{\int_{\theta}^{\infty} \sqrt{\alpha}(x - y(\xi))\phi[\sqrt{\alpha}(x - y(\xi))]d\xi}{\int_{\theta}^{\infty} \phi[\sqrt{\alpha}(x - y(\xi))]d\xi} \tag{37}$$

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

$$\varphi(\xi|x) := \frac{\phi[\sqrt{\alpha}(x - y(\xi))]}{\int_{\theta}^{\infty} \phi[\sqrt{\alpha}(x - y(\xi'))]d\xi'}, \qquad \xi \in [\theta, \infty) \tag{38}$$

Then after a slight rearrangement of terms in (37), $P_x < 0$ if and only if

$$H[\sqrt{\alpha}(x - \theta)] - \sqrt{\alpha}(x - \theta) > \sqrt{\alpha}\left[\theta - \int_{\theta}^{\infty} y(\xi)\varphi(\xi|x)d\xi\right] \tag{39}$$

24

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

$$\int_{\theta}^{\infty} y(\xi)\varphi(\xi|x)d\xi \geq \theta \tag{40}$$

But since $y(\xi) := \xi + a(\xi - \hat{x})$, $\xi \geq \theta$, and $a(\xi - \hat{x}) \geq 0$, condition (40) is always satisfied. Therefore $P_x < 0$. Finally, using the representation of P given in (35), we have that for $P(\theta, x) = P(\theta - \hat{x}, x - \hat{x})$ it is sufficient that $B(\theta, x) = B(\theta - \hat{x}, x - \hat{x})$. From (36) and using the linear change of variables $\xi' := \xi - \hat{x}$ we have

$$B(\theta, x) = \int_{\theta - \hat{x}}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x - \hat{x} - \xi' - a(\xi'))]d\xi' = B(\theta - \hat{x}, x - \hat{x}) \tag{41}$$

Therefore $P(\theta, x) = P(\theta - \hat{x}, x - \hat{x})$ too. □

PROOF OF LEMMA 2. Since $P(\theta, x)$ is continuously differentiable in x, an application of the implicit function theorem to (19) shows that ψ is continuous. Fixed points of ψ satisfy $x^* = \psi(x^*)$. Equivalently, by Lemma 1, they satisfy $P[\theta(x^*) - x^*, 0] = p$, where, in a slight abuse of notation, $\theta(\hat{x})$ is the critical state in the regime's problem (13)-(14). Since $P_\theta > 0$ with $P(-\infty, 0) = 0$ and $P(\infty, 0) = 1$, by the intermediate value theorem there is a unique $\Delta \in \mathbb{R}$ such that $P(\Delta, 0) = p$. Applying the implicit function theorem to (13)-(14) gives

$$\theta'(\hat{x}) = \frac{\sqrt{\alpha}\phi\{\sqrt{\alpha}[\hat{x} - \theta(\hat{x}) - a(\theta(\hat{x}) - \hat{x})]\}}{1 + \sqrt{\alpha}\phi\{\sqrt{\alpha}[\hat{x} - \theta(\hat{x}) - a(\theta(\hat{x}) - \hat{x})]\}} \in (0, 1) \tag{42}$$

Since $\theta(-\infty) = 0$ and $\theta(\infty) = 1$, there is a unique $x^* \in \mathbb{R}$ such that $\theta(x^*) - x^* = \Delta$, hence ψ has a unique fixed point, the same x^*. Now from (19), we have

$$\psi'(\hat{x}) = 1 + \frac{P_\theta[\theta(\hat{x}), \psi(\hat{x})]}{P_x[\theta(\hat{x}), \psi(\hat{x})]}[1 - \theta'(\hat{x})] \tag{43}$$

By Lemma 1, $P_\theta > 0$ and $P_x < 0$ and $\theta'(\hat{x}) \in (0, 1)$ from (42). Therefore $\psi'(\hat{x}) < 1$ for all $\hat{x}$. To see that $\psi'(x^*) > 0$, first notice that it is sufficient that $P_\theta(\theta^*, x^*)/P_x(\theta^*, x^*) \geq -1$ where $\theta^* := \theta(x^*)$. Calculating the derivatives shows that this is true if and only if

$$\phi[\sqrt{\alpha}(y(\theta^*) - x^*)] + \int_{\theta^*}^{\infty} \sqrt{\alpha}\phi'[\sqrt{\alpha}(y(t) - x^*)]dt \leq 0 \tag{44}$$

where again $y(t) = t + a(t)$ is the mean of the signal distribution from which a citizen is sampling if the regime has type $t \geq \theta^*$. To show that this condition always holds, we need to consider two separate cases: (i) where costs are linear, and (ii) where costs are strictly convex. If costs are linear, then $a(t) = \max[0, x^* + \gamma - t]$ where $\gamma > 0$ [as in equation (21)]. Calculating the integral and then simplifying shows that (44) holds if and only if $-\alpha\gamma\phi(\sqrt{\alpha}\gamma)a(\theta^*) \leq 0$ which is true because $a(\theta^*) \geq 0$. Now consider case (ii) where costs are strictly convex. From the optimality conditions for the regime's choice of action we have that $a(t) > 0$ for all $t \geq \theta^*$ and

$$\sqrt{\alpha}\phi[\sqrt{\alpha}(y(t) - x^*)] = C'[a(t)], \quad t \geq \theta^* \tag{45}$$

Differentiating with respect to t gives

$$\alpha\phi'[\sqrt{\alpha}(y(t) - x^*)]y'(t) = C''[a(t)]a'(t), \quad t \geq \theta^* \tag{46}$$

Using the associated second order condition shows that $y'(t) > 0$ for $t \geq \theta^*$. Since y is invertible, a change of variables shows that (44) holds if and only if

$$\int_{\theta^*}^{\infty} \phi'[\sqrt{\alpha}(y(t) - x^*)] \frac{a'(t)}{y'(t)} dt \geq 0 \tag{47}$$

Using (46) we equivalently have the condition

$$\int_{\theta^*}^{\infty} \frac{\phi'[\sqrt{\alpha}(y(t) - x^*)]^2}{C''[a(t)]} dt \geq 0 \tag{48}$$

which is true since the integrand is non-negative. Therefore, $P_\theta(\theta^*, x^*)/P_x(\theta^*, x^*) \geq -1$ and so from equation (43), $\psi'(x^*) > 0$.

Finally, $\psi(\hat{x}) < x^*$ for every $\hat{x} < x^*$ is proven by contradiction. Suppose not. Then by continuity of ψ there exists $\tilde{x} < x^*$ such that $\psi(\tilde{x}) = x^*$. By the definition of ψ in (19) this implies $P[\theta(\tilde{x}) - \tilde{x}, \psi(\tilde{x}) - \tilde{x}] = p$ and so by Lemma 1 and $\psi(\tilde{x}) = x^*$ simply $P[\theta(\tilde{x}), x^*] = p$. But by (42) and the hypothesis $\tilde{x} < x^*$ we must also have $\theta(\tilde{x}) < \theta(x^*)$ and since P is strictly increasing in θ we have

$$p = P[\theta(\tilde{x}), x^*] < P[\theta(x^*), x^*] = p \tag{49}$$

which is the needed contradiction. Therefore $\psi(\hat{x}) < x^*$ for every $\hat{x} < x^*$. A symmetric argument shows $\psi(\hat{x}) > x^*$ for every $\hat{x} > x^*$. $\qquad \square$

PROOF OF THEOREM 1. First, there is a unique equilibrium if strategies are restricted to be monotone. To show this, first solve (13) for arbitrary $\hat{x}, \hat{\theta}$. This provides the family $a(\theta - \hat{x})$ which is used to construct the P function in (15). Then from Lemma 1 there is a unique $\Delta \in \mathbb{R}$ such that $P(\Delta, 0) = p$. Plugging this value into (14) gives the unique $\theta^* \in [0, 1]$ such that the regime is overthrown if and only if $\theta < \theta^*$ and then the unique x^* such that a citizen subverts if and only if $x_i < x^*$ is $x^* = \theta^* - \Delta$. The equilibrium profile of hidden actions is then $a(\theta) := a(\theta - x^*)$. By Lemma 2, this x^* is the unique fixed point of the ψ function.

To show that there is no other equilibrium, we show that the only strategy profile for the citizens that survives the iterative elimination of (interim) strictly dominated strategies is the monotone strategy $s(x) = 1$ if and only if $x < x^*$. Recall that when almost all citizens subvert if their signal is less than $\hat{x}$, a citizen with x subverts if and only if $x < \psi(\hat{x})$. Let $\underline{x}_0 := \underline{x}$ and $\overline{x}_0 := \overline{x}$ [as in (18)]. Now generate $\{\underline{x}_n\}_{n=0}^{\infty}$ from $\underline{x}_{n+1} = \psi(\underline{x}_n)$ and generate $\{\overline{x}_n\}_{n=0}^{\infty}$ from $\overline{x}_{n+1} = \psi(\overline{x}_n)$. By Lemma 2, ψ is continuous, has a unique fixed point $x^* = \psi(x^*)$ with $\psi'(x^*) \in (0, 1)$ and $\psi(\underline{x}_n) < x^*$ for all $\underline{x}_n < x^*$. Therefore, 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 $\{\overline{x}_n\}_{n=0}^{\infty}$ is bounded below, strictly monotone decreasing and so converges $\overline{x}_n \searrow x^*$. After n iterations, the only candidates for an equilibrium strategy all have $s(x) = 1$ for $x < \underline{x}_n$ and $s(x) = 0$ for $x > \overline{x}_n$ with $s(x)$ arbitrary for $x \in [\underline{x}_n, \overline{x}_n]$. In the limit as $n \to \infty$, the only strategy that survives the elimination of strictly dominated strategies is the one with $s(x) = 1$ for $x < x^*$ and $s(x) = 0$ otherwise. Therefore the only equilibrium is the unique monotone equilibrium. $\qquad \square$

PROOF OF THEOREM 2. For each precision α, there is a unique equilibrium. I find a unique solution to a constrained problem consisting of the original system of nonlinear equations plus a set of constraints that govern the asymptotic behavior of the endogenous variables. But, 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

$$(1 - p)\Phi[\sqrt{\alpha}\,(\theta_\alpha^* - x_\alpha^*)] = p \int_{\theta_\alpha^*}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_\alpha^* - \theta - a_\alpha(\theta - x_\alpha^*))]d\theta \tag{50}$$

and

$$\theta_\alpha^* = \Phi[\sqrt{\alpha}(x_\alpha^* - \theta_\alpha^* - a_\alpha(\theta_\alpha^* - x_\alpha^*))] + C[a_\alpha(\theta_\alpha^* - x_\alpha^*)] \tag{51}$$

with actions characterized by

$$C'[a_\alpha(\theta - x_\alpha^*)] = \sqrt{\alpha}\phi[\sqrt{\alpha}(x_\alpha^* - \theta - a_\alpha(\theta - x_\alpha^*))], \qquad \theta \geq \theta_\alpha^* \tag{52}$$

Now let $\alpha \to \infty$. The auxiliary constraints that govern the asymptotic behavior of the endogenous variables are assumed to be

$$\lim_{\alpha \to \infty} \sqrt{\alpha}(x_\alpha^* - \theta_\alpha^* - a_\alpha(\theta_\alpha^* - x_\alpha^*)) = \lim_{\alpha \to \infty} \sqrt{\alpha}(\theta_\alpha^* - x_\alpha^*) = -\infty \tag{53}$$

If condition (53) holds, from (51) we have $\theta_\alpha^* = C[a_\alpha(\theta_\alpha^* - x_\alpha^*)]$. Similarly, if (53) holds, then $\Phi[\sqrt{\alpha}\,(\theta_\alpha^* - x_\alpha^*)] \to 0$ and the value of the integral on the right hand side of (50) converges to zero. From (50) and (52), this requires

$$\lim_{\alpha \to \infty} \int_{\theta_\alpha^*}^{\infty} C'[a_\alpha(\theta - x_\alpha^*)]d\theta = 0$$

Since $\theta_\alpha^* \in [0, 1]$ and $C'[a_\alpha(\theta - x_\alpha^*)] \geq 0$ and is uniformly continuous in α, this can only be true if $a_\alpha(\theta - x_\alpha^*) \to 0^+$ for all $\theta \geq \theta_\alpha^*$. But then if $a_\alpha(\theta_\alpha^* - x_\alpha^*) \to 0^+$, $C[a_\alpha(\theta_\alpha^* - x_\alpha^*)] \to 0^+$ and so $\theta_\alpha^* \to 0^+$ too. Finally, if both constraints are to hold simultaneously for large α, $x_\alpha^* - \theta_\alpha^*$ is positive and $x_\alpha^* - \theta_\alpha^* - a_\alpha(\theta_\alpha^* - x_\alpha^*)$ is negative. For both constraints to have the same sign, x_α^* can neither diverge nor converge to either a strictly positive or a strictly negative number. So $x_\alpha^* \to 0^+$. Hence we have found a solution to the constrained problem.

Now for the second part of the Theorem. Recall that for this part we assume strictly convex costs. Let $\alpha \to 0^+$ such that $\sqrt{\alpha}x_\alpha^* \to \infty$ holds. Then $x_\alpha^* \to \infty$. Since $\theta_\alpha^* \in [0, 1]$, we have $\sqrt{\alpha}\,(x_\alpha^* - \theta_\alpha^*) \to \infty$ and the integral on the right hand side of (50) must converge to zero. Hence, by (52), $a_\alpha(\theta - x_\alpha^*) \to 0^+$ for all $\theta \geq \theta_\alpha^*$ (the strict convexity of C is assumed here so that (52) holds for all θ even as $\alpha \to 0^+$; with constant marginal costs, this would not be true). But if $a_\alpha(\theta_\alpha^* - x_\alpha^*) \to 0^+$, $\theta_\alpha^* \in [0, 1]$, and $\sqrt{\alpha}x_\alpha^* \to \infty$, then (51) requires that $\theta_\alpha^* \to 1^-$. Once again we have found a solution to the constrained problem. $\square$

B   The role of coordination: regime vs. single agent

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 $\Pr(\theta < 1|x^*) = p$.

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

$$\int_{-\infty}^{\infty} \max[0, \theta - S(x)]f(x|\theta, a)dx - C(a) = -\min[\theta, 1]\Phi[\sqrt{\alpha}(x^* - \theta - a)] - C(a) \tag{54}$$

Regimes with $\theta < 0$ are necessarily overthrown in equilibrium and so never engage in costly manipulation. For simplicity, suppose that costs are strictly convex, $C''(a) > 0$. If so, all regimes with $\theta > 0$ will choose some positive manipulation, even regimes that ex post are overthrown. The key first order necessary condition for the regime's choice of action $a(\theta)$ is

$$\min[\theta, 1]\sqrt{\alpha}\phi[\sqrt{\alpha}(x^* - \theta - a) = C'(a), \quad \theta \geq 0 \tag{55}$$

(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 $\Pr(\theta < 1|x^*) = p$.

Standard arguments show that taking as given x^* the regime's optimal action $a(\theta) \to 0^+$ as $\alpha \to \infty$. Given this policy by the regime, the probability of overthrowing the regime $\Pr(\theta < 1|x) \to \mathbb{1}\{x < 1\}$ as $\alpha \to \infty$. Therefore, $x^* \to 1$. 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 $x < 1$. This leads to the downfall of any fragile regime $\theta \in [0, 1)$. By contrast, Theorem 2 tells us that if citizens are imperfectly coordinated then for precise information any fragile regime $\theta \in [0, 1)$ survives.

C Heterogeneous priors: uncontaminated private information

Individuals have a second idiosyncratically noisy observation $z_i := \theta + \varepsilon_{z,i}$ in addition to $x_i := \theta + a + \varepsilon_{x,i}$. An equilibrium consists of functions $a(\theta)$ and $x^*(z_i)$ and a scalar θ^* where a citizen subverts if $x_i \leq x^*(z_i)$. In what follows, let $f_X(x_i|\theta, a)$ and $f_Z(z_i|\theta)$ denote the signal densities and let $\hat{a}(\theta)$, $\hat{x}(z_i)$ and $\hat{\theta}$ denote candidates for the equilibrium objects.

The posterior beliefs of a citizen with signals x_i, z_i facing a threshold rule about the point $\hat{\theta}$ and manipulation policy $\hat{a}(\theta)$ are summarized by

$$\Pr(\theta < \hat{\theta}|x_i, z_i) = \frac{\int_{-\infty}^{\hat{\theta}} \hat{\pi}(\theta|x_i, z_i)d\theta}{\int_{-\infty}^{\infty} \hat{\pi}(\theta|x_i, z_i)d\theta}$$

where $\hat{\pi}(\theta|x_i, z_i) := f_X[x_i|\theta, \hat{a}(\theta)]f_Z(z_i|\theta)$. A citizen with signals x_i, z_i will then subvert if and only if $\Pr(\theta < \hat{\theta}|x_i, z_i) \geq p$. Given $\hat{\theta}$ and $\hat{a}(\theta)$, this implicitly defines a marginal citizen with signal z_i by

$$\Pr(\theta < \hat{\theta}|\hat{x}(z_i), z_i) = p \tag{56}$$

The hidden actions for the regime are $\hat{a}(\theta) = 0$ for $\theta < \hat{\theta}$ and otherwise

$$\hat{a}(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \left[\theta - \hat{S}(\theta, a) - C(a)\right], \qquad \theta \geq \hat{\theta} \tag{57}$$

where the mass of subversives is

$$\hat{S}(\theta, a) = \int_{-\infty}^{\infty} \int_{-\infty}^{\hat{x}(z_i)} f_X(x_i|\theta, a)f_Z(z_i|\theta)dx_i dz_i = \int_{-\infty}^{\infty} F_X(\hat{x}(z_i)|\theta, a)f_Z(z_i|\theta)dz_i \tag{58}$$

and the cutoff $\hat{\theta}$ is determined by the indifference condition

$$\hat{\theta} = \hat{S}[\hat{\theta}, \hat{a}(\hat{\theta})] + C[\hat{a}(\hat{\theta})] \tag{59}$$

I solve this model numerically. I first guess values $\hat{x}_n(z_i)$, $\hat{\theta}_n$ (say) and then solve the optimization problem in (57) for the associated hidden actions $\hat{a}_{n+1}(\theta)$. I then use $\hat{\theta}_n$ and the implied $\hat{a}_{n+1}(\theta)$ to compute revised estimates of the thresholds $\hat{x}_{n+1}(z_i)$ from (56) and use $\hat{a}_{n+1}(\theta)$ and $\hat{x}_{n+1}(z_i)$ to compute a new $\hat{\theta}_{n+1}$ from the single condition (59). I iterate until the process converges. More details and Matlab code are available on request.

References

ANGELETOS, G.-M., C. HELLWIG, AND A. PAVAN (2006): "Signaling in a global game: Coordination and policy traps," *Journal of Political Economy*, 114(3), 452–484. 1, 3, 5, 10, 11

——— (2007): "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks," *Econometrica*, 75(3), 711–756. 1

ANGELETOS, G.-M., AND I. WERNING (2006): "Crises and prices: Information aggregation, multiplicity and volatility," *American Economic Review*, 96(5), 1720–1736. 1, 20

ATKESON, A. (2000): "Discussion," in *NBER Macroeconomics Annual*, ed. by B. S. Bernanke, and K. Rogoff, pp. 162–171. MIT Press. 2

BANNIER, C. E., AND F. HEINEMANN (2005): "Optimal transparency and risk-taking to avoid currency crises," *Journal of Institutional and Theoretical Economics*, 161, 374–391. 23

CARLSSON, H., AND E. VAN DAMME (1993): "Global games and equilibrium selection," *Econometrica*, 61(5), 989–1018. 2, 7

CHAMLEY, C. P. (1999): "Coordinating regime switches," *Quarterly Journal of Economics*, 114(3), 869–905. 2

——— (2004): *Rational herds: Economic models of social learning.* Cambridge University Press. 2, 9

CHELI, B., AND P. DELLA-POSTA (2002): "Self-fulfilling currency attacks with biased signals," Universita di Pisa. 2

COOPER, R. W. (1999): *Coordination games: Complementarities and macroeconomics.* Cambridge University Press. 2

CRAWFORD, V. P., AND J. SOBEL (1982): "Strategic information transmission," *Econometrica*, 50(6), 1431–1451. 3, 11

FROMMER, B. (2005): *National cleansing: Retribution against Nazi collaborators in postwar Czechoslovakia.* Cambridge University Press. 6

HEINEMANN, F., AND G. ILLING (2002): "Speculative attacks: Unique equilibrium and transparency," *Journal of International Economics*, 58, 429–450. 23

HELLWIG, C. (2002): "Public information, private information, and the multiplicity of equilibria in coordination games," *Journal of Economic Theory*, 107, 191–222. 7, 20, 21

JACKSON, J. (2001): *France: The dark years, 1940-1944.* Oxford University Press. 6

KALATHIL, S., AND T. C. BOAS (2003): *Open networks, closed regimes: The impact of the internet on authoritarian rule.* Carnegie Endowment for International Peace. 2

MATTHEWS, S. A., AND L. J. MIRMAN (1983): "Equilibrium limit pricing: The effects of private information and stochastic demand," *Econometrica*, 51(4), 981–996. 3, 10, 11

METZ, C. (2002): "Private and public information in self-fulfilling currency crises," *Zeitschrift fur Nationalokonomie*, 76(1), 65–85. 20, 21, 22

MILGROM, P., AND J. ROBERTS (1982): "Limit pricing and entry under incomplete information: An equilibrium analysis," *Econometrica*, 50(2), 443–460. 10

MORRIS, S., AND H. S. SHIN (1998): "Unique equilibrium in a model of self-fulfilling currency attacks," *American Economic Review*, 88(3), 587–597. 2, 7, 9

——— (2000): "Rethinking multiple equilibria in macroeconomic modeling," in *NBER Macroeconomics Annual*, ed. by B. S. Bernanke, and K. Rogoff, pp. 139–161. MIT Press. 7, 20

——— (2003): "Global games: Theory and applications," in *Advances in Economics and Econometrics: Theory and Applications*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge University Press. 2, 7, 9, 20, 21

——— (2004): "Coordination risk and the price of debt," *European Economic Review*, 48(1), 133–153. 2, 7, 21

OBSTFELD, M. (1986): "Rational and self-fulfilling balance-of-payments crises," *American Economic Review*, 76(1), 72–81. 2, 6

——— (1996): "Models of currency crises with self-fulfilling features," *European Economic Review*, 40(3–5), 1037–1047. 2, 6

RILEY, J. G. (2001): "Silver signals: Twenty-five years of screening and signaling," *Journal of Economic Literature*, 39(2), 432–478. 11

TIROLE, J. (1988): *The theory of industrial organization.* MIT Press. 11

WILSON, R. (1992): "Strategic models of entry deterrence," in *Handbook of game theory*, ed. by R. J. Aumann, and S. Hart. North-Holland/Elsevier Science. 11

ZEMAN, Z. (1973): *Nazi propaganda.* Oxford University Press, second edn. 15