# Information Revolutions and the Overthrow of Autocratic Regimes*

Chris Edmond†

First draft: October 2003. This draft: September 2007

**Abstract**

This paper presents a model of information quality and political regime change. If enough citizens act against a regime, it is overthrown. Citizens are imperfectly informed about how hard this will be and the regime can, at a cost, engage in propaganda so that at face-value it seems hard. The citizens are rational and evaluate their information knowing the regime's incentives. The model makes three predictions. First, even rational citizens may not correctly infer the amount of manipulation. Second, as the intrinsic quality of information available becomes sufficiently high, the regime is more likely to survive. Third, the regime benefits from ambiguity about the amount of manipulation, and consequently, as it becomes cheaper to manipulate, the regime is also more likely to survive. Key results of the benchmark static model extend to a simple dynamic setting where there are waves of unrest.

*Keywords*: information, coordination, propaganda, regime change, global games.

*JEL classifications*: C7, D7, D8.

Will the information technology revolution make autocratic regimes easier to overthrow? Accounts of the fall of the communist Eastern European regimes often stress the key role played by the regimes' inability to control information (Kalathil and Boas, 2003, 1-2). Similarly, some argue, the information revolution of the 1990s will make the overthrow of regimes in China, Cuba and Saudi Arabia more likely.[1] But the relationship between information and autocracy has not always seemed benign: Nazi Germany and the Soviet Union both seem to be troubling cases. New infor-

*Both this paper and its companion, Edmond (2007), draw on a chapter of my dissertation which circulated under the title 'Information and the limits to autocracy'.

†Department of Economics, Stern School of Business, New York University. Email: ⟨cedmond@stern.nyu.edu⟩.
I would particularly like to thank Andrew Atkeson, Catherine de Fontenay, and Christian Hellwig for their encouragement and many helpful discussions. I also thank Marios Angeletos, Costas Azariadis, Stephen Morris, Alessandro Pavan, Hyun Shin, Iván Werning and seminar participants at the ANU, Boston University, the University of Chicago, the Federal Reserve Bank of Minneapolis, La Trobe University, the London School of Economics, the University of Melbourne, MIT, Northwestern University, Oxford University, the University of Rochester, UC Irvine, University College London and UCLA for their comments and suggestions.

[1] According to President Reagan in a 1989 speech: "Technology will make it increasingly difficult for the state to control the information its people receive ... The Goliath of totalitarianism will be brought down by the David of the microchip". In a 2000 speech, President Clinton said that China's efforts to heavily control internet use "is sort of like trying to nail Jell-O to the wall," while President G.W. Bush has said, in a 1999 campaign debate, "Imagine if the Internet took hold in China, imagine how freedom would spread" [in Kalathil and Boas (2003, 1,155,13)].

mation technologies like radio and film helped make the propaganda machinery of these regimes extraordinarily effective.[2]

So, should we be optimistic that recent breakthroughs in information technology will lead to the collapse of present-day autocratic regimes? To put structure on this question, I develop a formal model of information quality and regime change. It leads to the conclusion that perhaps we should not be optimistic. Section 1 outlines the model. A coordination game is played between an autocratic regime and a heterogeneous population of citizens. These citizens can either subvert the regime or not. Their actions are strategic complements and if enough of them subvert the regime it is overthrown. Citizens are imperfectly informed about how hard this will be, and, in making their decisions, must ask if others share their beliefs. The information available also depends on exogenous technological parameters and on actions taken by the regime in a deliberate attempt to manipulate information.

A regime manipulates information by taking a costly hidden action that shifts the distribution of signals from which citizens sample. This action lets the regime send signals that at face-value suggest it will be difficult to overthrow. The citizens are rational and form beliefs knowing that their information is contaminated by this *propaganda*.

In equilibrium, citizens' beliefs and the regime's manipulation must be mutually consistent. Section 2 characterizes the unique equilibrium of this model and explains why rational citizens are not able to infer the amount of manipulation. Intuitively, since individuals are imperfectly informed about the regime's 'type' they are also imperfectly informed about the amount of manipulation that has occurred. On average, individuals do discount their information because of the regime's misrepresentation, but they typically do not discount *enough*.

Section 3 shows the effects of changes in the information environment. As the intrinsic quality or *precision* of signals becomes sufficiently high, the regime is more likely to survive. The regime's propaganda apparatus is more effective when individuals are receiving, from a technological standpoint, intrinsically high quality signals. I interpret this as suggesting that the information revolution may not be as threatening to autocratic regimes as is sometime supposed. Loosely speaking, the regime is able to 'co-opt' the information revolution so that coordination against the regime becomes more difficult. More precisely, 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 this threshold. So if signals are precise, in this situation many citizens have signals suggesting the regime will survive. And consequently it is rational for any citizen, 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 citizen less likely to attack and so the aggregate mass who do is relatively small. This in turn makes it more likely that the regime does manipulate to create an artificially high signal mean thereby validating the original beliefs.

[2] On Nazi Germany and propaganda (especially by radio), see Zeman (1973, 34-62). More generally, see Arendt (1973, 341-364) and Friedrich and Brzezinski (1965, 129-147).

I then ask if a reduction in the cost of information manipulation will help regimes survive. In principle, a reduction in the cost of information manipulation has complicated effects because it increases the amount of manipulation a regime wants to do but this change in the regime's incentives is known to citizens and they may be able to offset it. It turns out that regimes *benefit from ambiguity* about the amount of manipulation. As costs fall, this ambiguity increases and so do a regime's chances of surviving. Lower costs of manipulation benefit the regime.

Section 4 uses the model to interpret research on the policies that China and Cuba have used to offset the effects of the information revolution. I argue that there is support for the model's prediction that regimes benefit because they are able to co-opt an increase in the intrinsic quality of information in a way that makes coordination more difficult. The model also helps reconcile the views of researchers who argue that these regimes are successfully exploiting new technologies to counter dissent [e.g., Kalathil and Boas (2003)] with those who argue that any success is due to more traditional authoritarian methods such as arrest and seizure [e.g., Chase and Malvenon (2002)]. The model implies that these policies for coping with dissent are complements.

In related work, Ginkel and Smith (1999) study a game between three unitary agents: a regime, a group of dissidents, and the mass public. They focus on dissidents' uncertainty about the regime surviving and consider *signaling* by the regime. There is no information heterogeneity in their model, however — dissidents and the public both receive the same noisy signal of the regime's type — and so there is no question of coordination failure or difficulty in forecasting-the-forecasts-of-others. Karklins and Petersen (1993) show how mass unrest is constructed by individuals playing a sequence of 'assurance games'. They provide a rich discussion of coalition formation, but do not consider information in a detailed way. Kuran (1991) and Lohmann (1994) are expressly concerned with the role of information and the sudden overthrow of the Eastern European communist regimes. Although details differ, both draw on the notion of an *information cascade* to generate sudden regime change. To capture the effects of information accumulating over time, Section 5 shows that the key results of the benchmark static model of this paper extend to a more dynamic setting with waves of unrest and accumulation of information about the regime.

This paper draws on the 'global games' approach to coordination games with imperfect information that was pioneered by Carlsson and van Damme (1993) and Morris and Shin (1998, 2000, 2003). Coordination games often have multiple self-fulfilling equilibria.[3] In the global games approach, imperfect information about the underlying fundamentals can serve to select a unique equilibrium. This paper differs from earlier global games papers because individuals' information is not exogenous but instead is endogenous through the regime's manipulation decision.[4]

**A motivating example: Ottoman Turkey.** Nineteenth century Ottoman Turkey was a brutal autocracy. But technological breakthroughs like railroads, telegraphs, steamships and mass newspapers

[3] See Cooper (1999) or Chamley (2004) for textbook discussions.

[4] Angeletos, Hellwig, and Pavan (2006) analyze a closely related global game where information is endogenous because of signaling by the regime and show that this leads back to equilibrium multiplicity. Edmond (2007) discusses in detail the relationship between their model and the one in this paper.

— printed in Arabic, Persian and Turkish — spurred interest in world developments and provided citizens of the Empire with a massive increase in the quality of their information. If technological and informational improvements of this kind help overthrow autocratic regimes, then one might think the citizens of the Ottoman Empire would be well placed to force change.

But no such change took place. The same improvements in information technologies also gave the regime more control over its subjects. The ability to monitor communications and to rapidly deploy troops both increased the power the regime could exercise directly and helped undermine traditional sources of countervailing power. For decades, until the end of the Empire in the aftermath of the First World War, improvements in information technology enabled more effective autocratic rule (Hodgson, 1974, 253-256).

By contrast, in his discussion of this period of Ottoman history, Lewis (2002) argues that recent technological developments lead in the opposite direction:

> Television and satellite, fax and internet, have brought and imposed a new openness, and are beginning to undermine the closed society and closed minds that sustain autocracy. Similarly, the spread of education or at least of literacy to much larger elements of the population has again imposed new limits on the autocracy of rulers ... (Lewis, 2002, 54).

But why? Why are *these* developments different; why have the late twentieth century's telegraphs and steamships imposed limits to autocracy when their precursors did not?

# 1 Model of information, coordination 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 decides whether to subvert the regime, $s_i = 1$, or not, $s_i = 0$. The population mass of subversives is $S := \int_0^1 s_i di$. Citizens expect to get a larger payoff if they are involved in the downfall of the regime. If the regime is overthrown a citizen gets random reward $w \in \{\underline{w}, \overline{w}\}$ with $\underline{w} < \overline{w}$ and $\Pr(w = \underline{w}) := \mu(s_i)$ with $0 \leq \mu(1) < \mu(0) \leq 1$, so there is more chance of getting $\overline{w}$ if $s_i = 1$. The utility cost of subverting is normalized to 1.

If a citizen believes the regime will be overthrown with probability $P_i$, she will subvert when the expected payoff from doing so is at least as large as that from not subverting, specifically if and only if

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

equivalently, if and only if

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

where $p$ is the *opportunity cost* of subverting. 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, this is an assumption about the attractiveness of 'free-riding'. Free-riding is less of a problem if the probability of being found out $\mu(0)$ is sufficiently high or if the penalty $\overline{w} - \underline{w}$ from being caught is sufficiently severe.[5] With Assumption 1, free-riding is possible in principle but will not occur in equilibrium.

The citizens face a regime indexed by a hidden state $\theta$ that is the regime's private information. The state $\theta$ is normalized so that the regime is overthrown if and only if $\theta < S$. The payoff to a citizen can therefore be summarized by

$$u(s_i, S, \theta) = s_i(\mathbb{1}\{\theta < S\} - p) \tag{2}$$

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 $\theta$, 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 $a$. The regime obtains a benefit $\theta - S$ from remaining in power and so has a direct aversion to $S$. Regimes prefer to avoid a Prague Spring or a Tiananmen Square. Suppressing a revolt is resource costly to the regime and this cost is increasing in the mass of rioters.[6] If $\theta < S$, the regime is overthrown and obtains an outside option with value normalized to zero. The payoff to a regime is therefore

$$B(S, \theta) - C(a) \tag{3}$$

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

Following a regime's hidden action $a$, each citizen simultaneously draws an idiosyncratic signal $x_i := \theta + a + \varepsilon_i$ where the noise $\varepsilon_i$ is independent of $\theta$ and is IID normally distributed with mean zero and *precision* $\alpha$ (that is, variance $\alpha^{-1}$). So the density of signals is $f(x_i|\theta, a) := \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a))$ where $\phi$ denotes the standard normal density. I begin by assuming that citizens have common priors for $\theta$ and that this prior is the (improper) uniform distribution over the whole real line. The realization of the signal $x_i$ is informative for both the type of the regime $\theta$ 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.

[5] 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.

[6] Also, and loosely speaking, 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. As emphasized by Skocpol (1979), deteriorating relations with foreign powers can provide crucial opportunities for social unrest.

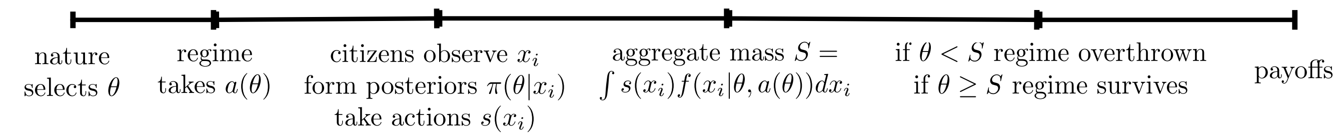


Figure 1: The timing of the model.

## 1.1 Equilibrium concept

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

$$\pi(\theta|x_i) = \frac{f(x_i|\theta, a(\theta))}{\int_{-\infty}^{\infty} f(x_i|\theta, a(\theta)) d\theta} \tag{4}$$

$$s(x_i) \in \underset{s_i \in \{0,1\}}{\operatorname{argmax}} \left\{ \int_{-\infty}^{\infty} u(s_i, S(\theta, a(\theta)), \theta) \pi(\theta|x_i) d\theta \right\} \tag{5}$$

$$S(\theta, a) = \int_{-\infty}^{\infty} s(x_i) f(x_i|\theta, a) dx_i \tag{6}$$

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \{B(S(\theta, a), \theta) - C(a)\} \tag{7}$$

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 Further discussion of model

Interpretation of hidden actions. The hidden action $a \geq 0$ of the regime gives it the potential to bias the information citizens receive. If $a > 0$ citizens draw from a signal distribution that at face value suggests the regime will be 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 directly and it instead must be inferred.

Nature of the regime. The regime is autocratic in that it is socially desirable for the regime to be removed and it is not opposed by competing political parties or engaged in strategic interactions with other *large* players. The one-dimensional $\theta$ summarizes many characteristics of regimes. More malevolent regimes have a high $\theta$ because they are willing to plunder society and have greater incentive to resist a given amount of dissent. And since the regime is treated as a unitary actor, $\theta$ also captures its degree of internal cohesion. If the regime is riven by factions or if the military is ambivalent in its support, the regime may be more vulnerable. Similarly, $\theta$ is also affected by the policies of foreign powers and their willingness to intervene for or against the regime in the event of an attempted overthrow. Critically, the regime's type captures those aspects of the society about which the regime has better information than do citizens.

Simultaneous moves and lack of communication. Every citizen receives her signal and makes her decision simultaneously. There is no communication between them. Any concerns about informers and the consequences of organizing against a regime also affect $\theta$. If a lot of credible communication is possible, $\theta$ is presumably lower. Section 5 below relaxes the assumption of simultaneous moves, allowing a sub-set of citizens to receive information from the collective behavior of other citizens.

## 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 $\theta$ is common knowledge, costly hidden actions are pointless and $a(\theta) = 0$ all $\theta$. The model reduces to a simple coordination game with multiple equilibria. 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 subvert, none do, and the regime survives. If $\theta \in [0, 1)$, the regime is 'fragile' and multiple self-fulfilling equilibria can be sustained. For example, if each individual believes that everyone else will subvert, it will be optimal for each citizen to do so and $S = 1 > \theta$ leads to the regime's overthrow and the vindication of initial expectations.

If hidden actions are prohibitively expensive, $a(\theta) = 0$ all $\theta$ 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.[7] In this equilibrium, strategies are *threshold rules*: there is a unique type $\theta^*$ 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 benchmark): The unique equilibrium thresholds $x^*_{\text{MS}}, \theta^*_{\text{MS}}$ simultaneously solve

$$\Phi\left[\sqrt{\alpha}\,(\theta^*_{\text{MS}} - x^*_{\text{MS}})\right] = p \tag{8}$$

$$\Phi\left[\sqrt{\alpha}\,(x^*_{\text{MS}} - \theta^*_{\text{MS}})\right] = \theta^*_{\text{MS}} \tag{9}$$

where $\Phi$ denotes the standard normal cumulative distribution. In particular, $\theta^*_{\text{MS}} = 1 - p$ independent of $\alpha$ and $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha}$.

If there are no hidden actions, a citizen with $x_i$ assigns $\Pr(\theta < \theta^*_{\text{MS}}|x_i) = \Phi[\sqrt{\alpha}(\theta^*_{\text{MS}} - x_i)]$ to the regime being overthrown and so the first condition says that if the regime's threshold is $\theta^*_{\text{MS}}$, a citizen with signal $x_i = x^*_{\text{MS}}$ will be indifferent between subverting or not. Given this, the mass of

[7]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.

subversives is $\Pr(x_i < x^*_{\text{MS}}|\theta) = \Phi[\sqrt{\alpha}(x^*_{\text{MS}} - \theta)]$ and a regime with type $\theta = \theta^*_{\text{MS}}$ will be indifferent. In the analysis below, I say that a regime's hidden action technology is *effective* if in equilibrium it does better than the Morris-Shin benchmark, $\theta^* < \theta^*_{\text{MS}} = 1 - p$.

As information becomes precise, some regimes are faced with a large incentive to shift the signal mean. To see this, notice that in the Morris-Shin benchmark, the mass of subversives is

$$S(\theta) = \Phi\left[\sqrt{\alpha}\left(x^*_{\text{MS}} - \theta\right)\right] = \Phi[\sqrt{\alpha}(1 - p - \theta) - \Phi^{-1}(p)] \tag{10}$$

As precision $\alpha \to \infty$, the mass $S(\theta) \to \mathbb{1}\{1 - p - \theta\}$, a step function. So if the regime has $\theta < \theta^*_{\text{MS}} = 1 - p$ it faces a unit mass of subversives, but if the regime has $\theta > \theta^*_{\text{MS}}$ it faces zero subversives. A small reduction in $\theta^*_{\text{MS}}$ would enable a regime with $\theta$ just smaller than $\theta^*_{\text{MS}}$ to switch from being overthrown to surviving. As information becomes precise, there is a large incentive for marginal regimes to shift the signal mean.

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 Equilibrium with endogenous information manipulation

This model has a unique perfect Bayesian equilibrium (Edmond, 2007). As in the Morris-Shin benchmark, the equilibrium is characterized by thresholds $x^*$ and $\theta^*$ so that for citizens $s(x_i) = 1$ for $x_i < x^*$ and zero otherwise while the regime is overthrown for $\theta < \theta^*$ and not otherwise. But with endogenous information the regime's hidden actions also need to be taken into account. Section 2.1 show how to compute the equilibrium. Section 2.2 sketches the properties of hidden actions. Section 2.3 explains why citizens have difficulty inferring the amount of manipulation.

## 2.1 Solving for the equilibrium

Let $\hat{x}, \hat{\theta}$ denote candidates for the equilibrium thresholds 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 $\theta$ 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 $\theta - \hat{x}$ and can be written $a(\theta|\hat{x}) = a(\theta - \hat{x})$.

**Regime's problem.** Taking $\hat{x} \in \mathbb{R}$ as given the mass of citizens facing the regime is

$$\hat{S}(\theta + a) := \int_{-\infty}^{\hat{x}} f(x_i|\theta, a)dx_i = \Phi[\sqrt{\alpha}(\hat{x} - \theta - a)] \tag{11}$$

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

$$V(\theta) := \max[0, W(\theta)] \tag{12}$$

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

$$W(\theta) := \max_{a\geq 0}\left[\theta - \hat{S}(\theta+a) - C(a)\right] \tag{13}$$

Using the envelope theorem and the definition of $\hat{S}(\theta+a)$ in equation (11) shows that $W'(\theta) > 1$ all $\theta$. 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 (12), the regime is overthrown if and only if $\theta < \hat{\theta}$. For $\theta \geq \hat{\theta}$, the actions of the regime solve

$$a(\theta - \hat{x}) \in \operatorname*{argmax}_{a\geq 0}\left[\theta - \hat{S}(\theta+a) - C(a)\right], \quad \theta \geq \hat{\theta} \tag{14}$$

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

$$W(\hat{\theta}) = \hat{\theta} - \hat{S}[\hat{\theta} + a(\hat{\theta} - \hat{x})] - C[a(\hat{\theta} - \hat{x})] = 0 \tag{15}$$

So taking $\hat{x} \in \mathbb{R}$ as given, equations (14)-(15) determine the threshold $\hat{\theta}$ and the hidden actions $a(\theta - \hat{x})$ that characterize the solution to the regime's problem.

Citizens' 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

$$P(\hat{\theta}, x_i) := \Pr(\theta < \hat{\theta}|x_i) = \frac{\int_{-\infty}^{\hat{\theta}} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_i - \theta)]d\theta}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_i - \theta - a(\theta - \hat{x}))]d\theta} \tag{16}$$

where the numerator uses the fact that the regime takes no action for $\theta < \hat{\theta}$. Clearly $P : \mathbb{R}\times\mathbb{R} \to [0,1]$ is strictly increasing in $\hat{\theta}$. An argument given in Edmond (2007) also shows that $P$ is strictly decreasing in $x_i$, and for any $\hat{x} \in \mathbb{R}$ satisfies

$$P(\hat{\theta}, x_i) = P(\hat{\theta} - \hat{x}, x_i - \hat{x}) \tag{17}$$

so that in fact $P$ depends only on the difference $\hat{\theta} - x_i$.

Now 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 (14)-(15), the signal threshold $\hat{x}$ solves

$$P(\hat{\theta}, \hat{x}) = P(\hat{\theta} - \hat{x}, 0) = p \tag{18}$$

To calculate an equilibrium, first solve the decision problem in (14) 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 (16). Since $P$ is continuous and strictly increasing in $\hat{\theta}$ with $P(-\infty, 0) = 0$ and $P(\infty, 0) = 1$, by the intermediate value theorem there is a unique difference $\theta^* - x^* \in \mathbb{R}$ such that $P(\theta^* - x^*, 0) = p$. This solution can be plugged into (15) to obtain a unique threshold $\theta^* \in [0,1)$ so that we know both $\theta^*$ and $x^* \in \mathbb{R}$ separately. The equilibrium hidden actions are then $a(\theta) := a(\theta - x^*)$. Moreover Edmond (2007) shows that only this equilibrium survives the iterative elimination of (interim) strictly dominated strategies and so it is the only perfect Bayesian equilibrium of the game.

## 2.2 Regime's hidden actions

In equilibrium, hidden actions $a(\theta)$ are characterized by the first order necessary condition[8]

$$C'(a) = \sqrt{\alpha}\phi[\sqrt{\alpha}(x^* - \theta - a)], \qquad \theta \geq \theta^* \tag{19}$$

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 manipulation to occur (meaning $a(\theta) > 0$ for at least some $\theta$), 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 $\theta^*$. As the fundamentals of the regime become strong, costly actions taken to generate a favorable signal distribution encounter diminishing returns and the action profile dies to zero. Figure 2 illustrates.

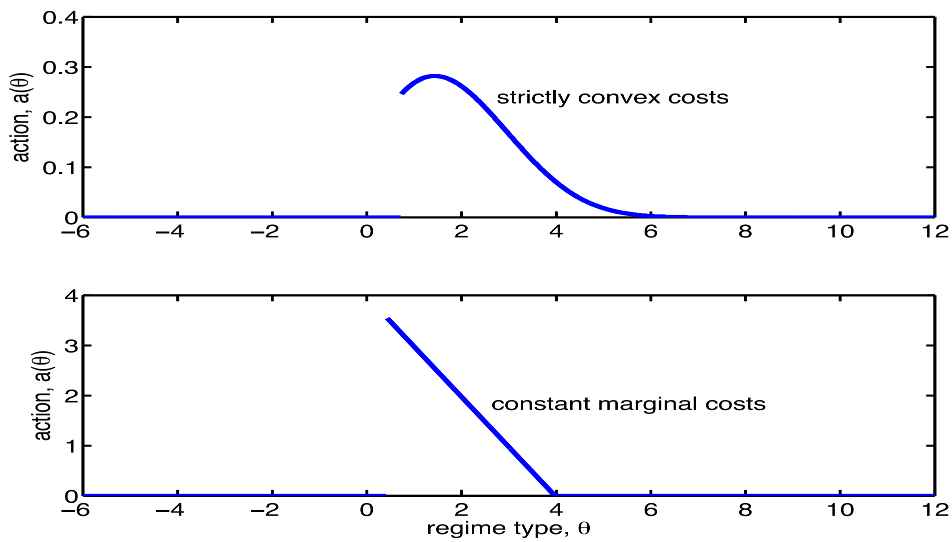


Figure 2: Equilibrium hidden actions $a(\theta)$.

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 $\theta$. Then manipulating equation (19) shows that interior solutions to the regime's problem are given by

$$a(\theta) = x^* + \gamma - \theta, \qquad \theta \in [\theta^*, \theta^{**}) \tag{20}$$

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

$$\gamma := \sqrt{\frac{2}{\alpha}\log\left(\frac{\sqrt{\alpha}\phi(0)}{c}\right)} > 0 \tag{21}$$

[8]The first order condition (19) 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.

This is an acute case of 'signal-jamming'. 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 $\theta$. If a regime manipulates, it generates signals with an artificially high mean $x^* + \gamma = \theta^{**} > \theta^*$ and as $\alpha$ becomes large these signals are tightly clustered around $\theta^{**}$.

## 2.3 Why can a regime manipulate beliefs in equilibrium?

If citizens are rational and know the regime's decision problem, shouldn't they be able to adjust their signal to account for the incentives of the regime so that *at equilibrium* information manipulation has no effect?

This intuition is wrong. If there is lack of common knowledge of the regime's type, a regime may manipulate information in equilibrium so long as different types of regimes would take different actions. To see this, suppose to the contrary that citizens 'know' any regime will take a constant action $\hat{a} > 0$ (say) irrespective of $\theta$. Then each citizen would adjust their signal up *one-for-one* with $\hat{a}$ so that the signal threshold would be $x^* + \hat{a}$ and the corresponding mass of subversives would be $S(\theta) = \Phi[\sqrt{\alpha}(x^* + \hat{a} - \theta - \hat{a})] = \Phi[\sqrt{\alpha}(x^* - \theta)]$ independent of $\hat{a}$. Since actions are costly, the regime would be better off with $a = 0$ and so this cannot be an equilibrium.

Any interesting candidate for an equilibrium action profile will not involve a constant $\hat{a}$ independent of $\theta$, because if actions are costly it is never optimal for a regime to take an action in a state that corresponds to it being overthrown in equilibrium. This means that each citizen is unsure about what action has been taken and this *ambiguity* allows the regime to manipulate information in equilibrium.

Example: ambiguity about regime's action. To see this, suppose $a(\theta) = 0$ for $\theta < \theta^*$, but $a(\theta) = \hat{a} > 0$ for $\theta \geq \theta^*$. Also, for simple calculations let $p = 1/2$. In the Morris-Shin benchmark, from (37), this would give $\theta^*_{\text{MS}} = 1 - p = 1/2$ and $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha} = 1/2$ too, since $\Phi^{-1}(1/2) = 0$. But manipulating the conditions (15)-(18) with $a(\theta) = \hat{a} > 0$ for $\theta \geq \theta^*$ gives

$$\theta^* = \Phi(-\sqrt{\alpha}\hat{a}/2) + C(\hat{a}) \quad \text{and} \quad x^* = \theta^* + \hat{a}/2 \tag{22}$$

So if $\hat{a} > 0$ and the action profile is not constant, the probability $\Phi(-\sqrt{\alpha}\hat{a}/2) < \Phi(0) = 1/2 = \theta^*_{\text{MS}}$ and the regime can achieve a lower threshold (an ex ante higher survival probability) and this can be an equilibrium if the cost $C(\hat{a})$ is not too large.[9] Moreover, holding fixed $\theta^*$, the signal threshold $x^*$ does not rise one-for-one with the level of the action $\hat{a}$: it only rises by $\hat{a}/2$. The marginal citizen discounts their signal on account of the manipulation, but not by *enough*.

Once we recognize that different types of regimes take different actions, there is no common discount factor that each individual can apply to her signal so as to 'undo' the manipulation. The right discount factor to use depends on $\theta$, but the citizens have heterogeneous beliefs about $\theta$. The ability of a regime to successfully manipulate information is inextricably linked to heterogeneity

[9] As shown in Section 3.2 below, if we restrict actions to $a \in [0, \overline{a}]$ then as the cost of information manipulation becomes small (but remains positive) the equilibrium hidden action profile is $a(\theta) = \overline{a}$ for $\theta \geq \theta^*$ and zero otherwise.

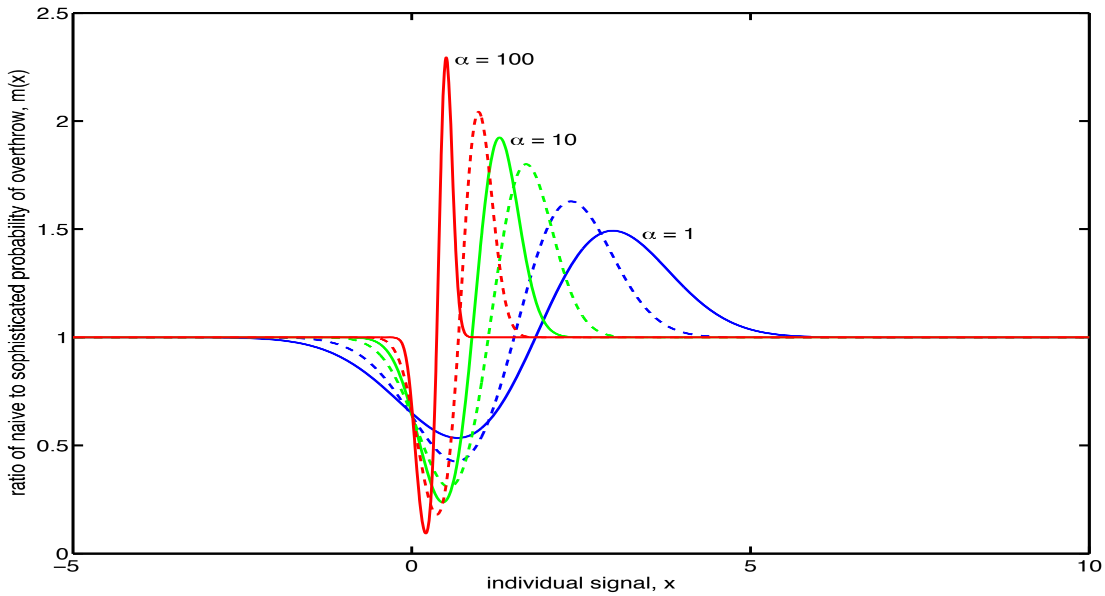


Figure 3: The cross-section of marginal factors $m(x_i)$ for various precisions $\alpha$. A citizen with $m(x_i) < 1$ assigns less probability to the regime's survival if she takes into account manipulation; a citizen with $m(x_i) > 1$ assigns more probability to the regime's survival.

in beliefs. If $\theta$ were common knowledge, costly information manipulation would be pointless. But with lack of common knowledge of $\theta$, there is no agreement on the appropriate discount factor to give to an individual's signal.

**Equilibrium cross-section of beliefs.** The marginal factor associated with each citizen's posterior provides a measure of the discount that different citizens want to apply to their signals. Suppose a citizen with $x_i$ was 'naive' and believed she lived in the benchmark Morris-Shin world. Then she would give probability $\Phi[\sqrt{\alpha}(\theta^* - x_i)]$ to the regime's overthrow. But a 'sophisticated' citizen who takes into account $a(\theta) > 0$ for $\theta \geq \theta^*$ will instead give this event probability

$$\Pr(\theta < \theta^* | x_i) = \frac{\Phi[\sqrt{\alpha}(\theta^* - x_i)]}{m(x_i)}$$

where

$$m(x_i) := \Phi[\sqrt{\alpha}(\theta^* - x_i)] + \int_{\theta^*}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x_i - \theta - a(\theta))]d\theta \tag{23}$$

The marginal $m(x_i)$ is a measure of the discrepancy between the sophisticated and the naive beliefs: a citizen with $m(x_i) < 1$ assigns less probability to the regime's survival if she acts 'sophisticated'. Figure 3 illustrates. For low $x_i$, the marginal is less than one and citizens discount their signals but for high $x_i$ the marginal is greater than one. Why does the marginal density have this shape? First, extreme signals result in marginals that are close to one since as $x_i \to -\infty$ or $x_i \to +\infty$ citizens assign almost the same probability to the regime's survival irrespective of concerns about manipulation. Second, the mere *existence* of the technology for manipulating signals is a form of aggregate information. Consider citizens with low signals. Since hidden actions are non-negative, if a citizen has a low signal it is likely that the regime's true type is low and that it did not manipulate. But at equilibrium, if a regime did no manipulation it must be because its survival probability is zero. Hence a citizen with a low signal will at equilibrium discount that signal yet further, $m(x_i) < 1$. The possibility of signal manipulation plus the low realized signal itself *reinforce* each other to convince a citizen that the regime's type must be low. Similarly, a high signal might be because of manipulation, but active manipulation means that at equilibrium the regime will survive with probability one. So if you have a high signal, you will rationally assign more probability to the regime's survival than you would if you were naive, $m(x_i) > 1$.

The existence of a technology for manipulating signals *amplifies* the ex ante heterogeneity in information, leading those with low signals to assign even less survival probability to the regime and leading those with high signals to assign even more survival probability to the regime.

Although regimes may be *able* to manipulate information in equilibrium it does not immediately follow that manipulation is *effective* in increasing the likelihood of the regime surviving. As shown next, manipulation is effective when $\alpha$ is sufficiently high.

# 3 Information revolutions

This section studies changes in the information environment. Section 3.1 shows that if signals are of high enough quality, manipulation is effective. Section 3.2 shows that if the costs of information manipulation fall, regimes will have higher survival probabilities despite the fact that citizens know they have an incentive to take larger actions.

Terminology. I measure the effectiveness of manipulation by its ability to reduce the threshold $\theta^*$ relative to the Morris-Shin benchmark of $\theta^*_{\text{MS}} = 1 - p$. A lower $\theta^*$ increases the regime's ex ante survival probability making it more likely that nature draws a $\theta \geq \theta^*$. I say that the regime *benefits* from lower $\theta^*$ even though this does not necessarily increase the regime's payoff. It might be that lower $\theta^*$ is achieved through large, costly actions that give the regime an overall lower payoff than they would get if hidden actions were impossible.

## 3.1 Increases in intrinsic signal quality

Let $\theta^*_\alpha$, $x^*_\alpha$, and $a_\alpha(\theta)$ denote the equilibrium thresholds and hidden action profile indexed by the precision $\alpha$. Then:

Proposition 2. As the signal precision $\alpha \to \infty$ the limiting thresholds and hidden actions 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 manipulation is strictly convex, $C''(a) > 0$ all $a$, then as $\alpha \to 0^+$ the limiting thresholds and hidden actions are

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

For high enough $\alpha$ information manipulation is *maximally* effective. And there is a partial converse.[10] If costs are strictly convex, then for low enough $\alpha$ hidden actions are ineffective in that $\theta^* > \theta^*_{\text{MS}} = 1 - p$. If so, regimes would want to credibly commit not to use them.

Example: constant marginal costs revisited. The special case of constant marginal costs is again useful for understanding what drives the result. We first need to solve for the thresholds. Use (20) and rearrange the indifference conditions (15) and (18) to get

$$\Phi[\sqrt{\alpha}(\theta^*_\alpha - x^*_\alpha)] = \frac{p}{1-p}\left[c(x^*_\alpha - \theta^*_\alpha + \gamma_\alpha) + \Phi(-\sqrt{\alpha}\gamma_\alpha)\right] \tag{24}$$

and

$$\theta^*_\alpha = c(x^*_\alpha - \theta^*_\alpha + \gamma_\alpha) + \Phi(-\sqrt{\alpha}\gamma_\alpha) \tag{25}$$

where writing $\gamma_\alpha$ acknowledges that coefficient defined in (21) also depends on the precision. For each $\alpha > 0$, these two equations uniquely determine $x^*_\alpha, \theta^*_\alpha$ [solve (24) for the unique difference $\theta^*_\alpha - x^*_\alpha$ and then plug into (25) to get $\theta^*_\alpha$]. The equilibrium mass of subversives that makes the regime indifferent is

$$S^*_\alpha := \Phi(-\sqrt{\alpha}\gamma_\alpha) = \Phi\left[-\sqrt{2\log\left(\frac{\sqrt{\alpha}\phi(0)}{c}\right)}\right] \tag{26}$$

This is strictly decreasing in $\alpha$ and $S^*_\alpha \to 0^+$ as $\alpha \to \infty$. High $\alpha$ 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. For large $\alpha$ solutions to equation (24) are approximately the same as solutions to

$$\mathbb{1}\{\theta^*_\alpha - x^*_\alpha \geq 0\} = -\frac{p}{1-p}c(\theta^*_\alpha - x^*_\alpha) \tag{27}$$

The only solution to equation (27) is $\theta^*_\alpha - x^*_\alpha = 0$. So as $\alpha \to \infty$, solutions to equation (24) approach zero too. From equation (25) we now know $\theta^*_\alpha \to 0^+$. Therefore, manipulation is effective when the precision $\alpha$ is large enough. For large $\alpha$ the threshold $\theta^*_\alpha$ is less than the Morris-Shin benchmark of $\theta^*_{\text{MS}} = 1 - p$ and the regime's survival probability is correspondingly higher.

Intuition for the result. Staying with the special case of constant marginal costs, if a regime manipulates, then from (20), 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 citizens have signals suggesting the regime will survive. And consequently it is rational for any citizen, 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 citizen less likely to attack and so the aggregate mass who do is relatively

[10] As discussed following equation (20) 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 Proposition 2 deals only with the case of strictly convex costs.

small. This in turn makes it more likely that the regime does manipulate and create an artificially high signal mean, thereby validating the original beliefs.

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 $\alpha$ 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 $\theta$ even when $\alpha$ 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 $\theta$.

**Locally uninformative signals in the general case.** To see how this extends to general cost functions, rearrange (19) to get an alternative implicit characterization of the hidden actions

$$a(\theta) = x^* - \theta + \sqrt{\frac{2}{\alpha}\log\left(\frac{\sqrt{\alpha}\phi(0)}{C'[a(\theta)]}\right)}, \qquad \theta \geq \theta^* \tag{28}$$

This is (20) generalized to arbitrary convex costs but at the expense of losing the closed-form solution. By implicitly differentiating with respect to $\theta$ 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 $\theta^*$ so that manipulation is effective as $\alpha$ 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 4 shows $\theta^*_\alpha$ as a function of precision $\alpha$ 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 $\alpha$.

**Discussion and 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 uninformative, but that is precisely because the regime is *co-opting* the technology to its own ends.[11] A regime will want to exert a strong influence over the media when the signal precision is high enough.

[11] 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).

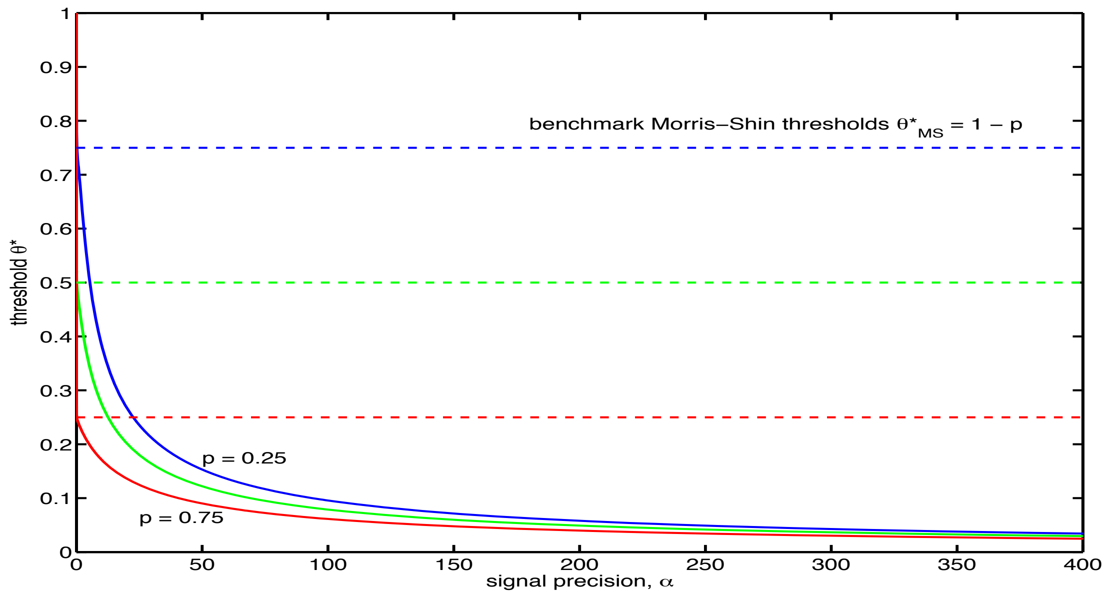


Figure 4: In the Morris-Shin game, thresholds are $\theta^*_{\mathsf{MS}} = 1 - p$ all $\alpha$. With information manipulation, large $\alpha$ implies $\theta^*_\alpha \to 0^+$, so in the limit all 'fragile' regimes with $\theta \in [0, 1)$ survive. Better information increases a regime's likelihood of survival. All calculations use $C(a) := 0.5a^2$.

Proposition 2 does not obtain because a regime is somehow able to get away with a big shift in the signal mean which, mechanically, would reduce the mass of subversives. Instead the actions are 'just big enough' and do not need to be large when precision $\alpha$ is high. As $\alpha$ increases the regime benefits from the technology for manipulation because it leads to a change in the composition of beliefs. Citizens receive signals through a high precision channel, but the regime is able to fine-tune the message that individuals receive in such a way that pivotal individuals — who might think the regime is fragile — do not act. Because of the coordination problem, this spills over to others and so the probability of the regime surviving increases.

### 3.2 Lower costs of manipulation

Consider a technological change that increases the costs of manipulation. In a perfect Bayesian equilibrium, this has two effects. Taking as given citizens' beliefs, an increase in costs will decrease the regime's desired hidden action. But taking as given the regime's incentives this will also lead citizens to draw different inferences about the true $\theta$. Which effect dominates?

Let $C_k(a)$ denote a family of cost functions that satisfies $C_{k'}(a) \geq C_k(a)$ for all $k' \geq k$ (with equality if and only if $k' = k$). Further, let $\lim_{k\to\infty} C_k(a) = \lim_{k\to\infty} C_k'(a) = \infty$ for all $a > 0$. Assume the costs of any hidden action are bounded below by an arbitrarily small but positive constant $\underline{c} > 0$ with $\lim_{k\to 0^+} C_k(a) = \lim_{k\to 0^+} C_k'(a) = \underline{c} > 0$ for all $a > 0$. This ensures the analysis is only concerned with *costly* hidden actions and that in the limit we do not have a model where manipulation is free. Finally, assume actions are chosen from $[0, \overline{a}]$ for some $\overline{a} < \infty$.

Let $\theta^*_k$, $x^*_k$ and $a_k(\theta)$ denote the equilibrium thresholds and hidden actions indexed by the cost of information manipulation. Then:

PROPOSITION 3. As costs of manipulation $k \to \infty$ the limiting thresholds and hidden actions are

$$\lim_{k\to\infty} \theta_k^* = \theta_{\text{MS}}^*, \qquad \lim_{k\to\infty} x_k^* = x_{\text{MS}}^*, \quad \text{and} \quad \lim_{k\to\infty} a_k(\theta) = 0^+ \quad \text{for all } \theta$$

Alternatively, as $k \to 0^+$ the limiting thresholds and hidden actions are

$$\lim_{k\to 0^+} \theta_k^* = \Phi[\sqrt{\alpha}(\nu - \overline{a})] + \underline{c} =: \theta_0^*, \qquad \lim_{k\to 0^+} x_k^* = \theta_0^* + \nu, \quad \text{and} \quad \lim_{k\to 0^+} a_k(\theta) = \begin{cases} 0^+ & \theta < \theta_0^* \\ \overline{a} & \theta \geq \theta_0^* \end{cases} \tag{29}$$

where $\nu \in \mathbb{R}$ is the unique solution of $(1-p)\Phi(-\sqrt{\alpha}\nu) = p\Phi[\sqrt{\alpha}(\nu - \overline{a})]$.

As $k \to \infty$, the costs of manipulation become extreme and no regime has $a(\theta) > 0$. The model reduces to the Morris-Shin benchmark. Alternatively, suppose $k \to 0^+$ so that manipulation becomes cheap (but not costless). Then all regimes that survive in equilibrium will take the same largest hidden action $\overline{a}$ while all regimes that are overthrown will, as usual, take no action. Each citizen knows that the regime has taken one of these two positions. With some probability citizens view themselves as living in the Morris-Shin world and with complementary probability they live in a world where the mass of subversives will always be lower (in amount determined by $\overline{a}$). Because of this, they are always more reluctant to subvert the regime and the equilibrium threshold is always less than the benchmark level $1-p$. So lower costs of information manipulation increase a regime's chances of survival. Figure 5 illustrates.

Moreover (29) shows that $\theta_0^*$ is strictly decreasing in the largest hidden action $\overline{a}$ and that as $\overline{a} \to \infty$ the state threshold $\theta_0^* \to \underline{c} > 0$. When manipulation is cheap and very large hidden actions are possible, regimes may have high ex ante survival probabilities. This shows again that regimes *benefit from ambiguity* about the nature of the hidden action they will take. If manipulation is sufficiently expensive this ambiguity disappears and the game being played reverts to the Morris-Shin benchmark with exogenous information.

# 4 How do autocratic regimes respond to the information revolution?

The model predicts that as the intrinsic quality of information improves, regimes may be more likely to survive because in equilibrium they are able to *co-opt* the technology. Drawing on the examples of China and Cuba, this section asks if there is evidence that regimes can benefit from co-opting an information revolution and concludes there is. The discussion focuses on the internet, but many points equally apply to technologies like satellite television and cell phones.

## 4.1 China

There is a degree of consensus that Chinese authorities, at least for the moment, have succeeded in countering the effects of the information revolution [Chase and Malvenon (2002); Kalathil and Boas (2003); Lynch (1999)]. Dissident groups — whether they be pro-democracy activists, Chinese

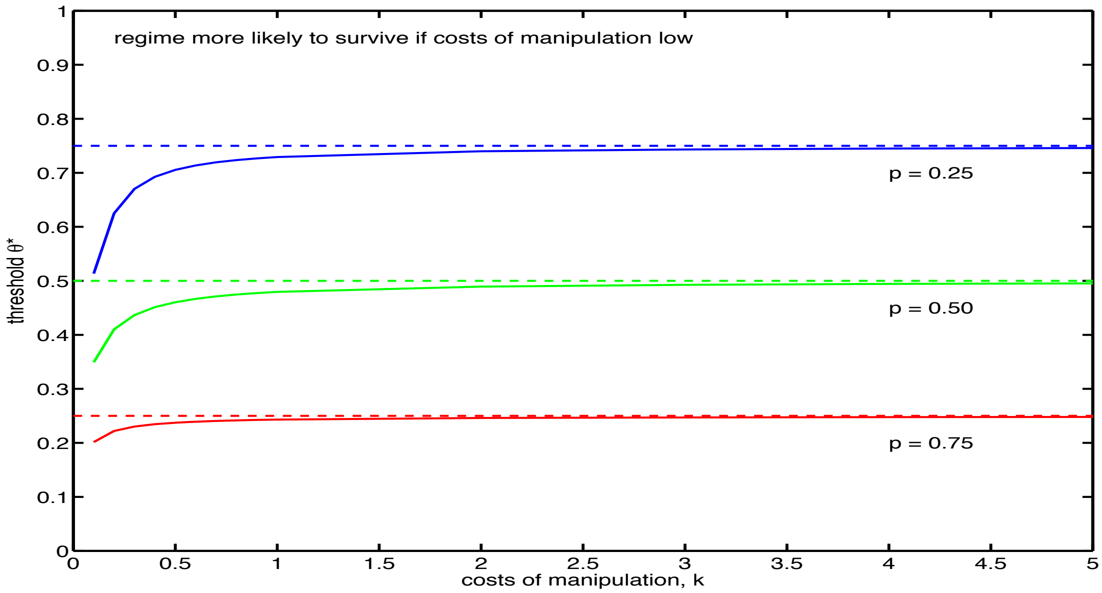


Figure 5: As costs of manipulation fall, $k \to 0^+$, the regime takes either no action, $a = 0$ or a very large action, $a = \bar{a}$. The regime benefits from this extreme ambiguity.

nationalists, Falungong spiritualists, Tibetan nationalists, and activists on various sides of the Taiwan question — have not been able to exploit new technologies to threaten the regime.

But there is less consensus as to how this successful off-setting of the information revolution has been achieved. On one side of the debate, Kalathil and Boas (2003, 25-29, 37-40) argue that Chinese authorities have indeed successfully co-opted developments in information technology. In particular, Chinese authorities: i) monitor, filter, and block access to sensitive overseas web-sites and email addresses,[12] ii) require that internet service providers report on the activities of users, iii) require censorship of bulletin boards and chat rooms, iv) use the internet to directly spread propaganda and v) attempt to retard the effectiveness of dissident activities by hacking web-sites and overloading email addresses.[13]

On the other side of the debate, Chase and Malvenon (2002) argue that the apparent success of the Chinese regime has come from traditional authoritarian methods demonstrating, through arrest, detention, and seizure, the supposed strength of the regime and the foolishness of engaging in subversive behavior.[14] For example, in December 1998, Wang Youcai, a leading member of the short-lived China Democracy Party (CDP) was sentenced to 11 years jail based on email communications with activists in Hong Kong and the US. In November 1999, other CDP members

[12] Access to the internet in China operates through a two-tier system. First, local service providers connect end-users to a backbone consisting of a small number of official and quasi-official networks. Second, international connections are made through the backbone networks. It is as if the country had a nationwide 'firewall' — albeit a somewhat porous one Kalathil and Boas (2003, 21).

[13] Chase and Malvenon (2002, 72) discuss examples of the hacking in July 1999 of Falungong sites by computers with the same IP (internet protocol) address as the Chinese Ministry of Public Security.

[14] Kalathil and Boas (2003) also discuss arrests and other direct demonstrations of state authority, such as the police sweep of tens of thousands of internet cafes in 2001, but do not argue for the superiority of traditional authoritarian methods.

received jail sentences of 5 to 10 years for communicating with organizations the regime disapproves of and for posting subversive messages on bulletin boards (Chase and Malvenon, 2002, 53). A school teacher Jiang Shihua was jailed for two years for posting "We all think about one sentence that none of us will say: overthrow the Communist Party," (Kalathil and Boas, 2003, 26).

But examples of apparent success using traditional authoritarian methods, as provided by Chase and Malvenon (2002) and Kalathil and Boas (2003), are not inconsistent with the model. If the regime has a high willingness to imprison and otherwise punish dissent, then individuals face a high individual opportunity cost of subverting $p$. And as discussed in Section 3.1 a high $p$ and improvements in the quality of information $\alpha$ are *complementary* in that the threshold $\theta^*$ decreases faster in $\alpha$ when $p$ is high (as illustrated in Figure 4). So, rather than being viewed as competing explanations, co-opting new technologies and traditional authoritarian practices should be seen as partial but mutually-reinforcing explanations of China's ability to offset the subversive effects of the information revolution.

## 4.2 Cuba

Both the US government and the Cuban exile community have attempted to use information provision to overthrow Castro's regime. Offsetting these efforts, the regime has engaged in relatively sophisticated attempts to jam US signals beamed into Cuba.

The CIA began covert broadcasts into Cuba as early as 1960, the year after Castro took power Soley (1987). This policy became more explicit with the beginning of broadcasts by *Radio Martí* (or 'Radio Free Cuba') in 1985 and *TV Martí* in 1990 (Kalathil and Boas, 2003, 48). Although radio broadcasts have had some success — at least in terms of attracting an audience, if not in terms of bringing about regime change — TV broadcasts by the US government have been notoriously unsuccessful due to Cuban jamming of the signal. In 2003, the US government began to upgrade *TV Martí* to a satellite broadcasting system in an attempt to strengthen the signal and beat the jamming. In addition to US government-sponsored broadcasts, Cuban exiles operating from the US have engaged in leaflet drops by plane.

Suppose we put aside any concerns that Cubans might have about the motives underlying external propaganda and agree to treat all extra information as a pure increase in the quality of their information. If we interpret an array of leaflet drops, satellite TV transmissions and sporadic internet access as an exogenous increase in the intrinsic quality of private information $\alpha$, then from Proposition 2 we ought to predict that such improvements act to *reduce coordination* to the benefit of otherwise relatively weak regimes.

This prediction seems to be born out in practice. So far, the regime has been relatively successful at blocking these sources of external propaganda. Foreign providers of information have had to rely on relatively diffuse means of communication to change the beliefs of the Cuban public. While this may improve the information available to Cubans, it may also suffer from an inability to generate a large mass of subversives willing to move against the regime. Improvements in information do not necessarily help solve the coordination problem.

## 5 Accumulating information and waves of unrest

In models of information and regime change based on *information cascades*, such as Kuran (1991) and Lohmann (1994), individuals' decisions are staggered and some get to learn from the decisions of others. But in the benchmark model of this paper, individuals simultaneously receive their signals and so cannot learn from each other. This section extends the model by confronting the regime with two 'waves' of unrest that cumulate into an aggregate attack. By observing the outcome of the first wave, individuals in the second learn more about the likelihood of overthrowing the regime. And a regime will take this into account when deciding how to manipulate information. The key results of the static model of this paper extend to this setting.

Two waves of unrest. Let citizens be exogenously divided into two waves: a 'leading' wave of size $\lambda \in (0,1)$ and a 'following' wave of size $1-\lambda$. Let $S_1$ denote the size of the first wave attack and let $S_2$ denote the size of the second wave attack. The cumulative attack is then $S := \lambda S_1 + (1-\lambda)S_2$ and the regime is overthrown if $S > \theta$.

Both waves get the idiosyncratic signal of the regime's type $x_i = \theta + a + \varepsilon_{x,i}$ with $\varepsilon_{x,i}$ IID normally distributed with mean zero and precision $\alpha_x > 0$. But in addition citizens in the second wave get endogenous idiosyncratic signals $y_i$ about the size of the first attack. Let $y_i := \Phi^{-1}(S_1) + \varepsilon_{y,i}$ where $\varepsilon_{y,i}$ is IID normally distributed with mean zero and precision $\alpha_y > 0$.[15] If citizens in the first wave attack when they have $x_i < x_1^*$ for some endogenous threshold $x_1^*$, the size of the first attack is

$$S_1(\theta, a) = \Phi[\sqrt{\alpha_x}(x_1^* - \theta - a)] \tag{30}$$

So the extra signals for the second wave are $y_i = \sqrt{\alpha_x}(x_1^* - \theta - a) + \varepsilon_{y,i}$. This is equivalent to giving them signals $z_i = \theta + a + \varepsilon_{z,i}$ where $\varepsilon_{z,i}$ is IID normally distributed with mean zero and endogenous precision $\alpha_z = \alpha_x \alpha_y$. If citizens in the second wave attack when they have $x_i < x_2^*(z_i)$ for some endogenous threshold function $x_2^*(z_i)$, the size of the second attack is

$$S_2(\theta, a) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2^*(z_i)} \sqrt{\alpha_x}\phi[\sqrt{\alpha_x}(x_i - \theta - a)]\sqrt{\alpha_z}\phi[\sqrt{\alpha_z}(z_i - \theta - a)]dx_i dz_i \tag{31}$$

$$= \int_{-\infty}^{\infty} \Phi[\sqrt{\alpha_x}(x_2^*(z_i) - \theta - a)]\sqrt{\alpha_z}\phi[\sqrt{\alpha_z}(z_i - \theta - a)]dz_i \tag{32}$$

The regime's hidden actions $a(\theta)$ and cutoff $\theta^*$ solve

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \left[\theta - S(\theta, a) - C(a)\right], \quad \theta \geq \theta^* \tag{33}$$

where $S(\theta, a) = \lambda S_1(\theta, a) + (1-\lambda)S_2(\theta, a)$ and

$$\theta^* = S[\theta^*, a(\theta^*)] + C[a(\theta^*)] \tag{34}$$

[15] The specification $\Phi^{-1}(S)$ for the signal mean was introduced (in a different context) by Dasgupta (2007) and gives rise to a simple information structure.

The single threshold $x_1^*$ for the first wave and the threshold function $x_2^*(z_i)$ for the second wave solve the indifference conditions

$$p = \Pr(\theta < \theta^* | x_1^*) \tag{35}$$

$$= \Pr(\theta < \theta^* | x_2^*(z_i), z_i) \tag{36}$$

where the posteriors are calculated using Bayes's rule. An equilibrium of this model consists of scalars $x_1^*$ and $\theta^*$ and functions $x_2^*(z_i)$ and $a(\theta)$ simultaneously satisfying (33)-(36). I solve for an equilibrium numerically, see Appendix B for details.

Discussion. The accumulation of information through waves of unrest means that a regime's manipulation has both a direct and an indirect effect on beliefs. As usual, there is a direct effect through the signals $x_i$. But there is also an indirect effect through the signals $z_i$ generated by the first attack. In choosing its policy $a(\theta)$ the regime has to take account of both channels.

This model of leaders and followers is based on Angeletos and Werning (2006) but differs in two respects. First, in Angeletos and Werning (2006) there is no information manipulation. And second, in their model, the information generated by the first-wave is *public* rather than idiosyncratic as it is here. This captures the idea that unrest may take place in many physically separated locations and individuals are likely to have idiosyncratically varying amounts of information about what has happened depending on how close they are to other centers of unrest.[16]

Numerical examples. Figure 6 shows $\theta^*$ as a function of the signal precision $\alpha_x$ for various sizes of the first wave attack $\lambda$ and levels of $p$. The left panel shows a low cost of subversion, $p = 0.25$ while the right panel shows a high cost of subversion, $p = 0.75$. As in the static model, for high enough signal precision $\alpha_x$ the threshold $\theta^*$ is less than the benchmark Morris-Shin level of $\theta^*_{\text{MS}} = 1 - p$ and the regime benefits from the increase in precision. One difference is that with high $p$ there is a non-monotonicity in $\theta^*$: for low $\alpha_x$, the thresholds are increasing and the regime is worse off than the benchmark Morris-Shin case but the thresholds reach a peak before falling for high enough $\alpha_x$. This monotonicity becomes less important when the size of the first wave is large. As $\lambda \to 1$ the model reduces to the static case where we know from Figure 4 that $\theta^*$ is monotonically declining in signal precision. More generally, when $\lambda$ is high so that most citizens are in the first wave the threshold $\theta^*$ is low and the regime's survival probability is high.

# 6 Skepticism

The model that generates these results is stylized and it's reasonable to question it. Moreover, even researchers who have concluded that autocratic regimes like China have so far been successful

[16] Angeletos and Werning (2006) are concerned with information generated by market prices and in this context it is natural to assume that the generated information is public. They show that if market prices aggregate idiosyncratic information, then the precision of the public information contained in prices is increasing in the precision of underlying private information. Following similar global games with exogenous public information, this can lead to 'approximate' common knowledge and reintroduce multiple self-fulfilling equilibria. See Hellwig (2002) and Morris and Shin (2000, 2003, 2004).

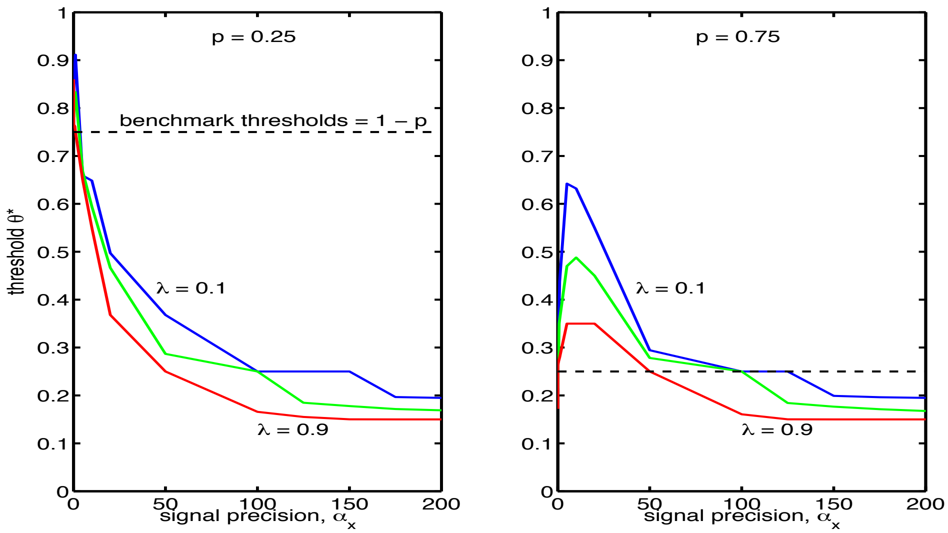


Figure 6: Thresholds $\theta^*$ as functions of precision $\alpha_x$ and size of first wave $\lambda$. With manipulation, a large enough $\alpha_x$ implies both that $\theta^* < 1-p$ and that $\theta^*$ is declining in $\alpha_x$ so that better information increases a regime's likelihood of survival. All calculations use $C(a) := 0.5a^2$ and $\alpha_z = \alpha_x$.

in offsetting the impact of the information revolution are skeptical that this success can continue. For example,

> While Beijing has done a remarkable job thus far of finding effective counterstrategies to what it perceives as the potential negative effects of the information revolution, the scale of China's information-technology modernization would suggest that time is eventually on the side of the regime's opponents (Chase and Malvenon, 2002, xiii).

One way to formally justify skepticism of this kind is to argue the main effect of the information revolution will be to help citizens realize the returns from overthrowing the regime are high.

**Higher returns from regime's overthrow.** Suppose that developments in information technology cause citizens in a closed autocratic society to learn that the return to overthrowing the regime is not $w \in \{\underline{w}, \overline{w}\}$ as they had thought, but instead $\xi w$ for $\xi > 1$. This just means their opportunity cost falls to $p/\xi < p$. The effects of such changes are straightforward:

Proposition 4. As the opportunity cost $p \to 0^+$ the limiting thresholds and hidden actions are

$$\lim_{p\to 0^+} \theta_p^* = 1^-, \qquad \lim_{p\to 0^+} x_p^* = +\infty, \quad \text{and} \quad \lim_{p\to 0^+} a_p(\theta) = 0^+, \quad \text{for all } \theta$$

But as $p \to 1^-$ the limiting thresholds and hidden actions are

$$\lim_{p\to 1^-} \theta_p^* = 0^+, \qquad \lim_{p\to 1^-} x_p^* = -\infty, \quad \text{and} \quad \lim_{p\to 1^-} a_p(\theta) = 0^+, \quad \text{for all } \theta$$

If citizens believed that the cost was $p$ with return $w$ and are surprised to learn it is $\xi w$, for big enough $\xi$ almost all citizens find it optimal to subvert, the ex ante survival probability of the regime falls, and no regime finds it worthwhile to take any hidden action.

Perhaps this is all that is meant when people talk enthusiastically of the role of information technologies in overthrowing autocracies. Maybe 'better information' means that citizens will learn the expected return to overthrowing the regime is larger than they had believed. This notion of an information revolution is quite Panglossian. As a partial corrective, in this paper I emphasize the more surprising result of Proposition 2, which suggest a dramatic increase in the intrinsic, technological, quality of information may be exploited by regimes to inhibit coordination and make their overthrow less likely. In reality, both notions of an information revolution are likely to be at play. Better information may help citizens in closed autocratic societies realize that their opportunities are not so limited, but it's important to realize that at the same time higher quality information may also help regimes inhibit coordination.

# 7 Conclusion

This paper presents a model of information quality and political regime change. If enough citizens act against a regime, it is overthrown. Citizens are imperfectly informed about how hard this will be and the regime can, at a cost, engage in propaganda so that at face-value it seems hard.

The most surprising result is that as the intrinsic quality of information becomes sufficiently high, the regime is more likely to survive. A regime can co-opt a technological improvement in the quality of information to make coordination against it more difficult. Perhaps information revolutions are not so threatening to autocratic regimes as is often supposed.

The model yields two additional insights. First, even rational citizens may find it difficult to filter their information appropriately when they are playing a coordination game and need to forecast not only the behavior of the regime but also the behavior of their fellow citizens. Knowing a regime's incentives, citizens discount their signals. But they may not discount enough. Second, a fall in the cost of influencing information benefits regimes. Regimes benefit from ambiguity about the amount of manipulation they do. As the costs of manipulation fall, this ambiguity increases and so does the regime's probability of surviving.

The coordination game in this paper is deliberately stylized so as to focus attention on how a regime can implicitly co-opt an information revolution. In keeping things simple, I have abstracted from issues that could play a role in a more complete theory. For example, in this paper I have assume free-riding is not a severe problem. But, rather than assuming it away, a more nuanced treatment would examine how citizens might alleviate the free-riding problem by using strategic communication to build credible coalitions.

# A Proofs

Proof of Proposition 1. (Morris-Shin benchmark): 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 $\theta$ 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 $\theta$, 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

$$\Phi\left[\sqrt{\alpha}\,(\theta^*_{\text{MS}} - x^*_{\text{MS}})\right] = p \tag{37}$$

$$\Phi\left[\sqrt{\alpha}\,(x^*_{\text{MS}} - \theta^*_{\text{MS}})\right] = \theta^*_{\text{MS}} \tag{38}$$

Since $\Phi(-w) = 1 - \Phi(w)$ for all $w \in \mathbb{R}$, adding these equalities gives $\theta^*_{\text{MS}} = 1 - p$. And plugging this solution for $\theta^*_{\text{MS}}$ back into (37) and rearranging gives $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha}$. □

The Proof of Proposition 2 is given in Edmond (2007) and is repeated here for completeness.

Proof of Proposition 2. For each precision $\alpha$, 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 $\alpha$, the solution to the original problem and to the constrained problem coincide.

The equilibrium conditions can be written

$$(1-p)\Phi[\sqrt{\alpha}\,(\theta^*_\alpha - x^*_\alpha)] = p\int_{\theta^*_\alpha}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x^*_\alpha - \theta - a_\alpha(\theta - x^*_\alpha))]d\theta \tag{39}$$

and

$$\theta^*_\alpha = \Phi[\sqrt{\alpha}(x^*_\alpha - \theta^*_\alpha - a_\alpha(\theta^*_\alpha - x^*_\alpha))] + C[a_\alpha(\theta^*_\alpha - x^*_\alpha)] \tag{40}$$

with actions characterized by

$$C'[a_\alpha(\theta - x^*_\alpha)] = \sqrt{\alpha}\phi[\sqrt{\alpha}(x^*_\alpha - \theta - a_\alpha(\theta - x^*_\alpha))], \qquad \theta \geq \theta^*_\alpha \tag{41}$$

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

$$\lim_{\alpha\to\infty} \sqrt{\alpha}(x^*_\alpha - \theta^*_\alpha - a_\alpha(\theta^*_\alpha - x^*_\alpha)) = \lim_{\alpha\to\infty} \sqrt{\alpha}(\theta^*_\alpha - x^*_\alpha) = -\infty \tag{42}$$

If condition (42) holds, from (40) we have $\theta^*_\alpha = C[a_\alpha(\theta^*_\alpha - x^*_\alpha)]$. Similarly, if (42) holds, then $\Phi[\sqrt{\alpha}\,(\theta^*_\alpha - x^*_\alpha)] \to 0$ and the value of the integral on the right hand side of (39) converges to zero. From (39) and (41), 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 $\alpha$ , 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 $\alpha$, $x^*_\alpha - \theta^*_\alpha$ is positive and $x^*_\alpha - \theta^*_\alpha - a_\alpha(\theta^*_\alpha - x^*_\alpha)$ is negative. For both constraints to have the same sign, $x^*_\alpha$ 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 (39) must converge to zero. Hence, by (41), $a_\alpha(\theta - x^*_\alpha) \to 0^+$ for all $\theta \geq \theta^*_\alpha$ (the strict convexity of $C$ is assumed here so that (41) holds for all $\theta$ 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 (40) requires that $\theta^*_\alpha \to 1^-$. Once again we have found a solution to the constrained problem. $\square$

The proofs of Propositions 3 and 4 are similar.

Proof of Proposition 3. Let $k \to \infty$. Then by assumption $C_k(a)$ and $C'_k(a) \to \infty$ for all $a > 0$. But the citizen's indifference condition [analogous to equation (39)] implies the bound

$$0 \leq \int_{\theta^*_k}^{\infty} C'_k[a_k(\theta - x^*_k)]d\theta = \frac{1-p}{p}\Phi[\sqrt{\alpha}\,(\theta^*_k - x^*_k)] \leq \frac{1-p}{p} < \infty$$

independent of $k$. Since $\theta^*_k \in [0,1]$, the only way this bound can be satisfied as $k \to \infty$ is if $a = 0$ for all $\theta \geq \theta^*_k$. Hence $\lim_{k\to\infty} a_k(\theta - x^*_k) = 0^+$ for all $\theta$. It is then immediate that $\lim_{k\to\infty}\theta^*_k = \theta^*_{\text{MS}} = 1 - p$ and $\lim_{k\to\infty} x^*_k = x^*_{\text{MS}}$. Similarly, let $k \to 0^+$. Then by assumption $C_k(a)$ and $C'_k(a) \to \underline{c} > 0$ for all $a > 0$. Since all positive actions cost the same amount in this limit, if a regime takes any positive action, it will take the biggest action, $\overline{a}$. Hence it is common knowledge that either the regime take no action (if $\theta < \lim_{k\to 0^+}\theta_k$) or the regime takes action $a = \overline{a}$ (if $\theta \geq \lim_{k\to 0^+}\theta_k$). To find the limiting thresholds $x^*_0, \theta^*_0$ follow the calculations leading to (22) but allow for arbitrary $p$ (instead of $p = 1/2$) and write $\overline{a} =: \hat{a}$ and $\underline{c} =: C(\hat{a})$. $\square$

Proof of Proposition 4. Let $p \to 1^-$. From the citizen's indifference condition [analogous to equation (39)], this requires that $\int_{\theta^*_p}^{\infty} C'[a_p(\theta - x^*_p)]d\theta \to 0^+$ and since $C' \geq 0$ and is continuous and $\theta^*_p \in [0,1]$, this can only be true if $a_p(t) \to 0^+$ for all $t \geq \theta^*_p$. According to the first order condition (19), this requires $x^*_p$ to diverge ($x^*_p \to \pm\infty$). But if $x^*_p \to +\infty$, all citizens engage in subversion which cannot be individually rational if $p \to 1^-$. Hence as $p \to 1^-$, $x^*_p \to -\infty$. Then according to the regime's indifference condition, we must also have $\theta^*_p - \Phi[\sqrt{\alpha}(x^*_p - \theta^*_p - a_p(\theta^*_p))] \to 0$, and so $\theta_p \to 0^+$. Similarly, let $p \to 0^+$. From the citizen's indifference condition, this requires that $\Phi[\sqrt{\alpha}\left(x^*_p - \theta^*_p\right)] \to 1^-$ and since $\theta^*_p \in [0,1]$, this requires $x^*_p \to +\infty$. Then the first order condition implies $a_p(t) \to 0^+$ for all $t \geq \theta^*_p$ and the regime's indifference condition implies $\theta^*_p \to 1^-$. $\square$

# B Computing an equilibrium with waves of unrest

Every citizen gets an exogenous idiosyncratically noisy signal $x_i = \theta + a + \varepsilon_{x,i}$. Fraction $\lambda \in (0,1)$ of the population is in the leading wave with decisions based only on $x_i$. The complementary $1 - \lambda$ fraction is in the following wave and get a second endogenous signal $z_i = \theta + a + \varepsilon_{z,i}$. An equilibrium of this model consists of scalars $x_1^*, \theta^*$ and functions $x_2^*(z_i)$ and $a(\theta)$ where citizens in the first wave subvert if and only if $x_i < x_1^*$ and citizens in the second wave subvert if and only if $x_i < x_2^*(z_i)$. In what follows, let $f_X(x_i|\theta, a)$ and $f_Z(z_i|\theta, a)$ denote the signal densities and let a circumflex above a variable denote a candidate equilibrium object.

The posterior beliefs of citizens who face a threshold rule about $\hat{\theta}$ and manipulation policy $\hat{a}(\theta)$ are summarized by

$$P_1(\hat{\theta}, x_i) := \Pr(\theta < \hat{\theta}|x_i) = \frac{\int_{-\infty}^{\hat{\theta}} \hat{\pi}_1(\theta|x_i)d\theta}{\int_{-\infty}^{\infty} \hat{\pi}_1(\theta|x_i)d\theta} \tag{43}$$

$$P_2(\hat{\theta}, x_i, z_i) := \Pr(\theta < \hat{\theta}|x_i, z_i) = \frac{\int_{-\infty}^{\hat{\theta}} \hat{\pi}_2(\theta|x_i, z_i)d\theta}{\int_{-\infty}^{\infty} \hat{\pi}_2(\theta|x_i, z_i)d\theta} \tag{44}$$

where $\hat{\pi}_1(\theta|x_i) := f_X[x_i|\theta, \hat{a}(\theta)]$ and $\hat{\pi}_2(\theta|x_i, z_i) := f_X[x_i|\theta, \hat{a}(\theta)]f_Z[z_i|\theta, \hat{a}(\theta)]$.

Given $\hat{\theta}$ and $\hat{a}(\theta)$ marginal citizens, $\hat{x}_1$ for the first wave and $\hat{x}_2(z_i)$ for the second wave, are determined using

$$p = P_1(\hat{\theta}, \hat{x}_1) \tag{45}$$

$$= P_2(\hat{\theta}, \hat{x}_2(z_i), z_i) \tag{46}$$

Given $\hat{x}_1$ and $\hat{x}_2(z_i)$, hidden actions for the regime are $\hat{a}(\theta) = 0$ for $\theta < \hat{\theta}$ and otherwise

$$a(\theta) \in \operatorname*{argmax}_{a \geq 0} \left[\theta - \hat{S}(\theta, a) - C(a)\right], \quad \theta \geq \hat{\theta} \tag{47}$$

where the aggregate mass of subversives is $\hat{S}(\theta, a) = \lambda \hat{S}_1(\theta, a) + (1 - \lambda)\hat{S}_2(\theta, a)$ with

$$\hat{S}_1(\theta, a) = \int_{-\infty}^{\hat{x}_1} f_X(x_i|\theta, a)dx_i = F_X(\hat{x}_1|\theta, a) \tag{48}$$

$$\hat{S}_2(\theta, a) = \int_{-\infty}^{\infty} \int_{-\infty}^{\hat{x}_2(z_i)} f_X(x_i|\theta, a)f_Z(z_i|\theta, a)dx_i dz_i = \int_{-\infty}^{\infty} F_X(\hat{x}_2(z_i)|\theta, a)f_Z(z_i|\theta, a)dz_i \tag{49}$$

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

$$\hat{\theta} = \hat{S}[\hat{\theta}, \hat{a}(\hat{\theta})] + C[\hat{a}(\hat{\theta})] \tag{50}$$

I solve this model numerically. I first guess values $\hat{x}_{1,n}, \hat{x}_{2,n}(z_i)$, $\hat{\theta}_n$ (say) and then solve the optimization problem in (47) for the associated hidden actions, call these $\hat{a}_{n+1}(\theta)$. I then use $\hat{\theta}_n$ and $\hat{a}_{n+1}(\theta)$ to compute revised estimates of the thresholds $\hat{x}_{1,n+1}$ and $\hat{x}_{2,n+1}(z_i)$ from (45)-(46) and use $\hat{a}_{n+1}(\theta)$ and the thresholds $\hat{x}_{1,n+1}, \hat{x}_{2,n+1}(z_i)$ to compute a new cutoff $\hat{\theta}_{n+1}$ from the single condition (50). I iterate until the process converges. Standard quadrature rules are used to calculate integrals numerically. 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. 3

Angeletos, G.-M., and I. Werning (2006): "Crises and prices: Information aggregation, multiplicity and volatility," *American Economic Review*, 96(5), 1720–1736. 21

Arendt, H. (1973): *The origins of totalitarianism*. André Deutsch, revised edn. 2

Carlsson, H., and E. van Damme (1993): "Global games and equilibrium selection," *Econometrica*, 61(5), 989–1018. 3, 7

Chamley, C. P. (2004): *Rational herds: Economic models of social learning*. Cambridge University Press. 3

Chase, M. S., and J. C. Malvenon (2002): *You've got dissent: Chinese dissident use of the internet and Beijing's counter-strategies*. RAND report MR-1543. 3, 17, 18, 19, 22

Cooper, R. W. (1999): *Coordination games: Complementarities and macroeconomics*. Cambridge University Press. 3

Dasgupta, A. (2007): "Coordination and delay in global games," *Journal of Economic Theory*, 134, 195–225. 20

Edmond, C. (2007): "Information manipulation, coordination and regime change," NYU working paper. 1, 3, 8, 9, 24

Friedrich, C. J., and Z. K. Brzezinski (1965): *Totalitarian dictatorship and autocracy*. Harvard University Press, 2 edn. 2

Frommer, B. (2005): *National cleansing: Retribution against Nazi collaborators in postwar Czechoslovakia*. Cambridge University Press. 5

Ginkel, J., and A. Smith (1999): "So you say you want a revolution: A game-theoretic explanation of revolution in repressive regimes," *Journal of Conflict Resolution*, 43(3), 291–316. 3

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

Hodgson, M. G. (1974): *The venture of Islam: Conscience and history in a world civilization*, vol. III. University of Chicago Press. 4

Jackson, J. (2001): *France: The dark years, 1940-1944*. Oxford University Press. 5

Kalathil, S., and T. C. Boas (2003): *Open networks, closed regimes: The impact of the internet on authoritarian rule*. Carnegie Endowment for International Peace. 1, 3, 17, 18, 19

Karklins, R., and R. Petersen (1993): "Decision calculus of protesters and regimes: Eastern Europe 1989," *Journal of Politics*, 55(3), 588–614. 3

Kuran, T. (1991): "Now out of never: The element of surprise in the European revolution of 1989," *World Politics*, 44(1), 7–48. 3, 20

LEWIS, B. (2002): *What went wrong? The clash between Islam and modernity in the Middle East.* HarperCollins. 4

LOHMANN, S. (1994): "The dynamics of information cascades: The Monday demonstrations in Leipzig, East Germany, 1989-91," *World Politics*, 47(1), 42–101. 3, 20

LYNCH, D. C. (1999): *After the propaganda state: Media, politics, and "thought work" in reformed China*. Stanford University Press. 17

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

——— (2000): "Rethinking multiple equilibria in macroeconomic modeling," in *NBER Macroeconomics Annual*, ed. by B. S. Bernanke, and K. Rogoff, pp. 139–161. MIT Press. 3, 7, 21

——— (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. 3, 7, 21

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

SKOCPOL, T. (1979): *States and social revolutions: A comparative analysis of France, Russia, and China*. Cambridge University Press. 5

SOLEY, L. C. (1987): *Clandestine radio broadcasting: A study of revolutionary and counterrevolutionary electronic broadcasting*. Praeger. 19

ZEMAN, Z. (1973): *Nazi propaganda*. Oxford University Press, second edn. 2, 15