NBER WORKING PAPER SERIES

INFORMATION MANIPULATION, COORDINATION, AND REGIME CHANGE

Chris Edmond

Working Paper 17395
http://www.nber.org/papers/w17395

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2011

This is a revised version of a chapter from my UCLA dissertation. Previous versions circulated under the title "Information and the limits to autocracy". I would particularly like to thank Andrew Atkeson and Christian Hellwig for their encouragement and many helpful discussions. I also thank Marios Angeletos, Costas Azariadis, Heski Bar-Isaac, Adam Brandenburger, Ethan Bueno de Mesquita, Ignacio Esponda, Catherine de Fontenay, Matias Iaryczower, Stephen Morris, Alessandro Pavan, Andrea Prat, Hyun Shin, Adam Szeidl, Laura Veldkamp, Iván Werning, seminar participants at the ANU, Boston University, University of Chicago (GSB and Harris), Duke University, FRB Minneapolis, La Trobe University, LSE, University of Melbourne, MIT, NYU, Northwestern University, Oxford University, University of Rochester, UC Irvine, University College London and UCLA and participants at the NBER political economy meetings for their comments. Vivianne Vilar provided excellent research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Information Manipulation, Coordination, and Regime Change
Chris Edmond
NBER Working Paper No. 17395
September 2011
JEL No. C7,D7,D8

ABSTRACT

This paper presents a model of information and political regime change. If enough citizens act against a regime, it is overthrown. Citizens are imperfectly informed about how hard this will be and the regime can, at a cost, engage in propaganda so that at face-value it seems hard. This coordination game with endogenous information manipulation has a unique equilibrium and the paper gives a complete analytic characterization of its comparative statics. If the quantity of information available to citizens is sufficiently high, then the regime has a better chance of surviving. However, an increase in the reliability of information can reduce the regime's chances. These two effects are always in tension: a regime benefits from an increase in information quantity if and only if an increase in information reliability reduces its chances. The model allows for two kinds of information revolutions. In the first, associated with radio and mass newspapers under the totalitarian regimes of the early twentieth century, an increase in information quantity coincides with a shift towards media institutions more accommodative of the regime and, in this sense, a decrease in information reliability. In this case, both effects help the regime. In the second kind, associated with diffuse technologies like modern social media, an increase in information quantity coincides with a shift towards sources of information less accommodative of the regime and an increase in information reliability. This makes the quantity and reliability effects work against each other. The model predicts that a given percentage increase in information reliability has exactly twice as large an effect on the regime's chances as the same percentage increase in information quantity, so, overall, an information revolution that leads to roughly equal-sized percentage increases in both these characteristics will reduce a regime's chances of surviving.

Chris Edmond
Department of Economics
University of Melbourne
Parkville VIC 3010
AUSTRALIA
chris.edmond@gmail.com

1 Introduction

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

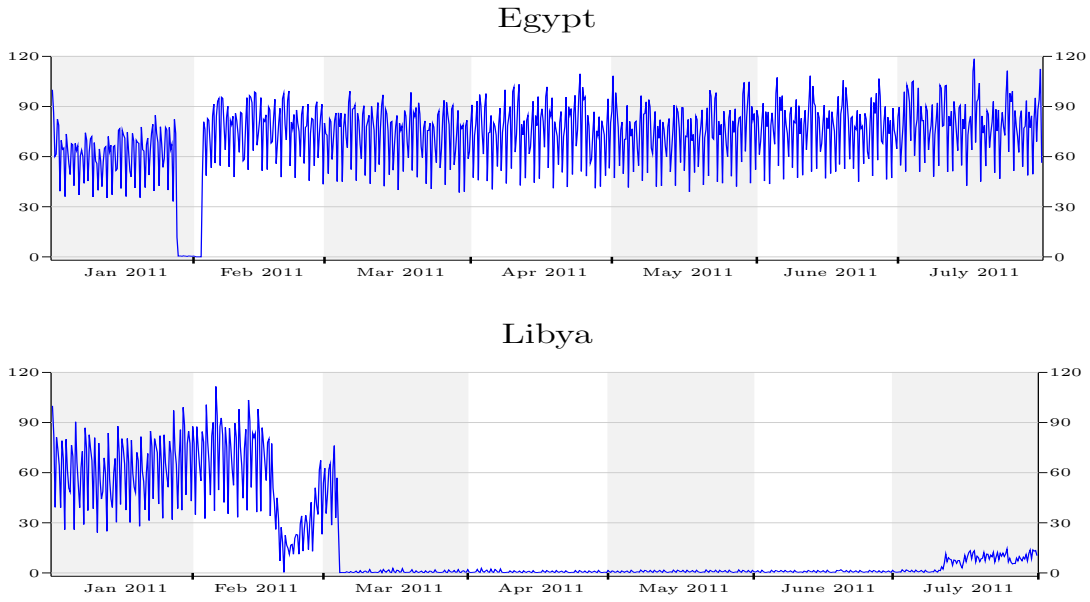


Figure 1: Recent disruptions to Google internet traffic under autocratic regimes.

Country traffic divided by worldwide traffic and normalized to 100 on January 2, 2011. (a) Egypt: starting on January 28, 2011, all Google services were inaccessible for 5 days during the height of protests against the Mubarak regime. (b) Libya: starting on March 4, 2011, all Google services became inaccessible as the civil war intensified. Source: Google Transparency project, August 2011.

Optimism about the use of new technologies in putting autocratic regimes under sustained pressure is hardly new; social media is only the latest technology to be viewed as a catalyst for regime change. Simple internet access, cell phones, satellite television, radio and newspaper have all been viewed as potential catalysts too. While information optimism has a long and somewhat mixed history, it is also worth bearing in mind that the relationship between new information technologies and autocratic regimes has a prominent dark side. Perhaps the most well known examples are the

[1] See, for instance, Kirkpatrick (2011) for an account of the use of social media during the Egyptian protests. Musgrove (2009) discusses the role of Twitter in the Iranian demonstrations. Optimistic sentiments have also been expressed in the context of Google's recent decision to cease its self-censorship of search results for Chinese audiences. See MacMillan (2010) for a round-up of reactions to Google's announcements.

use of mass media propaganda by totalitarian regimes like Nazi Germany and the Soviet Union. And even with more recent developments, it is clear that breakthroughs in information technology also provide opportunities for the regime. During the Iranian demonstrations, technologies like Twitter allowed the regime to spread rumors and disinformation (Esfandiari, 2010). Similarly, it is now well-known that the Chinese regime engages in efforts to counter online organization, efforts that make use of the exact same technologies that optimists hope will help in bringing regime change (Kalathil and Boas, 2003; Fallows, 2008; Morozov, 2011).

So, should we be optimistic that recent breakthroughs in information technology will lead to the collapse of present-day autocratic regimes? To help address this question, I develop a simple model of information and regime change. While stylized, this model provides a number of insights into ways in which a regime's chances of survival are affected by changes in information technology. The model predicts that (i) an increase in the *quantity* of information can increase the regime's chances of survival,[2] but (ii) an increase in the *reliability* of information can reduce the regime's chances. The model also predicts that these two effects are *always in tension*. The circumstances where an increase in the reliability of information works against a regime's interests are precisely the circumstances where an increase in the quantity of information is in the regime's interests.

The model clearly identifies situations where pessimism about the ability of new information technologies to threaten autocratic regimes is born out. The simplest example of a pessimistic situation is where an increase in the quantity of information is accompanied by a *decrease* in the reliability of information, e.g., if the media is increasingly cowed by and accommodative of the regime. Indeed, the model predicts that a regime will want to exert a strong influence over the media exactly when new technologies make the quantity of information high. In practice, this does seem to be a feature of autocratic regimes: a clear example is the heavily subsidized diffusion of radios in Nazi Germany (Zeman, 1973). More generally, even if an increase in the quantity of information is accompanied by an increase in information reliability, it can still be the case that the regime benefits if the size of the change in reliability is not large enough.

That said, the model also clearly gives some grounds for optimism that increases in information reliability can more than make up for increases in information quantity. The model predicts that a given percentage increase in information reliability has exactly *twice* as large an effect on a regime's chances of surviving as the same percentage increase in the quantity of information. Thus breakthroughs in information technology that lead to roughly equal-sized percentage increases in information quantity and reliability will reduce a regime's chances of surviving.

Section 2 outlines the model. There is a single regime and a large number of citizens with heterogeneous information. Citizens can either subvert the regime or not. If enough of them subvert the regime, it is overthrown. The ability of the regime to withstand an attempted overthrow is

[2]Moreover, the consequences of an increase in the quantity of information do *not* operate through the kinds of mechanisms identified by Mullainathan and Shleifer (2005), Sobbrio (2010), or Stone (2010) whereby increasing the number of media outlets induces greater polarization or segregation amongst consumers of information.

determined by a single parameter, the regime's type. Citizens are imperfectly informed about the regime's type and may coordinate either on overturning the regime or not. The regime is informed about its type and seeks to induce coordination on the status quo. Citizens receive information from a collection of "media outlets". These media outlets place some weight on reporting the regime's true type and some weight on accommodating the regime's preferred message, a message that depends on a costly *hidden action* taken by the regime. Balancing these considerations, each media outlet produces a report and citizens observe these reports with idiosyncratic noise. Effectively, this gives the regime a *signal-jamming* technology that influences the distribution of signals so that citizens receive information that suggests, at face-value, that the regime is difficult to overthrow. Citizens are rational and internalize the regime's incentives when forming their beliefs. Section 3 gives the first main result of the paper: this coordination game with endogenous information manipulation has a unique perfect Bayesian equilibrium (Proposition 1).

Section 4 turns to the question of whether more information can assist citizens in overthrowing the regime. The second main result of this paper is that the regime's information manipulation is *effective*, in the sense of increasing the regime's ex ante survival probability, when the quantity of information is sufficiently high (Proposition 2). In the model, the quantity of information is proportional to the number of media outlets. When this quantity of information is sufficiently high, *the regime survives in all states where it is possible for the regime to survive.*

The reason for this result is as follows. 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. Two effects then come into play: (i) when the quantity of information is high, citizens have signals that are precise (of low variance) and hence tightly clustered around the signal mean, and (ii) collectively the citizens are not, in equilibrium, able to completely infer the extent of the regime's manipulation. Because of the clustering around the mean, even quite a small increase in the signal mean can cause a large fall in the size of the aggregate attack on the regime. Some regimes can achieve precisely such an increase in the signal mean when there is a collective inability to infer the extent of manipulation. In turn, two features of the model account for the inability of citizens to correctly infer the manipulation: (a) different regime types take different actions so that there is uncertainty about the amount by which citizens should discount their signals, and (b) citizens are imperfectly coordinated. If all regimes took the same action or if citizens were perfectly coordinated, and hence able to completely pool their disparate information, then there would be no difficulty in inferring the extent of manipulation and any ability the regime has to use propaganda is rendered ineffectual.

While an increasing quantity of information can be a source of effective manipulation, the regime's chances of survival also depend on the reliability of information as determined by the media's willingness to accommodate the regime. If the media is unwilling to accommodate, then the regime's effective costs of manipulation are high. The third main result of this paper is that an increase in media reliability lowers a regime's chances of survival in exactly those situations where

an increase in the quantity of information raises a regime's chances (Proposition 3). In this sense, the two effects always pull in opposite directions.

Section 5 then uses the comparative static results to interpret historical and present-day examples of the relationships between information technologies, propaganda, and autocratic regimes. Section 6 provides various extensions of the model. These address questions such as:

- what if the regime is confronted by an increasing number of media outlets that are completely independent of its manipulation, will that necessarily undermine it?
- is manipulation more effective if it works through aggregate (i.e., common) information rather than individual information?
- how do outcomes change if the regime is challenged by a consolidated opposition that also tries to change beliefs?

Section 7 then concludes. All proofs and lengthy derivations are in the appendices.

Political economy of regime change and imperfect information. Political regime change is an important subject both in its own right and because the threat of regime change is an essential part of modern theories of democratization, the composition of civil society, economic and political redistribution, corruption, and a host of related topics. Acemoglu and Robinson (2006) and Bueno de Mesquita, Smith, Siverson, and Morrow (2003) provide recent introductions to this literature. To focus on the roles of information and coordination, this paper adopts a reduced form approach to the payoffs of the regime and citizens. It is taken as given that the regime prefers the status quo while citizens prefer regime change.

More specifically related are political economy models of coordination problems and/or imperfect information as barriers to regime change. Following the overthrow of the Eastern European communist regimes in 1989, Kuran (1989, 1991, 1995), Lohmann (1994b), Sandler (1992) and others adopted the use of models of *information cascades* to understand why regime change can occur seemingly spontaneously with no apparent change in economic or political fundamentals. Unlike this paper, in these contributions the regime is essentially passive and equilibrium outcomes do not depend on strategic interactions between the regime and the citizens.[3]

For simplicity, this paper adopts a static model with no cascades element. This makes the paper more closely related to Ginkel and Smith (1999) and Bueno de Mesquita (2010) who consider costly signaling by both a regime and a rival group of dissidents that each seek the support of a mass of citizens. In Ginkel and Smith there is no information heterogeneity.[4] By contrast, in Bueno de Mesquita, as in this paper, information heterogeneity plays a key role in determining equilibrium

[3] In an industrial organization context, however, see Bose, Orosel, Ottaviani, and Vesterlund (2006) for an information cascade problem where an informed monopolist seeks to control the ensuing herd behavior of consumers.

[4] See Baliga and Sjöström (2010) for a related model with cheap talk instead of costly signaling.

outcomes. In Bueno de Mesquita, heterogeneously informed citizens play a coordination game following the actions of the dissidents. The dissidents decide how much effort to expend on violent activities that send a noisy signal suggesting the regime is vulnerable. In this way, the dissidents seek to ensure that citizens coordinate on overthrowing the regime. My paper is complementary in that it also models regime change as a coordination game played by heterogeneous citizens, but focuses instead on the *regime's* efforts to ensure citizens coordinate on the status quo. Technically, however, the papers differ in several ways. Most importantly, in Bueno de Mesquita the dissidents are uninformed about the regime's type and so choose a single effort level (known in equilibrium). In my model, by contrast, the regime is informed and takes an action that depends on its type so that individual citizens have a genuine information filtering problem.

In other complementary work, Debs (2007) shows how a regime can use the media to implement *divide-and-rule* policies that may thwart regime change.

Media bias and media freedom. A recent literature determines the equilibrium degree of *media bias* emerging from competition between media outlets (e.g., Mullainathan and Shleifer, 2005; Baron, 2006; Gentzkow and Shapiro, 2006). Related work determines the equilibrium degree of *media freedom* from governmental influence (e.g., Besley and Prat, 2006; Egorov, Guriev, and Sonin, 2006; Gehlbach and Sonin, 2008). A common assumption in this literature is that some agents have an exogenous *preference* for information that is biased. This preference for bias affects the consumers in Mullainathan and Shleifer (2005), the journalists in Baron (2006), and the media outlets in Besley and Prat (2006). In my model citizens do prefer to know the truth, but cannot exactly infer the extent of manipulation and so some bias in their signals persists in equilibrium.

Coordination games with endogenous information. This paper draws on the *global games* approach to coordination games with imperfect information pioneered by Carlsson and van Damme (1993) and Morris and Shin (1998, 2000, 2003). While coordination games often have multiple equilibria, as is now widely known, the introduction of a small amount of idiosyncratic noise can ensure a unique equilibrium. In a political economy setting, Boix and Svolik (2009) have recently used the global games approach to study power-sharing arrangements in dictatorships while Chassang and Padro-i-Miquel (2010) have used the approach in work on strategic deterrence.[5]

The equilibrium uniqueness result in this paper contrasts with Angeletos, Hellwig, and Pavan (2006), who were the first to emphasize *endogenous* information in global games and who showed that this can lead to multiple equilibria.[6] In the version of their model closest to this paper, Angeletos, Hellwig, and Pavan let individuals receive *two* noisy signals, (i) a signal of the regime's

[5] Although a coordination game with heterogeneously informed citizens, Bueno de Mesquita (2010) is technically not a global game (because it lacks the *limit dominance* property). The threshold behavior in Persson and Tabellini (2009) is also motivated by a global game but that is not essential for their analysis.

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

action (an endogenous function of the underlying state), and (ii) a signal of the underlying state itself. By contrast, in my model individuals get *one* noisy signal of a function that depends on the underlying state both directly and indirectly through the regime's endogenous action. Loosely speaking, citizens in my model have less information about their strategic environment.

2 Model of information manipulation and regime change

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

Hidden actions. Given its θ, a regime may take a hidden action $\hat{a} \geq 0$ in an attempt to convey to citizens (via the media, as discussed below) that the regime's type is $\theta + \hat{a}$. Hidden actions incur a convex cost $C(\hat{a})$ where $C(0) = 0$, $C'(\hat{a}) > 0$ for $\hat{a} > 0$ and $C''(\hat{a}) \geq 0$ for all $\hat{a}$ with $C''(0) = 0$.

Regime payoffs. 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 minimize the costs of dealing with significant unrest and so wants S small even when it survives.[7] If $\theta < S$, the regime is overthrown and obtains an outside option with value normalized to zero. The payoff to a regime is therefore

$$B(S,\theta) - C(\hat{a}), \qquad B(S,\theta) := \max[0, \theta - S] \tag{1}$$

Media outlets. Citizens obtain information about the regime's type from N identical *media outlets*. Each media outlet $n = 1, ..., N$ chooses a signal mean y_n for the information it produces and each citizen $i \in [0,1]$ costlessly acquires a signal $x_{i,n}$, one from each of the N outlets.[8] Each signal is of the form $x_{i,n} = y_n + \varepsilon_{i,n}$ where the $\varepsilon_{i,n}$ are jointly IID normal across citizens and across media outlets with mean zero and precision $\hat{\alpha} > 0$ (that is, variance $1/\hat{\alpha}$). The owners of media outlets are assumed to have preferences that trade off a desire to *accommodate* the regime against a desire to provide a truthful, *reliable*, report of the regime's type. Each media outlet places a weight $r \in [0,1)$ on reporting the true type θ and weight $1 - r$ on accommodating the regime's preferred message $\theta + \hat{a}$. Each outlet chooses a signal mean y_n to minimize a quadratic loss function

$$r\,(y_n - \theta)^2 + (1 - r)\,(y_n - (\theta + \hat{a}))^2 \tag{2}$$

[7] The regime prefers to avoid events like suppressing a Prague Spring or a Tiananmen Square demonstration. The main results are unchanged if instead the regime cares only for survival and has no *direct* aversion to S.

[8] Following Mullainathan and Shleifer (2005), this can be interpreted as follows: the marginal cost of producing information is zero and Bertrand competition between symmetric media outlets has driven the price of information to zero. To be consistent with this interpretation, the number of symmetric media outlets should be $N \geq 2$.

with solution

$$y_n = \theta + (1 - r)\hat{a} \tag{3}$$

If the media is reliable, $r = 1$, then the signal mean is the true type θ while if the media is unreliable, $r = 0$, the signal mean is the regime's preferred report $\theta + \hat{a}$.

Citizen information. Citizens begin with common, uninformative, priors for the regime type θ. They then costlessly acquire their N signals from the media outlets

$$x_{i,n} := y_n + \varepsilon_{i,n} = \theta + (1 - r)\hat{a} + \varepsilon_{i,n}, \qquad n = 1, ..., N \tag{4}$$

Because the media outlets are symmetric, the information of citizen i can be represented by the average signal $x_i := \frac{1}{N}\sum_{n=1}^{N} x_{i,n}$ which satisfies[9]

$$x_i = \theta + (1 - r)\hat{a} + \varepsilon_i$$

where similarly $\varepsilon_i := \frac{1}{N}\sum_{n=1}^{N} \varepsilon_{i,n}$. Since the $\varepsilon_{i,n}$ are jointly IID normal across N with mean zero and precision $\hat{\alpha}$, a citizen's average noise ε_i is also normal with mean zero and precision $\alpha := N\hat{\alpha}$. Since it is proportional to N, I refer to the aggregate signal precision α as a measure of the *quantity* of information available to citizens. With this representation of citizen information, it is also natural to analyze the model in terms of the regime's *effective* hidden action $a := (1 - r)\hat{a}$. Since the qualitative properties of the model are the same for any fixed $r < 1$, I abuse notation slightly and generally write $C(a)$ rather than $C(a/(1 - r))$ for the regime's cost function. The density of x_i is

$$f(x_i|\theta, a) := \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a)) \tag{5}$$

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

Citizen payoffs. A citizen's payoffs depend on whether the regime is overthrown or not and on whether that individual participated or not. Let $p(S, \theta)$ denote the cost of subverting

$$p(S, \theta) := \begin{cases} \overline{p} & \text{if } \theta \geq S \\ \underline{p} & \text{if } \theta < S \end{cases}, \qquad 0 \leq \underline{p} \text{ and } \underline{p} \leq \overline{p}, \text{ strictly if } \underline{p} = 0 \tag{6}$$

so that an individual who subverts pays a higher price $\overline{p}$ if the regime survives and a lower price $\underline{p}$ if the regime is overthrown. This specification allows for the possibility that individual subversion is only costly if the regime survives (i.e., $\underline{p} = 0$ and $\overline{p} > 0$) or for the possibility that the cost of individual subversion does not depend on the regime outcome (i.e., $\underline{p} = \overline{p} > 0$). Similarly, let

[9] If different media outlets had different preferences for accommodating the regime, then citizens would not weigh them equally. An extension involving heterogeneous media outlets is given in Section 6 below.

$u(s_i, S, \theta)$ denote the benefit from the regime outcome

$$u(s_i, S, \theta) := \begin{cases} \overline{u} & \text{if } \theta < S \text{ and } s_i = 1 \\ \underline{u} & \text{if } \theta < S \text{ and } s_i = 0 \\ 0 & \text{otherwise} \end{cases}, \qquad 0 < \underline{u} \leq \overline{u} \tag{7}$$

so that if an individual subverts and the regime is overthrown, then that individual gets $\overline{u}$ while a citizen who "free-rides" on successful regime change gets $\underline{u} \leq \overline{u}$. Otherwise, if the regime survives, citizens get no benefit and pay costs according to (6) above. A citizen's net utility is

$$U(s_i, S, \theta) := u(s_i, S, \theta) - p(S, \theta)s_i \tag{8}$$

Or, in tabular form,

	subvert $s_i = 1$	not subvert $s_i = 0$
regime overthrown ($\theta < S$)	$\overline{u} - \underline{p}$	$\underline{u}$
not overthrown ($\theta \geq S$)	$0 - \overline{p}$	0

Citizens choose s_i to maximize expected utility.

2.1 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 regime hidden actions $a(\theta)$ such that

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

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

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

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

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 expected utility. 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 $\theta < S(\theta, a(\theta))$ and otherwise survives.

2.2 Further discussion of the model

Collective action and free-riding. This model involves a collective action problem. Overthrowing the regime requires *coordination* — the regime can only be overthrown if enough citizens act against it — but the benefits from regime change are a public good that can be enjoyed by all citizens.[10] As forcefully argued by Olson (1971), this creates an inventive for an individual to *free-ride* on the actions of others, an incentive that in turn undermines the prospects for successful regime change.

In this paper I impose a condition on citizen payoffs that prevents the incentive to free-ride from being "overwhelming" while still allowing this incentive to play a role in determining equilibrium outcomes. To derive this condition, let $P(x_i)$ denote the posterior probability assigned to the regime's overthrow for a citizen with signal x_i. The expected payoff from subverting the regime, $s(x_i) = 1$, is

$$(\overline{u} - \underline{p})P(x_i) + (0 - \overline{p})(1 - P(x_i))$$

while the expected payoff from not subverting the regime, $s(x_i) = 0$, is

$$(\underline{u} - 0)P(x_i) + (0 - 0)(1 - P(x_i))$$

Collecting terms and rearranging, this citizen will find subversion optimal if and only if

$$P(x_i) \geq \frac{\overline{p}}{(\overline{p} - \underline{p}) + (\overline{u} - \underline{u})} =: p \tag{9}$$

The difference $\overline{u} - \underline{u}$ measures the incentive to free-ride. A bad free-rider problem is *one* of the reasons why the effective opportunity cost of subversion, p, may be high. A sufficiently severe free-rider problem will make $p \geq 1$ in which case it is never rational for an individual to engage in subversion. To focus on the more interesting scenario where the free-rider problem is in *tension* with the coordination problem and the outcome of the game is not trivial, I assume parameters such that $p < 1$, specifically:

ASSUMPTION 1. The incentive to free-ride is not overwhelming, $\overline{u} - \underline{u} > \underline{p}$.

The existence of a differential gain to being part of a successful overthrow, $\overline{u} - \underline{u} > 0$, is necessary but not generally sufficient to ensure $p < 1$. In the important special case where $\underline{p} = 0$ so that citizens pay no price for subverting if the regime is successfully overthrown, however, then $\overline{u} - \underline{u} > 0$ is also sufficient to ensure $p < 1$. One straightforward interpretation of the differential gain $\overline{u} - \underline{u}$ is a higher probability of individual material rewards in the event of participating in successful regime change (more private consumption, preferential treatment, etc), but these considerations

[10] The use of a coordination game to model regime change is common in the political economy literature — see for example Kuran (1989, 1995) or more recently Fearon (2006) and Bueno de Mesquita (2010).

seem more appropriate for sustaining effective coordination by a small number of non-anonymous agents and less appropriate for a model of coordination by a large number of anonymous agents. Given this, it is important that the differential gains $\overline{u} - \underline{u}$ also capture non-material concerns such as individual *shame* from non-participation. Whenever Assumption 1 is satisfied, the individual s_i and the 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's best response is also to subvert.

Overcoming free-rider problems. A large literature in political economy discusses how free-rider problems can be mitigated in practice. Assumption 1 should be understood as a reduced form for these mechanisms. For example, Lohmann (1993, 1994a) considers a model where individuals participate in individually costly political action out of the desire to signal private information about a common fundamental.[11] In her model, individuals are heterogeneous with respect to their preferences over aggregate outcomes and thus, despite the fact that any individual is small relative to the population, some individuals — those with "moderate" preferences — have a disproportionate impact on the beliefs of others and so find it worthwhile to pay the individual cost of political action. Other theoretical approaches to the free-rider problem include Karklins and Petersen (1993) who consider a sequence of stag-hunt coordination games that capture the gradual building of a coalition against the regime. Fearon (2006) considers reputation-formation in a repeated game between a large number of citizens and a regime. In public choice theory, the literature on club goods as applied to social and political movements emphasizes the use of partial excludability to overcome free-rider problems, as in Tullock (1971, 1974), or for a recent application Berman and Laitin (2008). Another form of partial excludability is the threat of *reprisal* against individuals who collaborate with an overthrown regime.[12] Finally, from an empirical point of view, the evidence suggests that in practice it is hard for individuals to free ride on an insurgency against a regime (Kalyvas, 2007) and there is abundant historical evidence on the costs of collaboration, see Jackson (2001) and Frommer (2005) for instance.

Media outlets. Similar to the media bias model of Mullainathan and Shleifer (2005), where media outlets report an unbiased estimate of the truth plus some *slant*, here media outlets report the true θ plus the attempted manipulation of the regime $a = (1-r)\hat{a}$. In Mullainathan and Shleifer however, media outlets only add slant in equilibrium if citizens have an exogenous *preference* for

[11] Of these, Lohmann (1993) is most closely related to this paper. In that model, there is a large agent that takes a political action in response to the collective decisions of many small voters, but in her setting the large agent has preferences that align with the median voter whereas the large agent in this model, the regime, is diametrically opposed to the preferences of the citizens.

[12] In a binary action game like the one in this paper, individual citizens who do not subvert implicitly collaborate with the regime in that they make it harder to raise a mass of subversives S large enough to force regime change.

biased information.[13] In my model, information is biased in equilibrium without citizens having any preference for bias. Even though media outlets have some tolerance for manipulation, given by $1-r$, ultimately it is the underlying coordination game and incomplete information about the regime's type θ that allows bias in equilibrium, not the media's r. The regime is the ultimate source of any bias with the media outlets an essentially passive channel by which the regime's action is passed along. The *reliability* of information does change the regime's *effective costs* of manipulation but does not affect the citizens' ability to discard any bias that has been introduced.

Hidden actions and media influence. The hidden action $a = (1-r)\hat{a}$ represents the collective impact of all the regime's behind-the-scenes efforts at spinning, lobbying, bullying, and blackmailing media owners, editors and journalists. It is common knowledge that this influence *occurs*, but it is not possible for individual citizens to observe it directly and instead its *extent* must be inferred. The regime's cost function accounts for all the direct and indirect costs of exerting this behind-the-scenes influence over the media. To simplify the analysis and to focus on the effects of the regime's attempted manipulation in determining equilibrium outcomes, I treat r as a parameter. Besley and Prat (2006) and Gehlbach and Sonin (2008) provide models of the equilibrium extent of government influence over the media, but they do not share this paper's focus on coordination problems or regime change.

Public or private signals. An important issue in coordination games with incomplete information is the extent to which signals are *public* or *private*. A sufficiently precise public signal is often a source of multiple equilibria. More generally, changes in public information often have a disproportionate impact on equilibrium outcomes. In my setting, the term "public signal" is unfortunate in that it calls to mind *public media* and it is natural to think of the regime's manipulation operating through this channel (e.g., through broadcast television). In my model, the regime's action enters the citizens' individual signals $x_i = \theta + (1-r)\hat{a} + \varepsilon_i$ and I interpret the media, collectively, as determining the common or *systematic* component of the signal, $\theta + (1-r)\hat{a}$, while the *idiosyncratic* component, ε_i, captures all the cross-sectional variation in beliefs created by an individual citizen's haphazard media consumption. An individual's signal x_i does not reflect a private channel of communication from the regime to i but instead reflects individual i's idiosyncratic observation of a common channel of communication, namely the N media outlets.

[13] In Mullainathan and Shleifer (2005), if individuals have heterogeneous preferences — say some preferring slant one way, some the other — then competitive media outlets differentiate and the market for information is segmented in a manner that serves to align individuals' preference for biased information with the reports they actually receive. By contrast, in Gentzkow and Shapiro (2006) market competition serves to reduce the amount of bias.

2.3 Exogenous information benchmarks

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

Common knowledge. If θ is common knowledge, costly hidden actions are pointless and $a(\theta) = 0$ for all θ. The model reduces to a standard coordination game. If $\theta < 0$, any crowd $S \geq 0$ can overthrow the regime. It is optimal for any individual to riot, all do so, and the regime is overthrown. If $\theta \geq 1$, no crowd can overthrow the regime. It is optimal for any individual not to riot, 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 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.

Standard global game. If there are no hidden actions, $a(\theta) = 0$ for all θ, then each citizen has signal $x_i = \theta + \varepsilon_i$ and the analysis reduces to a standard global game. 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 much subsequent literature, this introduces the possibility of pinning down a unique equilibrium outcome.[14] 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 $x_i < x^*$. These thresholds are characterized by:

MORRIS-SHIN BENCHMARK. The unique equilibrium thresholds $x^*_{\text{MS}}, \theta^*_{\text{MS}}$ simultaneously solve

$$\Phi(\sqrt{\alpha}(\theta^*_{\text{MS}} - x^*_{\text{MS}})) = p \tag{10}$$

$$\Phi(\sqrt{\alpha}(x^*_{\text{MS}} - \theta^*_{\text{MS}})) = \theta^*_{\text{MS}} \tag{11}$$

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

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

As the signal precision α becomes high, some regimes are faced with a powerful incentive to shift the signal mean in their favor. Specifically, the equilibrium mass of subversives is

$$S^*_{\text{MS}}(\theta) := \Phi(\sqrt{\alpha}(x^*_{\text{MS}} - \theta)) = \Phi(\sqrt{\alpha}(1 - p - \theta) - \Phi^{-1}(p)) \tag{12}$$

[14] This result depends on a relatively diffuse common prior (Hellwig, 2002; Morris and Shin, 2000, 2003).

and as the precision $\alpha \to \infty$, the mass $S^*_{\text{MS}}(\theta) \to \mathbb{1}\{1-p > \theta\}$ where $\mathbb{1}\{\cdot\}$ denotes the indicator function. In this case, the equilibrium mass of subversives is a *step function* around the Morris-Shin benchmark. If the regime's type is $\theta < \theta^*_{\text{MS}} = 1-p$, it faces a unit mass of subversives and is overthrown. If the regime has $\theta > \theta^*_{\text{MS}}$ it faces zero subversives and survives. A *small* increase in the signal mean would suffice to enable a regime with θ just below θ^*_{MS} to achieve a *large* reduction in the size of the attack and to switch from being overthrown to surviving. In short, as information becomes precise, there is a large incentive for marginal regimes to shift the signal mean.

3 Unique equilibrium with information manipulation

When information depends on the the regime's manipulation, a citizen's signal x_i is informative for both the regime's θ and its hidden action. The hidden action is itself informative about θ and citizens take this into account when forming their beliefs. In equilibrium, the regime's action and the beliefs of citizens need to be mutually consistent. The first main result of this paper is that there is a unique equilibrium:

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

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

3.1 Equilibrium characterization

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

Regime problem. Since individual citizens subvert $s(x_i) = 1$ for $x_i < \hat{x}$, for any given $\hat{x}$ the aggregate mass of subversives facing the regime is

$$\int_{-\infty}^{\hat{x}} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a))\, dx_i = \Phi(\sqrt{\alpha}(\hat{x} - \theta - a)) \tag{13}$$

(using the expression for the signal density given in (5) above). And since the regime is overthrown for $\theta < \Theta(\hat{x})$, hidden actions are $a(\theta, \hat{x}) = 0$ for all $\theta < \Theta(\hat{x})$, otherwise the regime would be incurring a cost but receiving no benefit. For all $\theta \geq \Theta(\hat{x})$, the regime chooses hidden actions

$$a(\theta, \hat{x}) \in \underset{a \geq 0}{\text{argmin}} \left[\Phi(\sqrt{\alpha}(\hat{x} - \theta - a)) + C(a)\right] \tag{14}$$

A key step in proving equilibrium uniqueness is to recognize that hidden actions are given by $a(\theta, \hat{x}) = A(\theta - \hat{x})$, where the auxiliary function $A : \mathbb{R} \to \mathbb{R}_+$ is *exogenous* and in particular does *not* depend on the citizen threshold $\hat{x}$. Using (13), this function is defined by

$$A(t) := \underset{a \geq 0}{\operatorname{argmin}} \left[\Phi(\sqrt{\alpha}(-t - a)) + C(a) \right] \tag{15}$$

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

$$\Theta(\hat{x}) = \Phi[\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x}))] + C[A(\Theta(\hat{x}) - \hat{x})] \tag{16}$$

This condition requires that total costs equal total benefits at the extensive margin. For any given candidate citizen threshold $\hat{x}$, equations (15)-(16) determine the regime threshold $\Theta(\hat{x})$ and hidden actions $a(\theta, \hat{x}) = A(\theta - \hat{x})$ solving the regime's problem.

Citizen problem. Now given a candidate citizen threshold $\hat{x}$ and the solution to the regime's problem, an individual citizen with arbitrary signal x_i will subvert the regime if and only if $\Pr[\theta < \Theta(\hat{x}) \mid x_i,\, a(\cdot, \hat{x})] \geq p$, where p is the effective opportunity cost of subversion (as given in equation (9) above) and where the posterior probability assigned to the regime being overthrown is

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

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

$$\frac{\int_{-\infty}^{\Theta(\hat{x})} \sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \theta))\, d\theta}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(\hat{x} - \theta - A(\theta - \hat{x})))\, d\theta} = p \tag{17}$$

Monotone equilibrium. A monotone equilibrium is given by a pair of thresholds simultaneously solving the indifference conditions (16) and (17). As shown in Appendix A, there is a unique monotone equilibrium with thresholds denoted x^* and θ^*. The regime's equilibrium hidden actions are given by $a(\theta) = A(\theta - x^*)$ using the auxiliary function from (15). A key step in the proof is showing that the posterior probability on the left hand side of equation (17) depends only on the difference $\Theta(\hat{x}) - \hat{x}$ and is monotone increasing in this argument so that (17) can be solved for a unique difference $\theta^* - x^*$. Similarly, the right hand side of the regime indifference condition (16) only depends on the difference $\Theta(\hat{x}) - \hat{x}$ so we can take the unique solution $\theta^* - x^*$ from (17) and plug it into the right hand side of (16) to determine θ^* separately.

The appendix goes on to show that this unique monotone equilibrium is all that remains after

the iterative elimination of interim strictly dominated strategies. Thus, this monotone equilibrium is the only equilibrium.

3.2 Further discussion of equilibrium uniqueness

The uniqueness result in Proposition 1 contrasts with Angeletos, Hellwig, and Pavan (2006), who were the first to emphasize *endogenous* information in a global game. In their benchmark model, individuals get one noisy observation of θ plus one observation of a signal a chosen at cost $C(a)$ by the regime which may also be informative for θ. 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 the version of their model closest to this paper, Angeletos, Hellwig, and Pavan let individuals receive *two* noisy signals, (i) a signal of the regime's endogenous action $a(\theta)$, and (ii) a signal of θ itself. Multiple equilibria arise even in this scenario.[15] By contrast, in my model individuals get *one* noisy observation of one object, the sum $\theta + a(\theta)$, instead of separate signals for the two constituent parts. Since this is the only essential difference between the models, this suggests it is the restriction that citizens only see the sum that delivers equilibrium uniqueness in my setting.

3.3 Hidden actions

To characterize the regime's hidden actions, it is instructive to recast the regime's problem in terms of the *signal mean*. Let $V(y, \theta)$ denote the payoff to a regime of type θ if they choose signal mean $y = \theta + a$, that is

$$V(y, \theta) := \theta - \Phi(\sqrt{\alpha}(x^* - y)) - C(y - \theta), \qquad \theta \geq \theta^* \tag{18}$$

with $V(y, \theta) = 0$ for all $\theta < \theta^*$. The payoff $V(y, \theta)$ is *supermodular* in y and θ, in particular[16]

$$\frac{\partial^2}{\partial y \partial \theta} V(y, \theta) = C''(y - \theta) \geq 0 \tag{19}$$

Consequently, the signal mean $y(\theta) = \theta + a(\theta)$ is increasing in the regime's type θ, strictly increasing if the cost function is strictly convex. More specifically, for $\theta < \theta^*$ hidden actions are zero and the signal mean is just $y(\theta) = \theta$. At the threshold θ^* the mean jumps discretely to $y(\theta^*) = \theta^* + a(\theta^*)$. The signal mean $y(\theta)$ is thereafter increasing in θ, strictly if the cost function is strictly convex.

[15] The action $a(\theta)$ also has a payoff relevant effect in Angeletos, Hellwig, and Pavan (2006) but this is not essential.

[16] The objective function has a kink at the threshold, so at θ^* the expression (19) should be interpreted as the right-hand derivative (the left-hand derivative is zero at that point).

The hidden actions themselves are characterized by the first order necessary condition[17]

$$\frac{\partial}{\partial y}V(y,\theta)=0 \qquad \Leftrightarrow \qquad \sqrt{\alpha}\phi(\sqrt{\alpha}(x^*-\theta-a))=C'(a), \qquad \theta\geq\theta^* \tag{20}$$

The marginal benefit of an action is the associated reduction in the mass of subversives and at an interior solution this is equated to $C'(a)$.

4 Equilibrium information manipulation

The most interesting implication of this model is that the regime's information manipulation — or *signal-jamming* — is more effective if the quantity of information, as measured by the effective signal precision $\alpha = N\hat{\alpha}$, is sufficiently high. Section 4.1 gives an extended example of this effect using a specific case that is particularly easy to handle, namely the case where the regime's cost function is linear. Section 4.2 presents results for strictly convex cost functions. Section 4.3 provides further intuition and compares these results with other signal-jamming models in the literature. Finally, Section 4.4 shows that an increase in the *reliability* of information is in tension with an increase in the quantity of information in the sense that they always have opposite effects on the regime's chances of surviving.

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

4.1 Signal-jamming with linear costs

In the special case of a linear cost function, $C(a):=ca$ for some constant marginal cost c, the regime's hidden actions can be calculated explicitly.

Corner solutions. In this special case, the regime may be at a corner solution. In particular, the marginal benefit of an action a is the reduction in the mass of subversives, i.e., $\sqrt{\alpha}\phi(\sqrt{\alpha}(x^*-\theta-a))$. This marginal benefit is bounded above by $\sqrt{\alpha}\phi(0)$ where $\phi(0)=1/\sqrt{2\pi}\approx 0.399$ is the maximum

[17] The first order condition (20) may have zero, one or two solutions. In the event of two solutions, only the higher solution satisfies the second order condition.

value of the standard normal density. Consequently, if the signal precision is too low, namely

$$\alpha \leq \underline{\alpha} := \left(\frac{c}{\phi(0)}\right)^2 \tag{21}$$

then the marginal benefit from manipulation is too low to justify the cost and the regime is at a corner solution with $a(\theta) = 0$. Otherwise, if $\alpha > \underline{\alpha}$ then the regime may be at an interior solution.

Interior solutions. Manipulating the first order condition (20) shows that interior solutions to the regime's problem are given by

$$a(\theta) = \theta^{**} - \theta, \qquad \theta \in [\theta^*, \theta^{**}) \tag{22}$$

where $\theta^{**} := x^* + \gamma$ and where $\gamma > 0$ is defined by

$$\gamma := \sqrt{\frac{1}{\alpha} \log\left(\frac{\alpha}{\underline{\alpha}}\right)}, \qquad \alpha > \underline{\alpha} \tag{23}$$

In this case, the signal-jamming is *acute*. All regimes that manipulate information pool on the same distribution of signals. Since the signal mean is $y(\theta) = \theta + a(\theta)$, all regimes that manipulate, i.e., all $\theta \in [\theta^*, \theta^{**})$, generate a mean of θ^{**}. As shown in Figure 2, these regimes *mimic* the signal mean of a type θ^{**} that is intrinsically more difficult to overthrow (than they are) and generate signals for the citizens $x_i = x^* + \gamma + \varepsilon_i$ that are *locally completely uninformative* about θ. As a consequence of this signal-jamming, the equilibrium precision of a citizen's information is generally less than its intrinsic "fundamental" precision α. Appendix C provides a more detailed discussion of the implications of the signal-jamming for equilibrium beliefs.

Solving the indifference conditions. To complete the solution of the model, write the indifference condition of the marginal citizen (17) in terms of the equilibrium thresholds x^*, θ^* and hidden actions $a(\theta)$, as follows

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p} \int_{\theta^*}^{\infty} \sqrt{\alpha}\phi[\sqrt{\alpha}(x^* - \theta - a(\theta))]\, d\theta \tag{24}$$

Now use the first order condition (20) and $C'(a) = c$ to simplify the right hand side integral

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

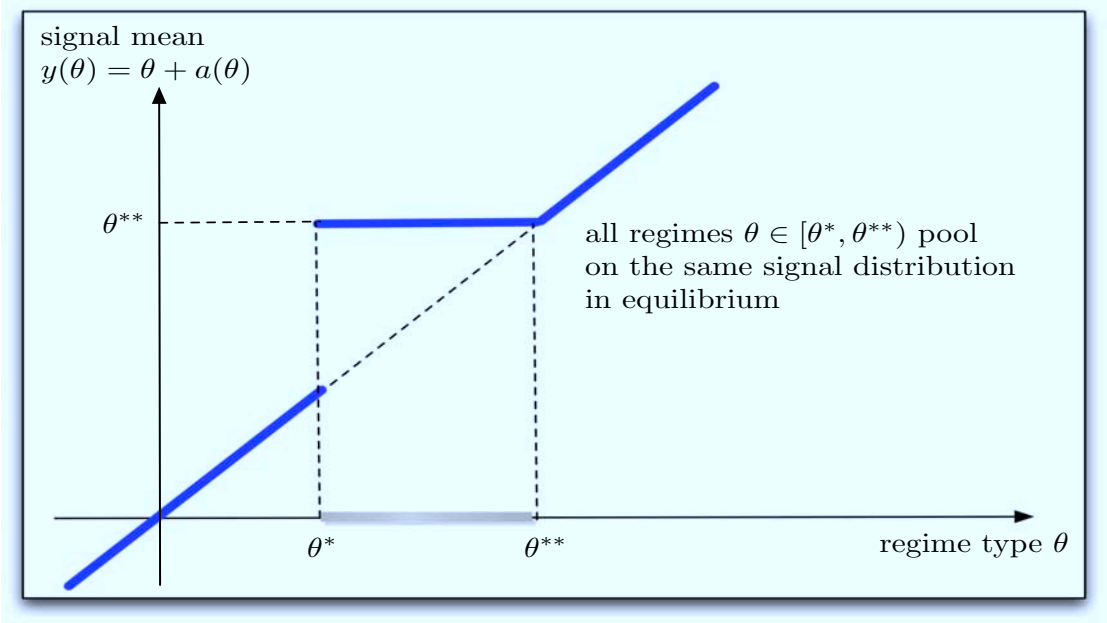


Figure 2: Signal-jamming with linear costs.

The equilibrium signal mean $y(\theta) = \theta + a(\theta)$ when the regime has linear costs of manipulation. All regimes with $\theta < \theta^*$ are overthrown. All regimes with $\theta \in [\theta^*, \theta^{**})$ generate the same signal distribution in equilibrium. They mimic a higher type of regime θ^{**} that will not be overthrown and generate signals for the citizens that are (locally) uninformative about θ.

where the first equality uses $a(\theta) = 0$ for $\theta \geq \theta^{**}$ and the second equality uses $\theta^{**} = x^* + \gamma$. Plugging this back into (24) gives us the first of two equations characterizing the two thresholds

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p}\left[(x^* - \theta^* + \gamma)c + \Phi(-\sqrt{\alpha}\gamma)\right] \tag{25}$$

As a function of the difference $\theta^* - x^*$, the left hand side is a continuous, strictly increasing one-to-one map from $\mathbb{R}$ to $[0, 1]$. Similarly, as a function of $\theta^* - x^*$ the right hand side is a continuous, strictly decreasing one-to-one map from $\mathbb{R}$ to $\mathbb{R}$ so, by the intermediate value theorem, there is a unique difference that solves this equation. As shown in Appendix A, the fact that the marginal citizen's indifference condition uniquely determines the threshold difference $\theta^* - x^*$ is true more generally and is not particular to the case of linear costs.

The regime threshold θ^* is then determined using the indifference condition (16) which, with linear costs, can be written

$$\theta^* = (x^* - \theta^* + \gamma)c + \Phi(-\sqrt{\alpha}\gamma) \tag{26}$$

where the difference $\theta^* - x^*$ on the right hand side is implicitly determined by (25) above.

Use of manipulation vs. effectiveness of manipulation. As in classic signaling games, the regime is able to send a (noisy) signal in equilibrium and this enables some weaker regime types to pool with stronger regime types. Whether this pooling behavior is *effective* in equilibrium is another matter. In principle, it might be true that the only regime types that are able to imitate

stronger regime types are those regimes that would have survived even in the absence of the signal-jamming technology. Moreover, it might also be true that some weak regimes that would survive if they could commit not to use information manipulation are overthrown because they cannot make that commitment. In short, regimes may manipulate information in equilibrium but it does not follow that manipulation is necessarily effective in increasing the likelihood of the regime surviving. It turns out that manipulation *is* effective when the signal precision α is high.

Effective manipulation when signal precision is high. In the special case of linear costs, it is possible to give a complete characterization of the effects of increases in α. The sensitivity of the regime threshold θ^* to the signal precision α is characterized by:

PROPOSITION 2. For each c there is a $\underline{\alpha}(c)$ such that for all $\alpha \leq \underline{\alpha}(c)$ all regimes are at a corner solution with $a(\theta) = 0$ for all θ and $\theta^* = \theta^*_{\text{MS}}$. Otherwise, for all $\alpha > \underline{\alpha}(c)$ regimes $\theta \in [\theta^*, \theta^{**})$ are at an interior solution and there is a critical precision $\alpha^*(c,p) \geq \underline{\alpha}(c)$ given by

$$\alpha^*(c,p) := \underline{\alpha}(c) \exp\left(\max\left[0, \Phi^{-1}(p)\right]^2\right) \tag{27}$$

such that

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \qquad \text{for all} \qquad \alpha > \alpha^*(c,p) \tag{28}$$

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

From equation (21), for $\alpha \leq \underline{\alpha}$ regimes are at a corner solution with $a(\theta) = 0$ for all θ and the regime threshold is $\theta^* = \theta^*_{\text{MS}} = 1 - p$. For $\alpha > \underline{\alpha}$ the regimes $\theta \in [\theta^*, \theta^{**})$ are at an interior solution with $a(\theta) > 0$ and the regime threshold θ^* is decreasing in α for all α greater than the critical precision α^*. In particular, if the opportunity cost of subverting is small, $p < 1/2$, then $\Phi^{-1}(p) < 0$ and so from (27) the critical precision is just $\alpha^* = \underline{\alpha}$ and the regime threshold is strictly decreasing in α for all $\alpha > \underline{\alpha}$. Hence, in this case, for all $\alpha > \underline{\alpha}$ the regime threshold is lower than the Morris-Shin benchmark $1 - p$. This is shown in the left panel of Figure 3. Alternatively, if the opportunity cost of subverting is large, $p > 1/2$, then $\Phi^{-1}(p) > 0$, the critical precision is $\alpha^* > \underline{\alpha}$, and the regime threshold is non-monotone in α. This is shown in the right panel of Figure 3. In this case, the regime threshold reaches a maximum at α^* and strictly decreases thereafter. Again, for high enough α it is the case that the regime threshold θ^* is lower than the Morris-Shin benchmark.

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

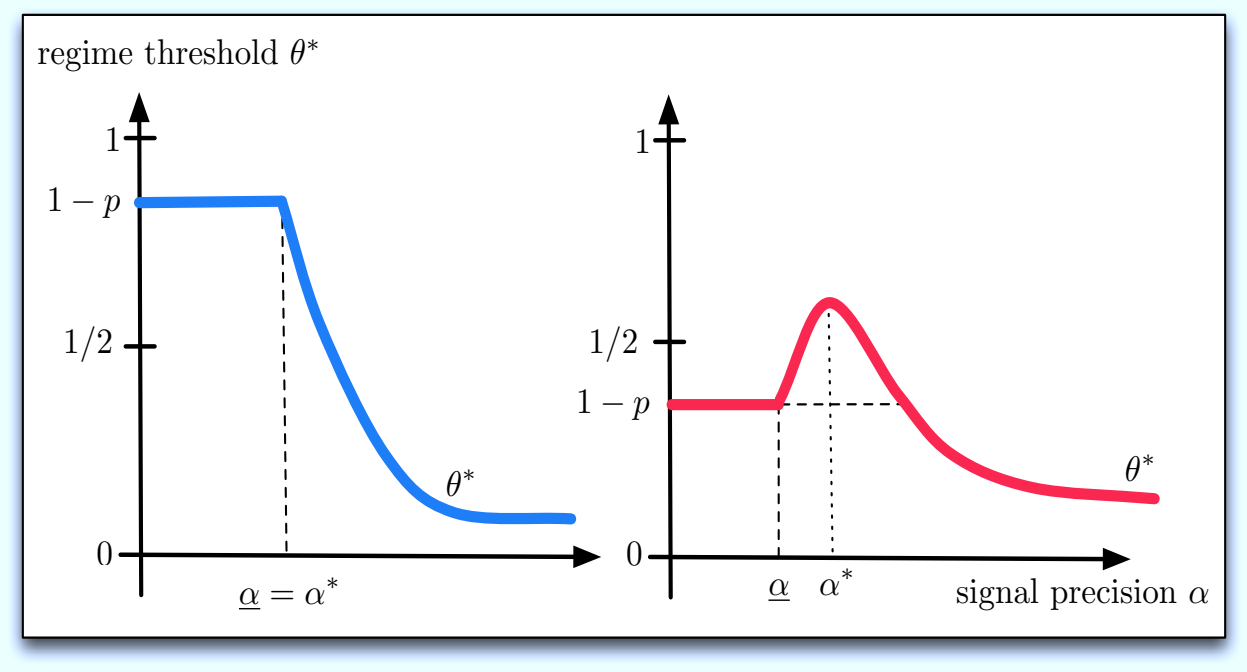


Figure 3: Information manipulation is effective when signal precision α is sufficiently high.

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

can be manipulated and signals are sufficiently precise, then even the very most fragile regimes can survive.

4.2 Strictly convex costs

The effectiveness of information manipulation when α is sufficiently high is true not just in the special case of linear costs. For any convex cost function $C(a)$ it can be shown that as the signal precision $\alpha \to \infty$ the limiting equilibrium thresholds and hidden action profile are

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

So for high enough α, the regime is able to reduce the threshold θ^* below the Morris-Shin benchmark. If in addition costs are *strictly* convex, $C''(a) > 0$ for all a, then as $\alpha \to 0^+$ the limiting equilibrium thresholds and hidden action profile are

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

In short, if costs are strictly convex then for low enough α, information manipulation *necessarily* backfires on the regime in the sense that $\theta^* > \theta^*_{\text{MS}} = 1 - p$. Regimes would want to be able to credibly commit to refrain from all media manipulation.

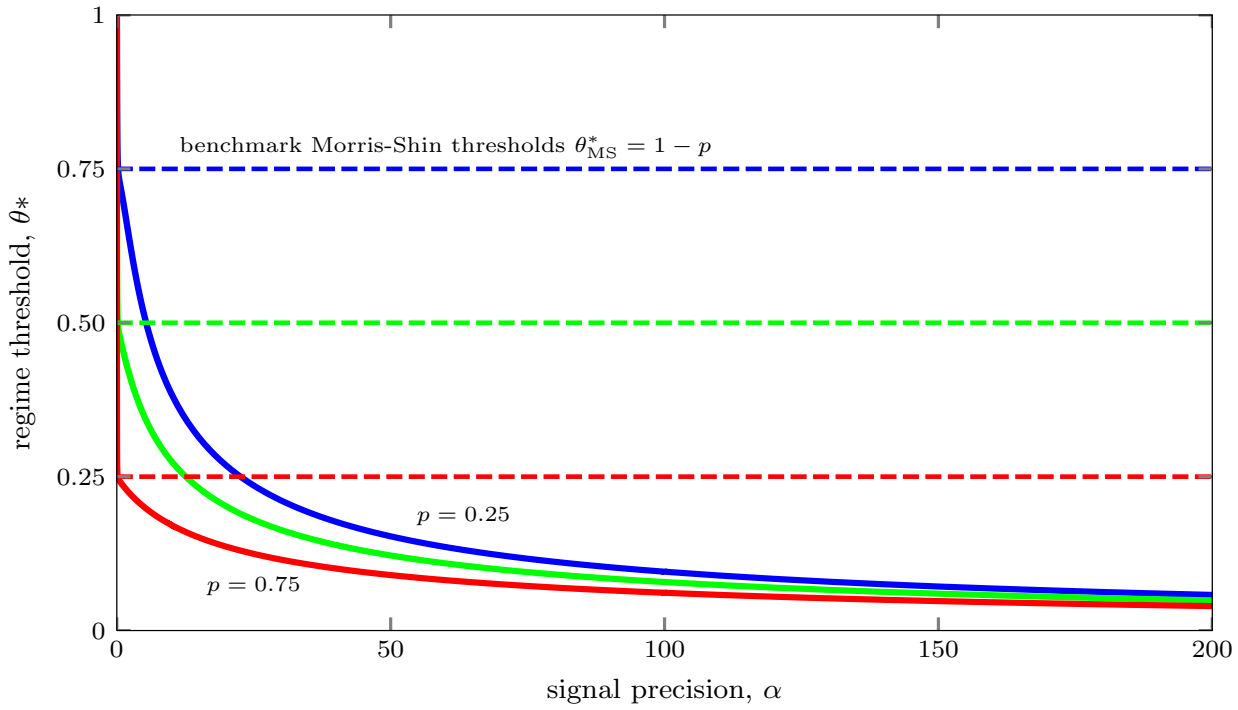


Figure 4: Information manipulation with strictly convex costs.

The Morris-Shin benchmark thresholds are $\theta^*_{\text{MS}} = 1 - p$ for all α. With information manipulation, the regime threshold θ^* is below $1 - p$ when α is sufficiently high. In this sense, more information increases a regime's chances of survival. These examples use strictly convex costs $C(a) = a^2/2$. The thresholds $\theta^* \to 0$ as $\alpha \to \infty$ so that in the limit all *fragile* regimes $\theta \in [0, 1)$ survive. Note that when α is sufficiently low information manipulation *backfires* in the sense that $\theta^* > 1 - p$ and moreover $\theta^* \to 1$ as $\alpha \to 0$.

Numerical examples. With general cost functions the model cannot be solved analytically. Figure 4 shows θ^* as a function of precision α under the assumption that $C(a) = a^2/2$ 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.[18]

4.3 Intuition for signal-jamming results

The result that information manipulation is effective when the signal precision α is sufficiently high depends on two crucial effects: (i) the increased density of signals near the signal mean when α is high, and (ii) the collective inability of the citizens to neutralize the increase in the signal mean via an increase in the signal threshold x^*. To see these two effects, consider first the Morris-Shin benchmark where, as a function of the regime type θ, the mass of subversives is

$$\begin{aligned} S^*_{\text{MS}}(\theta) &= \Phi(\sqrt{\alpha}(x^*_{\text{MS}} - \theta)) \\ &= \Phi(\sqrt{\alpha}(1 - p - \theta) - \Phi^{-1}(p)) \quad \longrightarrow \quad \mathbb{1}\{1 - p > \theta\} \qquad \text{as } \alpha \to \infty \end{aligned}$$

[18] If costs are linear $C(a) = ca$, then as discussed above, for any signal precision α below $\underline{\alpha}(c) := (c/\phi(0))^2$, the regime is at a corner solution such that the equilibrium threshold is the Morris-Shin benchmark $1 - p$. By contrast, with strictly convex costs the regime threshold can be driven all the way to 1 as α falls to zero.

In the Morris-Shin benchmark, as signals become precise the mass of subversives is a step function around $1-p$. The mass of subversives facing the marginal regime θ^*_{MS} is exactly $1-p$ for all α. Relative to this benchmark, a small exogenous "shock" that increased the signal mean from θ to $\theta+\tilde{a}$, say, would cause a shift of the step function so that the mass of subversives facing the marginal regime would fall from $1-p$ to 0 and more regimes would be able to survive. But this only delivers a reduction in the mass of subversives if the $\tilde{a}$ is *unanticipated*. If the citizens could correctly anticipate such an increase in the signal mean, then they would discount their signals accordingly and the signal threshold would rise to $x^*_{\text{MS}}+\tilde{a}$. If so, there would be no increase in the mass of subversives

$$\Phi(\sqrt{\alpha}(x^*_{\text{MS}}+\tilde{a}-(\theta+\tilde{a}))=\Phi(\sqrt{\alpha}(x^*_{\text{MS}}-\theta))=S^*_{\text{MS}}(\theta)$$

Thus there would be no effect on the regime's ex ante chances of surviving. For the regime to benefit from perturbing the signal mean when the density of signals near the mean is high, it must also be the case that the citizens are, collectively, somehow inhibited in their ability to correctly discount their signals.

Consider now the model with information manipulation. Some regimes are indeed able to achieve an *endogenous* shift in the signal mean that reduces the mass of subversives they face. All regimes with $\theta \geq \theta^*$ mimic the signal mean of a type $y(\theta)=\theta+a(\theta)\geq\theta$ that is harder to overthrow. And when signals are precise, there is a large density of citizens near this artificially high mean. But for the regime to benefit from this it must also be the case that the signal threshold x^* does not increase to neutralize the increase in the signal mean. In this model citizens know the regime's incentives, so why can they not correctly discount their signals?

Two features of the model account for the inability of citizens to correctly discount their signals on account of the regime's manipulation. First, it is important that different regime types θ generally take different actions $a(\theta)$ so citizens have imperfect information about the amount by which they should discount their signals. If all regimes took the same action, $\tilde{a}$, it would be easy to undo. Second, it is important that individual citizens be imperfectly coordinated.

Different regimes take different actions. To see why it is important that different regimes generally take different actions, consider by contrast a *career concerns* model, as in Holmström (1999) and Dewatripont, Jewitt, and Tirole (1999). A standard career concerns model has the same additive-normal setup for information manipulation $\theta+a+\varepsilon$. But in such models, a worker's costly effort to manipulate information *always* backfires in equilibrium: the firm receiving the signal is able to infer the worker's action and thereby correctly decompose their signal into its underlying components. Moreover, because the action is costly the worker is necessarily worse-off than they would be if no manipulation was possible. This happens because in a standard career concerns model the signal-jammer, the worker, is uninformed about its talent. The worker chooses

an action to maximize expected utility with expectations taken over the joint distribution of the worker's talent and the signals observed by the firm. So no matter what their underlying talent, all workers choose the same action. In equilibrium this means the firm can perfectly decompose their signal into the component due to talent and the component due to effort.

By contrast, in my model, the signal-jamming regime is informed about its type. The regime knows its θ and takes an action $a(\theta)$ contingent on it. While in equilibrium citizens know the *function* $a(\cdot)$, they do not know θ and so do not know the specific amount $a(\theta)$ by which they should discount their signals. In addition, because they each have different signals x_i, the citizens generally differ in their beliefs about the true θ and therefore differ in their beliefs about the hidden action that has been chosen. And finally, these citizens with different beliefs are playing a coordination game amongst themselves.

Imperfect coordination among signal receivers. In traditional models of strategic information transmission such as Crawford and Sobel (1982) and Holmström (1999) there is one sender and one signal receiver. But in this paper, and in the similar model of Angeletos, Hellwig, and Pavan (2006), there is instead a large cross-section of imperfectly coordinated receivers. This gives rise to effects absent from the traditional setup. Intuitively, since the incentives of the regime to manipulate information depend on the aggregate response of the citizens and since the aggregate response and an individual's decision are strategic complements, implicitly each citizen's information filtering problem depends simultaneously on the information filtering problems of all the other citizens. To see the role of imperfect coordination more formally, suppose to the contrary that citizens were perfectly coordinated and able to act as a single large agent who could force regime change for all $\theta < 1$. This agent receives one signal $x = \theta + a + \varepsilon$ with precision $\alpha \to \infty$. For simplicity, suppose also that costs are strictly convex. Then the single agent knows that $x \to y(\theta) = \theta + a(\theta)$ and $y(\theta)$ is strictly increasing and can be inverted to recover $\theta = y^{-1}(x)$. But knowing θ, the single agent can deduce any manipulation $a(\theta)$ and discard it so that the regime in fact has no incentive to undertake the costly manipulation. Therefore if citizens are perfectly coordinated, they know $x \to \theta$ and attack if and only if $x = \theta < 1$. In this case, all the fragile regimes with $\theta \in [0, 1)$ are wiped out. By contrast, if citizens are imperfectly coordinated all fragile regimes with $\theta \in [0, 1)$ survive (see Appendix B for more details).

4.4 Media reliability and the costs of manipulation

The regime chooses an effective action $a := (1-r)\hat{a}$ where the weight $r \in [0, 1)$ governs the media's willingness to accommodate the regime's preferred message. For any fixed signal precision α, a citizen's information is *more reliable* when $r = 0$ so that the media reports the true θ, and *less reliable* when $r = 1$ so that the media reports the regime's preferred message $\theta + \hat{a}$. If the cost of $\hat{a}$ is $C(\hat{a})$, the cost of the effective action a is $C(a/(1-r))$.

The reliability and quantity effects are always in tension. For simplicity, consider the case of linear costs $C(\hat{a}) = \hat{a}$ with constant marginal cost normalized to one. Then in terms of the effective action $a := (1-r)\hat{a}$ the cost function is ca, where the effective marginal cost is $c := 1/(1-r)$. Thus the effective marginal cost is in direct relationship to the reliability parameter r with $c = 1$ corresponding to least reliable and $c = \infty$ corresponding to most reliable. Qualitatively, the effects of an increase in r are the same as the effects of an increase in c and are given by:

PROPOSITION 3. The effect of a change in information reliability always has the opposite sign to the effect of a change in information quantity:

$$\frac{\partial \theta^*}{\partial c} = -\frac{2\alpha}{c}\frac{\partial \theta^*}{\partial \alpha} \tag{29}$$

If $\alpha \leq \underline{\alpha}(c)$, all regimes are at a corner solution with $a(\theta) = 0$ for all θ and $\theta^* = 1-p$ so that both effects are zero. Otherwise, for $\alpha > \underline{\alpha}(c)$, regimes $\theta \in [\theta^*, \theta^{**})$ are at an interior solution and the effects are determined by whether α is larger or smaller than the critical precision $\alpha^*(c,p)$.

In particular, for high enough α we have $\alpha > \alpha^*(c,p)$ and an increase in media reliability c would increase the regime's threshold θ^* and reduce the regime's ex ante chances of surviving while at the same time an increase in the quantity of information α would reduce the threshold and so increase its chances of surviving. In this region, $\alpha > \alpha^*(c,p)$, the *level* of the regime threshold is $\theta^* < \theta^*_{\text{MS}} = 1-p$ and the regime is benefitting from the ability to manipulate information — it is just that a marginal *increase* in media reliability would make the regime somewhat worse off.

Proposition 3 tells us that the effects of increases in information reliability and increases in information quantity work against each other. Since we would generally expect that dramatic changes in information technologies will have consequences for both these characteristics, this suggests that we will have to consider the *joint* effects of changes in information quantity and reliability if we are to make use of the model in interpreting the relationship between actual autocratic regimes and information technologies.

5 Implications of the model

5.1 Empirical predictions of the model

It may seem that the opposing quantity and reliability effects of changes in information undermine the model's empirical discipline. If "better information" can increase *or* decrease the regime's chances of survival depending on which effect dominates, then isn't the model unfalsifiable? To allay these concerns, in this section I explain in more detail the model's key empirical predictions.

Relative magnitudes of the quantity and reliability effects. The offsetting effects of a change in the quantity of information and a change in the reliability of information are related by

Proposition 3. Multiplying both sides of equation (29) by $c/\theta^* > 0$ and rearranging we have

$$\frac{\partial \log \theta^*}{\partial \log c} = -2\frac{\partial \log \theta^*}{\partial \log \alpha} \tag{30}$$

Thus the model predicts that the information reliability effect is *exactly twice as large* as the information quantity effect (though of the opposite sign). A breakthrough in information technology that gives rise to roughly equal-sized percentage increases in information quantity and reliability will, overall, increase the regime threshold θ^* and reduce the probability of the regime surviving.

Non-monotonicities and interaction effects. In addition, the sign of the marginal effect of α on the probability of regime change is predicted to depend both on the level of α itself and on interactions with the other explanatory variables. More specifically, an increase in α is predicted to decrease the probability of regime change only if α is larger than the critical precision $\alpha^*(c, p)$ given by equation (27). This critical precision is in turn increasing in the reliability measure c and (weakly) increasing in the opportunity cost measure p. If p is high enough ($p > 1/2$) then the effects of α on the probability of regime change are non-monotone and the size of the interval $(\underline{\alpha}(c), \alpha^*(c, p))$ is increasing in p. That is, the model predicts that if the average individual consequences of taking action against the regime, as measured by p, is large, then it is also more likely that we will find non-monotone effects of α on the probability of regime change.

In short, there is a well-defined set of circumstances for which the marginal effect of a greater quantity of information changes sign. Although the model admits the possibility of either a negative or a positive effect of the quantity of information on the probability of regime change, the fact that the positive effect can only arise under these narrow circumstances is an important source of empirical discipline. It would be clear evidence against this model, for example, if the quantity of information had a positive effect on regime change even when measures of the opportunity cost p are low. Similarly, it would be clear evidence against the model if the quantity of information had a negative effect on regime change when α is low while having a positive effect when α is high.

5.2 Information technologies and autocratic regimes

Together Proposition 2 and Proposition 3 suggest two different kinds of *information revolutions*. One kind, perhaps best associated with modern decentralized technologies like the internet and social media, involves channels of information that are harder to manipulate by the regime for a given cost. The second kind, perhaps best associated with relatively centralized technologies like radio and cinema, involves technologies that can be more easily manipulated for given cost.

So, will improvements in information technology help in overthrowing autocratic regimes? From an optimistic perspective, accounts of the 1989 collapse of the Eastern European communist regimes often stress changes in information technology and the increasing inability of these regimes

to control the information that people could access (Kalathil and Boas, 2003). But historically, the relationship between new information technologies and autocratic regimes has not always seemed benign. The use of propaganda by totalitarian regimes such as Nazi Germany and the Soviet Union is well known, and while the propaganda of these regimes operated through many overlapping channels, the key role played by then state-of-the-art technologies such as radio and cinema is troubling (Arendt, 1973; Friedrich and Brzezinski, 1965). Radio and cinema made the propaganda machinery of these regimes extraordinarily effective (Zeman, 1973). Another example serves to reinforce this concern. While not a totalitarian regime, nineteenth century Ottoman Turkey was nonetheless brutally autocratic. Towards the end of the nineteenth century, new information technologies like the telegraph and mass newspapers led to a pronounced increase in the quantity of available information available. But rather than foster regime change, as an optimist might have supposed, instead this period witnessed a dramatic consolidation of the regime's power.

My model suggests that regimes may particularly benefit when an increase in the quantity of information occurs at the same time as a decrease in the effective cost of information manipulation c (or equivalently r). If a particular technological breakthrough that increases the quantity of information also makes it easier to disseminate information through more centralized channels of communication, then it also seems likely that the media outlets that provide information through these new channels may increasingly fall under the influence of the regime. Indeed, in my model, a regime will *want* to exert a strong influence over the media precisely when α is sufficiently high. Consistent with this, from the moment the Nazi Party came to power in 1933 it sought ever-increasing influence over the German media establishment. Control over radio and film was easiest to establish and nearly total. Indeed, far from being threatened by new technologies, the regime actively subsidized the diffusion of the cheap radio sets. By 1939, 70% of households owned a radio, the highest proportion in the world at the time (Zeman, 1973).[19]

One response to these rather pessimistic views is that, precisely because they involve centralized technologies such as radio and cinema that can be easily influenced by a regime, these examples are basically uninformative about the prospects of using modern decentralized technologies to help in overthrowing regimes. In this more optimistic view, modern technologies like the internet are diffuse, less easily influenced by a regime, and so are more threatening to them.

My model suggests that an information revolution that leads to a surge in the quantity of information α at the same time as an increase in the costs of manipulation c will generally have ambiguous implications for regime change. But if the percentage change in the quantity of information and the costs of manipulation are roughly the same, then, because the magnitude of the cost of manipulation effect is twice that of the quantity of information effect, overall the chances of regime change will improve. In this sense, the model gives some grounds for optimism.

[19]Radios were also essential for propaganda because of their ability to deliver a simultaneous, mass audience and because of the regimes complete monopoly over the airwaves. Moreover, channels of communication like radio that make use of a shared experience can induce relatively high degree of common knowledge amongst the public (Chwe, 2001), thereby facilitating the regime's efforts at inducing coordination on its preferred outcomes.

In practice, the evidence on the effects of modern information technologies and autocratic regimes is mixed. On the one hand, and as is now well-known, some regimes have had considerable success in countering the effects of the internet and related technologies (Chase and Mulvenon, 2002; Kalathil and Boas, 2003; Fallows, 2008; Morozov, 2011). In particular, Kalathil and Boas (2003), emphasize the Chinese regime's use of the exact same breakthroughs in information technology to counter online dissent.[20] Similar examples of a regime's efforts to counteract modern technologies come from the 2009 presidential elections in Iran and subsequent demonstrations. In the run-up to the election, the regime suspended access to social media websites like Facebook (Ribeiro, 2009). On election day, mobile phone communications were interrupted and foreign news providers such as the BBC experienced jamming designed to impede their broadcasts. During the subsequent demonstrations, services like Twitter were used by the regime to provide disinformation (Esfandiari, 2010; Cohen, 2009). On the other hand, the wave of uprisings against autocratic regimes in Tunisia, Egypt, Libya, Syria and other Arab countries beginning in December 2010 has given renewed support to the idea that modern decentralized technologies can play a significant role in facilitating coordination against regimes even in the face of concerted efforts to disrupt such technologies (as illustrated in Figure 1, above).

6 Extensions

To illustrate the robustness of these results, in this section I consider alternative information structures. Section 6.1 presents a setting where citizens have additional *clean* information that is not affected by the regime's manipulation. Section 6.2 compares the effects of information manipulation that directly affects an aggregate signal as opposed to individual signals. Section 6.3 allows for a *struggle* over information as a rival opposition group attempts to shift information in the opposite direction to the regime. Section 6.4 considers a model where the regime's actions directly affect signal precision rather than the mean.

6.1 Heterogeneous media outlets

Suppose media outlets come in two types, some that are potentially amenable to the regime's message and others who resolutely report the truth. Specifically, let citizens have N_x reports from media outlets that each give weight $1-r$ to the regime's action $\hat{a}$, and N_z reports from media outlets that give *zero* weight to the regime's action. Citizens observe each of these reports with

[20] See also Fallows (2008) for an account of the surprising effectiveness of China's national firewall and the system of monitoring and censorship that the firewall interacts with. Chase and Mulvenon (2002) particularly emphasize the use of traditional authoritarian methods — arrest, detention, and seizure — in cracking down on online dissent. Kalathil and Boas (2003) also discuss related efforts by autocratic regimes to counteract the internet in Cuba, Saudi Arabia, and elsewhere. Similarly, Soley (1987) discusses earlier efforts by the Cuban regime to counteract US-based satellite radio and television broadcasts.

idiosyncratic noise that is jointly IID normal across them and across all media outlets with mean zero and precision $\hat{\alpha}$. Let $a = (1-r)\hat{a}$ again denote the regime's effective action. Then the information of a citizen can be represented by *two* representative signals

$$x_i = \theta + a + \varepsilon_{x,i}, \qquad \text{and} \qquad z_i = \theta + \varepsilon_{z,i} \tag{31}$$

where the noise terms $\varepsilon_{x,i}$ and $\varepsilon_{z,i}$ are independent, jointly normally distributed, both with mean zero and precisions $\alpha_x := N_x\hat{\alpha}$ and $\alpha_z := N_z\hat{\alpha}$ respectively. This gives citizens a *clean* source of information z_i not affected by the regime's manipulation. This setup is also equivalent to giving citizens noisy signals of the hidden action a itself. Subtracting z_i from x_i gives

$$(x_i - z_i) = a + (\varepsilon_{x,i} - \varepsilon_{z,i})$$

This is an unbiased signal of the regime's action a.

I consider a monotone equilibrium where the regime is overthrown for $\theta < \theta^*$ and citizens subvert, $s(x_i, z_i) = 1$, if their signals satisfy $x_i < x^*(z_i)$. Here θ^* is a single threshold and $x^* : \mathbb{R} \to \mathbb{R}$ is a threshold *function*, both to be determined endogenously. In this case, the mass of subversives facing a regime that takes action a is

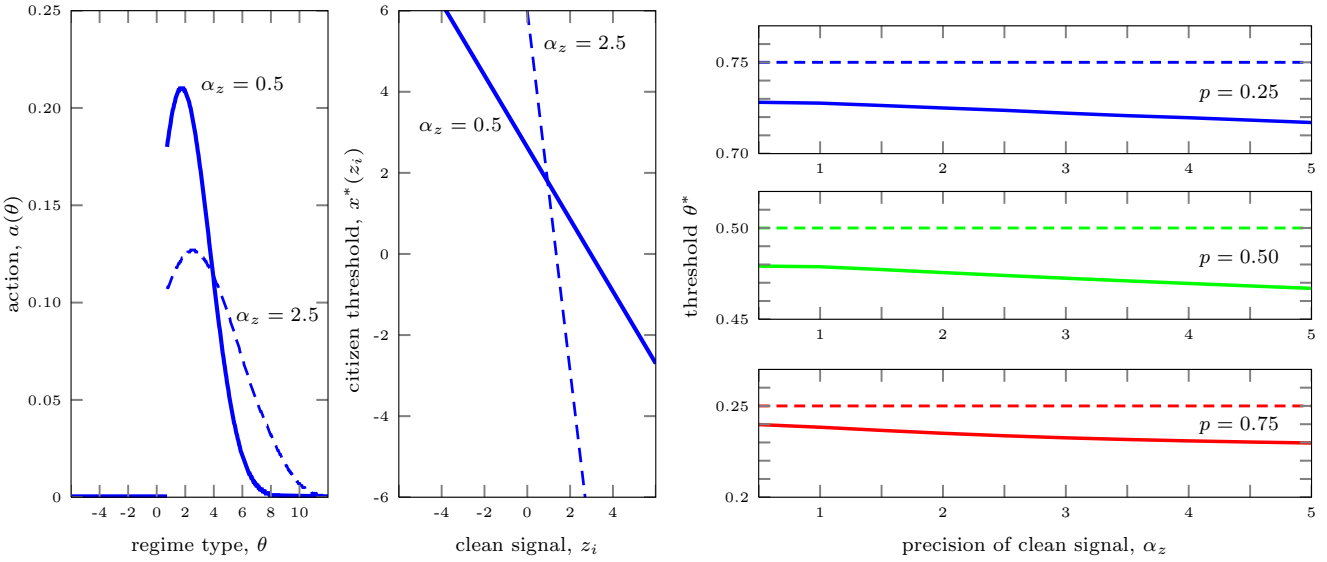
$$S(\theta, a) = \int_{-\infty}^{\infty} \Phi(\sqrt{\alpha_x}(x^*(z_i) - \theta - a))\sqrt{\alpha_z}\phi(\sqrt{\alpha_z}(z_i - \theta))\, dz_i$$

In general a citizen makes use of both types of information even though one is contaminated by a while the other is not. This is because, even considering the presence of manipulation, the x_i signals may still be more informative about θ than the z_i signals if the precision α_x is sufficiently high relative to α_z. Indeed, if $\alpha_z \to 0$, then we are back to the main model with only contaminated information since any uncontaminated information is too inherently noisy to be usable. Alternatively, as α_z increases, the z_i signals will be given more weight, and, as α_z becomes sufficiently large, the model reduces to the Morris-Shin benchmark where the only source of information is clean. For intermediate values of α_x and α_z, matters are more complex. And, unfortunately, it is not possible to give a simple analytic characterization of the equilibrium for general signal precisions. In Figure 5, I show several numerical examples.

The left panel shows the equilibrium hidden actions $a(\theta)$ and the citizen threshold function $x^*(z_i)$ for two cases, (i) with $\alpha_x = \alpha_z = .5$, so that the number of media outlets is the same for both kinds, and (ii) with $\alpha_x = .5$ but $\alpha_z = 2.5$, so that there are *five times* as many clean sources of information. In both cases the overall level of precision is relatively low, so even though the x_i signals are manipulated while the z_i signals are clean, citizens still draw on both kinds of information. The citizen threshold function $x^*(z_i)$ is decreasing in z_i because if a citizen gets a low z_i it takes a high x_i to induce subversion. And as α_z increases, the citizen threshold function

$x^*(z_i)$ becomes steeper so that the z_i are weighed more heavily and it takes an even bigger x_i to compensate for a low z_i. The right panel shows the regime threshold θ^* as a function of the clean precision α_z for various opportunity costs p and for fixed $\alpha_x = .5$ for the precision of the manipulated signal. In these examples, the θ^* are lower than the Morris-Shin benchmarks $1 - p$ and information manipulation is effective. Moreover, in this range the thresholds are decreasing in the precision α_z of the clean signal implying that, for these parameters, *even an increase in the quantity of clean information increases the regime's chances of surviving.*

These examples are only suggestive of what can happen in equilibrium. Still, it is clear that introducing clean information unaffected by the regime's manipulation does not by itself overturn the possibility that more information may increase the regime's chances of surviving.



(a) hidden actions $a(\theta)$ and citizen thresholds $x^*(z_i)$

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

Figure 5: Information manipulation still effective even though citizens have clean information.

Panel (a) shows that as the precision α_z of the clean signal information increases, regimes near θ^* take smaller actions $a(\theta)$ but θ^* hardly changes. Citizens give more weight to their clean signal so $x^*(z_i)$ is steeper, for low values of the clean signal z_i it takes a higher value of the manipulated signal x_i to induce subversion. In this example the opportunity cost of subversion is $p = .25$. Panel (b) shows the regime threshold θ^* as a function of the precision of the clean signal α_z. The regime still benefits from information manipulation in that $\theta^* < \theta^*_{\text{MS}} = 1 - p$. In all of these calculations, the manipulated signal has precision $\alpha_x = .5$ and the cost function is $C(a) = a^2/2$.

6.2 Manipulating aggregate information

To this point, information manipulation has entered through individual signals $x_i = \theta + a + \varepsilon_i$. But the competing roles of idiosyncratic and aggregate information is generally an important determinant of equilibrium outcomes in global games (Hellwig, 2002; Morris and Shin, 2003). In this section, I show that qualitatively similar results to those obtained for the main model can be obtained if information manipulation takes place through an aggregate signal. I contrast two setups, both with aggregate and idiosyncratic information but which differ in the channel by which manipulation enters citizens' information.

In the first setup, manipulation enters through the idiosyncratic information. That is, citizens have $x_i = \theta + a + \varepsilon_{x,i}$ as usual, but also have a common or *aggregate* signal $z = \theta + \varepsilon_z$ that is free from manipulation. Here $\varepsilon_{x,i}$ and ε_z are jointly normally distributed, both with mean zero and precisions α_x and α_z respectively. This provides an appropriate benchmark against which to judge the effects of manipulation that enters through aggregate information. The second setup has $x_i = \theta + \varepsilon_{x,i}$ but now the regime's manipulation enters the common signal $z = \theta + a + \varepsilon_z$.

Aggregate uncertainty. In both cases, I assume that the common signal z is realized after the regime chooses its action $a(\theta)$. Thus the regime faces *aggregate uncertainty* and can no longer perfectly anticipate play along the equilibrium path. I consider a monotone equilibrium where the regime is overthrown *ex post* for $\theta < \theta^*(z)$ and citizens subvert, $s(x_i, z) = 1$, if their signals satisfy $x_i < x^*(z)$. Here $\theta^* : \mathbb{R} \to [0, 1]$ and $x^* : \mathbb{R} \to \mathbb{R}$ are threshold *functions* to be determined.

Manipulation through individual signal. In this case, the *ex post* mass of subversives facing a regime that takes action a is

$$S(\theta, a, z) = \Phi(\sqrt{\alpha_x}(x^*(z) - \theta - a))$$

and the regime *ex ante* chooses $a(\theta)$ to maximize its expected payoff, namely

$$a(\theta) \in \operatorname*{argmax}_{a \geq 0} \left[-C(a) + \int_{-\infty}^{\infty} \max[0, \theta - S(\theta, a, z)] \sqrt{\alpha_z} \phi(\sqrt{\alpha_z}(z - \theta))\, dz \right]$$

The thresholds $\theta^*(z)$ and $x^*(z)$ are implicitly determined by indifference conditions for the regime and the citizens where the mass of subversives is $S(\theta, a, z)$ as above and where citizens' posterior densities are proportional to $\phi(\sqrt{\alpha_x}(x_i - \theta - a(\theta)))\phi(\sqrt{\alpha_z}(z - \theta))$.

Manipulation through aggregate signal. In this case, the *ex post* mass of subversives facing a regime is

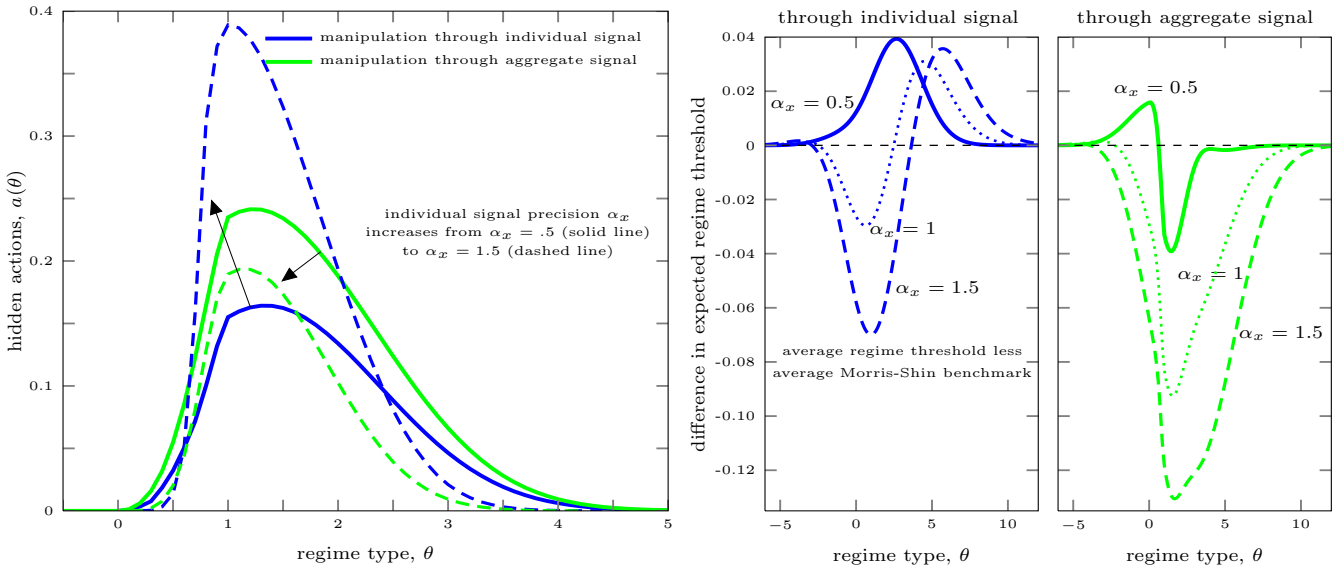
$$S(\theta, z) = \Phi(\sqrt{\alpha_x}(x^*(z) - \theta))$$

independent of the regime's hidden action a. Now the regime *ex ante* chooses $a(\theta)$ to maximize

$$a(\theta) \in \operatorname*{argmax}_{a \geq 0} \left[-C(a) + \int_{-\infty}^{\infty} \max[0, \theta - S(\theta, z)] \sqrt{\alpha_z} \phi(\sqrt{\alpha_z}(z - \theta - a))\, dz \right]$$

And again, the thresholds $\theta^*(z)$ and $x^*(z)$ are implicitly determined by indifference conditions for the regime and the citizens where the mass of subversives is $S(\theta, z)$ as above and where citizens' posterior densities are now proportional to $\phi(\sqrt{\alpha_x}(x_i - \theta))\phi(\sqrt{\alpha_z}(z - \theta - a(\theta)))$.

Numerical results. I solve these two models numerically.[21] Several examples are shown in Figure 6. The left panel shows the hidden action function $a(\theta)$ for the two models, each for two levels of signal precision, $\alpha_x = .5$ and three times higher at $\alpha_x = 1.5$. Notice that, because of the aggregate uncertainty, all regimes with $\theta > 0$ take hidden actions $a(\theta) > 0$. Even with aggregate uncertainty, regimes with $\theta < 0$ know they will be overthrown.



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

(b) difference in average thresholds $\mathbb{E}[\theta^*|\theta] - \mathbb{E}[\theta^*_{\text{MS}}|\theta]$

Figure 6: Manipulation through idiosyncratic vs. aggregate information.

Panel (a) shows the regime's hidden actions $a(\theta)$ taken to maximize its expected payoff. Since there is aggregate uncertainty, for all $\theta > 0$ regimes take positive actions. The darker lines show the case of manipulation through individual signals, the lighter lines show the case of manipulation through the aggregate signal. The solid lines show low signal precisions $\alpha_x = .5$ while the dashed lines show high signal precisions $\alpha_x = 1.5$. Panel (b) shows the difference between the average regime threshold and its Morris-Shin counterpart for the same specifications. For higher α_x, the average threshold tends to be lower than its Morris-Shin counterpart and the regime's gain is relatively larger when the manipulation takes place through aggregate information. In all these examples, $p = .25, \alpha_z = .5$ and the cost function is $C(a) = a^2/2$.

In these examples, the extent of manipulation is typically larger when the signal precision is at the higher level $\alpha_x = 1.5$ in the model where manipulation enters through the individual signal channel x_i. But the extent of manipulation is typically smaller when the signal precision is higher if the manipulation enters through the aggregate signal channel. In both cases, the ex post regime survival outcome depends on the realization of the aggregate signal z. The right panel shows the average regime threshold less the average regime threshold that would obtain in the absence of any manipulation (the corresponding Morris-Shin model with aggregate uncertainty) for each of the specifications. Here we see that for higher levels of signal precision, the average regime threshold tends to be lower than its Morris-Shin counterpart so that regimes expect to be better off. In this

[21] These calculations keep the precision α_z of aggregate information fixed and sufficiently low relative to the precision of idiosyncratic information that there is no multiplicity of monotone equilibria. As shown by Hellwig (2002) and Morris and Shin (2003), in global games multiple equilibria can be reintroduced if aggregate information is sufficiently precise compared to aggregate information (in which case there is "approximate" common knowledge).

sense, the results are qualitatively similar to those obtained for the main model without aggregate uncertainty. This specification of the model has the additional implication that the extent of the regime's expected gain is relatively larger if manipulation takes place through the aggregate information.

6.3 Struggles over information

Returning to the case of a single type of information, suppose now that there is an *opposition* and that if a regime is of type θ and takes action a while the opposition takes action e, then citizens draw signals

$$x_i = \theta + a - e + \varepsilon_i \tag{32}$$

where, as usual, ε_i is IID normal with mean zero and precision α. The regime's action a increases the signal mean but the opposition's action e decreases the signal mean. Both of these actions are unobserved by individual citizens.

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

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

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \{B(S(\theta, a, e(\theta)), \theta) - C(a)\} \tag{33}$$

$$e(\theta) \in \underset{e \geq 0}{\operatorname{argmax}} \{S(\theta, a(\theta), e) - C(\kappa e)/\kappa\} \tag{34}$$

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

Consider now a monotone equilibrium where the regime is overthrown if $\theta < \theta^*$ and citizens subvert $s(x_i) = 1$ if their signal is $x_i < x^*$ for thresholds x^*, θ^* to be determined. In this case, the mass of subversives is

$$S(\theta, a, e) = \Phi(\sqrt{\alpha}(x^* - \theta - a + e))$$

The first order necessary condition characterizing the regime's hidden action can be written, similar to (20) above,

$$C'(a) = \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - a + e(\theta))), \qquad \theta \geq \theta^* \tag{35}$$

with $a(\theta) = 0$ for all $\theta < \theta^*$. Similarly, for the opposition

$$C'(\kappa e) = \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - a(\theta) + e)), \qquad \text{all } \theta \tag{36}$$

Combining equations (35) and (36) we have that for all states $\theta \geq \theta^*$ where the regime survives

$$a(\theta) = \kappa e(\theta), \qquad \theta \geq \theta^* \tag{37}$$

And so for these states the actions of the regime are larger than those of the opposition if and only if $\kappa > 1$, i.e., when the regime has a cost advantage. For these states the equilibrium actions of the regime $a(\theta)$ implicitly solve

$$C'(a) = \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - a\,(\kappa - 1)/\kappa)), \qquad \theta \geq \theta^* \tag{38}$$

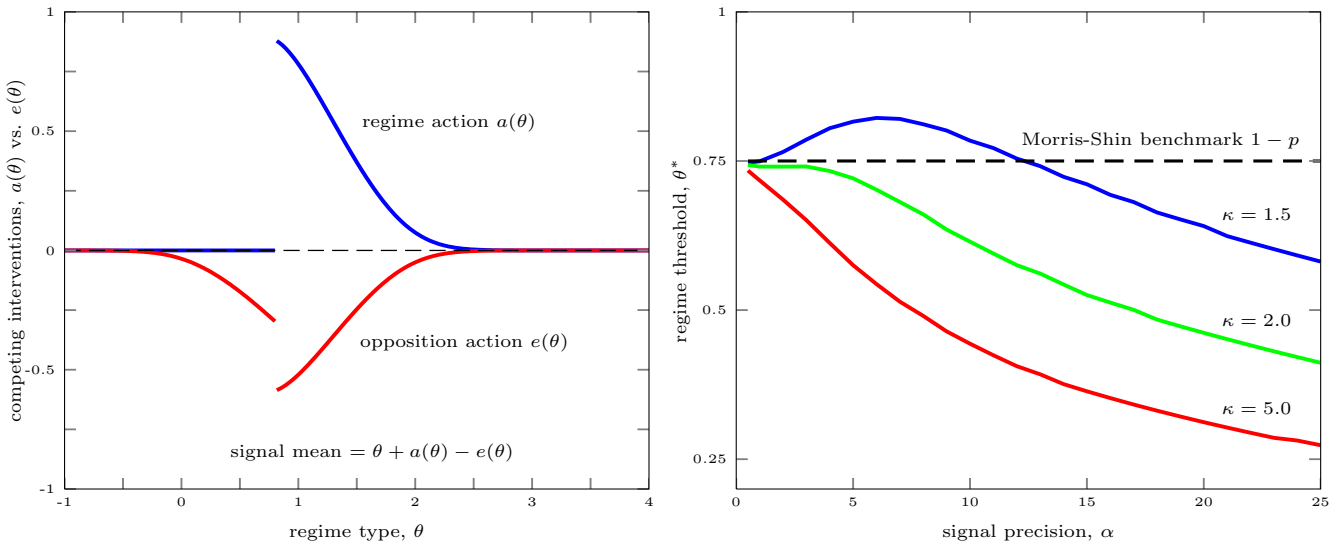
with the opposition's actions in these states then following from (37). Otherwise, when $\theta < \theta^*$ we have $a(\theta) = 0$ and the opposition's actions $e(\theta)$ implicitly solve

$$C'(\kappa e) = \sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta + e)), \qquad \theta < \theta^* \tag{39}$$

Notice that just as the regime's hidden actions jump discretely from $a = 0$ to $a(\theta^*) > 0$ at the threshold $\theta = \theta^*$, so too do the opposition's actions typically jump at the threshold (though their jump may be up *or* down, depending on parameters). The left panel of Figure 7 illustrates these action profiles $a(\theta)$ and $e(\theta)$ with $\kappa > 1$ so that the regime's actions $a(\theta)$ are larger than the opposition's actions $e(\theta)$ on $\theta \geq \theta^*$.

Does the opposition's action undo the regime's efforts? On the one hand, it is true that the presence of the opposition generally moves the threshold θ^* against the regime (it is higher than it would be in the model where there is no opposition, $\kappa = \infty$). On the other hand, the regime still manipulates information and for high enough signal precision α is still better off than it would be in the Morris-Shin benchmark where the threshold is $\theta^*_{\text{MS}} = 1 - p$. Figure 7 shows several numerical examples.

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



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

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

Figure 7: Hidden actions and regime threshold when there is an opposition.

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

6.4 Manipulating signal precision

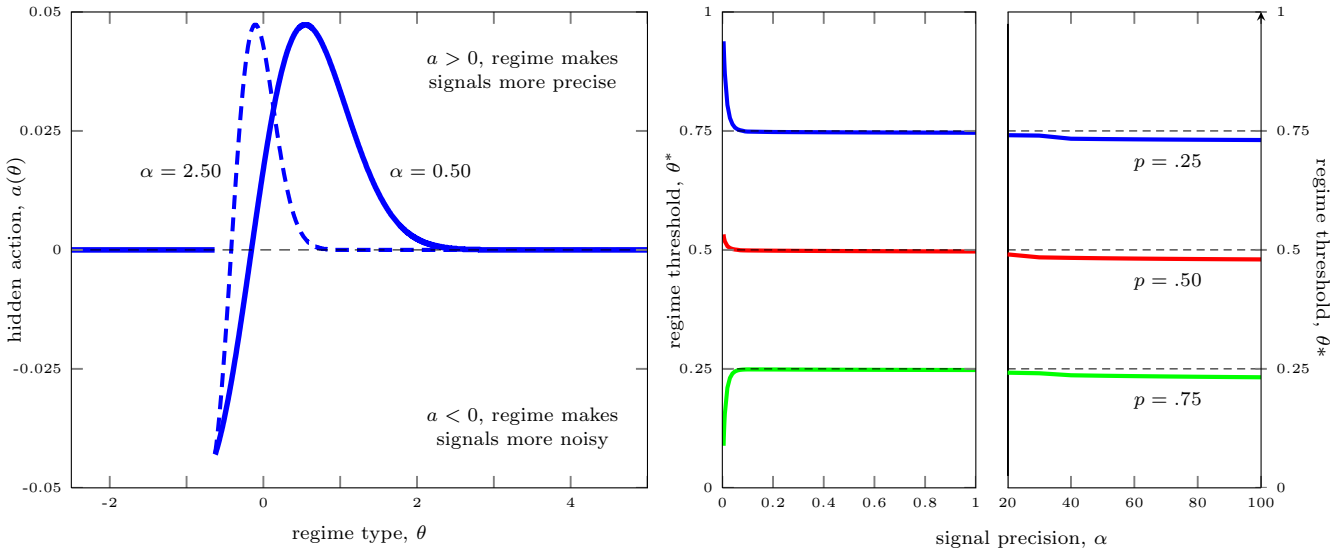
Until now, signal manipulation entered in an additive way, $x_i = \theta + a + \varepsilon_i$. With this specification the action shifts the signal mean and only *indirectly* influences the signal precision. In this section I consider an alternative approach where the regime can directly set the signal precision. In particular, let signals be $x_i = \theta + \varepsilon_i$ where the ε_i is IID normal with mean zero and precision $\beta(a) > 0$ that depends on the regime's hidden action a. I adopt the specification

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

The function $\beta : \mathbb{R} \to \mathbb{R}_+$ is strictly increasing in a with $\beta(-\infty) = \alpha/2$, $\beta(0) = \alpha$, and $\beta(\infty) = 3\alpha/2$. Thus when the regime takes no action, the precision is just α which in this context should be thought of as the *intrinsic* signal precision. Otherwise, by intervening, the regime can achieve a precision as much as 50% more or 50% less than this intrinsic level.

Again, I consider only a monotone equilibrium where the regime is overthrown for $\theta < \theta^*$ and citizens subvert $s(x_i) = 1$ for $x_i < x^*$ for thresholds x^*, θ^* to be determined. In this case, the mass of subversives is

$$S(\theta, a) = \Phi\left(\sqrt{\beta(a)}(x^* - \theta) \right)$$



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

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

Figure 8: Hidden actions and regime threshold when regime can manipulate signal precision.

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

Equilibrium hidden actions are characterized by the first order necessary condition

$$C'(a) = (\theta - x^*)\phi\left(\sqrt{\beta(a)}(x^* - \theta)\right)\frac{\partial}{\partial a}\sqrt{\beta(a)}, \qquad \theta \geq \theta^* \tag{40}$$

Actions are $a(\theta) = 0$ for $\theta < \theta^*$ before jumping discretely at θ^*. The sign of $a(\theta)$ is the same as the sign of $\theta - x^*$. If $\theta > x^*$, the hidden actions are positive. If $\theta < x^*$, the hidden actions are negative. This can only arise if $\theta^* < x^*$, in which case the hidden actions jump *down* at the threshold θ^*. Otherwise, if $\theta^* > x^*$, the hidden actions jump *up* at θ^*.

Intuitively, if the regime has intermediate type $\theta \in [\theta^*, x^*)$ then it makes the signal noisier than α (it *muddies* the signal) while if $\theta > x^*$, the regime makes the signal less noisy than α so as to *clarify* its position of strength. The left panel of Figure 8 illustrates the equilibrium hidden actions $a(\theta)$ with parameters chosen so that $x^* > \theta^*$, implying that the hidden actions jump down at θ^*. When α increases, the signal threshold x^* shifts down while the θ^* hardly moves so the interval of θ for which regimes degrade signal precision is smaller. Moreover, diminishing returns to information manipulation set in faster when α is high. An increase in α from .5 to 2.5 hardly shifts the state threshold θ^* at all. The right panel of Figure 8 confirms this general tendency. It shows θ^* as α varies from 0 to 100. The left panel shows the results for very low α. There is a brief interval of α where θ^* can initially rise. These effects play out very quickly and there is then a monotonic but *very gradual* decline in θ^*.

7 Conclusions

In this paper I develop a simple model of information and political regime change. Perhaps the most surprising result is that a regime's chances of surviving *increase* as the aggregate quantity of information available to individuals becomes sufficiently high. More specifically, I show how a regime's efforts to manipulate information to induce coordination on the status quo is effective, in the sense of increasing the regime's ex ante survival probability, when the quantity of information is sufficiently high (Proposition 2). In contrast with familiar signal-jamming games, the regime's manipulation presents citizens with a difficult signal-extraction problem and the manipulation is often payoff improving for the regime. This result suggests that breakthroughs in information technologies may not be as threatening to autocratic regimes as is often supposed.

Offsetting this pessimistic result, the model also predicts that in any circumstances where a regime is made better off by an increase in the quantity of information, it is also the case that the regime would be made worse off by an increase in the reliability of information (Proposition 3). In this sense, an increase in the quantity of information is in tension with an increase in the reliability of information; they always have opposite effects on a regime's ex ante chances of surviving.

Because of this tension, the model allows for two kinds of information revolutions. In the first kind, associated with the role of radio and mass newspapers under the totalitarian regimes of the early twentieth century, an increase in information quantity coincides with a shift towards media institutions more accommodative of the regime and hence a *decrease* in information reliability. Here the quantity and reliability effects both help the regime. In the second kind, perhaps best associated with the role of diffuse technologies like modern social media, an increase in information quantity coincides with a shift towards sources of information less accommodative of the regime and hence an increase in information reliability. Here the quantity and reliability effects work against each other.

Finally, the model goes beyond simply predicting that the reliability and quantity effects have opposite signs; it also predicts that the magnitude of the reliability effect is exactly *twice* as large as the magnitude of the quantity of information effect. Thus, if an information revolution gives rise to roughly equal-sized percentage changes in information quantity and reliability, the reliability effect will dominate so that overall the regime's chances of surviving are reduced.

The coordination game studied in this paper is deliberately stylized so as focus attention on the effectiveness of the regime's manipulation and its sensitivity to changes in the information environment. The results suggest several directions for future research. For example, this model takes as given the degree of influence the regime has over the media. But there are clear incentives for the regime to attempt to exert more media control when the quantity of information changes. It would be interesting to develop a richer model where the degree of influence over the media is itself an equilibrium outcome, in the spirit of Besley and Prat (2006) or Gehlbach and Sonin (2008), that needs to be determined simultaneously with the regime's manipulation and survival

probability. Another limitation of this paper is the reduced-form assumption that free-riding is not an overwhelming barrier to collective action against the regime. Following Lohmann (1993, 1994a), one way to endogenously mitigate the free-rider problem might be to focus on communication between individual citizens with ex ante heterogeneous preferences over the aggregate outcome. An alternative approach might be to follow Karklins and Petersen (1993) to address the building of credible coalitions against the regime. This would complement existing work such as Acemoglu, Egorov, and Sonin (2008), who study the formation of ruling coalitions within an autocratic regime. Finally, this paper has abstracted from all issues of individual and social learning. It would be interesting to develop a dynamic version of the model where both individual and social information can accumulate over time so that unrest against the regime builds (or dissipates). An extension along these lines would bring the analysis closer to the original information cascade models of regime change developed by Kuran (1991), Lohmann (1994b) and others, but would feature strategic interactions between the regime and citizens that are absent from their work.

Technical Appendix

A Proofs and omitted derivations

A.1 Morris-Shin Benchmark

Let $\hat{x}, \hat{\theta}$ denote candidates for the critical thresholds. The posterior beliefs of a citizen with x_i facing $\hat{\theta}$ are given by $\Pr[\theta < \hat{\theta} | x_i] = \Phi(\sqrt{\alpha}(\hat{\theta} - x_i))$. 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 strictly decreasing in x_i, so for each $\hat{\theta}$ there is a unique signal for which a citizen is indifferent. Similarly, if the regime faces threshold $\hat{x}$ the mass of subversives is $\Phi(\sqrt{\alpha}(\hat{x} - \theta))$. A regime θ will not be overthrown if and only if $\theta \geq \Phi(\sqrt{\alpha}(\hat{x} - \theta))$. The probability on the right hand side is continuous and strictly decreasing in θ, so for each $\hat{x}$ there is a unique state for which a regime is indifferent. The Morris-Shin thresholds $x^*_{\text{MS}}, \theta^*_{\text{MS}}$ simultaneously solve these best response conditions as equalities, as stated in equations (10)-(11) in the main text. It is then straightforward to verify that there is only one solution to these equations and that $\theta^*_{\text{MS}} = 1 - p$ independent of α and $x^*_{\text{MS}} = 1 - p - \Phi^{-1}(p)/\sqrt{\alpha}$.

A.2 Proof of Proposition 1

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

(i) *Unique equilibrium in monotone strategies*

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

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

Proof of Lemma 1. To begin, let

$$S(w) := \Phi(-\sqrt{\alpha}w) \tag{41}$$

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

$$\int_{-\infty}^{\hat{x}} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - \theta - a))\, dx_i = \Phi(\sqrt{\alpha}(\hat{x} - \theta - a)) = S(\theta + a - \hat{x}) \tag{42}$$

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

$$V(\theta, \hat{x}) := \max[0, W(\theta, \hat{x})] \tag{43}$$

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

$$W(\theta, \hat{x}) := \max_{a \geq 0} [\theta - S(\theta + a - \hat{x}) - C(a)] \tag{44}$$

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

$$a(\theta) = A(\theta - \hat{x}) \tag{45}$$

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

$$A(t) := \operatorname*{argmin}_{a \geq 0} [S(t + a) + C(a)] \tag{46}$$

The first order necessary condition for interior solutions can be written $C'(a) = -S'(t + a)$ and on using the formula for $S(\cdot)$ in equation (41) above,

$$C'(a) = \sqrt{\alpha}\phi(\sqrt{\alpha}(t + a))$$

where $\phi(w) := \exp(-w^2/2)/\sqrt{2\pi}$ for all $w \in \mathbb{R}$. This first order condition may have zero, one or two solutions for each t. If for a given t there are zero (interior) solutions, then $A(t) = 0$. If for given t there are two solutions, one of them can be ruled out by the second order sufficient condition $\alpha\phi'(\sqrt{\alpha}(t + a)) + C''(a) > 0$. Using the property $\phi'(w) = -w\phi(w)$ for all $w \in \mathbb{R}$ shows that if there are two solutions to the first order condition, only the "higher" of them satisfies the second order condition. Therefore for each t there is a single $A(t)$ that solves the regime's problem.

Making the substitution $t = \theta - \hat{x}$, the regime's threshold $\Theta(\hat{x})$ is then found from the indifference condition $W(\Theta(\hat{x}), \hat{x}) = 0$, or more explicitly

$$\Theta(\hat{x}) = S[\Theta(\hat{x}) - \hat{x} + A(\Theta(\hat{x}) - \hat{x})] + C[A(\Theta(\hat{x}) - \hat{x}))] \tag{47}$$

Taking $\hat{x}$ as given, equations (46) and (47) give the regime threshold $\Theta(\hat{x})$ and the hidden actions $a(\theta) = A(\theta - \hat{x})$ that solve the regime's problem. □

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

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

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

$$P(x_i, \hat{\theta} \,|\, a(\cdot)) := \Pr[\theta < \hat{\theta} \mid x_i, a(\cdot)] \geq p \tag{48}$$

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

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

$$K(x_i, \hat{x}) := \Pr[\theta < \Theta(\hat{x}) \mid x_i, A(\cdot)] \geq p \tag{49}$$

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

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

$$P(x, \theta) = \frac{\int_{-\infty}^{\theta} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - t))\, dt}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - t - a(t))]\, dt} \tag{50}$$

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

$$P(x, \theta) = \frac{N(\theta - x)}{N(\theta - x) + D(x, \theta)} \tag{51}$$

where

$$N(\theta - x) := \Phi(\sqrt{\alpha}(\theta - x)), \quad \text{and} \quad D(x, \theta) := \int_{\theta}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x - \xi - a(\xi)))\, d\xi \tag{52}$$

Differentiating (51) shows $P_x < 0$ if and only if $N'/N > -D_x/D$. Calculating the derivatives shows that this is equivalent to

$$H(\sqrt{\alpha}(x-\theta)) > -\frac{\int_\theta^\infty \phi'(\sqrt{\alpha}(x-y(\xi)))\,d\xi}{\int_\theta^\infty \phi(\sqrt{\alpha}(x-y(\xi)))\,d\xi} = \frac{\int_\theta^\infty \sqrt{\alpha}(x-y(\xi))\phi(\sqrt{\alpha}(x-y(\xi)))\,d\xi}{\int_\theta^\infty \phi(\sqrt{\alpha}(x-y(\xi)))\,d\xi} \tag{53}$$

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

$$\varphi(\xi|x) := \frac{\phi(\sqrt{\alpha}(x-y(\xi)))}{\int_\theta^\infty \phi(\sqrt{\alpha}(x-y(\xi')))\,d\xi'}, \qquad \xi \in [\theta,\infty) \tag{54}$$

Then after a slight rearrangement of terms in (53), $P_x < 0$ if and only if

$$H(\sqrt{\alpha}(x-\theta)) - \sqrt{\alpha}(x-\theta) > \sqrt{\alpha}\left[\theta - \int_\theta^\infty y(\xi)\varphi(\xi|x)d\xi\right] \tag{55}$$

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

$$\int_\theta^\infty y(\xi)\varphi(\xi|x)d\xi \geq \theta \tag{56}$$

But since $y(\xi) := \xi + a(\xi)$, $\xi \geq \theta$, and $a(\xi) \geq 0$, condition (56) is always satisfied. Therefore $P_x < 0$. Since $N' > 0$ and $D_\theta < 0$, $P_\theta > 0$ for all x, θ. Moreover, since $N(-\infty) = 0$ and $D > 0$ we have $P(x,-\infty) = 0$ for all x. Similarly, since $a(\xi) = 0$ for all $\xi < \theta$ as $\theta \to \infty$ we have $D(x,\theta) \to 1 - N(\theta - x)$ and since $N(+\infty) = 1$ this means $D(x,+\infty) = 0$ for all x. Therefore $P(x,+\infty) = 1$ for all x. The limit properties in x are established in parallel fashion.

(b) Fix a $\hat{x} \in \mathbb{R}$ and let $A(\theta - \hat{x})$ denote the associated hidden actions. Analogous to (51), write $P(x,\theta,\hat{x}) = N(\theta - x)/[N(\theta - x) + D(x,\theta,\hat{x})$ where $N : \mathbb{R} \to [0,1]$ is defined as in (52) above and where

$$D(x,\theta,\hat{x}) := \int_\theta^\infty \sqrt{\alpha}\phi(\sqrt{\alpha}(x - t - A(t-\hat{x})))\,dt \tag{57}$$

Now define $K(x,\hat{x}) := P(x,\Theta(\hat{x}),\hat{x})$. That $K(x,\hat{x})$ is continuous and decreasing in x is immediate from part (a) above. Finally, for $K(x,\hat{x}) = P(x-\hat{x},\Theta(\hat{x})-\hat{x},0)$ it is sufficient that $D(x,\Theta(\hat{x}),\hat{x}) = D(x-\hat{x},\Theta(\hat{x})-\hat{x},0)$. From (57) and using the change of variables $\xi := \theta - \hat{x}$ we have

$$D(x,\Theta(\hat{x}),\hat{x}) = \int_{\Theta(\hat{x})-\hat{x}}^\infty \sqrt{\alpha}\phi(\sqrt{\alpha}(x-\hat{x}-\xi-a(\xi))]\,d\xi = D(x-\hat{x},\Theta(\hat{x})-\hat{x},,0) \tag{58}$$

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

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

$$K(\psi(\hat{x}),\hat{x}) = p \tag{59}$$

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

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

Proof of Lemma 3. Since $K(x, \hat{x})$ is continuously differentiable in x, an application of the implicit function theorem to (59) shows that $\psi(\cdot)$ is continuous. Fixed points of $\psi(\cdot)$ satisfy $x^* = \psi(x^*)$. Equivalently, by part (b) of Lemma 2, they satisfy $K(x^*, x^*) = P(0, \Theta(x^*) - x^*, 0) = p$, where $\Theta(\hat{x})$ is the critical state in the regime's problem (46)-(47). By Lemma 2 and the intermediate value theorem there is a unique $z^* \in \mathbb{R}$ such that $P(0, z^*, 0) = p$. Then applying the implicit function theorem to (46)-(47) gives

$$\Theta'(\hat{x}) = \frac{\sqrt{\alpha}\phi[\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x}))]}{1 + \sqrt{\alpha}\phi[\sqrt{\alpha}(\hat{x} - \Theta(\hat{x}) - A(\Theta(\hat{x}) - \hat{x}))]} \in (0,1) \tag{60}$$

Since $\Theta(-\infty) = 0$ and $\Theta(+\infty) = 1$, there is a unique $x^* \in \mathbb{R}$ such that $\Theta(x^*) - x^* = z^*$, hence $\psi(\cdot)$ has a unique fixed point, the same x^*. Now using part (b) of Lemma 2 and implicitly differentiating (59) we have

$$\psi'(\hat{x}) = 1 + \frac{P_\theta[\psi(\hat{x}) - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0]}{P_x[\psi(\hat{x}) - \hat{x}, \Theta(\hat{x}) - \hat{x}, 0]}[1 - \Theta'(\hat{x})] \tag{61}$$

By Lemma 2, $P_\theta > 0$ and $P_x < 0$ and $\Theta'(\hat{x}) \in (0,1)$ from (60). Therefore $\psi'(\hat{x}) < 1$ for all $\hat{x}$. To see that $\psi'(x^*) > 0$, first notice that it is sufficient that $P_\theta / P_x \geq -1$ when evaluated at $\hat{x} = x^*$. Calculating the derivatives shows that this is true if and only if

$$\phi(\sqrt{\alpha}(y(\theta^*) - x^*)) + \int_{\theta^*}^{\infty} \sqrt{\alpha}\phi'(\sqrt{\alpha}(y(\theta) - x^*))\, d\theta \leq 0 \tag{62}$$

where $\theta^* := \Theta(x^*)$ and where $y(\theta) = \theta + a(\theta)$ is the mean of the signal distribution from which a citizen is sampling if the regime has type $\theta \geq \theta^*$. To show that this condition always holds, we need to consider the cases of linear costs and strictly convex costs separately. If costs are linear, $C(a) = ca$, then if $c \geq \bar{c} := \sqrt{\alpha}\phi(0)$ the result is trivial because $a(\theta) = 0$ for all $\theta \in \mathbb{R}$. So suppose $c < \bar{c}$. Then $a(\theta) = \max[0, x^* + \gamma - \theta]$ where $\gamma := \sqrt{2\log\left(\sqrt{\alpha}\phi(0)/c\right)/\alpha} > 0$. Calculating the integral and then simplifying shows that (62) holds if and only if $-\alpha\gamma\phi(\sqrt{\alpha}\gamma)a(\theta^*) \leq 0$ which is true because $a(\theta^*) \geq 0$. If costs are strictly convex, then from the optimality conditions for the regime's choice of action we have that $a(\theta) > 0$ for all $\theta \geq \theta^*$ and

$$\sqrt{\alpha}\phi(\sqrt{\alpha}(y(\theta) - x^*)) = C'(a(\theta)), \quad \theta \geq \theta^* \tag{63}$$

Differentiating with respect to θ gives

$$\alpha\phi'(\sqrt{\alpha}(y(\theta) - x^*))y'(\theta) = C''(a(\theta))a'(\theta), \quad \theta \geq \theta^* \tag{64}$$

Using the associated second order condition shows that $y'(\theta) > 0$ for $\theta \geq \theta^*$. Since $y(\cdot)$ is invertible, a change of variables shows that (62) holds if and only if

$$\int_{\theta^*}^{\infty} \phi'(\sqrt{\alpha}(y(\theta) - x^*)) \frac{a'(\theta)}{y'(\theta)} \, d\theta \geq 0 \tag{65}$$

Using (64) we equivalently have the condition

$$\int_{\theta^*}^{\infty} \frac{\phi'(\sqrt{\alpha}(y(\theta) - x^*))^2}{C''(a(\theta))} \, d\theta \geq 0 \tag{66}$$

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

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

$$\psi'(\tilde{x})[K_1(x^*, x^*) + K_2(x^*, x^*)] = 0 \tag{67}$$

where the hypothesis $\psi(\tilde{x}) = x^*$ is used to evaluate the partial derivatives K_1 and K_2. Since $\psi'(\tilde{x}) < 0$, this can only be satisfied if $K_1(x^*, x^*) + K_2(x^*, x^*) = 0$. But for any $\hat{x} \in \mathbb{R}$, the value $\psi(\hat{x})$ is implicitly defined by $K(\psi(\hat{x}), \hat{x}) = p$ so that by the implicit function theorem $\psi'(\hat{x}) = -K_2(\psi(\hat{x}), \hat{x}) / K_1(\psi(\hat{x}), \hat{x})$. From (61) we know $\psi'(\hat{x}) < 1$ for any $\hat{x}$ and since $K_1 < 0$ from Lemma 2 we conclude $K_1(\psi(\hat{x}), \hat{x}) + K_2(\psi(\hat{x}), \hat{x}) < 0$ for any $\hat{x}$. For $\hat{x} = x^*$ in particular, $K_1(x^*, x^*) + K_2(x^*, x^*) < 0$ so we have the needed contradiction. Therefore $\psi(\hat{x}) \leq x^*$ for every $\hat{x} < x^*$. A symmetric argument shows $\psi(\hat{x}) \geq x^*$ for every $\hat{x} > x^*$. □

Concluding that there is a unique equilibrium in monotone strategies. To conclude part (i) of the proof, we take an arbitrary $\hat{x} \in \mathbb{R}$ and solve the regime's problem to get $\Theta(\hat{x})$ and $a(\theta, \hat{x}) = A(\theta - \hat{x})$ using the auxiliary function from Lemma 1. We use these functions to construct $K(x_i, \hat{x})$ from (49) for each signal $x_i \in \mathbb{R}$ and use Lemma 2 to conclude that in particular $K(\hat{x}, \hat{x}) = P(0, \Theta(\hat{x}) - \hat{x}, 0)$. We then use the intermediate value theorem to deduce that there is a unique $z^* \in \mathbb{R}$ such that $P(0, z^*, 0) = p$. This gives a unique difference $z^* = \theta^* - x^*$ that can be plugged into the regime's indifference condition (47) to get the unique $\theta^* = \Theta(x^*) \in [0, 1)$ such that the regime is overthrown if and only if $\theta < \theta^*$. The unique signal threshold is then $x^* = \theta^* - z^*$ and the unique hidden action function is given by $a(\theta) := A(\theta - x^*)$.

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

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

Dominance regions. If the regime has $\theta < 0$, any mass $S \geq 0$ can overthrow the regime. Similarly, if the regime has $\theta \geq 1$ it can never be overthrown. Any regime that is overthrown

takes no action, since to do so would incur a cost for no gain. Similarly, any regime θ that is not overthrown takes an action no larger than the a such that $\theta = C(a)$. Any larger action must result in a negative payoff which can be improved upon by taking the outside option. Given this:

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

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

Iterative elimination. Starting from the dominance regions implied by $\underline{x}$ and $\overline{x}$ it is then possible to iteratively eliminate (interim) strictly dominated strategies. Recall that

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

and

$$A(t) := \operatorname*{argmin}_{a \geq 0} \left[S(t + a) + C(a)\right]$$

Again, these auxiliary functions do not depend on any endogenous variable and in particular do not depend on citizen thresholds.

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

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

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

$$\Theta(x) = S[\Theta(x) - x + A(\Theta(x) - x)] + C[A(\Theta(x) - x)] \tag{68}$$

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

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

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

$$\Phi(\sqrt{\alpha}(\underline{x}_n - \theta - a)) + \Delta(\theta + a) \tag{69}$$

for some function $\Delta : \mathbb{R} \to [0, 1]$. First consider the case $\Delta(\cdot) = 0$ where *only* citizens with $x_i < \underline{x}_n$ subvert the regime. From Lemma 1 there is a unique threshold $\underline{\theta}_n := \Theta(\underline{x}_n) \in [0, 1)$ sustained by hidden actions $a(\theta) = A(\theta - \underline{x}_n)$ solving (46)-(47) such that the regime is overthrown if $\theta < \underline{\theta}_n = \Theta(\underline{x}_n)$. Now consider the case $\Delta(\cdot) > 0$ where *some* citizens with signals $x_i \geq \underline{x}_n$ also subvert the regime. The proof that the regime is overthrown for at least all $\theta < \Theta(\underline{x}_n)$ is by contradiction. Suppose that when $\Delta(\cdot) > 0$ regime change occurs for all $\theta < \tilde{\theta}_n$ for some $\tilde{\theta}_n \leq \Theta(\underline{x}_n)$. A marginal regime $\tilde{\theta}_n$ must be indifferent between being overthrown and taking the outside option, so this threshold satisfies $\tilde{\theta}_n = S(\tilde{\theta}_n + \tilde{a}_n - \underline{x}_n) + C(\tilde{a}_n)$ where $\tilde{a}_n \geq 0$ is the optimal action for the marginal regime $\tilde{\theta}_n$. Then observe

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

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

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

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

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

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

$$K(\psi(\overline{x}_k), \overline{x}_k) = p \tag{70}$$

and

$$K(\psi(\underline{x}_k), \underline{x}_k) = p \tag{71}$$

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

A.3 Proofs of Proposition 2 and Proposition 3

Proof of Proposition 2

If $\alpha \leq \underline{\alpha} := (c/\phi(0))^2$, any regime is at a corner solution and has hidden actions $a(\theta) = 0$. In this case the regime threshold is the same as in the Morris-Shin benchmark economy, $\theta^* = 1 - p$ for all $\alpha \leq \underline{\alpha}$. If $\alpha > \underline{\alpha}$, then regimes $\theta \in [\theta^*, x^* + \gamma)$ take hidden actions $a(\theta) = x^* + \gamma - \theta$ where the coefficient $\gamma = \sqrt{\log(\alpha/\underline{\alpha})/\alpha} > 0$. For these interior solutions, substitute the regime indifference condition (26) into the citizen indifference condition (25) to obtain

$$\Phi[\sqrt{\alpha}(\theta^* - x^*)] = \frac{p}{1-p}\theta^* \qquad \Leftrightarrow \qquad \theta^* - x^* = \frac{1}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) \tag{72}$$

And now substitute this expression for the difference $\theta^* - x^*$ back into the regime indifference condition (26) to get a single equation characterizing the critical regime threshold θ^*, namely

$$T(\theta^*) := \theta^* + \frac{c}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = c\gamma + \Phi(-\sqrt{\alpha}\gamma) \tag{73}$$

Implicitly differentiating both sides with respect to α gives

$$T'(\theta^*)\frac{\partial \theta^*}{\partial \alpha} - \frac{c}{2\alpha\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = c\frac{\partial \gamma}{\partial \alpha} - \phi(\sqrt{\alpha}\gamma)\left(\sqrt{\alpha}\frac{\partial \gamma}{\partial \alpha} + \frac{1}{2\sqrt{\alpha}}\gamma\right) \tag{74}$$

Then using the fact that the coefficient γ satisfies $\sqrt{\alpha}\phi(\sqrt{\alpha}\gamma) = c$, we can simplify this to

$$T'(\theta^*)\frac{\partial \theta^*}{\partial \alpha} = \frac{c}{2\alpha}\left[\frac{1}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) - \gamma\right] \tag{75}$$

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

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \quad \Leftrightarrow \quad \theta^* < \theta_{\text{crit}} := \frac{1-p}{p}\Phi(\sqrt{\alpha}\gamma) \tag{76}$$

And because $T'(\theta) > 0$ for all θ we have $\theta^* < \theta_{\text{crit}}$ if and only if $T(\theta^*) < T(\theta_{\text{crit}})$. Applying $T(\cdot)$ to both sides of equation (76) and simplifying we have that the regime threshold θ^* is decreasing in α if and only if

$$p < \Phi(\sqrt{\alpha}\gamma) \tag{77}$$

Since $\sqrt{\alpha}\gamma > 0$ and $\Phi^{-1}(p) < 0$ for any $p < 1/2$, this condition is necessarily satisfied if $p < 1/2$. Using $\sqrt{\alpha}\gamma = \sqrt{\log(\alpha)/\underline{\alpha}}$ and rearranging we have the stated condition for the critical signal precision α^*, namely

$$\frac{\partial}{\partial \alpha}\theta^* < 0 \quad \Leftrightarrow \quad \alpha > \alpha^* = \underline{\alpha}\exp\left(\max\left[0, \Phi^{-1}(p)\right]^2\right) \tag{78}$$

For all $\alpha > \alpha^*$ the opportunity cost is $p < \Phi(\sqrt{\alpha}\gamma)$ so that the regime threshold θ^* is decreasing.

To establish that $\lim_{\alpha\to\infty}\theta^* = 0$, observe that for any $w \in \mathbb{R}$ the cumulative density $\Phi(\sqrt{\alpha}w) \to \mathbb{1}\{w > 0\}$ as $\alpha \to \infty$, i.e., to the indicator function that equals one if $w > 0$ and zero otherwise. Moreover, as $\alpha \to \infty$ the coefficient $\gamma = \sqrt{\log(\alpha/\underline{\alpha})/\alpha} \to 0$ and $\Phi(-\sqrt{\alpha}\gamma) \to 0$. Applying these to (25) we see that for large α solutions to the citizen's indifference condition are approximately the same as solutions to

$$\mathbb{1}\left\{\theta^* - x^* > 0\right\} = -\frac{p}{1-p}(\theta^* - x^*)c \tag{79}$$

The only solution to equation (79) is $\theta^* - x^* = 0$. So as $\alpha \to \infty$, solutions to (25) approach zero too. Then from the regime's indifference condition (26), if $\theta^* - x^* \to 0$ it must also be the case that $\theta^* \to 0$ as claimed. ■

Proof of Proposition 3

Following calculations similar to those in the proof of Proposition 2 above, we have

$$T'(\theta^*)\frac{\partial \theta^*}{\partial c} + \frac{1}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) = \gamma + c\frac{\partial \gamma}{\partial c} - \phi(\sqrt{\alpha}\gamma)\sqrt{\alpha}\frac{\partial \gamma}{\partial c} \tag{80}$$

Again using $\sqrt{\alpha}\phi(\sqrt{\alpha}\gamma) = c$ and rearranging we have

$$T'(\theta^*)\frac{\partial \theta^*}{\partial c} = \gamma - \frac{1}{\sqrt{\alpha}}\Phi^{-1}\left(\frac{p}{1-p}\theta^*\right) \tag{81}$$

Then using equation (81) to eliminate $T'(\theta^*)$ from (75) and simplifying we have

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

Since $\alpha/c > 0$, the two effects have the opposite sign as claimed. ■.

B Role of coordination

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

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

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \left[-C(a) + \int_{-\infty}^{\infty} \max[0, \theta - S(x)]\sqrt{\alpha}\phi(\sqrt{\alpha}(x - \theta - a))\, dx \right] \tag{83}$$

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

$$-C(a) - \min[\theta, 1]\Phi(\sqrt{\alpha}(x^* - \theta - a)) \tag{84}$$

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

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

$$\min[\theta, 1]\sqrt{\alpha}\phi(\sqrt{\alpha}(x^* - \theta - a)) = C'(a), \qquad \theta \geq 0 \tag{85}$$

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

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

$\theta < 1$. By contrast, if citizens are imperfectly coordinated then for precise information all regimes $\theta \geq 0$ survive.

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

C Equilibrium beliefs

C.1 Laplacian beliefs in the Morris-Shin benchmark

In this global game the marginal citizen with signal $x_i = x^*_{\text{MS}}$ believes the equilibrium mass of subversives $S^*_{\text{MS}}(\theta)$ is *uniformly distributed* on its support $[0, 1]$. That is, if the equilibrium mass of subversives is $S^*_{\text{MS}}(\theta) = \Phi[\sqrt{\alpha}(x^*_{\text{MS}} - \theta)]$, the marginal citizen assigns the event $S^*_{\text{MS}}(\theta) \leq k$ posterior probability

$$G^*_{\text{MS}}(k) := \Pr[S^*_{\text{MS}}(\theta) \leq k | x^*_{\text{MS}}] = k, \qquad k \in [0, 1]$$

i.e., the *uniform distribution* on $[0, 1]$. In a sense, the marginal citizen is agnostic about the mass of subversives. Morris and Shin (2003) refer to these as *Laplacian beliefs* (in reference to the principle of insufficient reason). Intuitively, since citizens have no prior information about the regime's θ, a citizen's signal x_i contain no information about that citizen's rank-order in the population and thus provides no information about the proportion of citizens who observe lower (or higher) signals.

C.2 Non-Laplacian beliefs if information is manipulated

Now let $S(\theta) = \Phi[\sqrt{\alpha}(x^* - \theta - a(\theta))]$ denote the equilibrium mass of subversives when the regime can manipulate. This function inherits a discontinuity at θ^* from the hidden actions $a(\theta)$.

For simplicity, consider the case of linear costs so that $a(\theta) = \theta^{**} - \theta$ on the interval $[\theta^*, \theta^{**})$ and zero elsewhere. Then it is straightforward to show that the marginal citizen with $x_i = x^*$ assigns the event $S(\theta) \leq k$ posterior probability

$$G^*(k) := \Pr[S(\theta) \leq k | x^*] = \begin{cases} k\dfrac{(1-p)}{\theta^*} & k \in [0, k^{**}) \\ 1-p & k \in [k^{**}, k^*] \\ (k-1)\dfrac{(1-p)}{\theta^*} + 1 & k \in (k^*, 1] \end{cases} \tag{86}$$

where the critical points $0 \leq k^{**} \leq k^* \leq 1$ are respectively the masses of subversives *just below* the regime threshold θ^* and at the regime type being imitated $\theta^{**} > \theta^*$, specifically

$$k^* := \lim_{\theta \uparrow \theta^*} S(\theta) = \Phi(\sqrt{\alpha}(x^* - \theta^*)) \tag{87}$$

and

$$k^{**} := S(\theta^{**}) = \Phi(\sqrt{\alpha}(x^* - \theta^{**})) \tag{88}$$

and where, from equations (25)-(26) above, k^* and the regime threshold θ^* are also related by

$$(1 - k^*)(1 - p) = p\theta^* \tag{89}$$

which from (86) implies that the posterior $G^*(k)$ is continuous at $k = k^*$. As illustrated in Figure 9, the posterior is piecewise linear in k with slope $(1 - p)/\theta^*$. At the lower point $k = k^{**}$ it jumps discretely and then takes the value $1 - p$ on the interval $[k^{**}, k^*]$ before rising again with slope $(1 - p)/\theta^*$ after it reaches the higher point k^*. In other words, the posterior probability has an *atom* at the point $k = k^{**}$. This comes from the discontinuity in the hidden actions $a(\theta)$ at the regime threshold θ^* and is a general feature of this model. By contrast, the constancy on $[k^*, k^{**}]$ is special to the case of linear costs where all regimes $\theta \in [\theta^*, \theta^{**}]$ produce exactly the same signal mean so that the mass of subversives $S(\theta)$ is the same for all such regimes. For strictly convex costs, the posterior is relatively flat on this interval without being exactly constant.

Key to the shape of the posterior is the ratio $\theta^*/(1 - p)$. Since the Morris-Shin benchmark is $\theta^*_{\text{MS}} = 1 - p$, the shape of the posterior is determined by whether θ^* is less than or more than the Morris-Shin benchmark $1 - p$. If the signal precision $\alpha \to \infty$ so that $\theta^* \to 0$, then in this limit the marginal citizen's posterior probability is the *binomial* distribution on $\{0, 1\}$ with probabilities $\{1-p, p\}$ respectively, as shown in the left panel of Figure 10. If, on the other hand, the parameters of the model are such that $\alpha < \underline{\alpha}$, then there is no manipulation in equilibrium, $a(\theta) = 0$ for all θ, so that $\theta^* = 1 - p$ and $k^* = k^{**}$ and the posterior $G^*(k)$ reduces to the uniform distribution on $[0, 1]$. This is shown in the right panel of Figure 10.

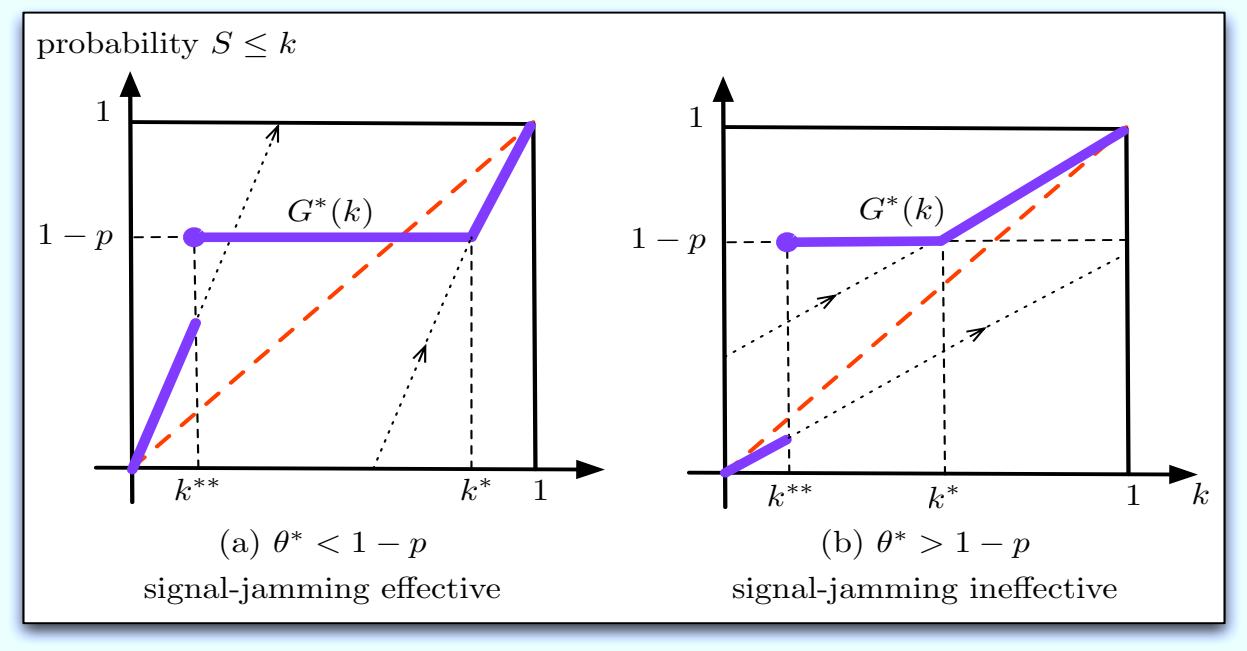


Figure 9: Non-Laplacian beliefs of the marginal citizen with signal $x_i = x^*$.

The shape of the posterior probability $G^*(k) := \Pr[S(\theta) \leq k | x^*]$ depends on whether the threshold θ^* is above or below the Morris-Shin benchmark $1-p$. If parameters are such that $\theta^* < 1-p$, then signal jamming is effective and the posterior is as in panel (a). Otherwise, if $\theta^* > 1-p$, signal-jamming is ineffective and the posterior is in panel (b). The *atom* at the point k^{**} comes from the discontinuity in the hidden actions $a(\theta)$ at the regime threshold θ^*. In either case, the posterior takes the value $1-p$ at this point. By contrast, in the Morris-Shin benchmark the marginal citizen has uniform beliefs over the mass of subversives (indicated by the 45° line).

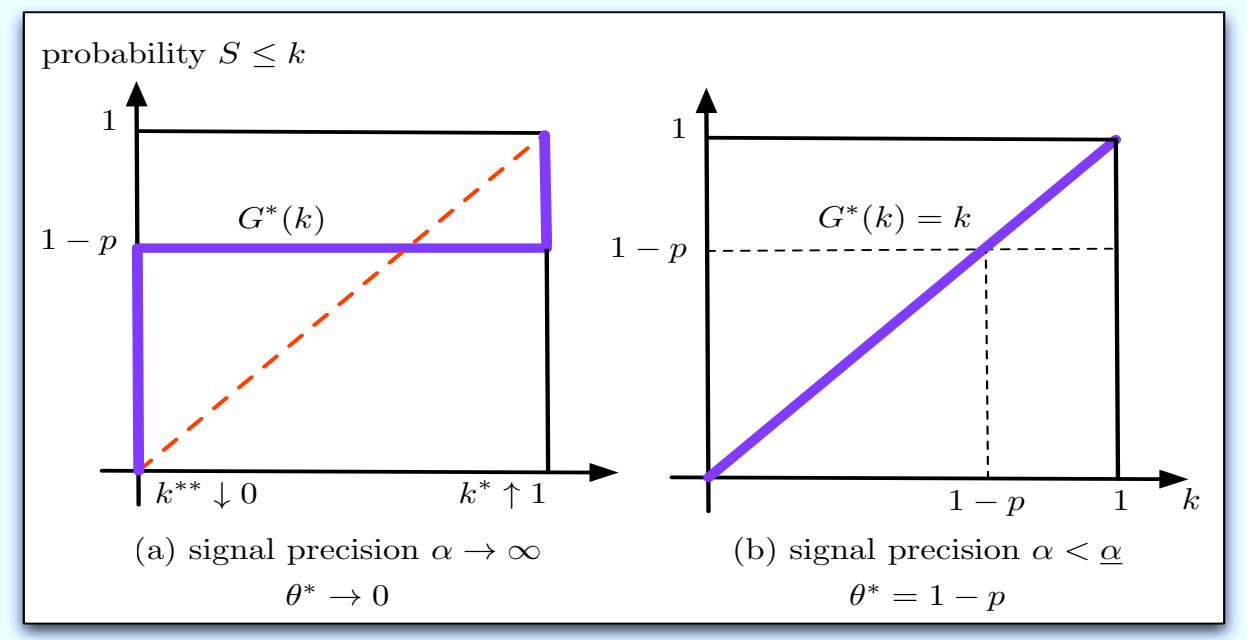


Figure 10: Binomial and uniform limiting cases.

As the signal precision $\alpha \to \infty$, the regime threshold $\theta^* \to 0$. In this case $k^{**} \to 0$ from above while $k^* \to 1$ from below. In the limit, the posterior for the mass of subversives is the *binomial* distribution on $\{0, 1\}$ with probabilities $\{1-p, p\}$ respectively. For low signal precision, $\alpha \leq \underline{\alpha}$, there is no information manipulation. In the limit, $k^{**} = k^* = 1-p$ and the posterior for the mass of subversives approaches is the *uniform* distribution on $[0, 1]$.

References

Daron Acemoglu and James A. Robinson. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press, 2006. 4

Daron Acemoglu, Georgy Egorov, and Konstantin Sonin. Coalition formation in non-democracies. *Review of Economic Studies*, 75(4):987–1009, 2008. 37

George-Marios Angeletos and Iván Werning. Crises and prices: Information aggregation, multiplicity and volatility. *American Economic Review*, 96(5):1720–1736, 2006. 5

George-Marios Angeletos, Christian Hellwig, and Alessandro Pavan. Signaling in a global game: Coordination and policy traps. *Journal of Political Economy*, 114(3):452–484, 2006. 5, 15, 23, 48

George-Marios Angeletos, Christian Hellwig, and Alessandro Pavan. Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756, 2007. 5

Hannah Arendt. *The Origins of Totalitarianism*. André Deutsch, revised edition, 1973. 26

Sandeep Baliga and Tomas Sjöström. The strategy of manipulating conflict. Northwestern University working paper, 2010. 4

David P. Baron. Perisistent media bias. *Journal of Public Economics*, 90(1-2):1–36, 2006. 5

Eli Berman and David D. Laitin. Religion, terrorism and public goods: Testing the club model. *Journal of Public Economics*, 92:1942–1967, 2008. 10

Timothy Besley and Andrea Prat. Handcuffs for the grabbing hand? The role of the media in political accountability. *American Economic Review*, 96(3):720–736, 2006. 5, 11, 36

Carles Boix and Milan Svolik. The foundations of limited authoritarian government: Institutions and power-sharing in dictatorships. Princeton University working paper, 2009. 5

Subir Bose, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund. Dynamic monopoly pricing and herding. *RAND Journal of Economics*, 37(4):910–928, 2006. 4

Bruce Bueno de Mesquita, Alastair Smith, Randolph M. Siverson, and James D. Morrow. *The Logic of Political Survival*. MIT Press, 2003. 4

Ethan Bueno de Mesquita. Regime change and revolutionary entrepreneurs. *American Political Science Review*, 104(3):446–466, 2010. 4, 5, 9, 32

Hans Carlsson and Eric van Damme. Global games and equilibrium selection. *Econometrica*, 61 (5):989–1018, 1993. 5, 12

Michael S. Chase and James C. Mulvenon. *You've Got Dissent: Chinese Dissident use of the Internet and Beijing's Counter-Strategies*. RAND report MR-1543, 2002. 27

Sylvain Chassang and Gerard Padro-i-Miquel. Conflict and deterrence under strategic risk. *Quarterly Journal of Economics*, 125(4):1821–1858, 2010. 5

Michael Suk-Young Chwe. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press, 2001. 26

Noam Cohen. Twitter on the barricades: Six lessons learned. *New York Times*, June 2009. 27

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6): 1431–1451, 1982. 23

Alexandre Debs. Divide-and-rule and the media. University of Rochester working paper, 2007. 5

Mathias Dewatripont, Ian Jewitt, and Jean Tirole. The economics of career concerns, part I: Comparing information structures. *Review of Economic Studies*, 66(1):183–198, 1999. 22

Georgy Egorov, Sergei Guriev, and Konstantin Sonin. Media freedom, bureaucratic incentives, and the resource curse. Harvard University working paper, 2006. 5

Golnaz Esfandiari. The Twitter devolution. *Foreign Policy*, June 2010. 2, 27

James Fallows. The connection has been reset. *Atlantic Monthly*, 301(2), 2008. 2, 27

James D. Fearon. Self-enforcing democracy. Stanford University working paper, 2006. 9, 10

Carl J. Friedrich and Zbigniew K. Brzezinski. *Totalitarian Dictatorship and Autocracy*. Harvard University Press, second edition, 1965. 26

Benjamin Frommer. *National Cleansing: Retribution Against Nazi Collaborators in Postwar Czechoslovakia*. Cambridge University Press, 2005. 10

Scott Gehlbach and Konstantin Sonin. Government control of the media. University of Wisconsin working paper, 2008. 5, 11, 36

Matthew Gentzkow and Jesse M. Shapiro. Media bias and reputation. *Journal of Political Economy*, 114(2):280–316, 2006. 5, 11

John Ginkel and Alastair Smith. So you say you want a revolution: A game-theoretic explanation of revolution in repressive regimes. *Journal of Conflict Resolution*, 43(3):291–316, 1999. 4

Christian Hellwig. Public information, private information, and the multiplicity of equilibria in coordination games. *Journal of Economic Theory*, 107:191–222, 2002. 12, 29, 31

Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182, 1999. 22, 23

Julian Jackson. *France: The Dark Years, 1940-1944*. Oxford University Press, 2001. 10

Shanthi Kalathil and Taylor C. Boas. *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule*. Carnegie Endowment for International Peace, 2003. 2, 26, 27

Stathis N. Kalyvas. How free is 'free riding' in civil wars? Violence, insurgency, and the collective action problem. *World Politics*, 59(2):1043–1068, 2007. 10

Rasma Karklins and Roger Petersen. Decision calculus of protesters and regimes: Eastern europe 1989. *Journal of Politics*, 55(3):588–614, 1993. 10, 37

David D. Kirkpatrick. Wired and shrewd, young Egyptians guide revolt. *New York Times*, February 2011. 1

Timur Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public Choice*, 61:41–74, 1989. 4, 9

Timur Kuran. Now out of never: The element of surprise in the European revolution of 1989. *World Politics*, 44(1):7–48, 1991. 4, 37

Timur Kuran. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press, 1995. 4, 9

Susanne Lohmann. A signaling model of informative and manipulative political action. *American Political Science Review*, 87(2):319–333, 1993. 10, 37

Susanne Lohmann. Information aggregation through costly political action. *American Economic Review*, 84(3):518–530, 1994a. 10, 37

Susanne Lohmann. The dynamics of information cascades: The monday demonstrations in leipzig, east germany, 1989-91. *World Politics*, 47(1):42–101, 1994b. 4, 37

Douglas MacMillan. Google's quixotic China challenge. *Bloomberg Businessweek*, March 2010. 1

Evgeny Morozov. *The Net Delusion*. Allen Lane, 2011. 2, 27

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, 88(3):587–597, 1998. 5, 12

Stephen Morris and Hyun Song Shin. Rethinking multiple equilibria in macroeconomic modeling. In Ben S. Bernanke and Kenneth Rogoff, editors, *NBER Macroeconomics Annual*, pages 139–161. MIT Press, 2000. 5, 12

Stephen Morris and Hyun Song Shin. Global games: Theory and applications. In Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics: Theory and Applications*. Cambridge University Press, 2003. 5, 12, 29, 31, 48

Sendhil Mullainathan and Andrei Shleifer. The market for news. *American Economic Review*, 95 (4):1031–1053, 2005. 2, 5, 6, 10, 11

Mike Musgrove. Twitter is a player in Iran's drama. *Washington Post*, June 2009. 1

Mancur Olson. *The Logic of Collective Action*. Harvard University Press, revised edition, 1971. 9

Torsten Persson and Guido Tabellini. Democratic capital: The nexus of political and economic change. *American Economic Journal: Macroeconomics*, 1(2):88–126, 2009. 5

John Ribeiro. Facebook blocked in Iran ahead of elections. *PC World*, May 2009. 27

Todd Sandler. *Collective Action: Theory and Applications*. University of Michigan Press, 1992. 4

Francesco Sobbrio. A citizen-editors model of news media. IMT Lucca working paper, 2010. 2

Lawrence C. Soley. *Clandestine Radio Broadcasting: A Study of Revolutionary and Counterrevolutionary Electronic Broadcasting*. Praeger, 1987. 27

Daniel F. Stone. Ideological media bias. Oregon State University working paper, 2010. 2

Gordon Tullock. The paradox of revolution. *Public Choice*, 11(Fall):89–100, 1971. 10

Gordon Tullock. *The Social Dilemma: The Economics of War and Revolution*. Blacksburg: Center for the Study of Public Choice, 1974. 10

Z.A.B. Zeman. *Nazi Propaganda*. Oxford University Press, second edition, 1973. 2, 26