Information Manipulation, Coordination, and Regime Change:

Supplementary Online Appendix

Chris Edmond*

June 2013

Contents

*University of Melbourne. Email: ⟨cedmond@unimelb.edu.au⟩.

This supplementary appendix is organized as follows. Appendix I provides further discussion of the related literature and the model assumptions. Appendix II outlines results for several alternative information structures. In particular, Appendix II.1 discusses common informative priors, and Appendix II.2 considers heterogeneous informative priors. Both of these setting can be interpreted as giving individuals two types of information, only one of which is contaminated by the regime's manipulation. I emphasize this interpretation in Appendix II.2 since there all information is idiosyncratic (and in that sense is on the same footing). Appendix II.3 then compares the benchmark model to an alternative setup where the regime's manipulation enters via *public* information. Finally, Appendix II.4 explains how the effects of changing precision can be disentangled from changes in the amount of *correlation* in individual signals.

I Further discussion of the model

I.1 Related literature

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 (1994a), 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.[1]

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.[2]

[1] 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.

[2] See Baliga and Sjöström (2012) for a related model with cheap talk instead of costly signaling.

By contrast, in Bueno de Mesquita, as in this paper, information heterogeneity plays a key role in determining equilibrium 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.

I.2 Model assumptions

Media outlets and citizen information. Citizens obtain information about the regime's type from n identical *media outlets*. Each media outlet $j = 1, ..., n$ chooses a signal mean y_j for the information it produces and each citizen $i \in [0, 1]$ costlessly acquires a signal x_{ij}, one from each of the n outlets.[3] Each signal is of the form $x_{ij} = y_j + \varepsilon_{ij}$ where the ε_{ij} 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

[3]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 $n \geq 2$.

true type θ and weight $1-r$ on accommodating the regime's preferred message $\theta+\hat{a}$. Each outlet chooses a signal mean y_j to minimize a quadratic loss function

$$L(y,\theta,\hat{a}) := r\,(y-\theta)^2 + (1-r)\,(y-(\theta+\hat{a}))^2, \qquad 0<r<1 \tag{1}$$

with solution

$$y = \theta + (1-r)\hat{a} \tag{2}$$

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

Because the media outlets are symmetric, the information of citizen i can be represented by the *average* signal $x_i := \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ which satisfies[4]

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

where similarly $\varepsilon_i := \frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij}$. Since the ε_{ij} are jointly IID normal with mean zero and precision $\hat{\alpha}$, a citizen's average noise ε_i is also normal with mean zero and precision $\alpha := n\hat{\alpha}$. With this representation of citizen information, it is also natural to analyze the model in terms of the regime's *effective* hidden action $a := (1-r)\hat{a}$. In terms of the effective hidden action a, the regime's cost function is $C(a/(1-r))$, and hence the regime's costs are increasing in r for any $a>0$. For example, if $C(a) = a/(1-r)$, the regime's marginal cost is $c := 1/(1-r)$ and Proposition 5 from the main text can be immediately restated in terms of media reliability (with an increase in reliability r monotonically increasing the regime's marginal cost, etc).

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 biased information.[5] In my model, information is biased in equilibrium without citizens having any preference for bias. Nothing predisposes citizens to have biased beliefs in equilibrium. They still use the decision problem of the regime to draw inferences about the mapping between the regime's type and its action. In principle, they could still in equilibrium completely negate the regime's manipulation. That this does not happen is because of the underlying coordination game plus heterogeneous information, not because of 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 $C(a/(1-r))$ but does not affect the citizens' ability to discard any bias that has been introduced.

[4] 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 Appendix II.2 below.

[5] 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.

Citizen payoffs: general version. Suppose that 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 state-contingent cost of attacking

$$p(S,\theta) := \begin{cases} \overline{p} & \text{if } \theta \geq S \\ \underline{p} & \text{if } \theta < S \end{cases}, \qquad 0 \leq \underline{p} \text{ and } \underline{p} \leq \overline{p}, \text{ strictly if } \underline{p} = 0 \tag{3}$$

so that an individual who attacks 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 participation 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 participation does not depend on the regime outcome (i.e., $\underline{p} = \overline{p} > 0$). Similarly, let $u(s_i, S, \theta)$ denote the benefit from the regime outcome

$$u(s_i, S, \theta) := \begin{cases} \overline{u} & \text{if } \theta < S \text{ and } s_i = 1 \\ \underline{u} & \text{if } \theta < S \text{ and } s_i = 0 \\ 0 & \text{otherwise} \end{cases}, \qquad 0 < \underline{u} \leq \overline{u} \tag{4}$$

so that if an individual attacks 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 (3) above. A citizen's net utility is

$$U(s_i, S, \theta) := u(s_i, S, \theta) - p(S, \theta)s_i \tag{5}$$

Or, in tabular form,

	attack $s_i = 1$	not attack $s_i = 0$
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.

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.[6] 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 attacking the regime, $s(x_i) = 1$, is

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

[6] 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).

while the expected payoff from not attacking 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 participating in the attack optimal if and only if

$$P(x_i) \geq \frac{\overline{p}}{(\overline{p} - \underline{p}) + (\overline{u} - \underline{u})} =: p \tag{6}$$

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, 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 participate in the attack. 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, in the main text I always 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 attacking 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 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 attack 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 to also participate in the attack.

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, 1994b) considers a model where individuals participate in individually costly political action out of the desire to signal private information about a common fundamental.[7] In her model, individuals are heterogeneous with respect to their preferences over aggregate outcomes and thus, despite

[7] 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.

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.[8] 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.

II Alternative information structures

II.1 Common informative priors

In addition to their distorted signal x_i, let citizens have a public signal z that is also informative for θ but that is uncontaminated by the regime's action

$$x_i = \theta + a + \varepsilon_{x,i}, \qquad \text{and} \qquad z = \theta + \varepsilon_z \tag{7}$$

where the noise terms $\varepsilon_{x,i}$ and ε_z are independent, jointly normally distributed, both with mean zero and precisions α_x and α_z, respectively. Equivalently, citizens have a common informative prior that is normal with mean z and precision α_z.

No manipulation benchmark. If no hidden actions are possible, the model reduces to a setup studied by Angeletos and Werning (2006), Hellwig (2002), Metz (2002), and Morris and Shin (2000, 2003) and others. For each z, a monotone equilibrium is a threshold $x^*(z)$ such that individuals attack if $x_i < x^*(z)$ and a $\theta^*(z)$ such that regimes are overthrown if $\theta < \theta^*(z)$. It is well known that if public information is too precise relative to private information, there may be multiple monotone equilibria [see, e.g., Hellwig (2002) and Morris and Shin (2003, 2004)]. If public information is too precise, there is "approximate" common knowledge of θ. Hellwig (2002) derived a sufficient condition for a unique monotone equilibrium in a game of this kind, namely $\alpha_z/\sqrt{\alpha_x} < \sqrt{2\pi}$. In the discussion that follows, I assume that this condition is satisfied.

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

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

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

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

Figure 1 confirms that this intuition is correct. When primitives are relatively favorable to the regime, the coordination effect and the information manipulation effect pull in opposing directions. The coordination effect tends to drive the equilibrium threshold up and to reduce the survival probability of the regime. But for high α_x the coordination effect is irrelevant while the information manipulation effect is powerful, as in the benchmark model. In this case, the thresholds are "hump-shaped". But if primitives are unfavorable to the regime, both the coordination and information manipulation effects are in the regime's favor and reinforce each other. In this case, the regime threshold is monotone decreasing and asymptotes to zero. If primitives are unfavorable to the regime, there is a large benefit from information manipulation (this contrasts with the benchmark model where regimes with favorable primitives benefitted more from a given increase in α_x).

Thus informative priors do not undo the central message of the benchmark model. For

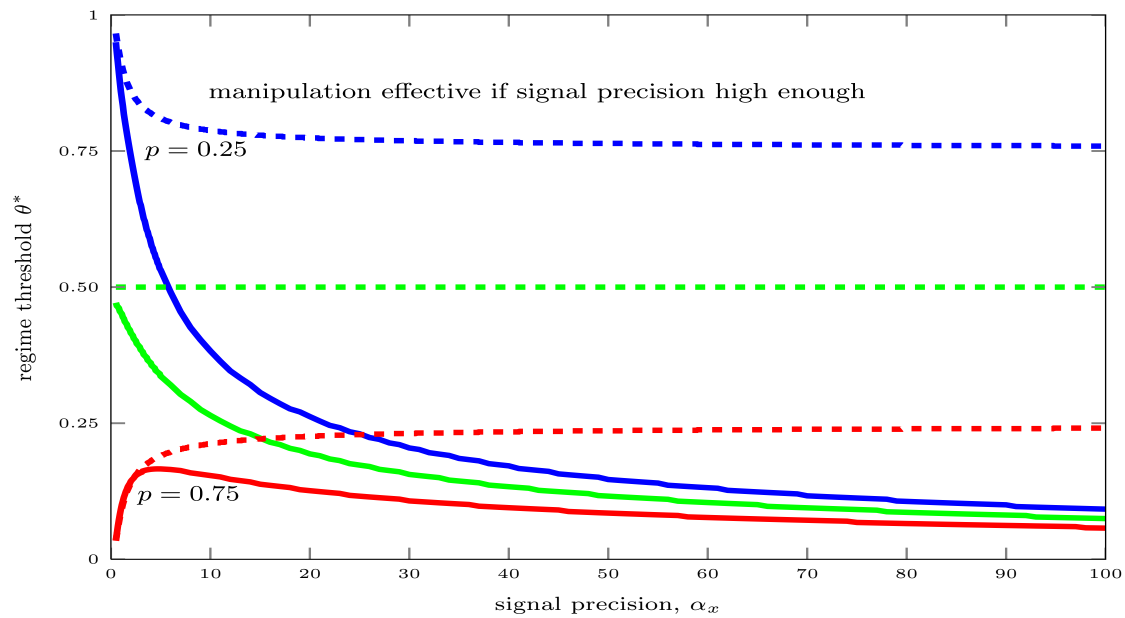


Figure 1: Information manipulation with informative common priors.

Shows the regime threshold θ^* as functions of α_x. Solid lines show the regime thresholds when there is information manipulation, dashed lines show the thresholds when there is no information manipulation. If primitives are relatively favorable to the regime (say $p = 0.75$), the coordination effect identified by Metz (2002) and the information manipulation effect pull in opposite directions, giving rise to a "hump-shaped" function. If primitives are unfavorable to the regime (say $p = 0.25$), both forces drive θ^* down.

high α_x signal-jamming is effective, $\theta^* < \theta^*_{\text{MS}} = 1 - p$. High signal precision may increase the regime's ex ante survival probability. But perhaps this is not the most interesting comparative static. What if the precision of the *public* signal increases? Then we run into the problem of multiplicity. For given α_x, the inequality $\alpha_z/\sqrt{\alpha_x} < \sqrt{2\pi}$ will eventually be violated and there are multiple monotone equilibria. In this case, we lose the ability to draw sharp conclusions. What if both the precisions of the public and private information increase together? Then if the bound on the relative precision of the public information is satisfied the analysis goes through essentially as above. At a given level of the relative precision $\alpha_z/\sqrt{\alpha_x}$, the threshold may be increasing or decreasing (depending on the relative strengths of the coordination effect and the information manipulation effect), but as the ratio $\alpha_z/\sqrt{\alpha_x}$ becomes small the hidden action effect dominates.

II.2 Heterogeneous informative priors

In this section I consider an alternative and perhaps more interesting way to give citizens an uncontaminated source of information. In particular, I suppose that media outlets come in two types, some that are amenable to the regime's message and others who resolutely report the truth. Specifically, let citizens have n_x reports from media outlets that contain the regime's action and n_z reports from media outlets that do not contain the regime's action. Citizens observe each of these reports with idiosyncratic noise that is jointly IID normal across them and across all media outlets with mean zero and precision $\hat{\alpha}$. Then the information of a citizen can be represented by *two* representative signals

$$x_i = \theta + a + \varepsilon_{x,i}, \qquad \text{and} \qquad z_i = \theta + \varepsilon_{z,i} \tag{8}$$

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 is equivalent to giving citizens heterogeneous informative priors. It 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.

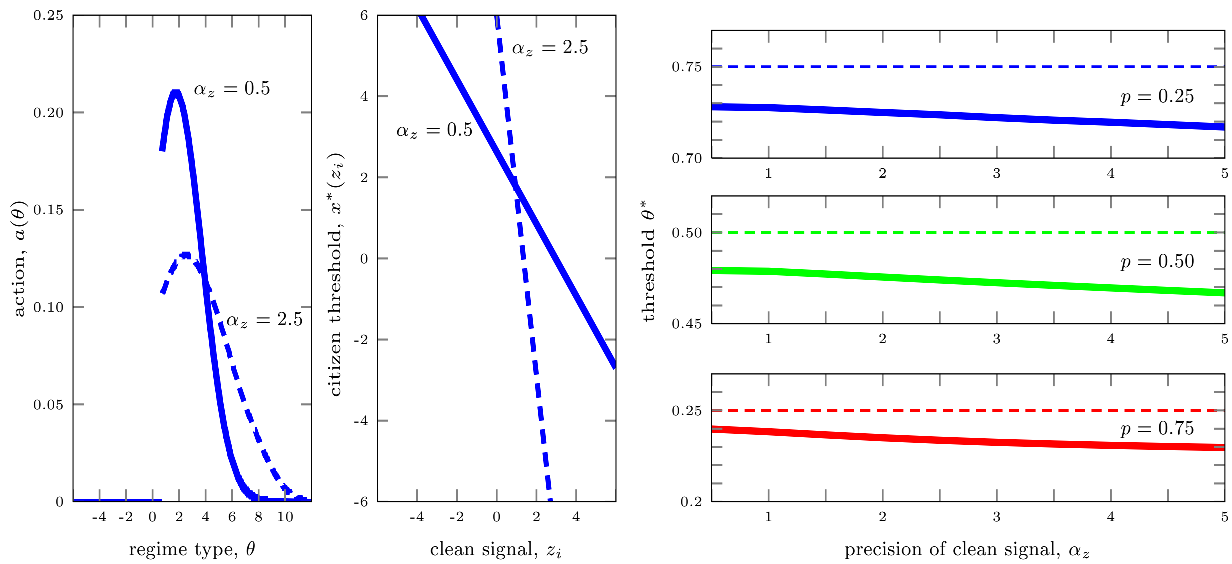
I consider a monotone equilibrium where the regime is overthrown for $\theta < \theta^*$ and citizens attack, $s(x_i, z_i) = 1$, if their signals satisfy $x_i < x^*(z_i)$. Here θ^* is a single threshold and $x^* : \mathbb{R} \to \mathbb{R}$ is a threshold *function*, both to be determined endogenously. In this case, the aggregate attack 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 2, 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 participation in an attack. 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.*



(a) hidden actions $a(\theta)$ and citizen thresholds $x^*(z_i)$ (b) regime threshold θ^* as function of clean signal precision α_z

Figure 2: 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 participation in the attack. In this example the opportunity cost of participating 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 $C(a) = a^2/2$.

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.

II.3 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. In this appendix, 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 idiosyncratic and aggregate information but which differ in the channel by which manipulation enters.

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 attack, $s(x_i, z) = 1$, 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* aggregate attack 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

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

The thresholds $\theta^*(z)$ and $x^*(z)$ are implicitly determined by indifference conditions for the regime and the citizens where the aggregate attack 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* aggregate attack facing a regime is

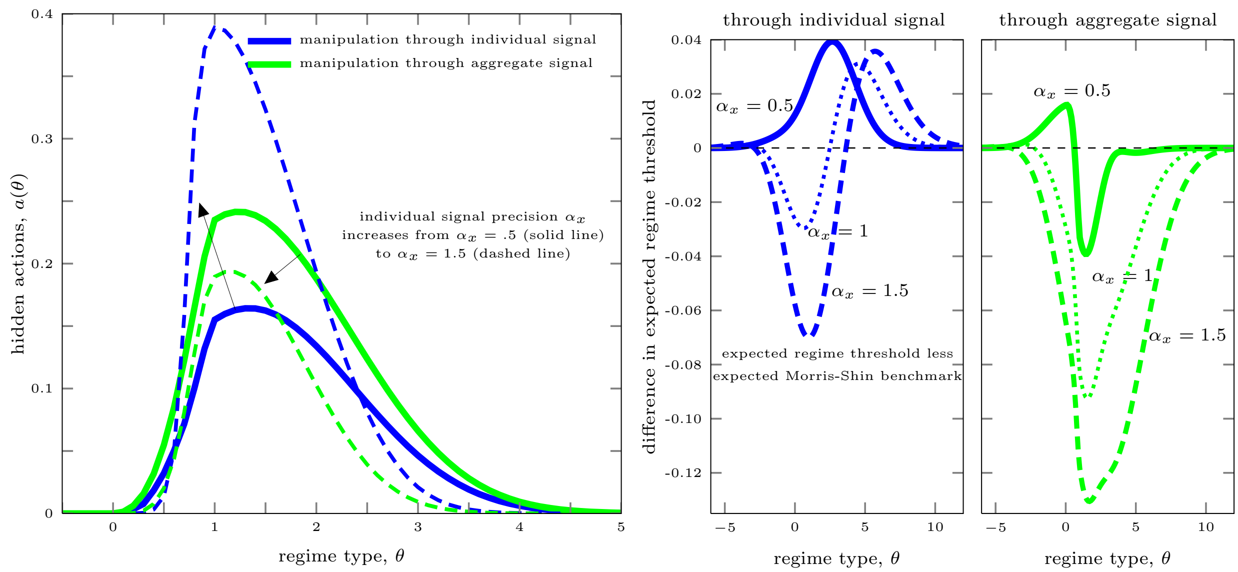
$$S(\theta, z) = \Phi(\sqrt{\alpha_x}(x^*(z) - \theta))$$

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

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

And again, the thresholds $\theta^*(z)$ and $x^*(z)$ are implicitly determined by indifference conditions for the regime and the citizens where the aggregate attack 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.[9] Several examples are shown in Figure 3. 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 expected thresholds $\mathbb{E}[\theta^*|\theta] - \mathbb{E}[\theta^*_{\text{MS}}|\theta]$

Figure 3: 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 sense, the results are qualitatively similar to those obtained for the main model without aggregate uncertainty. This specification of

[9] 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.

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.

II.4 Correlation vs. precision

In the main model a single parameter, α, determines both the precision of signals and, implicitly, the correlation of signals across citizens. To see this, suppose citizens have signals

$$x_i = y + \varepsilon_i$$

where the noise is independent of y and is IID with mean zero and precision α in the population of citizens. In order to make statements about the *unconditional* distribution of signals, I need to fix a prior.[10] For the purposes of this appendix, there is no loss of generality in setting the prior mean to zero. I set the prior precision to $\alpha_0 > 0$. Then the unconditional correlation of any pair of signals x_i and x_j is

$$\text{Corr}[x_i, x_j] = \frac{\text{Cov}[x_i, x_j]}{\text{Var}[x_i]} = \frac{\text{Var}[y]}{\text{Var}[y] + \text{Var}[\varepsilon_i]} = \frac{\alpha}{\alpha + \alpha_0}$$

Thus as $\alpha \to \infty$, the correlation $\to 1$ for any prior precision $\alpha_0 > 0$. This raises the question of whether the fall in the regime threshold θ^* as $\alpha \to \infty$ is driven by the increase in signal precision *per se* or by the implied increase in correlation.

Separating correlation from precision. To address this issue, I solve a version of the model where the signal precision can be held fixed while the signal correlation is varied. Specifically, I suppose citizens have signals

$$x_i = y + \xi_i$$

where the noise ξ_i has both a common and an idiosyncratic component

$$\xi_i = \sqrt{\rho} z + \sqrt{1-\rho}\varepsilon_i, \qquad 0 < \rho < 1 \tag{9}$$

The common component z is normal with mean zero and precision $\alpha > 0$ while the idiosyncratic component ε_i is independent of z and IID across citizens also with mean zero and *the same precision* α. As the correlation coefficient ρ approaches zero, this reduces to the main model of the text with uncorrelated noise $\xi_i = \varepsilon_i$, and hence signals $x_i = y + \varepsilon_i$. As ρ approaches one, the noise is perfectly correlated across citizens $\xi_i = z$, and hence signals are $x_i = y + z$ for all i. This corresponds to the model with perfect coordination, as in Appendix B of the main text.

[10] In the model I assumed an improper uniform prior on $\mathbb{R}$ for which the unconditional correlation is not defined.

For intermediate levels of ρ, the correlation of any pair of signals x_i and x_j is

$$\text{Corr}[x_i, x_j] = \frac{\text{Var}[y] + \rho\text{Var}[z]}{\text{Var}[y] + \text{Var}[\xi_i]}$$

The precision of the signal noise is

$$\text{Var}[\xi_i] = \rho\text{Var}[z] + (1-\rho)\text{Var}[\varepsilon_i] = \frac{1}{\alpha}$$

If we allowed the common and idiosyncratic components to have different precisions then any change in ρ would change the precision of the overall signal noise ξ_i so that we would be back to conflating the two issues. Given $\text{Var}[\xi_i] = \text{Var}[z] = 1/\alpha$, the correlation simplifies to

$$\text{Corr}[x_i, x_j] = \frac{\alpha + \rho\alpha_0}{\alpha + \alpha_0} \tag{10}$$

Again, as $\alpha \to \infty$ we will have the the correlation $\to 1$ for any prior precision $\alpha_0 > 0$. Highly precise information will always be correlated across citizens, but now we can fix a particular signal precision α and ask what happens as the correlation coefficient ρ changes.

Monotone equilibrium with latent correlation. Suppose the signal mean is $y = \theta + a$ so that citizens have signals $x_i = \theta + a + \xi_i$ where ξ_i is given by (9) above. Since the regime does not know which value of the common component z will realize, this version of the model has aggregate uncertainty. In the limit as the correlation coefficient $\rho \to 0$ this aggregate uncertainty disappears and we return to the main model.

I consider a monotone equilibrium where the regime is overthrown *ex post* for $\theta < \theta^*(z)$ and citizens attack, $s(x_i) = 1$, if their signals satisfy $x_i < x^*$. Here $x^* \in \mathbb{R}$ is a single threshold and $\theta^* : \mathbb{R} \to [0, 1]$ is a threshold *function*, both to be determined endogenously.

Regime problem. In this case, the density of citizen signals x_i conditional on the signal mean $y = \theta + a$ and on the common shock z is

$$f(x_i \,|\, y, z) := \sqrt{\frac{\alpha}{1-\rho}}\phi\left(\sqrt{\frac{\alpha}{1-\rho}}\,(x_i - y - \sqrt{\rho}z)\right)$$

where $\phi(\cdot)$ is the standard normal PDF. Thus, in a monotone equilibrium, if the realized common shock is z then ex post the aggregate attack facing a regime of type θ that takes action a is

$$S(\theta, a, z) := \int_{-\infty}^{\infty} s(x_i) f(x_i \,|\, \theta + a, z)\, dx_i = \Phi\left(\sqrt{\frac{\alpha}{1-\rho}}\,(x^* - \theta - a - \sqrt{\rho}z)\right)$$

where $\Phi(\cdot)$ is the standard normal CDF. A high realization of the common shock z will cause citizens to observe high values of x_i and this will reduce the size of the aggregate attack. The sensitivity to z depends on the correlation coefficient ρ and is larger the more correlation there is.

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

$$a(\theta) \in \underset{a \geq 0}{\operatorname{argmax}} \left[-C(a) + \int_{-\infty}^{\infty} \max[0, \theta - S(\theta, a, z)] \sqrt{\alpha}\phi(\sqrt{\alpha}z)\, dz \right] \tag{11}$$

Ex post the regime's threshold $\theta^*(z)$ is determined by the indifference condition

$$\theta^*(z) = S(\theta^*(z), a(\theta^*(z)), z) + C(a(\theta^*(z))), \qquad \text{for each } z \in \mathbb{R} \tag{12}$$

Both $a(\theta)$ and $\theta^*(z)$ implicitly depend on the citizen threshold x^*.

Citizen problem. Conditional on θ, the signals x_i are normally distributed with mean $y(\theta) = \theta + a(\theta)$ and precision α. Thus the citizens' posterior for θ is

$$\pi(\theta \mid x_i) = \frac{\sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - y(\theta)))}{\int_{-\infty}^{\infty} \sqrt{\alpha}\phi(\sqrt{\alpha}(x_i - y(\theta)))\, d\theta}$$

A citizen attacks, $s(x_i) = 1$, if and only if the probability of the regime being overthrown is at least p. If the common shock z was known, the posterior probability of the regime being overthrown would be

$$\int_{-\infty}^{\theta^*(z)} \pi(\theta \mid x_i)\, d\theta$$

but, because z is not known, citizens also have to integrate with respect to the density of z so that the indifference condition characterizing the citizen threshold x^* is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\theta^*(z)} \pi(\theta \mid x^*)\, d\theta \sqrt{\alpha}\phi(\sqrt{\alpha}z)\, dz = p \tag{13}$$

The correlation coefficient ρ matters for individual citizens' decisions through the hidden actions $a(\theta)$ and the regime thresholds $\theta^*(z)$, not in any direct fashion.

Numerical examples. I compute numerical solutions for this model with correlation by solving equations (11), (12) and (13) simultaneously for the functions $a(\theta)$ and $\theta^*(z)$ and the scalar x^*. Figure 4 shows the ex ante *expected* regime threshold as a function of the correlation coefficient ρ for three different levels of the signal precision.

In these examples the individual opportunity cost is $p = .25$ so that the Morris-Shin benchmark is $\theta^*_{\text{MS}} = 1 - p = .75$. As $\rho \to 0$ we have the main model with information manipulation and with correlation in signals only because of increasing precision. For these parameters the equilibrium threshold θ^* (which is deterministic when $\rho = 0$, since then there is no aggregate uncertainty) is less than the Morris-Shin benchmark and is decreasing in the signal precision α. Given that $p < .50$, this is exactly what we expect from Proposition 3 in the main text.

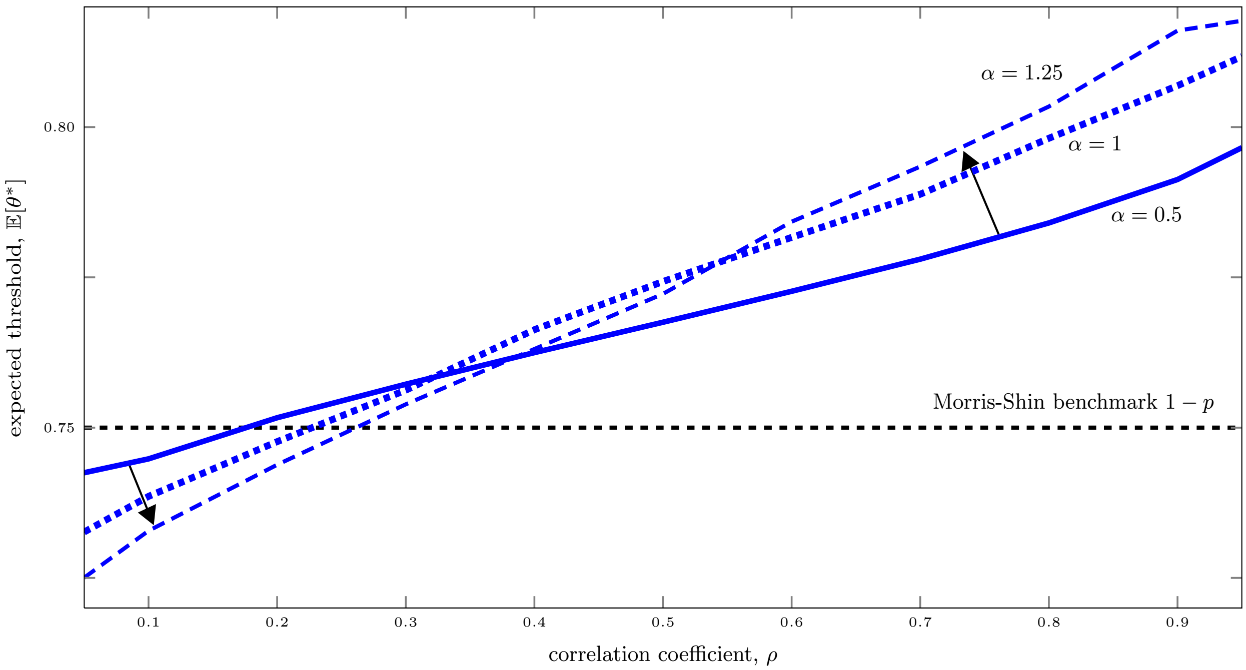


Figure 4: For fixed signal precision α, regime disadvantaged by higher correlation ρ.

The expected regime threshold $\mathbb{E}[\theta^*]$ as a function of the correlation coefficient ρ for various levels of the signal precision α. For fixed α, the expected regime threshold is increasing in ρ, suggesting that increasing correlation works against the regime (it is more likely to be overthrown). At low levels of ρ, an increase in α reduces the expected regime threshold, as in the main model. At high levels of ρ, an increase in α increases the expected regime threshold, as in the perfect coordination model in Appendix B in the main text. In all these examples, $p = .25$ and the cost function is $C(a) = a^2/2$.

Now fix a particular level of α. As the correlation coefficient ρ increases we have that the expected regime threshold *rises*. More underlying *latent* correlation for a given level of signal precision means that the regime expects to be worse off in the sense that it is more likely to be overthrown. As ρ increases the expected regime threshold is driven above the Morris-Shin benchmark $1 - p$ for a fixed level of α. The *total* correlation in signals depends on both the latent correlation in the noise governed by ρ and on the level of signal precision α. As α increases, the total correlation increases for any level of ρ, as in equation (10). Here we see that, for high levels of ρ, an increase in signal precision α continues to shift the expected regime threshold even higher.

These examples also agree with the analytic results in Appendix B in the main text where the regime is confronted by a single perfectly coordinated agent (i.e., $\rho = 1$) for some arbitrary level of signal precision α. In that perfect correlation case, as the signal precision becomes arbitrarily high, $\alpha \to \infty$, the regime threshold $\theta^* \to 1$ so that all the fragile regimes $\theta \in [0, 1)$ are overthrown. This is the exact opposite of the result with uncorrelated signals (i.e., $\rho = 0$) where $\alpha \to \infty$ implies the regime threshold $\theta^* \to 0$ so that all the fragile regimes $\theta \in [0, 1)$ survive. Overall, this suggests that it is the increase in signal precision rather than the implied increase in total correlation that is responsible for the regime threshold being driven to zero in the main model. Correlation per se is not in a regime's interest. At any given level of signal precision, the regime would prefer it if its opponents were less coordinated.

References

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

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

Baliga, Sandeep and Tomas Sjöström, "The Strategy of Manipulating Conflict," *American Economic Review*, 2012, *102* (6), 2897–2922.

Baron, David P., "Perisistent Media Bias," *Journal of Public Economics*, 2006, *90* (1-2), 1–36.

Berman, Eli and David D. Laitin, "Religion, Terrorism and Public Goods: Testing the Club Model," *Journal of Public Economics*, 2008, *92*, 1942–1967.

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

Bose, Subir, Gerhard Orosel, Marco Ottaviani, and Lise Vesterlund, "Dynamic Monopoly Pricing and Herding," *RAND Journal of Economics*, 2006, *37* (4), 910–928.

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

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

Debs, Alexandre, "Divide-and-Rule and the Media," 2007. University of Rochester working paper.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin, "Media Freedom, Bureaucratic Incentives, and the Resource Curse," 2006. Harvard University working paper.

Fearon, James D., "Self-Enforcing Democracy," 2006. Stanford University working paper.

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

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

Gentzkow, Matthew and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

Ginkel, John and Alastair Smith, "So You Say You Want a Revolution: A Game-Theoretic Explanation of Revolution in Repressive Regimes," *Journal of Conflict Resolution*, 1999, *43* (3), 291–316.

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

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

Kalyvas, Stathis N., "How Free is 'Free Riding' in Civil Wars? Violence, Insurgency, and the Collective Action Problem," *World Politics*, 2007, *59* (2), 1043–1068.

Karklins, Rasma and Roger Petersen, "Decision Calculus of Protesters and Regimes: Eastern Europe 1989," *Journal of Politics*, 1993, *55* (3), 588–614.

Kuran, Timur, "Sparks and Prairie Fires: A theory of Unanticipated Political Revolution," *Public Choice*, 1989, *61*, 41–74.

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

_ , *Private Truths, Public Lies: The Social Consequences of Preference Falsification*, Harvard University Press, 1995.

Lohmann, Susanne, "A Signaling Model of Informative and Manipulative Political Action," *American Political Science Review*, 1993, *87* (2), 319–333.

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

_ , "Information Aggregation Through Costly Political Action," *American Economic Review*, 1994, *84* (3), 518–530.

Metz, Christina, "Private and Public Information in Self-Fulfilling Currency Crises," *Zeitschrift fur Nationalokonomie*, 2002, *76* (1), 65–85.

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

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

_ **and** _ , "Coordination Risk and the Price of Debt," *European Economic Review*, 2004, *48* (1), 133–153.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

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

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

Tullock, Gordon, "The Paradox of Revolution," *Public Choice*, 1971, *11* (Fall), 89–100.

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