Transparency and Stability:

On Hollyer, Rosendorff, and Vreeland's Models of Economic Transparency and Political Stability

Mehdi Shadmehr[1] Dan Bernhardt[2]

[1]Harris School of Public Policy, University of Chicago, and Department of Economics, University of Calgary. E-mail: mshadmeh@gmail.com

[2]Department of Economics, University of Illinois, and Department of Economics, University of Warwick. E-mail: danber@illinois.edu

Abstract

This paper revisits the theoretical foundations that Hollyer, Rosendorff, and Vreeland use in their research program on transparency and political stability. We show: (1) the payoff structure of their core model is substantively unsound; (2) they misidentify the mechanism driving their main result; (3) they get the model's implied direction of causality backwards, from more transparency to more instability. Their model implies that more transparency increases the likelihood of revolution if this likelihood is sufficiently small, but reduces the likelihood of revolution if it is sufficiently large. This reflects the well-known logic of "gambling for resurrection": when you're ahead, don't give information, but when you're behind, gamble for resurrection by providing more information. Their model suggests that protest risk drives transparency: regimes facing a low likelihood of revolution reduce transparency, while those facing a high likelihood of revolution raise transparency, generating a positive correlation between transparency and instability.

Keywords: Transparency, Stability, Protest, Revolution

In a series of influential papers, Hollyer, Rosendorff, and Vreeland explore the relationship between transparency and political stability. These papers build on their 2015 APSR paper, "Transparency, Protest and Autocratic Instability", and have been collected and expanded into a book, *Information, Democracy and Autocracy: Economic Transparency and Political (In)Stability* (Hollyer et al. 2018a), by Cambridge University Press. The authors develop a model to argue autocratic regimes that increase the transparency of their economic data face a higher probability of protest and regime change. Their notion of transparency is given by the variance of noise in economic data, which they empirically measure based on how much economic data are not reported. Our paper revisits the theoretical foundations of their model, highlighting issues with the interpretations and conclusions drawn.

We identify three broad issues. First, the payoff structure of their core model is substantively unsound. We show that neither the public aggregate economic situation nor a citizen's private economic well-being affects a citizen's net payoff from revolting. Thus, economic interest, either self-interest or sociotropic interest, is not itself an incentive for individual action. Instead, public economic data are *assumed* to be focal points for coordination, making them indistinguishable from any other public data. This payoff structure results in a continuum of equilibria, of which Hollyer et al. select one. In this equilibrium, for some range of public economic data, citizens' actions depend on their private economic well-being, while for other ranges of public economic data, private economic well-being does not influence behavior. We show that other, simpler, equilibria include a class in which all citizens revolt only when public data are bad, and a class in which all citizens revolt only when public data are good.

This issue is starkly displayed in Hollyer et al. (2018b), which extends their model of protest to analyze transparency and stability in democracies by adding a voting stage before the protesting stage. Citizens vote sincerely based on the incumbent's economic performance. Voting perfectly reveals whether the incumbent is good or bad, but the same continuum of equilibria exists at the protest stage regardless of his revealed type. To generate reasonable results, Hollyer et al. *assume* that all citizens protest and remove an incumbent who wins if and only if the election reveals that he is bad. However, there is an otherwise identical equilibrium in which citizens only remove good incumbents.

Second, Hollyer et al. misidentify the key underlying mechanism that drives their results.

Their key empirical prediction is that, when revolution is sufficiently unlikely, more precise public data increases the likelihood of regime change. They fail to recognize that when revolution is sufficiently likely, more precise public data reduces the likelihood of regime change—instead, they assert that the relationship between transparency and stability is ambiguous in this case. Critically, they incorrectly attribute these results to coordination incentives or to differences between public and private signals. Underscoring this mis-attribution, we show that the same predictions obtain in a representative citizen setting—having multiple citizens is unnecessary. The true causal mechanism is gambling for resurrection (Downs and Rocke 1994): when you're ahead, don't give any information; but when you're behind, "gamble for resurrection" and provide more information. In the context of regime change, this logic implies that stable regimes are safe as long as accurate bad information does not arrive, so they should send noisy, uninformative signals—if it's not broken, don't fix it. In contrast, unstable regimes likely collapse unless accurate good information changes citizens' behavior, so they should gamble to resurrect. This result appears in many settings, including revolutions and media freedom (Shadmehr and Bernhardt 2011, 2015; Edmond 2013; Gehlbach and Sonin 2014), electoral competition (Gul and Pesendorfer 2012), grading standards (Boleslavsky and Cotton 2015), and Bayesian persuasion models (Kamenica and Gentzkow 2011).

Third, confusion over this mechanism results in a mistaken theoretical attribution of causality between transparency and threat of revolution. In their model, the risk of protest drives transparency not, as they claim, the reverse. When the risk of revolution is high, regimes optimally provide informative economic data, "gambling for resurrection", implying a positive empirical correlation between protest and economic transparency. Hollyer et al. ignore the fact that transparency is an endogenous function of protest risk. They argue that (i) the relevant parameter region is where the prior likelihood of revolution is low, thereby excluding the logic of "gambling for resurrection," and (ii) regimes use high transparency for some exogenous reason—ignoring that their model's logic says that regimes facing low risk should reduce transparency—thereby reducing their own stability. To resolve the dilemma embedded in (ii), Hollyer et al. (2019) develop a different model in which higher transparency raises the likelihood of regime change, but this, in turn, reduces the probability of a coup because, *by assumption*, greater transparency subjects coup leaders to even greater risks of revolt. They argue that regimes use higher transparency to destabilize themselves and thereby discourage coups. This convoluted combination of assumptions and models lets them reconcile the positive correlation between transparency and political stability. We observe that the logic of "gambling for resurrection" and accounting for prior protest risk already does this.

One can also approach their model's empirical implication from a different angle. Hollyer et al.'s *implicit* argument is that regimes can *commit* to a level of transparency, for example, ex-ante stable regimes can commit to low transparency. Then, if the likelihood of regime change rose, a regime cannot increase transparency—due to commitment—even if later adverse events would make it want to gamble for resurrection. Assuming such commitment is strong, especially given their measure of transparency, which is based on how much economic data is reported and how much is missing (Hollyer et al. 2014, 2018a). When regimes control reporting agencies, they can order agencies to report more economic data when it is in their interest. This brings us to a mismatch between their measurement of transparency and its conceptualization in the model. Their measurement is based on data disclosure and missing economic data,[1] but their formalization is based on noise variance in the data. Models of censoring public data (Shadmehr and Bernhardt 2015) are very different from models of making public data noisier (Edmond 2013): a regime censors by not reporting bad news, which is not the same as adding noise to the original distribution, as in their formulation.

1 Transparency and Stability

Consider a simple model with a single representative citizen who must decide whether to revolt against a ruler whose type is unknown to the citizen. Revolting costs $k > 0$. If the citizen revolts, the revolution succeeds with probability $p \in (0, 1)$. Some rulers are better than others. The ruler's type is given by $\theta \in \mathbb{R}$: higher θs are better rulers. The citizen receives a net payoff of 1 from revolting and removing a ruler whose type is below a threshold, $\theta \leq T$, and he receives a net payoff of 0 from removing better rulers with types $\theta > T$. We assume $k/p \in (0, 1)$, so that revolting is sometimes optimal. Let $a = 1$ indicate that the citizen revolts and

[1] "Our empirical measure of this concept is a function of the missingness/nonmissingness of data from the WDI [World Development Indicators from the World Bank]" (Hollyer et al. 2014, p. 426).

$a = 0$ indicate that he does not revolt. The citizen's payoff from action $a \in \{0, 1\}$ in state θ is:

$$u(a, \theta) = a \cdot (p \cdot \mathbf{1}_{\{\theta \leq T\}} - k),$$

where $\mathbf{1}_{\{.\}}$ is the indicator function. The ruler's type is uncertain, with $\theta \sim N(0, \rho)$. The citizen does not see θ, but he sees a signal $y = \theta + \nu$, where $\nu \sim N(0, \tau)$ is distributed independently. We derive when, from an ex-ante perspective, lower noise (higher transparency) increases the likelihood of revolution.

Given signal y, the citizen wants to revolt if and only if $Pr(\theta \leq T|y) > k/p$, i.e.,

$$\Phi\left(\frac{T - \frac{\rho^2}{\rho^2+\tau^2} y}{\sqrt{\frac{\rho^2\tau^2}{\rho^2+\tau^2}}}\right) > k/p.$$

A lower signal y means a ruler is more likely to have a lower type. Thus, the citizen revolts if and only if:

$$y < y^* \equiv \frac{\rho^2 + \tau^2}{\rho^2} \left(T - \sqrt{\frac{\rho^2\tau^2}{\rho^2 + \tau^2}}\ \Phi^{-1}(k/p)\right).$$

This implies that ex-ante likelihood of regime change is:

$$P(\tau) \equiv Pr(y < y^*) = \Phi\left(\frac{y^*}{\sqrt{\rho^2 + \tau^2}}\right) = \Phi\left(\frac{\sqrt{1 + (\tau/\rho)^2}\ T - \tau\ \Phi^{-1}(k/p)}{\rho}\right).$$

Differentiating with respect to τ yields:

$$\frac{\partial P}{\partial \tau} > 0 \text{ if } \frac{\tau/\rho^2}{\sqrt{1 + (\tau/\rho)^2}} \cdot T > \Phi^{-1}(k/p), \quad \text{but} \quad \frac{\partial P}{\partial \tau} < 0 \text{ if } \frac{\tau/\rho^2}{\sqrt{1 + (\tau/\rho)^2}} \cdot T < \Phi^{-1}(k/p).$$

Thus, when the normalized costs of revolt k/p is high or the threshold T is low, so that revolution is unlikely, lower noise (higher transparency), raises the likelihood of revolution. In contrast when k/p is low or the threshold T is high, so that revolution is likely, lower noise reduces the likelihood of revolution. Summarizing,

Proposition 1 *In this representative citizen game, more transparency increases the likelihood of regime change if this likelihood is below a threshold, but reduces the likelihood of regime change if this likelihood is above a threshold.*

Proposition 1 states that transparency in the sense of lower noise variance hurts stable regimes but helps unstable regimes. The intuition is simple: when you're ahead, don't give information, but when you're behind, gamble for resurrection. To see this, suppose the citizen does not receive any signal y, or equivalently, the variance τ is almost infinite so the signal y is almost pure noise and hence irrelevant. Now, the citizen acts only based on his prior, revolting if and only if $Pr(\theta \leq T) > k/p$. As a result, if T is small or k/p is large, the citizen never revolts. Therefore, the ruler would not want to give the citizen any information—greater transparency can only increase the risk of revolution. In contrast, if T is large or k/p is small, the citizen always revolts. Now, absent additional information, the citizen is sure to revolt, so giving information can only help the ruler—transparency can only reduce the risk of revolution. This logic does not hinge on coordination, different effects of public or private signals, higher order beliefs, or the methodology of global games, beauty contests, or sun spot games. Because it is so basic, it holds in this simple single representative citizen model.

This logic suggests an empirical implication: the likelihood of protest and transparency should be positively correlated. To illustrate, suppose we begin from a setting where all regimes are equally transparent, but some face a high likelihood of protest while others face a low risk, for example due to differences in T or k/p. Now, regimes with a high likelihood of protest should raise transparency, and those with a low likelihood should reduce it. Let $(\text{protest risk}, \text{transparency})$ denote a combination of the likelihood of protest and transparency. The logic says that if, for example, we begin with $\{(\text{low}, \text{medium}), (\text{high}, \text{medium})\}$, we may end up with $\{(\text{very low}, \text{low}), (\text{medium}, \text{high})\}$. That is, the model suggests strategic decisions by regimes generate a positive correlation between protest and transparency, with higher protest risk causing more transparency. We do not take a stand about the plausibility of this mechanism; our goal is only to deduce the model's empirical implication.

2 The Hollyer et al. Model

In Hollyer et al.'s core model a continuum of citizens decide whether to revolt. The revolution succeeds if and only if the fraction of revolters, l, exceeds a threshold $T \in (0, 1)$. There is a binary state of world $\theta \in \{0, 1\}$. Citizens share a prior that $Pr(\theta = 1) = p$. Citizens observe a public signal, $y = \theta + \nu$. Each citizen i also sees a private signal, $x_i = \theta + \epsilon_i$, where $\nu \sim N(0, \tau)$ and $\epsilon_i \sim N(0, \sigma)$, and θ, ν and ϵ_is are independently distributed. Figure 1 presents citizen payoffs, where $k > 0$ is the cost of revolt, and $\beta > k$ is the benefit of participating in a successful revolt.

		outcome	
		$l \geq T$	$l < T$
Citizen i	Protest	$E[\theta] + \beta - k$	$E[\theta\|x_i, y] - k$
	No Protest	$E[\theta]$	$E[\theta\|x_i, y]$

Figure 1: Hollyer et al.'s Citizen Game

Hollyer et al. interpret the public signal y as public economic data, x_is as citizens' income, and θ as the ruler's type. After a regime change, a random ruler is selected, whose type θ becomes the average income. Because citizens lack information about the new ruler other than their prior, their expectation of his type is $E[\theta]$. Without regime change, the current ruler remains in power, whose type θ, is again the average income. Given public signal y and private signal x_i, citizen i's expectation about the current ruler's type is $E[\theta|x_i, y]$.

Their model has an additional stage, in which the ruler moves before citizens, deciding whether to provide a public good. If he provides the public good, average income is $g > 0$; if he does not provide it, average income is 0. A type θ ruler's payoff is $\mathbf{1}_{\{provide\ public\ good\}} \cdot \theta$. Thus, a $\theta = 1$ ruler provides the public good, and a $\theta = 0$ ruler does not. We set $g = 1$ to save on notation without losing content, so a ruler's type equals average income.

The first key point to glean from citizen payoffs in Figure 1 is that $E[\theta]$ and $E[\theta|x_i, y]$ are irrelevant for individual decision making. To see this, observe that citizens base decisions to protest on comparisons of net payoffs. When at least measure T revolt, the net payoff from protest is $(E[\theta] + \beta - k) - E[\theta] = \beta - k$, and when fewer than measure T revolt, the net payoff from protest is $(E[\theta|x_i, y] - k) - E[\theta|x_i, y] = -k$. Thus, strategic behavior and the set of equilibria are exactly the same as when citizen payoffs are given by Figure 2 below, in which all parameters, including the threshold T, are known.

		outcome	
		$l \geq T$	$l < T$
Citizen i	Protest	$\beta - k$	$-k$
	No Protest	0	0

Figure 2: Hollyer et al.'s Citizen Game

From a substantive perspective, this payoff structure does not make sense: A citizen's incentive to revolt does not depend on whether the ruler is good or bad (θ), the citizen's private economic well-being (x_i), or the aggregate economic data y. Hollyer et al. (2015) claim that "incentives to engage in unrest...are highest when...economic performance is poor" (p. 766). However, in their model, a poor citizen does not revolt because he has lower *opportunity costs* of revolt, or because he is more frustrated with the status quo (*expressive motives*), or because he believes a new ruler is likely to be better, and wants to participate in a movement that replaces a bad ruler with a better one (*pleasure-in-agency motives*). Rather, a citizen revolts only because he somehow believes others are likely to revolt. That is, the Hollyer et al. model is a sun-spot coordination game, with economic data playing the role of sun spots, which act as focal points for coordination. This is problematic: in their theory, economic well-being is a key citizen concern, so aspirations to improve own economic well-being or that of the country should underlie decisions to revolt. Then, of course, a citizen must estimate the chances of success and take it into account. However, in the model, how a citizen acts is just based on an *exogenous social norm*.

An immediate implication is that there is a continuum of equilibria. For example, it is an equilibrium when citizens revolt if and only if aggregate economic data are neither too good nor too bad, e.g., if and only if $y \in [-1, 1]$. To see this, note that if everyone revolts, the revolution succeeds, in which case a citizen who revolted receives $\beta - k > 0$ by assumption. Thus, if a citizen believes that all others revolt after seeing economic data $y \in [-1, 1]$, then he revolts. Beliefs and equilibrium strategies are consistent, so this is an equilibrium. There is also an equilibrium in which citizens revolt if and only if the second digit of the aggregate economic data is odd: beliefs are so unrestricted that even implausible outcomes are not precluded.

From a methodological perspective, this is a sun-spot game (Cass and Shell 1983), in which the public signal acts as a coordinating device. Given any public signal, there are two equilibria, one in which everyone revolts and the regime changes, and one in which no one revolts and the regime survives: nothing links survival to fundamentals, so absent an arbitrary equilibrium selection criteria, there are no empirical restrictions. Because uncertainty, whether public or private, has nothing to do with citizens' payoffs, this game has little to do with global games or beauty contests (Morris and Shin 2002, 2003) or their applications (Bueno de Mequita 2010; Boix and Svolik 2013; Casper and Tyson 2014; Rundlett and Svolik 2016; Tyson and Smith 2018), save for being a subset of games with strategic complementarity. Nonetheless, Hollyer et al. (2015) relate their model to global games, "We depart from global games literature in a technical assumption: Classical formulations of global games exhibit the property of two-sided limit dominance" (p. 768). Such statements mislead by focusing on secondary issues and technical terms (e.g., limit dominance), while missing fundamental issues.

2.1 Analysis

We next reproduce the equilibrium and analysis of Hollyer et al. (2015, 2018a). Our simpler presentation and proofs do away with redundant algebra and conditions (e.g., $\sigma < \tau$) used in their proofs, while highlighting undesirable properties of that equilibrium. The properties include discontinuities in the equilibrium thresholds set by citizens as a function of the public signal, and the insensitivity of citizens' equilibrium behavior to model parameters.

Hollyer et al. focus on a particular equilibrium selection, a "responsive equilibrium" (responsive to private signals) in which given the public signal, citizens with sufficiently low private signals revolt, while those with higher signals do not. That is, for some realizations of the public signal y, a citizen i revolts whenever his signal is below a threshold $x^*(y) \in \mathbb{R}$. A responsive equilibrium exists whenever the public signal falls between two thresholds ($\underline{y} < y < \overline{y}$), neither too low, nor too high. To close equilibrium selection for other public signals, Hollyer et al. *assume* that when the public signal is below the lower threshold ($y < \underline{y}$), everyone revolts, and when it exceeds the upper threshold ($y > \overline{y}$), no one revolts. A proof is in Online Appendix A.

Proposition 2 ***(Hollyer et al. 2015, 2018a)*** *Let $s(x_i, y)$ be the strategy of a citizen who observes private signal x_i and public signal y, where $s = 1$ indicates revolt, and $s = 0$ indicates no revolt. Equilibrium is described by three thresholds $(x^*(y), \underline{y}, \overline{y})$:*

$$s(x_i, y) = \begin{cases} 1 & ; y \leq \underline{y} \\ 1 & ; \underline{y} < y \leq \overline{y} \text{ and } x_i \leq x^*(y) \\ 0 & ; \underline{y} < y \leq \overline{y} \text{ and } x_i > x^*(y) \\ 0 & ; \overline{y} < y \end{cases} \quad \text{where } Pr(\theta = 0 | x_i = x^*(y), y) = \frac{k}{\beta},$$

and $\underline{y} < \overline{y}$ solve:

$$Pr(\theta = 0 | x_i = 1 + \sigma\Phi^{-1}(T), \underline{y}) = Pr(\theta = 0 | x_i = \sigma\Phi^{-1}(T), \overline{y}) = \frac{k}{\beta}. \tag{1}$$

Regime change occurs if $y \leq \underline{y}$, but not if $y > \overline{y}$, and if $y \in (\underline{y}, \overline{y}]$, regime change occurs if and only if $\theta = 0$.

When $y < \underline{y}$, Hollyer et al. choose an equilibrium in which citizens always revolt. This amounts to setting $x^*(y) = \infty$ for $y < \underline{y}$. Similarly, they choose an equilibrium for $y > \overline{y}$ that corresponds to setting $x^*(y) = -\infty$ for $y > \overline{y}$. Because $x^*(y)$ remains finite when $y \in (\underline{y}, \overline{y})$, this implies that $x^*(y)$ is discontinuous at the thresholds $\underline{y}$ and $\overline{y}$, reflecting that the behavior of citizens inside the boundary $(\underline{y}, \overline{y})$ is independent of their assumed behavior outside of it.

From Proposition 2, the ex-ante likelihood of regime change given public signal y is:

$$Pr(\text{regime change}|y) = \begin{cases} 1 & ; y \leq \underline{y} \\ Pr(\theta = 0|y) & ; \underline{y} < y \leq \overline{y} \\ 0 & ; \overline{y} < y. \end{cases}$$

Just as $x^*(y)$ was discontinuous at the boundaries $\underline{y}$ and $\overline{y}$, so is $Pr(\text{regime change}|y)$. For almost all public signal realizations, citizens do not respond to marginal variations in the environment:

$$\frac{\partial Pr(\text{regime change}|y)}{\partial(\text{any model parameter})} = 0, \ \forall y \notin \{\underline{y}, \overline{y}\}.$$

Similarly, for any $y \in (\underline{y}, \overline{y})$, a citizen's behavior does not change if he believes others are marginally more or less likely to revolt. This is because, when $y \in (\underline{y}, \overline{y})$, by construction of the equilibrium, the citizen only cares about estimating the likelihood that the state is $\theta = 0$. Hollyer et al. analyze the effect of variations in the variance of public signal noise τ on the ex-ante likelihood of revolt. Within any region ($y < \underline{y}$, $\underline{y} < y < \overline{y}$, $\overline{y} < y$), a citizen's behavior and likelihood of regime change for the equilibrium of that region do *not* change with variations in model parameters, including τ. However, variations do change the unconditional ex-ante likelihood of regime change by altering the thresholds $\underline{y}$ and $\overline{y}$ that separate regions with different selected equilibrium. Here we present the result that covers both the case where revolution is unlikely and where it is likely. A proof is in Online Appendix A.

Corollary 1 *More transparency* <u>*increases*</u> *the likelihood of regime change if this likelihood is below a threshold, but* <u>*reduces*</u> *the likelihood of regime change if the likelihood exceeds a threshold.*

The intuition is the same as Proposition 1: when you're ahead, don't give information, but when you're behind, gamble for resurrection by giving more information. Hollyer et al.'s (2015, Proposition 3) focus on the first part of Corollary 1, ignoring the second part that reflects the logic of gambling for resurrection. They argue that "since incidences of successful mass protest are relatively rare," only the first part is relevant "in the vast majority of cases" (p. 772). This implies that regimes that choose greater transparency face a higher likelihood of revolution, generating a positive correlation between protest and transparency, which they find in their empirical analysis. However, focusing only on the first part of Corollary 1 creates a dilemma for the theory: if higher transparency causes instability, why don't regimes *reduce* transparency?

"Why Do Autocrats Disclose?" (Hollyer et al. 2019) addresses this. They argue that autocrats increase transparency (disclose) to reduce the likelihood of coups. To get this result, they *assume* (1) higher transparency increases the likelihood of protest, but (2) it increases this likelihood even more after coups. Thus, higher transparency hurts an autocrat, but it hurts those who coup even more. Some regimes then choose higher transparency than others for reasons that are exogenous to the model of Hollyer et al.'s (2015), i.e., due to the exogenous probability of coups in Hollyer et al. (2019). This juxtaposition delivers the direction of causality from higher transparency to higher protest, justifying the regression of measures of stability on transparency in Hollyer et al. (2015, 2018a). However, as we showed in the previous section, accounting for the logic of gambling for resurrection (which they exclude) can naturally generate the positive correlation between transparency and instability in their initial framework. This logic does not hinge on coordination, or the difference between public and private signals.

3 Conclusion

We revisit the theoretical foundations of Hollyer et al.'s research program on transparency and stability, focusing primarily on their core model (2015, 2018a, 2018b), first published in *APSR*. We show that the payoff structure of their model is substantively unsound. In particular, none of the key variables that they emphasize (income, aggregate economic conditions, ruler quality) affect a citizen's net payoff from revolting. A symptom of this flawed payoff structure is the presence of a continuum of equilibria, of which they select one.

Their model has two implications for the relationship between transparency and stability: (1) when revolution is unlikely, more precise public data *increases* the likelihood of regime change, but (2) when revolution is likely, more precise public data *decreases* the likelihood of regime change. They misidentify the mechanism deriving this result, mis-attributing it to coordination incentives or to differences between public and private signals. In fact, the mechanism reflects the known logic of "gambling for resurrection", which holds in a single agent setting: when you're ahead, don't give information (underlying part 1), but when you're behind, gamble for resurrection by providing more information (underlying part 2). They argue that only the first part is relevant, excluding the logic of gambling for resurrection.

This confusion causes them to get causality backwards. In their model, protest risk drives transparency: regimes facing a low likelihood of regime change reduce transparency, while those facing a high likelihood of regime change raise transparency. This simple logic can generate the positive correlation between transparency and instability found in the data. However, they exclude the logic of gambling for resurrection, interpreting the model's empirical implication as more transparency causing instability. This creates an artificial dilemma:

why do regimes choose high transparency if it harms them? To resolve this, they develop a separate model (2019) in which rulers raise transparency to *cause* instability, because instability is assumed to reduce the risk of coup.

4 References

Boix, Carles, and Milan Svolik. 2013. "The Foundations of Limited Authoritarian Government: Institutions and Power-sharing in Dictatorships." *Journal of Politics* 75 (2): 300-16.

Boleslavsky, Raphael, and Christopher Cotton. 2015. "Grading Standards and Education Quality." *American Economic Journal: Microeconomics* 7 (2): 248-79.

Bueno de Mesquita, Ethan. 2010. "Regime Change and Revolutionary Entrepreneurs." *American Political Science Review* 104 (3): 446-66.

Casper, Brett, and Scott Tyson. 2014. "Popular Protest and Elite Coordination in a Coup Detat." *Journal of Politics* 76 (2): 548-64.

Cass, David, and Karl Shell. 1983. "Do Sunspots Matter?" *Journal of Political Economy* 91 (2): 193-227.

Downs, George, and David Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War." *American Journal of Political Science* 38 (2): 362-80.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies* 80 (4): 1422-58.

Gehlbach, Scott, and Konstantin Sonin. 2014. "Government Control of the Media." *Journal of Public Economics* 118: 163-71.

Gul, Faruk, and Wolfgang Pesendorfer. 2012. "The War of Information." *Review of Economic Studies* 79: 707-34.

Hollyer, James, Peter Rosendorff, and James Vreeland. 2014. "Measuring Transparency." *Political Analysis* 22 (4): 413-34.

Hollyer, James, Peter Rosendorff, and James Vreeland. 2015. "Transparency, Protest, and Autocratic Instability." *American Political Science Review* 109 (4): 764-84.

Hollyer, James, Peter Rosendorff, and James Vreeland. 2018a. *Information, Democracy and Autocracy: Economic Transparency and Political (In)Stability.* New York: Cambridge University Press.

Hollyer, James, Peter Rosendorff, and James Vreeland. 2018b. *Transparency, Protest and Democratic Stability. British Journal of Political Science. Forthcoming.*

Hollyer, James, Peter Rosendorff, and James Vreeland. 2019. "Why Do Autocrats Disclose? Economic Transparency and Inter-Elite Politics in the Shadow of Mass Unrest." *Journal of Conflict Resolution* 63(6): 1488-1516.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590-2615.

Morris, Stephen, and Hyun Song Shin. 2002. "The Social Value of Public Information." *American Economic Review* 92 (5): 1521-34.

Morris, Stephen, and Hyun Song Shin. 2003. "Global Games: Theory and Application." In *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Vol. I.* Ed. Dewatripont, Hansen, and Turnovsky. New York: Cambridge University Press.

Rundlett, Ashlea, and Milan Svolik. 2016. "Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud." *American Political Science Review* 110 (1): 180-97.

Shadmehr, Mehdi, and Dan Bernhardt. 2011. "Collective Action with Uncertain Payoffs: Coordination, Public Signals and Punishment Dilemmas." *American Political Science Review* 105 (4): 829-51.

Shadmehr, Mehdi, and Dan Bernhardt. 2015. "State Censorship." *American Economic Journal: Microeconomics* 7 (2): 280-307.

Smith, Alastair, and Scott Tyson. 2018. "Dual-Layered Coordination and Political Instability." *Journal of Politics* 80 (1): 44-58.

Online Appendix

A Proofs

Proof of Proposition 2: We seek equilibria in which for some realizations of the public signal y, a citizen i revolts whenever his signal is below a threshold $x^*(y) \in \mathbb{R}$. For such realizations, the fraction of revolters $l(\theta)$ is:

$$l(\theta; x^*(y)) = \begin{cases} \Phi\left(\frac{x^*(y)}{\sigma}\right) & ; \theta = 0 \\ \Phi\left(\frac{x^*(y)-1}{\sigma}\right) & ; \theta = 1. \end{cases}$$

Clearly, $l(1; x^*(y)) < l(0; x^*(y))$. Thus, if $l(0; x^*(y)) < T$, then the revolution surely fails, so no one revolts, and hence there is no finite-threshold equilibrium. Similarly, if $l(1; x^*(y)) \geq T$, then the revolution surely succeeds, and hence everyone revolts, and, again, there is no finite-threshold equilibrium. Thus, a necessary condition for a finite-threshold equilibrium is $l(1; x^*(y)) < T \leq l(0; x^*(y))$. In this equilibrium, the regime collapses if and only if $\theta = 0$. Thus, given public signal y, a citizen i with private signal x_i revolts if and only if $Pr(\theta = 0 | x_i, y) > \frac{k}{\beta}$; the marginal citizen with threshold signal $x_i = x^*(y)$ must be indifferent between revolting and not. In sum, given a public signal y, the threshold $x^*(y)$ describes an equilibrium if and only if:

$$\Phi\left(\frac{x^*(y)-1}{\sigma}\right) < T \leq \Phi\left(\frac{x^*(y)}{\sigma}\right) \quad \text{(belief consistency)}$$
$$Pr(\theta = 0 | x_i = x^*(y), y) = \frac{k}{\beta} \quad \text{(indifference condition)}$$

These equations can be satisfied by an x^* if and only if the public signal is neither too large, nor too small. To see this, suppose the public signal is very low. Then, indifference requires that $x^*(y)$ be very large: when the public signal is very low, a citizen believes that $\theta = 0$ is very likely, so he almost always revolts. But if everyone almost always revolts, the revolution succeeds even when $\theta = 1$, contradicting the equilibrium premise that the revolution succeeds if and only if $\theta = 0$. In particular, the public signal cannot be so low that $T \leq \Phi\left(\frac{x^*(y)-1}{\sigma}\right)$.

The public signal at which this happens makes $x^*(y)$ just large enough to hit $\Phi\left(\frac{x^*(y)-1}{\sigma}\right) = T$, i.e., $x^*(y) = 1 + \sigma\Phi^{-1}(T)$. From the indifference condition, it is the public signal y that satisfies $Pr(\theta = 0|x_i = 1 + \sigma\Phi^{-1}(T), y) = \frac{k}{\beta}$. Letting $\underline{y}$ be this unique solution, if $y \leq \underline{y}$, then the indifference condition and belief consistency cannot simultaneously be satisfied. After any given $y \leq \underline{y}$, there is an equilibrium in which players revolt, and one in which they do not revolt. Similarly, letting $\overline{y}$ be the unique public signal realization that satisfies $Pr(\theta = 0|x_i = \sigma\Phi^{-1}(T), y) = \frac{k}{\beta}$, if $y > \overline{y}$, the indifference condition and belief consistency cannot simultaneously be satisfied. After any given $y > \overline{y}$, multiple equilibria again exist.

To deal with this multiplicity of equilibria, Hollyer et al. assume that citizens revolt after any given $y \leq \underline{y}$ and do not revolt after any given $y > \overline{y}$. □

Proof of Corollary 1: Given state $\theta = 0$, the regime changes whenever $y \leq \overline{y}$; and given state $\theta = 1$, the regime changes whenever $y \leq \underline{y}$, yielding the unconditional ex-ante likelihood of regime change:

$$Pr(\text{regime change}) = (1-p)\ \Phi(\overline{y}/\tau) + p\Phi((\underline{y}-1)/\tau),$$

where $1 - p$ is the prior probability that $\theta = 0$. Differentiating with respect to τ yields:

$$\frac{\partial Pr(\text{regime change})}{\partial \tau} = (1-p)\ \phi\left(\frac{\overline{y}}{\tau}\right)\frac{\partial}{\partial \tau}\left(\frac{\overline{y}}{\tau}\right) + p\ \ \phi\left(\frac{\underline{y}-1}{\tau}\right)\frac{\partial}{\partial \tau}\left(\frac{\underline{y}-1}{\tau}\right). \tag{2}$$

Using Bayes rule and the Normal distribution functional form,

$$Pr(\theta = 0|x_i, y) = \frac{1}{1 + \frac{p}{1-p}\frac{\phi\left(\frac{x_i-1}{\sigma}\right)}{\phi\left(\frac{x_i}{\sigma}\right)}\frac{\phi\left(\frac{y-1}{\tau}\right)}{\phi\left(\frac{y}{\tau}\right)}} = \frac{1}{1 + \frac{p}{1-p}\ e^{\frac{2x_i-1}{2\sigma^2}}\ e^{\frac{2y-1}{2\tau^2}}}. \tag{3}$$

Substituting from (1) into (3) and simplifying algebra yields:

$$\begin{aligned}
\underline{y} &= \frac{1}{2}\left(1 - \frac{\tau^2}{\sigma^2}\right) + \tau^2 Log\left(\frac{\beta - k}{k}\frac{1-p}{p}\right) - \frac{\tau^2}{\sigma}\Phi^{-1}(T) \\
\overline{y} &= \frac{1}{2}\left(1 + \frac{\tau^2}{\sigma^2}\right) + \tau^2 Log\left(\frac{\beta - k}{k}\frac{1-p}{p}\right) - \frac{\tau^2}{\sigma}\Phi^{-1}(T).
\end{aligned}$$

Differentiating with respect to τ yields:

$$\begin{aligned}
\frac{\partial}{\partial \tau}\frac{\overline{y}}{\tau} &= +\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\tau^2}\right) + Log\left(\frac{1-p}{p}\right) - \left(Log\left(\frac{k}{\beta - k}\right) + \frac{\Phi^{-1}(T)}{\sigma}\right) = \frac{\overline{y}-1}{\tau^2} \\
\frac{\partial}{\partial \tau}\frac{\underline{y}-1}{\tau} &= -\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\tau^2}\right) + Log\left(\frac{1-p}{p}\right) - \left(Log\left(\frac{k}{\beta - k}\right) + \frac{\Phi^{-1}(T)}{\sigma}\right) = \frac{\underline{y}}{\tau^2}.
\end{aligned}$$

Combining this with equation (2) yields:

$$\frac{\partial Pr(\text{regime change})}{\partial \tau} = (1-p)\ \phi\left(\frac{\overline{y}}{\tau}\right)\ \frac{\overline{y}-1}{\tau^2} + p\ \ \phi\left(\frac{\underline{y}-1}{\tau}\right)\ \frac{\underline{y}}{\tau^2}. \tag{4}$$

Thus, when k/β or T are sufficiently large, so that the revolution is sufficiently unlikely, $\max\{\overline{y}-1, \underline{y}\} < 0$, and hence $\frac{\partial Pr(\text{regime change})}{\partial \tau} < 0$. In contrast, when k/β or T are sufficiently small, so that revolution is sufficiently likely, $\min\{\overline{y}-1, \underline{y}\} > 0$, and hence $\frac{\partial Pr(\text{regime change})}{\partial \tau} > 0$. That is, when regime change is sufficiently unlikely, transparency raises the likelihood of regime change; when regime change is sufficiently likely, transparency reduces the likelihood of regime change. □

B Getting a Unique Responsive Equilibrium

To eliminate the need to select equilibria where responsive equilibria fail to exist, we modify the Hollyer et al. model so that rather than citizens receiving signals about a binary state (ruler's type), they receive signals about a continuously (normally) distributed state. Then, for any public signal, a responsive equilibrium exists in which a citizen revolts whenever his signal is below a (public-signal contingent) threshold. This enrichment eliminates the range in which responsive equilibria fail to exist, simplifying the equilibrium. However, we emphasize that there is still a continuum of equilibria due to the state-independent payoff structure of the game. Moreover, this equilibrium is unstable, and has perverse comparative statics—e.g., higher revolution costs increase the likelihood of revolution. This model allows for clean comparisons between marginal increases in the noise of public and private signals. We show that when the noise levels in the two signals are equal, the magnitude of the marginal impact of increased noise (higher variance) on a regime's probability of survival is *lower* for the public signal than for the private signals. That is, the publicness of the public signal, which one might interpret as facilitating coordination, does not amplify the impact of higher noise on the regime's stability.

In particular, unlike Hollyer et al. who assume that the ruler's type is either good or bad, $\theta \in \{0, 1\}$, suppose the ruler's type is continuous, $\theta \in \mathbb{R}$, and citizens share a prior that $\theta \sim N(0, \rho)$. We focus on equilibria in which, for a given public signal y, a citizen i revolts if and only if his signal is below a (finite) threshold $x^*(y)$.

Proposition 3 *Suppose the ruler's type is continuous with $\theta \sim N(0, \rho)$. For any public signal y, there is a unique equilibrium in which a citizen revolts if and only if his signal is below* *a threshold, $x_i < x^*(y)$, and the regime changes if and only if $\theta < \theta^*(y)$. In equilibrium, more transparency increases the likelihood of regime change if the likelihood of regime change is below a threshold, but reduces the likelihood of regime change if the likelihood of regime change is above a threshold.*

We show in the proof that the equilibrium is unstable. In particular, if others raise their cutoff by δ, a citizen's best response is to raise his cutoff by more that δ. Because the equilibrium is unstable, making regime change more difficult (higher T) or increasing the costs of revolting (k/β) *increase* the likelihood of regime change. However, stability and such perverse comparative statics do not change the basic logic of gambling for resurrection. The proof of Proposition 3 shows that when T or k/β are sufficiently high, so that the likelihood of regime change is sufficiently high, lower transparency (higher τ) increases the likelihood of regime change, hurting the regime. In contrast, when T or k/β are sufficiently low, so that the likelihood of regime change is sufficiently low, lower transparency (higher τ) reduces the likelihood of regime change, helping the regime.

A natural question is the role of the public-ness of the public signal on the effect of changes in its variance. That is, would it be different if we changed the variance of private signals instead of the variance of the public signal? This question is intimately linked to the effect of coordination. If there was only one player, then all signals would be on equal standing, and there would be no difference between y and x_i when $\tau = \sigma$. To capture the effect of coordination and the role of public-ness of the public signal we calculate the ratio of the marginal impact of noise in the public signal on the probability of regime change to the marginal impact of noise in the private signal on this probability at $\tau = \sigma$:

Proposition 4 *The magnitude of the marginal effect of increases in the variance of the public signal is lower than the magnitude of the marginal effect of increases in the variance of private signals:*

$$\left| \frac{\frac{\partial Pr(regime\ change)}{\partial \tau}}{\frac{\partial Pr(regime\ change)}{\partial \sigma}} \right|_{\tau=\sigma} = \frac{\rho^2}{\rho^2 + \tau^2}.$$

B.1 Proofs for Online Appendix B

Proof of Proposition 3: Equilibria are symmetric, and are characterized by cutoff pairs (θ^*, x^*) that satisfy:

$$\begin{aligned} Pr(x_i < x^*(y)|\theta^*(y)) &= T \quad \text{(belief consistency)} \\ Pr(\theta < \theta^*(y)|x_i = x^*(y), y) &= \frac{k}{\beta} \quad \text{(indifference condition)} \end{aligned}$$

That is,

$$\Phi\left(\frac{x^* - \theta^*}{\sigma}\right) = T \quad \text{(belief consistency)}$$

$$\Phi\left(\frac{\theta^* - \frac{\tau^2\rho^2 x^* + \sigma^2\rho^2 y}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}}{\sqrt{\frac{\tau^2\sigma^2\rho^2}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}}}\right) = \frac{k}{\beta} \quad \text{(indifference condition)} \tag{5}$$

The equilibrium is unstable: if others raise their cutoff by δ, a citizen's best response is to raise his cutoff by more that δ. To see this, suppose all citizens $j \neq i$ raise their threshold from x^* to $x^* + \delta$. Then, the belief consistency condition implies that θ^* increases to $\theta^* + \delta$. But then, the indifference condition implies that player i raises his threshold by $(1 + \sigma^2 \frac{\rho^2+\tau^2}{\rho^2\tau^2})\delta > \delta$. Thus, as we will see, the equilibrium features perverse comparative statics.

From (5),

$$x^* = \theta^* + \sigma\Phi^{-1}(T) \quad \text{(belief consistency)}$$

$$\theta^* - \frac{\tau^2\rho^2 x^* + \sigma^2\rho^2 y}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2} = \sqrt{\frac{\tau^2\sigma^2\rho^2}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}}\ \Phi^{-1}\left(\frac{k}{\beta}\right) \quad \text{(indifference condition)}$$

Substitute belief consistency into the indifference condition and group the θ^* terms to obtain:

$$\frac{\sigma^2\rho^2 + \tau^2\sigma^2}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}\theta^* = \frac{\tau^2\rho^2}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}\sigma\Phi^{-1}(T) + \frac{\sigma^2\rho^2 y}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2} + \sqrt{\frac{\tau^2\sigma^2\rho^2}{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}}\ \Phi^{-1}\left(\frac{k}{\beta}\right).$$

Solving for θ^* yields:

$$\theta^* = \frac{\tau^2\rho^2}{(\rho^2 + \tau^2)\sigma^2}\sigma\Phi^{-1}(T) + \frac{\rho^2 y}{\rho^2 + \tau^2} + \frac{\sqrt{\tau^2\sigma^2\rho^2(\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2)}}{(\rho^2 + \tau^2)\sigma^2}\ \Phi^{-1}\left(\frac{k}{\beta}\right).$$

Re-arranging to put in standard normal form:

$$\frac{\theta^* - \frac{\rho^2 y}{\rho^2+\tau^2}}{\sqrt{\frac{\rho^2\tau^2}{\rho^2+\tau^2}}} = \sqrt{\frac{\tau^2\rho^2}{\rho^2 + \tau^2}}\frac{1}{\sigma}\Phi^{-1}(T) + \sqrt{\frac{\tau^2\rho^2 + \sigma^2\rho^2 + \tau^2\sigma^2}{\rho^2 + \tau^2}}\ \frac{1}{\sigma}\Phi^{-1}\left(\frac{k}{\beta}\right). \tag{6}$$

Thus, substituting from equation (6),

$$\begin{aligned} P \equiv Pr(\text{regime change}) &= \int Pr_{\theta|y}(\theta < \theta^*(y)) pdf(y) dy \\ &= \int \Phi\left(\frac{\theta^* - \frac{\rho^2 y}{\rho^2+\tau^2}}{\sqrt{\frac{\rho^2\tau^2}{\rho^2+\tau^2}}}\right) pdf(y)dy \\ &= \Phi\left(\sqrt{\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}\ \Phi^{-1}(T) + \sqrt{1+\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}\ \Phi^{-1}\left(\frac{k}{\beta}\right)\right). \end{aligned}$$

To compare the effect of τ and σ, when differentiating with respect to τ, it is convenient to differentiate with respect to (τ/σ) while holding τs outside (τ/σ) fixed, and then do the opposite. Thus, abusing notation,

$$\frac{dP}{d\tau} = \phi(Z)\ \left(\frac{\partial Z}{\partial(\tau/\sigma)}\frac{1}{\sigma} + \frac{\partial Z}{\partial \tau}\right),$$

where Z is the argument of Φ:

$$Z = \sqrt{\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}\ \Phi^{-1}(T) + \sqrt{1+\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}\ \Phi^{-1}\left(\frac{k}{\beta}\right).$$

Calculating $\frac{\partial Z}{\partial(\tau/\sigma)}$ and $\frac{\partial Z}{\partial \tau}$ yields:

$$\begin{aligned} \frac{\partial Z}{\partial(\tau/\sigma)} &= \sqrt{\frac{\rho^2}{\rho^2+\tau^2}}\ \left(\Phi^{-1}(T) + \frac{\tau}{\sigma}\sqrt{\frac{\frac{\rho^2}{\rho^2+\tau^2}}{1+\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}}\ \Phi^{-1}\left(\frac{k}{\beta}\right)\right) \\ \frac{\partial Z}{\partial \tau} &= -\frac{\tau}{\sigma}\ \frac{\tau}{\rho^2+\tau^2}\ \sqrt{\frac{\rho^2}{\rho^2+\tau^2}}\ \left(\Phi^{-1}(T) + \frac{\tau}{\sigma}\sqrt{\frac{\frac{\rho^2}{\rho^2+\tau^2}}{1+\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}}\ \Phi^{-1}\left(\frac{k}{\beta}\right)\right). \end{aligned}$$

Substituting for $\frac{\partial Z}{\partial(\tau/\sigma)}$ and $\frac{\partial Z}{\partial \tau}$ into $\frac{dP}{d\tau}$ yields:

$$\frac{dP}{d\tau} = \phi(Z)\ \frac{1}{\sigma}\ \left(\frac{\rho^2}{\rho^2+\tau^2}\right)^{3/2}\ \left(\Phi^{-1}(T) + \frac{\tau}{\sigma}\sqrt{\frac{\frac{\rho^2}{\rho^2+\tau^2}}{1+\frac{\rho^2}{\rho^2+\tau^2}\left(\frac{\tau}{\sigma}\right)^2}}\ \Phi^{-1}\left(\frac{k}{\beta}\right)\right).$$

□

Proof of Proposition 4: Using the same notation as in Proposition 3,

$$\frac{dP}{d\sigma} = \phi(Z)\ \frac{\partial Z}{\partial(\tau/\sigma)}\ \left(-\frac{\tau}{\sigma^2}\right). \tag{7}$$

Substituting the equations (7) and (7) for $\frac{dP}{d\tau}$ and $\frac{dP}{d\sigma}$ to calculate their ratio yields

$$\left|\frac{\frac{dP}{d\tau}}{\frac{dP}{d\sigma}}\right|_{\sigma=\tau} = \frac{\frac{1}{\tau}\,\left(\frac{\rho^2}{\rho^2+\tau^2}\right)^{3/2}}{\frac{1}{\tau}\sqrt{\frac{\rho^2}{\rho^2+\tau^2}}} = \frac{\rho^2}{\rho^2+\tau^2} < 1.$$

□