

Censorship and Repression^{*}

A. Arda Gitmez[†]

Pooya Molavi[‡]

Konstantin Sonin[§]

October 22, 2026

We study the interplay between censorship and repression, two instruments of authoritarian control. A regime suppresses adverse information about its behavior, while repression increases citizens' cost of protesting. Censorship involves a frequency–credibility trade-off: suppressing adverse reports makes silence more frequent but less credible, inducing skeptical citizens to protest even without adverse news. Higher repression shifts this trade-off toward more intense censorship. Whenever optimal censorship is interior, a small increase in repression induces more censorship. Censorship intensity also increases following the selective repression of less supportive citizens and when citizens' beliefs become more concentrated.

Keywords: Authoritarian regimes, repression, censorship, heterogeneous beliefs.

^{*}This paper supersedes “The Dictator’s Dilemma: A Theory of Propaganda and Repression” by Gitmez and Sonin and “Informational Autocrats, Diverse Societies” by Gitmez and Molavi. We are grateful to Daron Acemoglu, Raphael Boleslavsky, Georgy Egorov, Scott Gehlbach, Emin Karagözoğlu, Anton Kolotilin, Aleksandr Levkun, Elliot Lipnowski, Zhaotian Luo, Roger Myerson, Tymofiy Mylovanov, Nicola Persico, Torsten Persson, Emilie Sartre, Mehdi Shadmehr, Denis Shishkin, Maria Titova, Alex Wolitzky, Andy Zapechelnyuk, and Leifu Zhang for many helpful comments.

[†]Bilkent University & CEPR. E-mail: arda.gitmez@bilkent.edu.tr.

[‡]Northwestern University. E-mail: pmolavi@kellogg.northwestern.edu.

[§]University of Chicago. E-mail: ksonin@uchicago.edu.

1 Introduction

After the collapse of the Soviet Union and other socialist dictatorships in the late 1980s–early 1990s, authoritarianism as a form of government appeared to be on its way out (Fukuyama, 1992). Three decades later, authoritarian regimes have proven remarkably durable, adapting to the challenges of globalization, technological progress, and increased awareness of civil rights (Applebaum, 2024). A critical element of this adaptation has been the novel use of information manipulation and repression—the two primary tools of authoritarian control.

Information manipulation and repression have always played an important role in keeping autocrats in power (Machiavelli, 1532; Wintrobe, 1990; Svoboda, 2012). In the 20th century, information manipulation was a cornerstone of totalitarian rule in Hitler's Germany, Stalin's Russia, and Mao's China (Arendt, 1951; Friedrich and Brzezinski, 1956; Dimitrov, 2023; Harrison, 2023). With the demise of totalitarian dictatorships, information manipulation has been used not only for ideological indoctrination, but also to maintain the leader's reputation as a strong and competent hand (Treisman, 2011; Lorentzen, 2013; Huang, 2018; Rozenas and Stukal, 2019; Guriev and Treisman, 2019; Gratton and Lee, 2024). In the 21st century, leaders manipulate information by controlling the state media (Rozenas and Stukal, 2019), co-opting or pressuring independent media outlets (McMillan and Zoido, 2004; Szeidl and Szucs, 2021), and censoring unfavorable news (Lorentzen, 2014; Shadmehr, 2014).¹ Instead of arresting millions as Stalin, Hitler, or Mao, or carrying out public executions as Pol Pot, Hussein, or Nguema, they primarily rely on selective censorship, digital surveillance, and sophisticated propaganda. They are, as Guriev and Treisman (2019, 2020, 2022) put it, *informational autocrats*.

Yet *repression* remains a critical instrument in the autocrat's arsenal (Tyson, 2018; Montagnes and Wolton, 2019; Rozenas, 2020). Following the 2020 protests, Belarus' Alexander Lukashenko had more than 30,000 people arrested and hundreds given long jail terms, a more than ten-fold increase over the average number of political prisoners during the previous decade (Way and Tolvin, 2023; Anisin, 2024). After the start of Russia's full-scale invasion of Ukraine in February 2022, Putin's regime has doubled down on repression: tens of thousands of people opposing the war and Putin's rule were arrested or forced out of the country; hundreds if not thousands received sentences in the range of 8–25 years, numbers unheard of since Stalin's

¹ Attempts to control media and manipulate information extend beyond autocracies. Examples in democracies abound, from Argentina (Di Tella and Franceschelli, 2011) to Mexico (Stanig, 2015) to Italy (Durante and Knight, 2012) to the United States (Qian and Yanagizawa-Drott, 2017; Gentzkow and Shapiro, 2008).

years (Treisman, 2022; Stoner, 2023). At the same time, information manipulation, which has long been Putin’s weapon of choice against the domestic opposition (Gehlbach, 2010; Treisman, 2011; Rozenas and Stukal, 2019; Guriev and Treisman, 2022), increased dramatically after the invasion.

In this paper, we study the relationship between the two main tools of authoritarian control: repression and information manipulation. We consider a regime that can act pro-socially or anti-socially on an issue of public interest. Citizens have heterogeneous priors about whether the regime has acted pro-socially and rely on media reporting to find out about the regime’s behavior. Each citizen protests against the regime if she is sufficiently confident that the regime has acted anti-socially; otherwise, she acquiesces. The regime uses repression to discourage citizens from protesting and uses censorship to manipulate information. Repression is modeled as any action by the regime that increases the cost of protesting. Censorship is modeled as an institution that suppresses adverse information about the regime’s behavior with some probability, thereby generating media silence. The regime faces a fundamental trade-off when choosing the censorship intensity that minimizes the expected size of protests: More intense censorship makes adverse reports less frequent but makes silence less credible.

We have two main findings. First, when citizens are more heavily repressed or expect a harsher punishment for opposing the regime, the regime engages in more censorship. In other words, repression and censorship are *complements*. This complementarity survives when protest is instrumental—because sufficiently large protests can topple the regime and citizens value participating in the overthrow of an anti-social regime. Second, when the distribution of citizens’ prior beliefs is less concentrated, the regime censors less. For single-peaked distributions, lower concentration corresponds to greater polarization, so greater polarization is associated with less intense censorship.

The basic logic behind the complementarity result is as follows. For a given censorship intensity, sufficiently skeptical citizens protest even after media silence. If the regime increases the intensity of censorship, adverse reports are suppressed more often, and silence becomes more frequent and therefore a less reliable signal of pro-social behavior. This induces more citizens to protest after silence. The regime thus trades off the frequency of mass protests triggered by adverse reports against the size of smaller-scale protests following media silence. Harsher repression changes the terms of this trade-off: By raising the cost of protesting, it reduces the size of protests after silence. The regime can therefore increase the censorship

intensity—and hence reduce the frequency of mass protests—without inducing as many citizens to protest following media silence. This shifts the regime’s optimal strategy towards more intense censorship.

Our complementarity result is local. When the regime does not engage in full censorship, a sufficiently small increase in repression leads to more intense censorship. Without further restrictions on the distribution of citizens’ prior beliefs, complementarity need not hold globally. In particular, when the regime initially engages in full censorship, an increase in repression may induce a discontinuous fall in the optimal censorship intensity. We illustrate this possibility through an example that clarifies the scope of our complementarity result.

A more extreme form of repression is the liquidation of those critical of the regime.² **Montagnes and Wolton (2019)** and **Rozenas (2020)** use communist purges in Stalin’s Russia and Mao’s China to demonstrate the compositional effects of repression. Physical elimination as in **Esteban, Morelli and Rohner (2015)** changes the composition of society; other forms of political disenfranchisement might also have the same effect.³ In addition to mass executions, Stalin relocated hundreds of thousands from places where they were a political threat to distant regions of Russia. In most cases, Stalin’s mass repression campaigns were organized around broad ethnic or social categories (**Gregory, Schröder and Sonin, 2011**). We model this form of repression as the regime purging citizens based on imperfect information about their initial beliefs. We find that, after a purge, the regime censors the remaining population more intensively. Thus, censorship is also complementary to this more extreme form of repression.

Our second result establishes that when citizens’ beliefs are less concentrated, the regime finds it optimal to decrease the censorship intensity. The intuition follows from the same trade-off between the frequency of mass protests triggered by adverse reports and the size of smaller-scale protests following media silence. Consider a single-peaked distribution. When beliefs are tightly concentrated, the regime can choose a censorship intensity that ensures that the large cluster of citizens around the mode acquiesces after silence. This strategy keeps protests following silence small while reducing the frequency of mass protests triggered by adverse reports. The regime therefore has little incentive to preserve enough credibility to persuade the remaining, more skeptical citizens. A decrease in the concentration of beliefs

²**Arendt (1951)** makes a distinction between *dictatorial* terror, aimed against well-identified opponents of the regime, and an all-pervasive *totalitarian* terror of purges, mass executions, and concentration camps. Modern theories of repression with strategic targeting and selection include **Myerson (2015)**, **Tyson (2018)**, and **Dragu and Przeworski (2019)**.

³In the realm of democratic politics, **Glaeser and Shleifer (2005)** show that the incumbent politician could deliberately choose policies that drive voters who oppose him out of the district.

increases the mass of these skeptics and, consequently, the size of protests following media silence. The regime therefore has a stronger incentive to persuade these skeptics to acquiesce after silence. Doing so requires silence to be more credible, which the regime achieves by reducing censorship. Our theoretical contribution is to formalize this result and extend it beyond single-peaked distributions using a novel partial order that captures the concentration of beliefs.

Our baseline model takes censorship as a primitive information control technology rather than allowing the regime to choose an arbitrary information structure. This allows us to characterize the complementarity between censorship and repression without imposing any substantive assumptions on the distribution of citizens' priors. In the appendix, we embed our baseline model in a standard Bayesian persuasion problem with heterogeneous priors (Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016; Bergemann and Morris, 2019) and show that, when the distribution of priors is strictly log-concave, a censorship strategy is optimal even within the larger class of unrestricted persuasion strategies. We also show that our main insights remain robust when the regime can only partially commit to a censorship intensity, when its behavior can be partly pro-social, and when it controls multiple media sources with varying censorship intensities.

Our paper contributes to the vast theoretical literature in political economy on authoritarian rule. Like Gehlbach, Luo, Shirikov and Vorobyev (2026), we examine the interplay between information manipulation and repression in autocratic regimes. However, our analysis differs from theirs along two important dimensions. First, in their model, repression and censorship are technologically linked: Increasing repressive capacity reduces the cost of suppressing protests, and directly increases the frequency of favorable news about the regime. In our model, by contrast, repression and censorship are distinct instruments, and the link between them is behavioral: Repression changes citizens' incentives to protest, which in turn changes the regime's chosen censorship intensity. Second, the underlying trade-off is different. In our model, the key trade-off is between the frequency and credibility of silence: More intense censorship makes silence more frequent but less credible, and greater repression changes the terms of this trade-off by reducing citizens' inclination to protest. In Gehlbach et al. (2026), repression has a built-in cost of more frequent and less credible favorable news, and elections can weaken this tight link.

Our analysis is also closely related to Guriev and Treisman (2020), which emphasizes the

substitution between repression and information manipulation in autocracies. In [Guriev and Treisman \(2020\)](#), the equilibrium use of repression perfectly reveals the state, leaving no role for further information manipulation. The regime therefore chooses between overt repression and information manipulation (a combination of propaganda, censorship, and co-optation of elites, which [Guriev and Treisman \(2020\)](#) call an *informational autocracy*). In our model, repression does not reveal the underlying state; instead, it raises citizens' cost of opposing the regime. Information manipulation can therefore coexist with repression. This allows us to study how the regime's chosen censorship intensity varies with the repression level. We show that, through this incentive channel, greater repression leads to more intense censorship. Thus, our analysis identifies a distinct force through which repression and information control can operate as complements.

Our comparative static results are related to those in [Curello and Sinander \(2025\)](#). However, our analysis differs from theirs along two important dimensions. First, [Curello and Sinander \(2025\)](#) consider a general Bayesian persuasion setting and perform comparative statics with respect to the sender's *value function*.⁴ In contrast, we restrict our attention to censorship strategies and perform comparative statics directly with respect to primitives such as repression and the distribution of receiver types. Second, we derive local comparative statics results without imposing any substantive restrictions on the distribution of citizens' priors. In particular, we establish local complementarity between censorship and repression even in environments where global complementarity may fail. Our results therefore characterize how optimal censorship changes with a small change in primitives, even in cases where [Curello and Sinander \(2025\)](#)'s results do not apply.⁵

The rest of the paper is organized as follows. Section 2 sets up our model. Section 3 studies the complementarity of repression and censorship. Section 4 discusses how changes in the distribution of views in the population affect the regime's optimal choice of censorship. Section 5 discusses an alternative model where sufficiently large protests can lead to the toppling of the regime, thus creating a coordination motive for citizens. Section 6 concludes. Extensions and robustness results, as well as omitted proofs, are relegated to three appendices.

⁴Similarly, Proposition 2 of [Kolotilin, Mylovanov and Zapechelnnyuk \(2022\)](#) presents a comparative static result based on the *shifts* of the value functions.

⁵To be more precise, we compare the optimal censorship intensities even when the resulting sender value functions are *not* ranked by [Curello and Sinander \(2025\)](#)'s *coarse convexity* partial order—see the bimodal distribution in Example 1.

2 Setup

2.1 The regime and citizens

We consider a regime that tries to dissuade citizens from protesting against it using the tools of censorship and repression. The regime faces many issues of public interest and, on each issue, acts either *pro-socially* (i.e., in the public interest) or *anti-socially* (i.e., against the public interest). For example, citizens may be uncertain about whether the regime has acted in the public interest in its handling of a public health crisis, its management of a natural disaster, or its allocation of public resources. We let $\omega \in \{0, 1\}$ denote the regime's behavior on a given issue, with $\omega = 1$ corresponding to pro-social behavior and $\omega = 0$ to anti-social behavior. If the regime is behaving pro-socially, there is no benefit from agitating against it, but if the regime is misbehaving, citizens derive an expressive benefit from protesting.

Absent media reporting, citizens do not know whether the regime is acting in the public interest. We assume that citizens have *heterogeneous priors* about whether the regime is acting pro-socially, with the prior of citizen $i \in I = [0, 1]$ given by

$$\mu_i = \Pr_i(\omega = 1).$$

The heterogeneity in priors reflects both citizens' varying degrees of prior knowledge about the regime's behavior and their genuine disagreement about what constitutes pro-social behavior. The distribution of priors among citizens is given by F . We assume that F has a density function f , which satisfies the following assumption:

Assumption 1. The density function f is continuously differentiable and strictly positive on $(0, 1)$, and $f(0) = f(1) = 0$.

2.2 Protests and repression

Each citizen decides whether to *protest* against the regime or to *acquiesce*. We let $a_i \in \{\text{protest}, \text{acquiesce}\}$ denote this action. Protests represent any agitation against the regime that weakens it politically. The regime uses repression to increase the cost of protests to citizens and reduce the size of protests. We take the regime's repressive capacity as given and characterize the regime's optimal censorship conditional on the level of repression.

Table 1 describes citizens' payoff given the regime's *repression* level $r \in (0, 1)$. The payoff from acquiescing is normalized to zero. Citizens who protest face a cost r that is independent

	$\omega = 0$	$\omega = 1$
$a_i = \text{protest}$	$1 - r$	$-r$
$a_i = \text{acquiesce}$	0	0

Table 1. Citizen payoffs given repression level r .

of whether the regime is behaving pro-socially and get a benefit (normalized to one) only if the regime is misbehaving (i.e., $\omega = 0$). We let $u(a_i, \omega)$ denote citizen i 's payoff when she chooses action a_i and the regime's behavior is given by ω . Given these payoffs, citizens protest if and only if they believe the regime to be acting anti-socially with high enough probability. Since the benefit from protesting is purely expressive, citizens' protest decisions are strategically independent; in Section 5, we consider a coordination game in which the benefit from protesting depends on whether protests succeed in toppling the regime.

Repression is any act by the regime that increases the cost of opposing it. This includes harassment, intimidation, imprisonment, torture, or execution. The incidence of repression is determined by a citizen's choice of action and therefore can be thought of as the expected cost of opposing the regime. That is important as even in totalitarian dictatorships only a small share of the population is actually punished. The Great Terror, 1937–1939, the most intensive years of Stalin's repression, had about 650,000 people executed, and 800,000 more sent to jails and labor camps. These are staggering numbers and, yet, only about 1% of the USSR population at the time (Conquest, 2008).

The regime incurs a cost for every citizen who protests. Specifically, the regime's payoff is given by

$$u_s(\{a_i\}_{i \in I}) = - \int_{i \in I} \mathbb{1}_{\{a_i = \text{protest}\}} di. \quad (1)$$

The regime's payoff decreases continuously with the number of citizens who protest. This assumption captures the idea that protests weaken the regime but do not topple it. In Section 5, we consider a more drastic form of protest where the regime may be toppled if the share of citizens who protest against the regime is sufficiently high. In such an environment, the size of protest has a discontinuous impact on the regime's payoff. This naturally generates a *coordination problem* among the citizens. We use the global game approach to analyze the resulting game and show that the insights of our baseline model continue to hold.

2.3 Censorship

While in an open society, the media or the rest of the political establishment would disseminate full information about government activities, authoritarian regimes engage in *censorship* to prevent the media or others from exposing any activity that is against the public interest. Examples include bullying or prosecuting reporters, shutting down media organizations, denying media access to evidence, and building state-owned media that suppress adverse reporting about the regime.

We assume that the regime establishes a censorship institution rather than deciding what to censor on a case-by-case basis. The institution is characterized by a *censorship intensity* $\beta \in [0, 1]$, which is the probability that adverse information is suppressed when the regime behaves anti-socially on a given issue. Citizens observe either an adverse report, $m = \text{report}$, or silence, $m = \text{silence}$. When the regime behaves pro-socially, there is no adverse information to report, so citizens observe silence with probability one. When the regime behaves anti-socially, an adverse report is suppressed with probability β :

$$\beta \equiv \Pr(m = \text{silence} \mid \omega = 0) \in [0, 1].$$

Thus, $\beta = 0$ corresponds to free reporting of anti-social behavior, whereas higher values of β correspond to more intense censorship. $\beta = 1$ is full censorship, where citizens never learn about regime misbehavior. For intermediate values of β , receiving a report is conclusive evidence that the regime has behaved anti-socially. We assume that the regime faces no explicit cost of choosing a high censorship intensity. However, as we discuss later, there is still an implicit cost of choosing a high β .

The regime chooses the censorship intensity before knowing the exact issues that will arise or how it will behave on any given issue. We take the regime's behavior as exogenous to its censorship policy: censorship affects what citizens learn about the regime's behavior, not the behavior itself. We let $\mu_s = \Pr_s(\omega = 1) \in (0, 1)$ denote the regime's prior belief that it will behave pro-socially on any given issue when it sets up the institution of censorship.

We assume that both the regime's prior, μ_s , and the distribution of priors among citizens, f , are common knowledge. We also assume that citizens know the censorship institution's reporting rule, including the probability that adverse information is suppressed. This assumption does not require citizens to observe individual acts of censorship or to infer the reporting rule from a single message.

3 Censorship and repression

We start our analysis by characterizing the regime's choice of optimal censorship intensity. We then analyze how that choice varies with the repression level.

3.1 Optimal censorship

Given a repression level $r \in (0, 1)$, what is the regime's optimal censorship intensity? Consider citizen i with posterior $\mu'_i = \Pr_i(\omega = 1|m)$. Given the payoff in Table 1, she protests if and only if

$$(1 - \mu'_i) \cdot (1 - r) + \mu'_i \cdot (-r) \geq 0,$$

which simplifies to $\mu'_i \leq 1 - r$.⁶ Fix a censorship intensity $\beta \in [0, 1]$.

- If citizens observe $m = \text{report}$, they all conclude $\omega = 0$. In that case, $\mu'_i = 0$ for all i and all citizens protest.⁷ Thus, the regime's payoff is -1 .
- If citizens observe $m = \text{silence}$, by Bayes' rule, citizen i 's posterior is:

$$\begin{aligned} \mu'_i &= \Pr_i(\omega = 1|m = \text{silence}) \\ &= \frac{\Pr_i(\omega = 1)}{\Pr_i(\omega = 1) + \Pr_i(\omega = 0) \Pr_i(m = \text{silence}|\omega = 0)} \\ &= \frac{\mu_i}{\mu_i + (1 - \mu_i)\beta}, \end{aligned} \tag{2}$$

which is decreasing in β : More intense censorship reduces the credibility of $m = \text{silence}$.

Given (2), $\mu'_i \leq 1 - r$ if and only if

$$\mu_i \leq \frac{\beta}{\beta + \frac{r}{1-r}}.$$

Thus, following $m = \text{silence}$, citizens with priors below $\frac{\beta}{\beta + \frac{r}{1-r}}$ protest, leading to a payoff of $-F\left(\frac{\beta}{\beta + \frac{r}{1-r}}\right)$ for the regime.

Now, consider a regime choosing a censorship intensity β before ω is realized. From the regime's perspective, $m = \text{report}$ will be realized with probability $(1 - \mu_s)(1 - \beta)$ and $m = \text{silence}$ will be realized with the complementary probability. Therefore, the regime's (subjec-

⁶We assume that a citizen protests when indifferent. Since f is continuously differentiable, it has no mass points, so the measure of citizens with posteriors exactly equal to $1 - r$ is zero. Therefore, this tie-breaking rule is inconsequential for the analysis.

⁷When $\beta = 1$, $m = \text{report}$ is never observed by the citizens and the beliefs following $m = \text{report}$ are free. Consequently, even in this case we can choose the posterior following $m = \text{report}$ to be zero.

tive) expected payoff from choosing β is:

$$\begin{aligned} V(\beta; r) &\equiv (1 - \mu_s)(1 - \beta) \cdot (-1) + (\mu_s + (1 - \mu_s)\beta) \cdot \left(-F\left(\frac{\beta}{\beta + \frac{r}{1-r}}\right) \right) \\ &= -1 + (\mu_s + (1 - \mu_s)\beta) \cdot \left(1 - F\left(\frac{\beta}{\beta + \frac{r}{1-r}}\right) \right). \end{aligned} \quad (3)$$

The expression in (3) exhibits the main trade-off in the regime's problem. A higher level of β leads to a higher frequency of $m = \text{silence}$. This can be seen through the first appearance of β in (3). On the other hand, a higher frequency of $m = \text{silence}$ means, by definition, that the absence of an adverse report is less surprising, and so, moves the citizens' beliefs by less—this is seen through the second appearance of β in Equation (3). Therefore, more intense censorship is inevitably less *credible*. The regime chooses β to optimally resolve the trade-off between frequency and credibility. The optimal censorship intensity chosen by the regime is

$$\beta^*(r) \equiv \arg \max_{\beta \in [0,1]} V(\beta; r).$$

The existence of a maximizer follows from a standard continuity and compactness argument, but $\beta^*(r)$ can be multi-valued. In Lemma 1 below, we show that if f is strictly log-concave, $\beta^*(r)$ is guaranteed to be single-valued. It also turns out that when f is strictly log-concave, the censorship strategy is optimal among a broader class of information manipulation strategies. In Appendix A.6, we show this by modeling the regime's problem as an information manipulation problem and using tools from the Bayesian persuasion literature.

A useful reformulation of the regime's problem relies on the following observation. Under the optimal strategy, the regime chooses a cutoff citizen with belief μ^* such that (i) citizens with priors $\mu_i > \mu^*$ protest only upon observing $m = \text{report}$ and (ii) citizens with priors $\mu_i \leq \mu^*$ protest regardless of m . A higher β increases μ^* (and hence the fraction of citizens who always protest), which is bad for the regime, but it decreases the probability of protest for those with $\mu \geq \mu^*$, which is good for the regime. The optimal censorship intensity resolves this trade-off.

To formalize this intuition, we adopt a change of variables $\mu = \frac{\beta}{\beta + \frac{r}{1-r}}$ and obtain the following optimization problem:

$$\begin{aligned} \mu^*(r) &= \arg \max_{\mu \in [0, 1-r]} \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) \frac{1 - F(\mu)}{1 - \mu} \\ &= \arg \max_{\mu \in [0, 1-r]} \Phi(\mu; r), \end{aligned} \quad (4)$$

where

$$\Phi(\mu; r) \equiv \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \frac{1 - F(\mu)}{1 - \mu}. \quad (5)$$

Here, $\mu^*(r)$ is the prior belief of the marginal citizen who is indifferent between protesting and acquiescing following $m = \text{silence}$ under the optimal censorship intensity $\beta^*(r)$. Note that $\mu^*(r)$ can be multi-valued. Since μ is a strictly monotonic transformation of β , there is a bijection between $\mu^*(r)$ and $\beta^*(r)$:

$$\mu = \frac{\beta}{\beta + \frac{r}{1-r}} \in \mu^*(r) \iff \beta = \frac{\mu}{1-\mu} \frac{r}{1-r} \in \beta^*(r).$$

This reformulation is simpler to work with because $\Phi(\mu; r)$ is *strictly supermodular* in (μ, r) . ($V(\beta; r)$ is not necessarily strictly supermodular in (β, r) .) Note, however, that the maximization problem in (4) does not satisfy the standard monotone comparative statics properties of [Milgrom and Shannon \(1994\)](#) because the feasible set $[0, 1 - r]$ is shrinking in r .

Though $\mu^*(r)$ can be multi-valued, in the case of strictly log-concave distributions, the regime's problem has a unique solution. Even though none of our main results (except Corollary 1) rely on log-concavity, this is a useful case and is of general interest.⁸

Lemma 1. *Suppose f is strictly log-concave, that is, $\frac{\partial^2}{\partial \mu^2} \log f(\mu) < 0$ for all $\mu \in (0, 1)$. Then, $\mu^*(r)$, and hence $\beta^*(r)$, is single-valued.*

3.2 Complementarity between censorship and repression

Having characterized the optimal censorship intensity, we can now focus on our main question: How does repression complement censorship? In this section we describe comparative statics with respect to r , the repression level. Our first proposition establishes that if the optimizers are interior, the optimal censorship intensity is increasing in r in a very strong sense.

Proposition 1. *Consider $r \in (0, 1)$, and suppose that all optimal censorship intensities are interior, i.e., $\beta^*(r) \subset (0, 1)$. There exists a neighborhood R of r such that, for every $r_1, r_2 \in R$ with $r_1 < r_2$,*

$$\max \mu^*(r_1) < \min \mu^*(r_2), \quad \text{and} \quad \max \beta^*(r_1) < \min \beta^*(r_2).$$

Proposition 1 shows that small increases in the repressive capacity of the regime will be accompanied by an increase in censorship level. Thus, repression and censorship are complements.

⁸Log-concavity is satisfied by many standard distribution families. For example, the following distributions are log-concave: uniform, (truncated) normal, and beta (with both parameters ≥ 1), among others—see [Bagnoli and Bergstrom \(2005\)](#) for a list.

The economic intuition behind Proposition 1 can be understood through the trade-off between the frequency and credibility of silence. Increasing the marginal citizen's prior μ allows the regime to censor more intensely, making silence more frequent. However, this comes at the cost of inducing more citizens to protest even after observing silence. Higher repression changes the terms of this trade-off. For a given marginal citizen μ , higher repression allows the regime to increase censorship while keeping that citizen indifferent. Consequently, the marginal benefit of increasing μ , through a higher frequency of silence, increases with repression, whereas the marginal cost, through a reduction in the fraction of citizens who acquiesce following silence, remains unchanged. This shifts the regime's incentives toward more intense censorship. To see this explicitly, consider the effect of increasing μ , which amounts to increasing the intensity of censorship. By (5),

$$\frac{\partial \Phi(\mu; r)}{\partial \mu} = \underbrace{\frac{\gamma}{(1 - \mu)(1 - \mu + \gamma\mu)}}_{\text{marginal benefit}} - \underbrace{\frac{f(\mu)}{1 - F(\mu)}}_{\text{marginal cost}}, \quad \text{where } \gamma \equiv \frac{r}{1 - r} \frac{1 - \mu_s}{\mu_s}.$$

The first term here corresponds to the marginal benefit of increasing μ , defined as the proportional increase in the probability of silence. The second term is marginal cost, defined as the proportional decrease in the fraction of citizens who acquiesce following silence. Crucially, higher repression increases the former while leaving the latter unchanged. Consequently, the regime can attain a larger increase in the frequency of silence while suffering the same loss of citizens who acquiesce following silence.

Mathematically, the proof of Proposition 1 relies on a straightforward complementarity argument: The key stage of the proof uses the fact that $\frac{\partial^2 \Phi(\mu; r)}{\partial \mu \partial r} > 0$. It is precisely this strong feature of the transformed problem that allows us to derive a very strong comparison between $\mu^*(r_1)$ and $\mu^*(r_2)$ when optimizers are interior: *every* optimizer in $\mu^*(r_2)$ is strictly larger than *every* optimizer in $\mu^*(r_1)$. This result closely echoes Theorem 1 of Edlin and Shannon (1998).⁹ It is worth repeating that neither the optimization problem characterizing $\beta^*(r)$ nor the one characterizing $\mu^*(r)$ satisfies the standard monotone comparative statics conditions of Milgrom and Shannon (1994). This is also evident from the fact that, when $\beta^*(r)$ is not interior, it may not be increasing in r in the sense of *strong set order*.¹⁰ See the following counterexample.

Example 1 (Failure of global monotonicity). Let $\mu_s = 0.08$, and let $\text{Beta}(a, b)$ denote the density

⁹The only difference here is that in Edlin and Shannon (1998)'s setting the feasible set does not vary with r .

¹⁰For any two sets $A, B \subseteq [0, 1]$, A is higher than B in the strong set order, $A \succeq_{SSO} B$, if the following property holds: For any $a \in A$ and $b \in B$, $\max\{a, b\} \in A$ and $\min\{a, b\} \in B$.

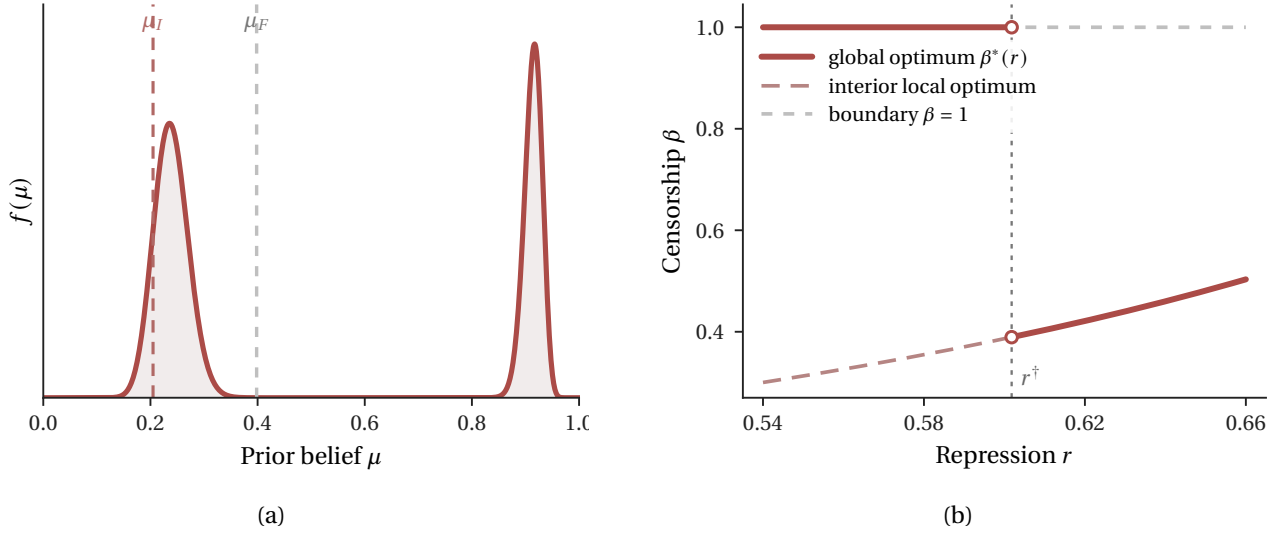


Figure 1. Failure of global monotonicity with a bimodal distribution. Panel (a) shows the density in Example 1. At $r = r^\dagger$, the marginal citizen has prior $\mu_I \approx 0.2049$ under the interior policy and $\mu_F = 1 - r^\dagger \approx 0.3982$ under full censorship. Panel (b) shows the two relevant local-optimum branches around $r^\dagger$. The dark line is the global optimum; it follows the full-censorship branch below $r^\dagger$ and the interior branch above $r^\dagger$.

of a Beta distribution with shape parameters $a, b > 0$. Suppose the distribution of citizens' prior beliefs has density

$$f = 0.6 \cdot \text{Beta}(42, 134) + 0.4 \cdot \text{Beta}(255, 24).$$

As panel (a) of Figure 1 shows, the resulting distribution is bimodal and satisfies Assumption 1. A direct computation of $V(\beta; r)$ shows that there is a repression level $r^\dagger \approx 0.6018$ at which the regime has two optimal censorship intensities: full censorship, $\beta = 1$, and an interior policy, $\beta_I \approx 0.3894$. For r just below $r^\dagger$, full censorship is uniquely optimal; for r just above $r^\dagger$, the interior policy is uniquely optimal. Thus, an increase in repression can cause optimal censorship to fall discontinuously.

The two optimal policies at $r^\dagger$ target different parts of the population. Under full censorship, the marginal citizen has prior $\mu_F = 1 - r^\dagger \approx 0.3982$, which lies in the low-density region between the two modes. The regime therefore effectively gives up on the lower-prior cluster of citizens: Approximately 60 percent of citizens protest even after silence. Under the interior policy, silence is more credible and the marginal prior falls to $\mu_I \approx 0.2049$, so most citizens in the lower-prior cluster also acquiesce following silence. At $r^\dagger$, the increase in credibility exactly offsets the lower frequency of silence, and the two policies generate the same expected protest size.

The example also clarifies the scope of Proposition 1. In the region of r where there is an

interior optimum, censorship increases with repression, exactly as the proposition predicts. However, global monotonicity fails because the identity of the global optimum changes: As repression crosses $r^\dagger$, the regime switches from full censorship to partial censorship. The failure of global monotonicity is therefore caused by a discrete switch between local optima, rather than by a reversal of the local complementarity between repression and censorship.

Example 1 uses a bimodal distribution f , where there are multiple locally optimal censorship strategies. When f is strictly log-concave, $\beta^*(r)$ is guaranteed to be single-valued for any r , and the complementarity argument globalizes: $\beta^*(r)$ is single-valued and increasing in r .

Remark 1. It is straightforward to augment the model by adding an earlier stage in which the regime chooses the repression level r , subject to a standard increasing and convex cost of repression. A reduction in the marginal cost of repression then raises the regime's optimal choice of r . Proposition 1 implies that a sufficiently small resulting increase in repression also raises censorship. Thus, making repression cheaper induces the regime to choose both more repression and more censorship, so the two instruments are complements in the standard consumer-theory sense.

4 Censorship and the distribution of views

In this section, we study how changes in the distribution of citizens' views change the regime's choice of optimal censorship intensity. We first consider changes that make citizens uniformly more pro-regime. This can be the result of propaganda or a purge of skeptics by the regime. We then consider the effect of changes in the concentration of views among citizens.

4.1 Ex ante repression and censorship

An important feature of our model of repression so far is that the punishment was only applied to those who actually acted against the regime. We now consider another type of repression, in which citizens are targeted *ex ante*, rather than *ex post*. For instance, the regime might identify, perhaps with some noise, those who are hostile to the regime and eliminate them. In other words, the distribution of opinions in society is no longer fixed. Repression alters the distribution of attitudes towards the regime, rather than the citizens' incentive to protest. We refer to such strategies as *ex ante repression*. Our analysis below demonstrates that the effect

of ex ante repression is qualitatively the same: The more “pro-regime” or “conformist” the resulting distribution, the higher the intensity of censorship for those who are not eliminated.

Consider a regime that can purge its opponents at will. Let f_0 be the initial density of priors. The regime has access to some informative signal about the citizens’ priors. Suppose that there is an institution, for example, a secret police, that assigns a label $\ell_i \in \{\text{skeptic}, \text{supporter}\}$ to each citizen i with prior μ_i . The secret police’s labeling technology is noisy:

$$\begin{aligned}\Pr(\ell_i = \text{supporter} \mid \mu_i) &= \rho(\mu_i), \\ \Pr(\ell_i = \text{skeptic} \mid \mu_i) &= 1 - \rho(\mu_i),\end{aligned}$$

where $\rho(\mu_i) \in [0, 1]$. Suppose $\rho(\mu)$ is continuously differentiable and increasing in $\mu \in [0, 1]$, with $\rho(\mu) > 0$ for all $\mu \in (0, 1)$. This formulation allows for a variety of labeling technologies, such as the linear ($\rho(\mu) = \mu$) or sigmoid ($\rho(\mu) = \frac{1}{1+e^{-\mu}}$) functions.

Suppose the regime purges citizens labeled $\ell_i = \text{skeptic}$, so that only those labeled $\ell_i = \text{supporter}$ remain. Denote the distribution of priors for the remaining citizens by f_1 :

$$f_1(\mu) = \frac{\rho(\mu)f_0(\mu)}{\int_0^1 \rho(\tilde{\mu})f_0(\tilde{\mu})d\tilde{\mu}} \implies \frac{f_1(\mu)}{f_0(\mu)} = \frac{\rho(\mu)}{\int_0^1 \rho(\tilde{\mu})f_0(\tilde{\mu})d\tilde{\mu}}.$$

Because ρ is continuously differentiable, f_1 is continuously differentiable, has full support on $[0, 1]$, and satisfies $f_1(0) = f_1(1) = 0$. Hence, f_1 also satisfies Assumption 1. More notably, because $\rho(\mu)$ is increasing in μ , so is $\frac{f_1(\mu)}{f_0(\mu)}$. That is, f_1 is larger than f_0 in the *likelihood ratio order*.

The following proposition shows that the regime engages in more intense censorship following a purge of skeptics. Because $\beta^*(r)$ may be multi-valued, the comparison is in the strong set order.

Proposition 2. *Consider two distributions of priors, f_0 and f_1 , which both satisfy Assumption 1 and where f_1 is larger than f_0 in the likelihood ratio order:*

$$\frac{f_1(\mu)}{f_0(\mu)} \text{ is increasing in } \mu.$$

Then, the optimal censorship intensity under f_1 is higher than the optimal censorship intensity under f_0 in the strong set order, that is, for any $r \in (0, 1)$,

$$\beta_1^*(r) \geq_{SSO} \beta_0^*(r).$$

Proposition 2 shows that *ex ante* repression and censorship complement each other as well. With skeptics purged, the remaining citizens face more intense censorship. Intuitively, the

regime facing the distribution f_1 is a “more-universally-loved” regime than a regime facing the distribution f_0 . In equilibrium, a more popular regime *rides its popularity* and provides less information to citizens, resulting in a higher censorship intensity.

More broadly, the result suggests that censorship is complementary to any tool that shifts citizens’ views in a pro-regime direction. Consider propaganda, for instance. It can be understood as the deliberate, systematic manipulation of public opinion and perceptions; it is commonly used by authoritarian regimes to bolster public support (Guriev and Treisman, 2022). To the extent that propaganda increases support for the regime (in the likelihood-ratio sense of Proposition 2), it also increases the regime’s incentive to censor. Thus, propaganda and censorship can be complementary.

4.2 Polarization and censorship

We next consider how changes in the concentration of citizens’ prior views affect the regime’s optimal choice of censorship. We introduce a novel partial order on densities that captures whether probability mass is more heavily concentrated in already high-density regions. For single-peaked distributions, lower concentration has a natural interpretation as greater polarization.

Definition 1. Consider two probability density functions f_0 and f_1 supported on a common compact set. f_1 is *more concentrated* than f_0 if

$$f_1(x) = \alpha(f_0(x)) \quad \text{for all } x, \quad (6)$$

for some strictly increasing and convex function $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $\alpha(0) = 0$.

This partial order has an intuitive interpretation. Since α is strictly increasing, it preserves the ranking of points by density: $f_1(x) > f_1(y)$ if and only if $f_0(x) > f_0(y)$. In particular, the two densities have the same set of modes. Since α is convex and f_0 and f_1 both have to integrate to one, transforming f_0 by α magnifies the parts of f_0 with higher values and shrinks the parts with lower values. Moving from f_0 to f_1 thus moves the mass from parts of the distribution that initially have a smaller mass to parts with a larger initial mass. In other words, f_1 looks like f_0 , but with higher peaks and deeper troughs. This is the sense in which f_1 is more concentrated than f_0 .¹¹

¹¹ Johnson and Myatt (2006) introduce a related *rotation order* on cumulative distribution functions, which requires cdf’s of two distributions to cross once. When f_0 is single-peaked, our concentration order implies a contraction of the density toward its mode, and hence a rotation of

Members of many parametric families of distributions can be ordered by the concentration of their density functions. Figure 2 illustrates the probability density functions for a set of single-peaked beta distributions that are ranked in concentration order.

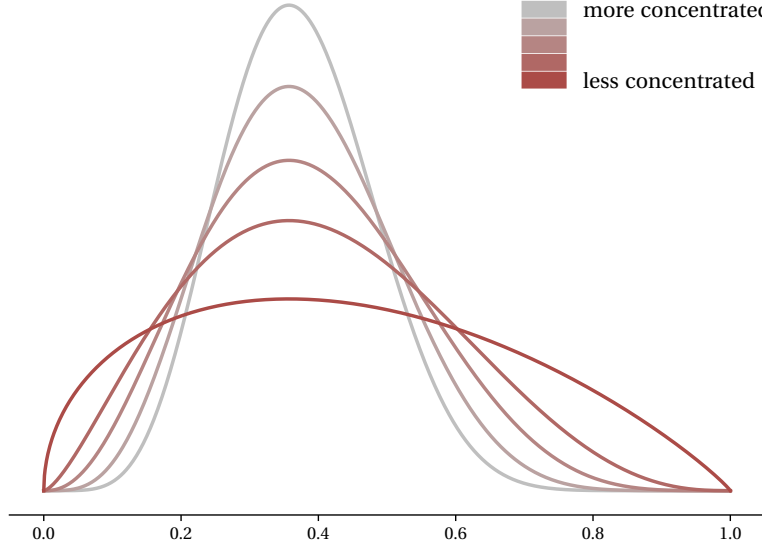


Figure 2. The concentration partial order.

Example 2. Consider two single-peaked beta probability density functions

$$f_0 = \text{Beta}(a_0, b_0),$$

$$f_1 = \text{Beta}(a_1, b_1),$$

where $\frac{a_0-1}{a_0+b_0-2} = \frac{a_1-1}{a_1+b_1-2}$. If $a_0 \geq a_1$, then f_0 is more concentrated than f_1 , while if $a_0 \leq a_1$, then f_1 is more concentrated than f_0 . In particular, any two single-peaked beta density functions with the same mode are ranked according to the concentration partial order.

Example 3. Consider the probability density functions for two truncated normal distributions on $[0, 1]$ with the same mode m :

$$f_0 = \text{TruncatedNormal}(m, \sigma_0^2),$$

$$f_1 = \text{TruncatedNormal}(m, \sigma_1^2).$$

If $\sigma_0^2 \geq \sigma_1^2$, then f_1 is more concentrated than f_0 .

the cdf in the sense of [Johnson and Myatt \(2006\)](#). If f_0 is multimodal, the corresponding cdf's may cross multiple times, so our concentration order does not necessarily imply the rotation order. The additional structure imposed by the convex transformation is crucial for the comparative statics in Lemma 2, which relies on Jensen's inequality.

When f_0 is single-peaked, the concentration order has a natural interpretation in terms of *polarization*. A single-peaked density has relatively more mass near its mode than in the periphery. Therefore, f_1 has more mass around the mode and less mass in the periphery relative to f_0 ; in this sense, f_1 represents a less polarized society than f_0 . Neither the concentration order nor our main comparative static result in this section requires single-peakedness. However, for the common case of unimodal distributions, we interpret lower concentration as greater polarization.

We are now ready to examine how a change in concentration affects censorship. We begin by presenting a key result that drives the comparative statics.

Lemma 2. *Let f_0 and f_1 be two densities that satisfy Assumption 1. For $k \in \{0, 1\}$, let $\Phi_k(\mu; r)$ be the function defined in (5):*

$$\Phi_k(\mu; r) = \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \frac{1 - F_k(\mu)}{1 - \mu}.$$

Suppose $r \leq \mu_s$. If f_1 is more concentrated than f_0 , then, for any $\mu \in [0, 1 - r]$,

$$\frac{\partial \Phi_0(\mu; r)}{\partial \mu} \geq 0 \implies \frac{\partial \Phi_1(\mu; r)}{\partial \mu} \geq 0.$$

Proof of Lemma 2. Note that

$$\frac{\partial \Phi_k(\mu; r)}{\partial \mu} = \frac{\gamma}{(1 - \mu)(1 - \mu + \gamma\mu)} - \frac{f_k(\mu)}{1 - F_k(\mu)}, \quad \text{where } \gamma \equiv \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} > 0,$$

and therefore,

$$\frac{\partial \Phi_k(\mu; r)}{\partial \mu} \geq 0 \iff \frac{1 - F_k(\mu)}{1 - \mu} \geq f_k(\mu) \frac{1 - \mu + \gamma\mu}{\gamma}. \quad (7)$$

Take any μ with $\frac{\partial \Phi_0(\mu; r)}{\partial \mu} \geq 0$. Then,

$$\begin{aligned} \frac{1 - F_1(\mu)}{1 - \mu} &= \frac{\int_{\mu}^1 f_1(x) dx}{1 - \mu} = \frac{\int_{\mu}^1 \alpha(f_0(x)) dx}{1 - \mu} \underbrace{\geq}_{\text{Jensen's}} \alpha \left(\frac{\int_{\mu}^1 f_0(x) dx}{1 - \mu} \right) = \alpha \left(\frac{1 - F_0(\mu)}{1 - \mu} \right) \\ &\underbrace{\geq}_{(7)} \alpha \left(f_0(\mu) \frac{1 - \mu + \gamma\mu}{\gamma} \right) \underbrace{\geq}_{\text{Convexity of } \alpha(\cdot)} \alpha(f_0(\mu)) \frac{1 - \mu + \gamma\mu}{\gamma} = f_1(\mu) \frac{1 - \mu + \gamma\mu}{\gamma}, \end{aligned}$$

which, by (7), implies $\frac{\partial \Phi_1(\mu; r)}{\partial \mu} \geq 0$. Note that in the first inequality we use the integral form of Jensen's inequality (e.g., [Dragomir, Adil Khan and Abathun 2016](#)), and the last inequality relies on $r \leq \mu_s$ (which is equivalent to $\gamma \leq 1$). $\square$

We interpret Lemma 2 as a *local* single-crossing condition: whenever Φ_0 is increasing in μ , so is Φ_1 .¹² This condition constrains the location of the *local* maximizers. Comparing the location of global maximizers requires imposing more structure on the optimization problem. For instance, if either f_0 or f_1 is strictly log-concave, then the corresponding objective function is single-peaked by Lemma 1. The following result then follows as a direct corollary of Lemma 2.¹³

Corollary 1. *Let f_0 and f_1 be two densities that satisfy Assumption 1. Suppose $r \leq \mu_s$ and f_1 is more concentrated than f_0 . If either f_0 or f_1 is strictly log-concave, then,*

$$\max \mu_0^*(r) \leq \min \mu_1^*(r), \quad \text{and} \quad \max \beta_0^*(r) \leq \min \beta_1^*(r).$$

The next result shows that a *sufficiently small* increase in concentration of opinions leads to an increase in censorship. We consider this the counterpart of Proposition 1 for the concentration order. Because we have a *local* single-crossing condition here, the result is also weaker: A small increase in concentration leads to an increase in the *smallest* optimal censorship intensity.

Proposition 3. *Let f_0 and f_1 be two densities that satisfy Assumption 1. Suppose $r \leq \mu_s$, f_1 is more concentrated than f_0 , and the smallest optimal censorship intensity under f_0 is interior, i.e., $\min \beta_0^*(r) \in (0, 1)$. For $\varepsilon \in [0, 1]$, let*

$$f_\varepsilon(x) = (1 - \varepsilon)f_0(x) + \varepsilon f_1(x) \quad \text{for all } x,$$

and let $\mu_\varepsilon^*(r)$ and $\beta_\varepsilon^*(r)$ denote the corresponding sets of optimal cutoff beliefs and censorship intensities. Then, there exists $\bar{\varepsilon} > 0$ such that, for every $\varepsilon \in (0, \bar{\varepsilon})$,

$$\min \mu_0^*(r) \leq \min \mu_\varepsilon^*(r), \quad \text{and} \quad \min \beta_0^*(r) \leq \min \beta_\varepsilon^*(r).$$

Corollary 1 and Proposition 3 demonstrate the same insight: An increase in concentration leads to more intense censorship. To gain some intuition, recall that the regime's censorship intensity targets a marginal citizen who, upon observing $m = \text{silence}$, barely acquiesces. Increasing censorship makes silence more frequent but less credible, thus raising the marginal citizen's prior belief and, consequently, the mass of citizens who protest despite observing silence. To illustrate how concentration affects this trade-off, consider a single-peaked distribution of beliefs. When beliefs are tightly clustered around the mode, the regime can induce most

¹²It is not a *global* single-crossing condition, which would require: whenever $\Phi_0(\mu'; r) \geq \Phi_0(\mu; r)$ for some $\mu' > \mu$, then $\Phi_1(\mu'; r) \geq \Phi_1(\mu; r)$. Instead, Lemma 2 imposes this property to hold for μ' close to μ .

¹³In Section A of the Appendix, where we discuss the robustness of our results, we use the results of Kolotilin, Mylovannov and Zapechelnyuk (2022) to demonstrate that the main idea in Section 4.2 naturally extends to continuous state spaces (see Lemma A-1).

citizens to acquiesce by targeting a citizen just below this cluster. It therefore has little incentive to make silence sufficiently credible to persuade the remaining skeptics. When beliefs are less concentrated, however, reaching these skeptics becomes more valuable. Persuading them requires a more credible silence, which the regime achieves by reducing censorship.

Lemma 2 requires $r \leq \mu_s$, that is, the regime is relatively confident in the alignment of the policies it pursues with the public interest. Having $r \leq \mu_s$ allows us to compare societies only through the concentration of f , as we did in Definition 1. If $r > \mu_s$, it is still possible to define a measure of concentration and prove that an increase in concentration of opinions leads to an increase in censorship. However, the concentration order will also involve the values of r and μ_s . We consider this extension in Appendix B.

Remark 2. Our results suggest that greater polarization is associated with less intense censorship. Independent media and online media are two forces that may produce such changes in the distribution of views. In a recent working paper, [Enikolopov, Rochlitz, Schoors and Zakharov \(2025\)](#) demonstrate that access to independent online TV in Russia before the 2016 elections had asymmetric effects on individuals who rely on news from social media. Specifically, it boosted support among supporters of the regime while leading to a decline in support among those who opposed the regime. To the extent that such asymmetric effects make the distribution of views less concentrated in the sense of Definition 1, our results imply that they also reduce the regime’s incentive to censor. Thus, online media need not reach every citizen to temper the effects of state censorship—as long as they influence the opinions of *some* citizens, they could reduce the regime’s incentive to engage in censorship.

5 Coordination and regime change

In this section, we model citizens’ protest decisions as a *coordination game* in which sufficiently large protests can topple the regime. As in the baseline model, each citizen chooses $a_i \in \{\text{protest}, \text{acquiesce}\}$. However, here, the regime is toppled if the size of protests $A \equiv \int_0^1 \mathbf{1}_{\{a_i = \text{protest}\}} di$ exceeds a parameter $\theta \in [\underline{\theta}, \bar{\theta}]$, which represents the *regime strength*. To isolate the role of strategic coordination, we abstract from heterogeneity in prior beliefs and assume that citizens share a common prior.

The timing of the game is as follows. The regime and citizens start with a common prior $\mu_0 = \Pr(\omega = 1) \in (0, 1)$ about how the regime will behave on a given issue. Given this prior, the

regime commits to a censorship strategy (as in the benchmark model). Then, ω is realized, and citizens observe $m \in \{\text{report}, \text{silence}\}$ given the regime's censorship policy. The regime strength $\theta \sim U[\underline{\theta}, \bar{\theta}]$ is then drawn, independently of ω ; and each citizen i observes a private signal of the regime strength, $x_i = \theta + \rho v_i$, where $\rho > 0$, $v_i \sim U[-1, 1]$, and v_i 's are iid across citizens. Lastly, each citizen chooses $a_i \in \{\text{protest}, \text{acquiesce}\}$. The regime is toppled if and only if the size of protests exceeds the regime strength, i.e., $A \geq \theta$.

	$A \geq \theta$	$A < \theta$		$A \geq \theta$	$A < \theta$
$a_i = \text{protest}$	$-(1 + g) - r$	$1 - r$	$a_i = \text{protest}$	$(1 + g) - r$	$-1 - r$
$a_i = \text{acquiesce}$	-1	1	$a_i = \text{acquiesce}$	1	-1
(a) Payoffs when $\omega = 1$.			(b) Payoffs when $\omega = 0$.		

Table 2. Citizens' payoffs in different states.

Citizens' payoffs are detailed in Table 2. Citizens want to topple a regime if and only if it misbehaves. In particular, each citizen receives a payoff that is equal to 1 if either (i) the regime remains in place when it is acting pro-socially or (ii) the regime is toppled when it behaves anti-socially; otherwise, they get a payoff of -1 . In addition, each participant in a protest that leads to the toppling of the regime gets a warm-glow payoff that is equal to $g > 0$ when the regime acts anti-socially and is equal to $-g$ when the regime acts pro-socially. Finally, participating in a protest lowers a citizen's payoff by $r > 0$, where r is the repression level.

	$A \geq \theta$	$A < \theta$
$a_i = \text{protest}$	$(1 - 2\mu)(1 + g) - r$	$(2\mu - 1) - r$
$a_i = \text{acquiesce}$	$1 - 2\mu$	$2\mu - 1$

Table 3. Citizens' expected payoffs.

We can express citizens' expected payoffs in terms of their posterior belief about the regime behavior. Let $\mu_m = \Pr(\omega = 1|m) \in [0, 1]$ denote the posterior belief following m ; for brevity, we drop the subscript and write $\mu \in [0, 1]$ for a generic posterior belief. Then, citizen i 's expected payoffs are given in Table 3. We assume that $r < g$, $\bar{\theta} > 1 + 2\rho$, and $\underline{\theta} < -2\rho$. These assumptions ensure that protesting is not a dominated strategy and that there are *dominance regions*, i.e., there are some private signals that allow the citizens to conclude that the regime will surely fail or surely survive.

We characterize the citizens' decision to protest using the machinery of global games (Carlsson and Van Damme, 1993; Morris and Shin, 1998, 2003). We consider a symmetric cutoff

strategy equilibrium of the coordination game where $a_i = \text{protest}$ if and only if $x_i \leq x^*$. The following lemma characterizes the equilibrium using standard arguments from the global games literature.

Lemma 3. *Given $\mu \in [0, 1]$,*

- *If $\mu > \frac{1}{2} \left(1 - \frac{r}{g}\right)$, in any equilibrium, the regime remains in power if and only if $\theta > 0$.*
- *If $\mu \leq \frac{1}{2} \left(1 - \frac{r}{g}\right)$, in any symmetric cutoff strategy equilibrium, the regime remains in power if and only if $\theta > \theta^*$, where*

$$\theta^* = 1 - \frac{r}{g(1 - 2\mu)}.$$

The regime designs the institution of censorship to maximize its likelihood of survival. Specifically, we assume that the regime's payoff is given by $-\mathbf{1}_{A \geq \theta}$, i.e., the regime gets a payoff of zero if it stays in place and a payoff of -1 if there is a successful protest and it is toppled. Given Lemma 3, the regime's expected payoff when citizens hold posterior μ is given by

$$\pi(\mu; r) \equiv \begin{cases} -H\left(1 - \frac{r}{g(1-2\mu)}\right) & \text{if } \mu \leq \frac{1}{2} \left(1 - \frac{r}{g}\right), \\ -H(0) & \text{if } \mu > \frac{1}{2} \left(1 - \frac{r}{g}\right), \end{cases} \quad (8)$$

where H is the cdf of $U[\underline{\theta}, \bar{\theta}]$. This function is continuous in μ ; it is strictly convex in μ when $\mu \leq \frac{1}{2} \left(1 - \frac{r}{g}\right)$. Thus, the regime's optimal censorship intensity is given by

$$\beta^*(r) = \arg \max_{\beta \in [0,1]} (1 - \mu_0)(1 - \beta) \cdot \pi(0; r) + (\mu_0 + (1 - \mu_0)\beta) \cdot \pi\left(\frac{\mu_0}{\mu_0 + (1 - \mu_0)\beta}; r\right). \quad (9)$$

Proposition 4. *In the coordination game, $\beta^*(r)$ is single-valued with*

$$\beta^*(r) = \min \left\{ \frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r}, 1 \right\}.$$

Moreover, $\beta^*(r)$ is increasing in r .

The last part of Proposition 4 parallels Proposition 1 and shows that the key insight on the complementarity of censorship and repression carries over to this setting.

Let $\hat{\mu} = \frac{1}{2} \left(1 - \frac{r}{g}\right)$ denote the cutoff belief from Lemma 3. This is the lowest posterior at which citizens are unwilling to protest even if they are certain that the protest will succeed. Thus, once the posterior following silence reaches $\hat{\mu}$, greater confidence that the regime has behaved pro-socially cannot further improve its survival probability. If $\mu_0 \leq \hat{\mu}$, Proposition 4 shows that

the regime chooses the highest censorship intensity consistent with inducing precisely this posterior following silence.

Why does the regime not censor even more and accept a posterior below $\hat{\mu}$? Below this threshold, citizens' willingness to protest depends on their expectations about whether the protest will succeed. A less favorable posterior makes protest more attractive, which increases participation and, through the coordination motive, makes a successful protest more likely. This feedback makes the loss from reducing the credibility of silence particularly costly. In our specification, that loss outweighs the benefit from making silence more frequent. Higher repression weakens this feedback; it lowers $\hat{\mu}$, so citizens require less reassurance about the regime's behavior before acquiescing. The regime can therefore make silence less credible and censor more intensely without inducing coordinated protests.

6 Conclusion

In this paper, we offer a model of censorship and repression, two instruments of authoritarian control. We model censorship as a form of information manipulation whereby the regime controls the disclosure of adverse information. We consider various types of repression. In our baseline setting, repression determines the citizens' incentives to protest against the regime, but we also consider forms of repression that alter the distributions of citizens' beliefs.

Our main finding is that increases in repression tend to be accompanied by more intense censorship. Censorship involves a trade-off between the frequency and credibility of silence: suppressing adverse reports makes silence more frequent but less credible. When repression increases, citizens become more predisposed to acquiesce given a certain level of skepticism against the regime. The regime can therefore increase censorship intensity, making silence more frequent, while facing less severe consequences from the associated loss of credibility. A society where people have more dispersed views poses a problem for an authoritarian regime as censorship is less effective in such circumstances, so interventions that increase the concentration of beliefs also tend to increase censorship. The complementarity of censorship and repression survives when protest is a coordination game capable of toppling the regime. These findings all illustrate how censorship and repression can reinforce one another rather than serve as alternative instruments of authoritarian control.

In George Orwell's totalitarian dystopia, citizens of Oceania are forced to use the *newspeak*,

a special language designed to limit their ability to articulate antigovernment concepts, cannot switch off the radio that transmits propaganda, and are forced to participate in ideological indoctrination meetings. However, the ultimate message of *1984* is that it is physical torture, applied to some, that makes citizens believe what the government wants them to believe.

References

- Alonso, Ricardo and Odilon Câmara. 2016. "Bayesian Persuasion with Heterogeneous Priors." *Journal of Economic Theory* 165:672–706.
- Alyukov, Maxim. 2022. "Propaganda, Authoritarianism and Russia's Invasion of Ukraine." *Nature Human Behaviour* 6(6):763–765.
- Anisin, Alexei. 2024. "Fraternisation and Repression During the 2020–2021 Attempted Revolution in Belarus." *Contemporary Politics* pp. 1–21.
- Applebaum, Anna. 2024. *Autocracy, Inc.: The Dictators Who Want to Run the World*. Diversified Publishing.
- Arendt, Hannah. 1951. *The Origins of Totalitarianism*. New York: Harcourt and Brace.
- Bagnoli, Mark and Ted Bergstrom. 2005. "Log-Concave Probability and Its Applications." *Economic Theory* 26(2):445–469.
- Barbera, Salvador and Matthew O. Jackson. 2020. "A Model of Protests, Revolution, and Information." *Quarterly Journal of Political Science* 15(3):297–335.
- Bergemann, Dirk and Stephen Morris. 2019. "Information Design: A Unified Perspective." *Journal of Economic Literature* 57(1):44–95.
- Bueno de Mesquita, Ethan. 2010. "Regime Change and Revolutionary Entrepreneurs." *American Political Science Review* 104(3):446–466.
- Carlsson, Hans and Eric Van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61(5):989–1018.
- Conquest, Robert. 2008. *The Great Terror: A Reassessment*. Oxford University Press on Demand.
- Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):1431–1451.
- Curello, Gregorio and Ludvig Sinander. 2025. "The Comparative Statics of Persuasion." *Working Paper*.
- Di Tella, Rafael and Ignacio Franceschelli. 2011. "Government Advertising and Media Coverage of Corruption Scandals." *American Economic Journal: Applied Economics* 3(4):119–51.

- Dimitrov, Martin K. 2023. *Dictatorship and Information: Authoritarian Regime Resilience in Communist Europe and China*. Oxford University Press.
- Dorsch, Michael and Paul Maarek. 2018. "Rent Extraction, Revolutionary Threat, and Coups in Non-democracies." *Journal of Comparative Economics* 46(4):1082–1103.
- Dragomir, Silvestru Sever, Muhammad Adil Khan and Addisalem Abathun. 2016. "Refinement of the Jensen Integral Inequality." *Open Mathematics* 14(1):221–228.
- Dragu, Tiberiu and Adam Przeworski. 2019. "Preventive Repression: Two Types of Moral Hazard." *American Political Science Review* 113(1):77–87.
- Durante, Ruben and Brian Knight. 2012. "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy." *Journal of the European Economic Association* 10(3):451–481.
- Edlin, Aaron S. and Chris Shannon. 1998. "Strict Monotonicity in Comparative Statics." *Journal of Economic Theory* 81(1):201–219.
- Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies* 80(4):1422–1458.
- Egorov, Georgy and Konstantin Sonin. 2021. "Elections in Non-Democracies." *Economic Journal* 131(636):1682–1716.
- Egorov, Georgy and Konstantin Sonin. 2024. "The Political Economics of Non-democracy." *Journal of Economic Literature* 62(2):594–636.
- Enikolopov, Ruben, Michael Rochlitz, Koen J.L. Schoors and Nikita Zakharov. 2025. "Polarize and Rule: Independent Media in Autocracy and the Role of Social Media." *Working Paper, Available at SSRN 4131355*.
- Esteban, Joan, Massimo Morelli and Dominic Rohner. 2015. "Strategic Mass Killings." *Journal of Political Economy* 123(5):1087–1132.
- Friedrich, Carl J. and Zbigniew K. Brzezinski. 1956. *Totalitarian Dictatorship and Autocracy*. Cambridge, MA: Harvard University Press.
- Fukuyama, Francis. 1992. *The End of History and the Last Man*. Penguin Adult.

- Gehlbach, Scott. 2010. "Reflections on Putin and the Media." *Post-Soviet Affairs* 26(1):77–87.
- Gehlbach, Scott and Philip Keefer. 2011. "Investment without Democracy: Ruling-Party Institutionalization and Credible Commitment in Autocracies." *Journal of Comparative Economics* 39(2):123–139.
- Gehlbach, Scott, Zhaotian Luo, Anton Shirikov and Dmitriy Vorobyev. 2026. "Is There Really a Dictator's Dilemma? Information and Repression in Autocracy." *American Journal of Political Science* 70(1):381–395.
- Gentzkow, Matthew and Jesse M. Shapiro. 2008. "Competition and Truth in the Market for News." *Journal of Economic Perspectives* 22(2):133–154.
- Glaeser, Edward L. and Andrei Shleifer. 2005. "The Curley Effect: The Economics of Shaping the Electorate." *Journal of Law, Economics, and Organization* 21(1):1–19.
- Gratton, Gabriele and Barton E. Lee. 2024. "Liberty, Security, and Accountability: The Rise and Fall of Illiberal Democracies." *Review of Economic Studies* 91(1):340–371.
- Gregory, Paul R., Philipp J. H. Schröder and Konstantin Sonin. 2011. "Rational Dictators and the Killing of Innocents: Data from Stalin's Archives." *Journal of Comparative Economics* 39(1):34–42.
- Guriev, Sergei and Daniel Treisman. 2019. "Informational Autocrats." *Journal of Economic Perspectives* 33(4):100–127.
- Guriev, Sergei and Daniel Treisman. 2020. "A Theory of Informational Autocracy." *Journal of Public Economics* 186:104158.
- Guriev, Sergei and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century*. Princeton, New Jersey: Princeton University Press.
- Harrison, Mark. 2023. *Secret Leviathan: Secrecy and State Capacity Under Soviet Communism*. Stanford University Press.
- Heo, Kun and Antoine Zerbini. 2025. "Segment and Rule: Modern Censorship in Authoritarian Regimes." *Working Paper*.
- Huang, Haifeng. 2018. "The Pathology of Hard Propaganda." *Journal of Politics* 80(3):1034–1038.

- Johnson, Justin P. and David P. Myatt. 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review* 96(3):756–784.
- Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101(6):2590–2615.
- King, Gary, Jennifer Pan and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2):326–343.
- Kolotilin, Anton, Tymofiy Mylovanov and Andriy Zapechelnyuk. 2022. "Censorship as Optimal Persuasion." *Theoretical Economics* 17(2):561–585.
- Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk and Ming Li. 2017. "Persuasion of a Privately Informed Receiver." *Econometrica* 85(6):1949–1964.
- Kuran, Timur. 1989. "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice* 61(1):41–74.
- Lipnowski, Elliot, Doron Ravid and Denis Shishkin. 2022. "Persuasion via Weak Institutions." *Journal of Political Economy* 130(10):2705–2730.
- Little, Andrew T. 2012. "Elections, Fraud, and Election Monitoring in the Shadow of Revolution." *Quarterly Journal of Political Science* 7(3):249–283.
- Lohmann, Susanne. 1993. "A Signaling Model of Informative and Manipulative Political Action." *American Political Science Review* 87(2):319–333.
- Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2):402–414.
- Lorentzen, Peter L. 2013. "Regularizing Rioting: Permitting Public Protest in an Authoritarian Regime." *Quarterly Journal of Political Science* 8(2):127–158.
- Machiavelli, Niccolo. 1532. [2019]. *The Prince*. Cambridge Texts in the History of Political Thought 2nd ed. Cambridge: Cambridge University Press.
- McMillan, John and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives* 18(4):69–92.

- Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62(1):157–180.
- Montagnes, Pablo and Stephane Wolton. 2019. "Mass Purges: Top-Down Accountability in Autocracy." *American Political Science Review* 113(4):1045–1059.
- Morris, Stephen and Hyun Song Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* pp. 587–597.
- Morris, Stephen and Hyun Song Shin. 2003. Global Games: Theory and Applications. In *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Vol. I*, ed. Mathias Dewatripont, Lars Peter Hansen and Stephen J. Turnovsky. Cambridge: Cambridge University Press.
- Myerson, Roger B. 2015. "Moral Hazard in High Office and the Dynamics of Aristocracy." *Econometrica* 83(6):2083–2126.
- Persson, Torsten and Guido Tabellini. 2009. "Democratic Capital: The Nexus of Political and Economic Change." *American Economic Journal: Macroeconomics* 1(2):88–126.
- Qian, Nancy and David Yanagizawa-Drott. 2017. "Government Distortion in Independently Owned Media: Evidence from US News Coverage of Human Rights." *Journal of the European Economic Association* 15(2):463–499.
- Rozenas, Arturas. 2020. "A Theory of Demographically Targeted Repression." *Journal of Conflict Resolution* 64(7-8):1254–1278.
- Rozenas, Arturas and Denis Stukal. 2019. "How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television." *The Journal of Politics* 81(3):982–996.
- Shadmehr, Mehdi. 2014. "Why Don't Regimes Publicize Repression?" *Working Paper*.
- Shadmehr, Mehdi. 2019. "Multiplicity and Uniqueness in Regime Change Games." *The Journal of Politics* 81(1):303–308.
- Shadmehr, Mehdi and Dan Bernhardt. 2011. "Collective Action with Uncertain Payoffs: Coordination, Public Signals, and Punishment Dilemmas." *American Political Science Review* 105(4):829–851.

- Shaked, Moshe and J. George Shanthikumar. 2007. *Stochastic Orders*. New York, NY: Springer Science & Business Media.
- Stanig, Piero. 2015. "Regulation of Speech and Media Coverage of Corruption: An Empirical Analysis of the Mexican Press." *American Journal of Political Science* 59(1):175–193.
- Stoner, Kathryn. 2023. "The Putin Myth." *Journal of Democracy* 34(2):5–18.
- Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. New York: Cambridge University Press.
- Szeidl, Adam and Ferenc Szucs. 2021. "Media Capture through Favor Exchange." *Econometrica* 89(1):281–310.
- Taubman, William. 2003. *Khrushchev: The Man and His Era*. WW Norton & Company.
- Treisman, Daniel. 2011. "Presidential Popularity in a Hybrid Regime: Russia Under Yeltsin and Putin." *American Journal of Political Science* 55(3):590–609.
- Treisman, Daniel. 2022. "Putin Unbound: How Repression at Home Presaged Belligerence Abroad." *Foreign Affairs* 101(3):40–53.
- Tyson, Scott A. 2018. "The Agency Problem Underlying Repression." *The Journal of Politics* 80(4):1297–1310.
- Tyson, Scott and Alastair Smith. 2018. "Dual-Layered Coordination and Political Instability: Repression, Co-optation, and the Role of Information." *Journal of Politics* 80(1):44–58.
- Way, Lucan and Amelie Tolvin. 2023. "Why the 2020 Belarusian Protests Failed to Oust Lukashenka." *Nationalities Papers* 51(4):787–802.
- Wintrobe, Ronald. 1990. "The Tinpots and the Totalitarian: An Economic Theory of Dictatorship." *American Political Science Review* 84(3):849–872.

Appendices

A Robustness and extensions

In this appendix, we discuss alternatives to our main assumptions and the robustness of our main results. Specifically, we focus on the assumptions of full commitment by the regime, the binary action space, and heterogeneous priors. Moreover, in Section A.6, we allow the regime to choose among a richer set of information manipulation strategies and show that, as long as the distribution of priors is strictly log-concave, a censorship strategy remains optimal.

A.1 Partial commitment

In this appendix, we consider a more general model that does not assume full commitment on the regime's part. Suppose the regime can censor an adverse report *ex post* with probability $p \in [0, 1)$. The citizens do not observe whether the regime received the *ex post* opportunity to censor. The $p = 0$ case corresponds to the model studied so far. When $p > 0$, we have a more general model of censorship without full commitment, as in Lipnowski, Ravid and Shishkin (2022). The $p = 1$ limit corresponds to the no-commitment protocol, or the cheap talk benchmark of Crawford and Sobel (1982).¹⁴

Our qualitative insights extend to this model straightforwardly. Given censorship intensity β and *ex post* censorship probability p , the *ex ante* probability that the regime assigns to citizens observing $m = \text{silence}$ is

$$\mu_s + (1 - \mu_s) (\beta + (1 - \beta)p).$$

Citizen i 's posterior that $\omega = 1$ after observing $m = \text{silence}$ is

$$\frac{\mu_i}{\mu_i + (1 - \mu_i) (\beta + (1 - \beta)p)},$$

and the posterior that $\omega = 1$ after observing $m = \text{report}$ is zero. Therefore, the regime's optimal censorship intensity is given by

$$\begin{aligned} \beta_p^*(r) = \arg \max_{\beta \in [0,1]} & (1 - \mu_s)(1 - \beta)(1 - p) \cdot (-1) \\ & + (\mu_s + (1 - \mu_s) (\beta + (1 - \beta)p)) \cdot \left(-F \left(\frac{\beta + (1 - \beta)p}{\beta + (1 - \beta)p + \frac{r}{1-r}} \right) \right). \end{aligned}$$

¹⁴In the cheap talk benchmark of this model, there is no information transmission in any equilibrium. If there were, the regime would strictly prefer to send some messages over others in every state, and it would have a profitable deviation. Throughout this section, we ignore the $p = 1$ case.

The regime's objective function is a modified version of (3), and the two functions coincide when $p = 0$. Indeed, we can write the regime's problem as that of choosing the *effective censorship intensity* $\hat{\beta} = \beta + (1 - \beta)p \in [p, 1]$. The regime's optimization problem can then be written as follows:

$$\hat{\beta}_p^*(r) = \arg \max_{\hat{\beta} \in [p, 1]} V(\hat{\beta}; r).$$

Thus, partial commitment does not change the regime's objective as a function of the effective censorship intensity $\hat{\beta}$. It only imposes a lower bound on effective censorship, requiring it to be at least p . For any optimal effective censorship intensity level $\hat{\beta} \in \hat{\beta}_p^*(r)$, the corresponding censorship intensity is optimal: $\frac{\hat{\beta}-p}{1-p} \in \beta_p^*(r)$. Note that $\frac{\hat{\beta}-p}{1-p} \leq \hat{\beta}$. As in the running example of [Lipnowski, Ravid and Shishkin \(2022\)](#), the regime sets up a less stringent censorship institution (a more informative structure) to compensate for the fact that the beliefs following $m = \text{silence}$ are distorted downward because citizens realize that the regime may have censored the message ex post.

This change of variables allows us to compare $\beta^*(r)$, the censorship intensity under full commitment, and $\beta_p^*(r)$, the committed censorship intensity under partial commitment. Consider some $\hat{\beta} \in \beta^*(r)$. If $\hat{\beta} \geq p$, we directly conclude that $\frac{\hat{\beta}-p}{1-p} \in \beta_p^*(r)$. If $\hat{\beta} < p$, however, one needs to impose extra assumptions on the shape of $V(\beta; r)$ to ensure that the solution remains well-behaved. Recall that when f is strictly log-concave, $V(\beta; r)$ is single-peaked by the proof of Lemma 1. The following result is then obtained as a direct corollary.

Corollary A-1. *Given $p \in [0, 1)$, if f is strictly log-concave, the optimal committed censorship intensity $\beta_p^*(r)$ is single-valued and satisfies:*

$$\beta_p^*(r) = \max \left\{ \frac{\beta^*(r) - p}{1 - p}, 0 \right\}.$$

The preceding discussion and Corollary A-1 reveal that the overarching insights under full commitment carry over to the case of partial commitment. That is, even with partial commitment, the censorship intensity is increasing in repression and concentration of opinions.

A.2 Non-binary action spaces

Our choice of a binary action space is guided by the common assumption in the political economy literature that authoritarian regimes maximize their chances of staying in power—even if

they also have other, secondary motives. They do so by trying to prevent coups, revolutions, or massive protests that would lead to the ouster of the regime. [Svolik \(2012\)](#) provides data showing that these types of exit cover the overwhelming majority of autocrats' exits; see [Dorsch and Maarek \(2018\)](#), and [Egorov and Sonin \(2024\)](#), for updated statistics. Following the early formal theories of autocrats' critical moments ([Kuran, 1989](#); [Lohmann, 1993](#)), nearly all modern models assume a binary choice, "support" ("abstain") versus "no support" ("rebel"), for citizens ([Persson and Tabellini, 2009](#); [Bueno de Mesquita, 2010](#); [Shadmehr and Bernhardt, 2011](#); [Little, 2012](#); [Edmond, 2013](#); [Tyson and Smith, 2018](#); [Shadmehr, 2019](#); [Barbera and Jackson, 2020](#); [Egorov and Sonin, 2021](#)).

Although the binary action space for citizens is a standard assumption in models of authoritarian control, it greatly simplifies our analysis of optimal censorship intensity. Specifically, the proof of Lemma A-2, which deals with private versus public censorship, does not readily extend to a larger action space.¹⁵ Still, the underlying frequency–credibility trade-off is not inherently tied to binary decisions. More intense censorship makes adverse reports less frequent but reduces the credibility of silence, potentially inducing citizens to take actions less favorable to the regime. Repression may mitigate this cost by discouraging such actions.

A.3 Heterogeneous payoffs and common priors

The choice of heterogeneous priors in our model is guided by the goal of being as realistic as possible. As previously discussed, in an authoritarian society, a change at the top requires collective action by citizens, which justifies the assumption of a binary action space. From a theoretical standpoint, assuming heterogeneous priors with a binary action space is essentially equivalent to assuming heterogeneous payoff functions for citizens.

In the next subsection, we present a version of the model featuring heterogeneous preferences rather than heterogeneous priors, and we show that our main insights carry over. With heterogeneous payoffs, citizens have the same priors about the regime's alignment yet have different individual payoffs from protesting. With heterogeneous priors, the payoffs are the same, but the subjective probabilities about the regime's alignment differ. The empirical literature on authoritarian transitions points to the heterogeneity of people's beliefs about the leader's quality (e.g., [Kuran, 1989](#)). In a recent study on the impact of propaganda amid

¹⁵Also see a discussion of the difference between binary and non-binary action spaces in [Kolotilin et al. \(2017\)](#).

increased repression on Russian attitudes towards the Russia-Ukraine war, [Alyukov \(2022\)](#) notes the heterogeneity of attitudes.

A.4 Continuous state spaces

We opted for a binary state space (rather than a continuous one) for tractability. Our qualitative insights nevertheless extend to continuous state spaces. Suppose the state of the world is distributed over $\Omega = [0, 1]$ with some continuous density. The belief of each citizen $i \in I$ is a distribution $F_i \in \Delta(\Omega)$, and a distribution of priors is an object in $\Delta(\Delta(\Omega))$. Generalizing our results to such a setting requires defining a measure of concentration as in Definition 1 on $\Delta(\Delta(\Omega))$, a much more difficult task. With a binary state, each citizen is represented by a single number μ_i ; thus, there is a natural order of citizens based on their priors and a simple transformation of the distribution of priors that captures the idea of more concentration.

However, one can adopt a setting with heterogeneous *preferences* and a continuous state space, which is the model in Section 5.1 of [Kolotilin, Mylovanov and Zapechelnyuk \(2022\)](#), as follows: There is a state of the world, denoted by $\omega \in \Omega = [0, 1]$. Here, ω is a measure of the extent to which the regime's behavior aligns with the citizens' interests, with higher numbers corresponding to more pro-social behavior. Citizens and the regime share a common prior that ω is distributed with a density $g(\omega)$. Citizens' payoff from acquiescing is normalized to zero, whereas citizen i 's payoff from protesting is $(1 - \omega) - z_i - r$, where $z_i \in (-1, 1)$ is citizen i 's protest cost. Citizens have heterogeneous payoffs, with $f(z_i)$ denoting the density of z_i 's in the population. We assume that z_i is the private information of citizen i and that f is common knowledge, continuously differentiable and bounded over its support.

There is a continuum of media outlets indexed with $c \in [0, 1]$. Each media outlet c perfectly observes ω and issues an adverse report on the regime's behavior if and only if $\omega \leq c$. Therefore, a media outlet with high c is more likely to report the regime's anti-social behavior. The regime engages in *censorship* by shutting down media sources with $c \geq \omega^*$.¹⁶ The remaining media sources follow their reporting strategies. Note that if $\omega < \omega^*$, citizens can perfectly deduce the value of ω by observing an adverse report from outlets with $c \in [\omega, \omega^*]$. Otherwise, citizens can only infer that $\omega \geq \omega^*$. In other words, a censorship strategy with cutoff ω^* reveals the states

¹⁶In the terminology of [Kolotilin, Mylovanov and Zapechelnyuk \(2022\)](#), this is the *upper censorship* strategy. It is indeed optimal in a broader class of mechanisms in an information design problem. Suppose the regime can commit to any message distribution $\sigma : \Omega \rightarrow \Delta(M)$. Without loss of generality, assume that $M = [0, 1]$ and messages are *direct*, i.e., $\mu = \mathbb{E}[\omega|\mu]$ for each $\mu \in M$. Then, following a message $\mu \in M$, citizen i protests if and only if $z_i \leq 1 - \mu - r$. The regime's payoff from generating posterior μ is $v(\mu; r) \equiv -F(1 - \mu - r)$. If f is single-peaked, $v(\mu; r)$ is strictly S-shaped. By Theorem 1 of [Kolotilin, Mylovanov and Zapechelnyuk \(2022\)](#), for any g , the unique optimal strategy is an upper censorship.

below ω^* and censors information regarding states above ω^* . Consequently, a low value of ω^* corresponds to more intense censorship. The regime's expected payoff from censorship cutoff ω^* is

$$v(\omega^*; r) = - \int_0^{\omega^*} F(1 - \omega - r) g(\omega) d\omega - \int_{\omega^*}^1 F(1 - \mu^* - r) g(\omega) d\omega, \quad (\text{A-1})$$

where

$$\mu^* \equiv \mathbb{E}[\omega | \omega \geq \omega^*] = \frac{\int_{\omega^*}^1 \omega g(\omega) d\omega}{1 - G(\omega^*)}.$$

Suppose f is strictly single-peaked. Then, $-F(1 - \omega - r)$ is strictly S-shaped in ω and, by Lemma 1 of [Kolotilin, Mylovanov and Zapechelnyuk \(2022\)](#), $v(\omega^*; r)$ is strictly quasi-concave in ω^* . Therefore, there is a unique optimal censorship cutoff. Moreover, Proposition 2 of [Kolotilin, Mylovanov and Zapechelnyuk \(2022\)](#) implies that the optimal ω^* decreases in r . That is, as in Proposition 1, higher repression leads to more intense censorship, and hence our insights in Section 3.2 continue to hold.

Regarding the relationship between censorship and concentration in Section 4.2, we have the following generalization of Lemma 2.

Lemma A-1. *Let f_0 and f_1 be two continuously differentiable densities supported on a common interval. For $k \in \{0, 1\}$, let $v_k(\omega^*; r)$ be as defined in (A-1):*

$$v_k(\omega^*; r) = - \int_0^{\omega^*} F_k(1 - \omega - r) g(\omega) d\omega - \int_{\omega^*}^1 F_k(1 - \mu^* - r) g(\omega) d\omega.$$

If f_1 is more concentrated than f_0 , then, for any $\omega^* \in (0, 1)$,

$$\frac{\partial v_0(\omega^*; r)}{\partial \omega^*} \leq 0 \implies \frac{\partial v_1(\omega^*; r)}{\partial \omega^*} \leq 0.$$

Lemma A-1 reveals that in a model with a continuum of states, more concentrated preferences give stronger incentives for reducing ω^* (i.e., increasing censorship intensity) to the regime. It can be coupled with, for instance, a strictly single-peaked density assumption to obtain a ranking of the global maximizers as in Corollary 1.

A.5 Targeted censorship

In the analysis of repression in Section 4.1, the crucial condition for the complementarity of censorship and ex ante repression was the reliance on repression being *targeted*. A natural question is whether the regime also benefits from targeted censorship. In this section, we argue

that the answer, again, depends on the ability of the regime to determine the citizens' beliefs. If the regime knows the private beliefs of citizens, then it can benefit from this knowledge, targeting information that citizens receive. However, if the regime has to rely on citizens' self-selection into a unique source of information and design a type-specific censorship strategy—that is, allowing for *private censorship*—then the regime's toolkit is not actually expanded. We show that it is never (strictly) optimal for the regime to create different information sources that adopt different censorship strategies. Therefore, under the optimal policy, the regime continues to use public censorship, supplemented by repression as in Section 3.

The first observation is almost trivial from a theoretical point of view: If the regime can address each citizen individually, then it can achieve the maximum possible level of persuasion and ultimately minimize the size of protests. Critically, this requires the regime to keep each citizen in the dark about the messages that other citizens receive. However, sending individualized messages to citizens would require the regime to know each individual's belief. In authoritarian regimes, dissidents have a strong incentive to conceal their type out of the fear of being singled out for repression. The regime then may need to rely on each citizen to self-select into receiving the message intended for her type. To analyze such a situation, we need to introduce the formalism of *incentive compatible censorship*, which is the adaptation of incentive compatible persuasion (Kolotilin et al., 2017; Bergemann and Morris, 2019) into our setting.

Consider the setup in Section 2, but suppose that instead of relying on a public message m , the regime can use *private censorship* as follows. Each citizen reports her prior $\tilde{\mu} \in [0, 1]$. The regime then assigns a censorship intensity $\sigma(\tilde{\mu})$ to the citizen, and the citizen observes message $m_{\tilde{\mu}} \in \{\text{report}, \text{silence}\}$ generated according to the censorship intensity $\sigma(\tilde{\mu}) = \Pr(m_{\tilde{\mu}} = \text{silence} \mid \omega = 0)$. Let $\sigma = \{\sigma(\tilde{\mu})\}_{\tilde{\mu} \in [0,1]}$ denote a **private censorship mechanism**.

Consider a receiver i with prior $\mu_i = \mu$. Her (subjective) payoff from reporting a prior $\tilde{\mu}$ is

$$\begin{aligned} U_{\sigma}(\mu, \tilde{\mu}) &\equiv (1 - \mu) (1 - \sigma(\tilde{\mu})) (1 - r) + (\mu + (1 - \mu) \sigma(\tilde{\mu})) \max \left\{ 1 - r - \frac{\mu}{\mu + (1 - \mu) \sigma(\tilde{\mu})}, 0 \right\} \\ &= (1 - \mu) (1 - \sigma(\tilde{\mu})) (1 - r) + \max \{ (1 - \mu) \sigma(\tilde{\mu}) (1 - r) - \mu r, 0 \} \\ &= \max \{ 1 - \mu - r, (1 - \mu) (1 - \sigma(\tilde{\mu})) (1 - r) \} . \end{aligned}$$

Incentive compatibility is defined in the usual way: In an incentive compatible mechanism, no citizen could benefit from misreporting her belief. Formally, a private censorship mechanism σ is **incentive compatible** if, for all $\mu, \tilde{\mu} \in [0, 1]$, $U_{\sigma}(\mu, \mu) \geq U_{\sigma}(\mu, \tilde{\mu})$. Our next result shows that,

for any incentive compatible private censorship mechanism, there exists a public censorship strategy that achieves the same outcome.

Lemma A-2. *Consider an incentive compatible private censorship mechanism σ . Public censorship at intensity*

$$\hat{\beta} = \inf_{\tilde{\mu} \in [0,1]} \sigma(\tilde{\mu})$$

induces the same probability of protesting in each state for almost every citizen.

Lemma A-2 implies that the regime can achieve the same outcome through public censorship, where it imposes the same censorship intensity on each citizen. This justifies our focus on public censorship in Section 3 and implies that the optimal strategy we characterized yields a higher payoff to the regime than any incentive compatible private censorship mechanism. Intuitively, this is because the incentive compatibility constraints are extremely tight for the regime, to the extent that public censorship (which trivially satisfies incentive compatibility) can yield the same payoff. This result is a restatement of the result in Kolotilin et al. (2017) on the “impossibility of private persuasion,” adapted to a setting with heterogeneous priors.

Lemma A-2 provides a rationale for why authoritarian regimes might prefer a standardized approach to censorship. The reality is, of course, more complex. Vladimir Lenin, the founder of the communist state after the Russian Revolution, stated in 1921: “We need full and truthful information. And the truth should not depend on who it has to serve. We can accept only the division between the unofficial information (for the Comintern Executive Committee only) and official information (for everybody).”¹⁷ In a famous example of information presented differently to different audiences, the 1956 speech by Soviet leader Nikita Khrushchev denouncing Stalin’s repressions and cult of personality was distributed, in written form, to local communist party secretaries, read aloud to ordinary party members at specially arranged meetings, and kept secret from the rest of the population (Taubman, 2003). In effect, different groups of citizens were given access to information tailored to their classification. Critically, the Soviet leader relied on the existing designation of citizen types, with the communist party members presumably being more loyal than citizens at large. Lemma A-2 suggests that, absent the existing division of citizens into observable types, Khrushchev could not do better than to rely on public censorship.

¹⁷Lenin’s formula was employed by other communist regimes as well; see Gehlbach and Keefer (2011), King, Pan and Roberts (2013) and Lorentzen (2014) for the analysis of China’s information policy.

An intermediate approach between observable and non-observable types would be introducing a cost or benefit of accessing different information sources. If benefits of accessing a certain information source (say, due to its entertainment content) are sufficiently correlated with types, the regime can do better by choosing a cost of access so that citizens self-select.¹⁸

A.6 Information design

In this section, we allow the regime to choose any information manipulation strategy, rather than being restricted to choose among censorship strategies. We show that when the distribution of priors is strictly log-concave, the regime's optimal strategy turns out to be a censorship strategy.

Let the citizens' decisions and all payoff functions be as described in the main text. The only difference is as follows: Suppose, to persuade citizens to acquiesce, the regime uses a public persuasion mechanism. That is, the regime commits to an information structure $\{\sigma(\cdot|\omega)\}_{\omega \in \{0,1\}}$, where

$$\sigma(\cdot|\omega) \in \Delta(M) \quad \text{for all } \omega \in \{0, 1\},$$

and the realized message, $m \in M$, is publicly observable to each citizen and the regime. We assume that $|M|$ is large enough that there are sufficiently many action recommendations for each receiver.

To characterize the regime's optimal information structure given repression level $r \in (0, 1)$, we follow a standard approach to analyzing models of information manipulation (Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016). For each message $m \in M$, let $\mu_m = Pr_s(\omega = 1 | m)$ denote the regime's posterior belief following message m . We construct the **value function** of the regime as a function of its own posterior and then use the concavification approach of Kamenica and Gentzkow (2011). In what follows, we write $\mu \in [0, 1]$ for a generic realization of μ_m .¹⁹ Because the regime and citizens observe (and update their beliefs based on) the same public message, we can back out the posterior of citizen i from the regime's posterior. Using

¹⁸See Heo and Zerbini (2025) for a model with a single state-controlled sender and a single opposition media whose strategy the regime cannot control.

¹⁹Note that the regime perfectly learns the realization of the state. μ is what the regime's posterior would be had it not learned the state and instead only observed the message sent by the mechanism. For us, μ simply serves as a theoretical device that helps characterize the optimal persuasion mechanism.

the Bayes formula, citizen i has the following posterior:²⁰

$$\mu'_i = \frac{\mu \frac{\mu_i}{\mu_s}}{\mu \frac{\mu_i}{\mu_s} + (1 - \mu) \frac{1 - \mu_i}{1 - \mu_s}}. \quad (\text{A-2})$$

Consider citizen i with posterior μ'_i . Given the discussion in the main text, she protests if and only if $\mu'_i \leq 1 - r$. Substituting (A-2) into this inequality and rearranging terms, we conclude that any citizen i with prior

$$\mu_i \leq \frac{1 - \mu}{1 - \mu + \mu \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s}}$$

protests. Given (1), the regime's value function as a function of its own posterior, μ , is

$$v(\mu; r) \equiv -F\left(\frac{1 - \mu}{1 - \mu + \mu \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s}}\right) = -F\left(\left[1 + \frac{\mu}{1 - \mu} \frac{r}{1 - r} \frac{1 - \mu_s}{\mu_s}\right]^{-1}\right). \quad (\text{A-3})$$

The optimal solution relies on the characterization of the concave closure of $v(\mu; r)$. An inspection of (A-3) immediately reveals that $v(\mu; r)$ is strictly increasing in μ , with $v(0; r) = -1$ and $v(1; r) = 0$. The next result characterizes the shape of the value function when the distribution of priors is strictly log-concave.

Lemma A-3. *Suppose f is strictly log-concave. Then, $v(\mu; r)$ is strictly S-shaped, that is, there is some $\tilde{\mu} \in [0, 1]$ such that $v(\mu; r)$ is strictly convex for $\mu \in [0, \tilde{\mu}]$ and strictly concave for $\mu \in [\tilde{\mu}, 1]$.*

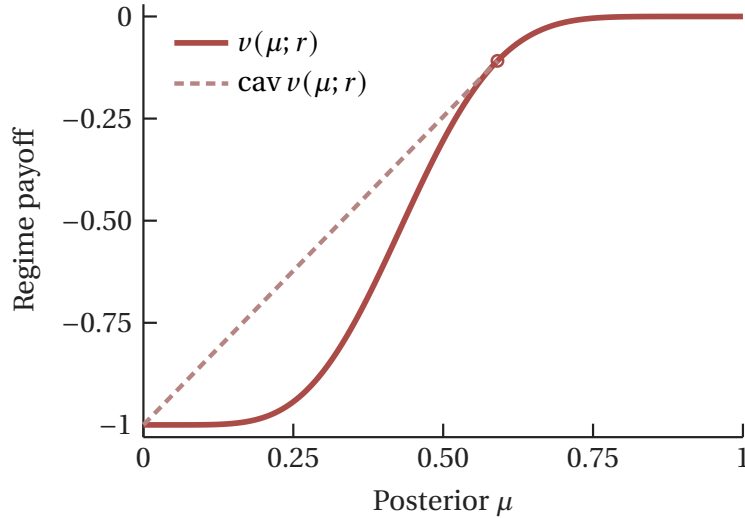


Figure A-1. The regime's value function and its concavification.

Figure A-1 depicts the regime's payoff as a function of its posterior. Visual inspection reveals that, from the regime's standpoint, the optimal information structure invokes two posteriors.

²⁰This expression is derived in [Alonso and Câmara \(2016\)](#).

This can be achieved by sending two messages about the regime's alignment: $m \in \{\text{report}, \text{silence}\}$. Moreover, one of the posteriors in the support is $\mu = 0$, i.e., one of the messages perfectly reveals that $\omega = 0$.²¹ This can be achieved by setting $\sigma(m = \text{silence}|\omega = 1) = 1$ in the optimal policy. Therefore, the optimal policy is characterized by a single-dimensional object:

$$\beta = \sigma(m = \text{silence}|\omega = 0) \in [0, 1].$$

The optimal information structure with a strictly log-concave prior distribution, therefore, ends up being a censorship strategy analyzed in the main text. Consequently, all our main findings regarding $\beta^*(r)$ apply to this more general problem.

B A general measure of concentration

In this section, we define a more general measure of concentration and show that the result of Lemma 2 holds even when $r > \mu_s$ under this more general measure.

Definition 2. Consider two distributions with densities f_0 and f_1 supported on a common compact set, and let $\gamma = \frac{r}{1-r} \frac{1-\mu_s}{\mu_s}$. f_1 is *generally more concentrated* than f_0 given γ if

$$f_1(x) \frac{(1-x+\gamma x)^2}{\gamma} = \alpha \left(f_0(x) \frac{(1-x+\gamma x)^2}{\gamma} \right) \quad \text{for all } x \quad (\text{A-4})$$

for some strictly increasing and convex function $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $\alpha(0) = 0$.

The generally more concentrated order is a generalization of the more concentrated order defined in Definition 1. When $r = \mu_s$ (i.e. $\gamma = 1$), the two definitions are equivalent. The interpretation also carries the same flavor: An adjusted version of $f_1(x)$ is obtained by transforming an adjusted version of $f_0(x)$ through a convex function. The adjustment terms include r and μ_s . Thus, Definition 2 can be interpreted as f_1 being more concentrated than f_0 after taking into account r and μ_s .

Our next result shows that Lemma 2 extends to the generally-more-concentrated order.

Lemma A-4. Let f_0 and f_1 be two densities that satisfy Assumption 1. For $k \in \{0, 1\}$, let $\Phi_k(\mu; r)$ be as defined in (5):

$$\Phi_k(\mu; r) = \log \left(1 - \mu + \frac{r}{1-r} \frac{1-\mu_s}{\mu_s} \mu \right) + \log \frac{1 - F_k(\mu)}{1 - \mu},$$

²¹When $v(\mu; r)$ is strictly S-shaped, there exists some $\hat{\mu} \in [0, 1]$ such that, for all $\mu \geq \hat{\mu}$, $\text{cav } v(\mu; r) = v(\mu; r)$. If $\mu_s \geq \hat{\mu}$, the optimal policy does not reveal any information. In this case, any policy satisfying $\sigma(m = \text{silence}|\omega = 0) = \sigma(m = \text{silence}|\omega = 1)$ is optimal. Among the many optimal policies, we choose the one where both probabilities equal one, so that $m = \text{report}$ is never observed. Consequently, the beliefs following $m = \text{report}$ are free, and we assign them the value zero as in footnote 7.

If f_1 is generally more concentrated than f_0 , then, for any $\mu \in [0, 1 - r]$,

$$\frac{\partial \Phi_0(\mu; r)}{\partial \mu} \geq 0 \implies \frac{\partial \Phi_1(\mu; r)}{\partial \mu} \geq 0.$$

C Omitted proofs

Proof of Lemma 1. As f is continuously differentiable by Assumption 1, $\Phi'(\mu; r) = \frac{\partial \Phi(\mu; r)}{\partial \mu}$ and $\Phi''(\mu; r) = \frac{\partial^2 \Phi(\mu; r)}{\partial \mu^2}$ exist. We show that, for any $\hat{\mu} \in (0, 1 - r)$,

$$\Phi'(\hat{\mu}; r) = 0 \implies \Phi''(\hat{\mu}; r) < 0. \quad (\text{A-5})$$

By (5),

$$\Phi'(\mu; r) = \frac{\gamma}{(1 - \mu)(1 - \mu + \gamma\mu)} - \frac{f(\mu)}{1 - F(\mu)}, \quad \text{where } \gamma \equiv \frac{r}{1 - r} \frac{1 - \mu_s}{\mu_s} > 0. \quad (\text{A-6})$$

$$\Phi''(\mu; r) = \frac{1}{(1 - \mu)^2} - \left(\frac{\gamma - 1}{1 - \mu + \gamma\mu} \right)^2 - \frac{f'(\mu)}{1 - F(\mu)} - \left(\frac{f(\mu)}{1 - F(\mu)} \right)^2. \quad (\text{A-7})$$

Suppose $\Phi'(\hat{\mu}; r) = 0$ for some $\hat{\mu} \in (0, 1 - r)$. By (A-6),

$$\frac{f(\hat{\mu})}{1 - F(\hat{\mu})} = \frac{\gamma}{(1 - \hat{\mu})(1 - \hat{\mu} + \gamma\hat{\mu})} = \frac{\gamma - 1}{1 - \hat{\mu} + \gamma\hat{\mu}} + \frac{1}{1 - \hat{\mu}}.$$

Substituting this into (A-7) yields:

$$\begin{aligned} \Phi''(\hat{\mu}; r) &= \frac{1}{(1 - \hat{\mu})^2} - \left(\frac{f(\hat{\mu})}{1 - F(\hat{\mu})} - \frac{1}{1 - \hat{\mu}} \right)^2 - \frac{f'(\hat{\mu})}{1 - F(\hat{\mu})} - \left(\frac{f(\hat{\mu})}{1 - F(\hat{\mu})} \right)^2 \\ &= -2 \left(\frac{f(\hat{\mu})}{1 - F(\hat{\mu})} \right)^2 + 2 \frac{f(\hat{\mu})}{1 - F(\hat{\mu})} \frac{1}{1 - \hat{\mu}} - \frac{f'(\hat{\mu})}{1 - F(\hat{\mu})} \\ &= - \underbrace{\frac{f(\hat{\mu})}{1 - F(\hat{\mu})}}_{>0 \text{ by Assumption 1}} \left(2 \frac{f(\hat{\mu})}{1 - F(\hat{\mu})} - \frac{2}{1 - \hat{\mu}} + \frac{f'(\hat{\mu})}{f(\hat{\mu})} \right) \end{aligned} \quad (\text{A-8})$$

Since f is strictly log-concave, $\log f$ is strictly concave; therefore, for any $x \in (\hat{\mu}, 1)$,

$$\log f(x) < \log f(\hat{\mu}) + (x - \hat{\mu}) \left. \frac{d \log f(\mu)}{d \mu} \right|_{\mu=\hat{\mu}} = \log f(\hat{\mu}) + (x - \hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})}.$$

Rearranging,

$$\log \frac{f(x)}{f(\hat{\mu})} < (x - \hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})} \implies \frac{f(x)}{f(\hat{\mu})} < e^{(x - \hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})}}.$$

Integrating both sides from $x = \hat{\mu}$ to $x = 1$,

$$\frac{1 - F(\hat{\mu})}{f(\hat{\mu})} < \int_{\hat{\mu}}^1 e^{(x - \hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})}} dx \implies \frac{f(\hat{\mu})}{1 - F(\hat{\mu})} > \left[\int_{\hat{\mu}}^1 e^{(x - \hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})}} dx \right]^{-1}.$$

We consider two cases.

i. Suppose $f'(\hat{\mu}) \neq 0$. Then,

$$\begin{aligned} \frac{f(\hat{\mu})}{1-F(\hat{\mu})} &> \left[\int_{\hat{\mu}}^1 e^{(x-\hat{\mu})\frac{f'(\hat{\mu})}{f(\hat{\mu})}} dx \right]^{-1} \implies \frac{f(\hat{\mu})}{1-F(\hat{\mu})} > \frac{f'(\hat{\mu})}{f(\hat{\mu})} \frac{1}{e^{(1-\hat{\mu})\frac{f'(\hat{\mu})}{f(\hat{\mu})}} - 1} \\ &\implies (1-\hat{\mu}) \frac{f(\hat{\mu})}{1-F(\hat{\mu})} > (1-\hat{\mu}) \frac{f'(\hat{\mu})}{f(\hat{\mu})} \frac{1}{e^{(1-\hat{\mu})\frac{f'(\hat{\mu})}{f(\hat{\mu})}} - 1}. \end{aligned}$$

Since $\frac{z}{e^z-1} \geq 1 - \frac{z}{2}$ for all $z \in \mathbb{R}$, this implies

$$(1-\hat{\mu}) \frac{f(\hat{\mu})}{1-F(\hat{\mu})} > 1 - \frac{1-\hat{\mu}}{2} \frac{f'(\hat{\mu})}{f(\hat{\mu})} \iff 2 \frac{f(\hat{\mu})}{1-F(\hat{\mu})} - \frac{2}{1-\hat{\mu}} + \frac{f'(\hat{\mu})}{f(\hat{\mu})} > 0.$$

ii. Suppose $f'(\hat{\mu}) = 0$. Then,

$$\begin{aligned} \frac{f(\hat{\mu})}{1-F(\hat{\mu})} &> \left[\int_{\hat{\mu}}^1 e^{(x-\hat{\mu})\frac{f'(\hat{\mu})}{f(\hat{\mu})}} dx \right]^{-1} \implies \frac{f(\hat{\mu})}{1-F(\hat{\mu})} > \frac{1}{1-\hat{\mu}} \\ &\implies 2 \frac{f(\hat{\mu})}{1-F(\hat{\mu})} - \frac{2}{1-\hat{\mu}} + \underbrace{\frac{f'(\hat{\mu})}{f(\hat{\mu})}}_{=0} > 0. \end{aligned}$$

In either case, we find $2 \frac{f(\hat{\mu})}{1-F(\hat{\mu})} - \frac{2}{1-\hat{\mu}} + \frac{f'(\hat{\mu})}{f(\hat{\mu})} > 0$. Substituting this into (A-8) yields $\Phi''(\hat{\mu}; r) < 0$ and (A-5) follows.

Note that $\Phi'(0; r) = \gamma > 0$. By (A-5), every zero of $\Phi'(\mu; r)$ is a crossing from positive to negative, and there can be at most one. If $\Phi'(\mu; r)$ has no zeroes, $\Phi(\mu; r)$ is strictly increasing in μ and $\mu^*(r) = 1 - r$ is the unique maximizer. In any case, $\mu^*(r)$ is single-valued. Then, single-valuedness of $\beta^*(r)$ follows since $\beta^*(r) = \left\{ \beta = \frac{\mu}{1-\mu} \frac{r}{1-r} : \mu \in \mu^*(r) \right\}$. $\square$

Proof of Proposition 1. Note that, since $\mu^*(r) = \left\{ \mu = \frac{\beta}{\beta + \frac{r}{1-r}} : \beta \in \beta^*(r) \right\}$, interiority of $\beta^*(r)$ implies that $\mu^*(r)$ is interior. That is, $\beta^*(r) \subset (0, 1)$ implies $\mu^*(r) \subset (0, 1 - r)$. Then, there exist $\underline{\mu}$ and $\bar{\mu}$ such that

$$0 < \underline{\mu} < \min \mu^*(r) \leq \max \mu^*(r) < \bar{\mu} < 1 - r.$$

Since $\mu^*(\cdot)$ is upper hemicontinuous at r by Berge's maximum theorem, there is an open interval neighborhood R of r such that, for all $\tilde{r} \in R$,

$$0 < \underline{\mu} < \min \mu^*(\tilde{r}) \leq \max \mu^*(\tilde{r}) < \bar{\mu} < 1 - \tilde{r}.$$

Moreover, for any $\tilde{r} \in R$, by (5),

$$\frac{\partial^2 \Phi(\mu; \tilde{r})}{\partial \mu \partial \tilde{r}} = \frac{1}{\left(1 - \mu + \frac{\tilde{r}}{1-\tilde{r}} \frac{1-\mu_s}{\mu_s} \mu\right)^2} \frac{1-\mu_s}{\mu_s} \frac{1}{(1-\tilde{r})^2} > 0. \quad (\text{A-9})$$

Take any $r_1, r_2 \in R$ with $r_1 < r_2$. Since $\mu^*(r_1)$ is compact, $\mu_1 := \max \mu^*(r_1)$ is well-defined. Similarly, since $\mu^*(r_2)$ is compact, $\mu_2 := \min \mu^*(r_2)$ is well-defined.

We first show that $\mu_1 \leq \mu_2$. Suppose, towards a contradiction, that $\mu_1 > \mu_2$. Then,

$$(\Phi(\mu_1; r_2) - \Phi(\mu_2; r_2)) - (\Phi(\mu_1; r_1) - \Phi(\mu_2; r_1)) = \int_{r_1}^{r_2} \underbrace{\int_{\mu_2}^{\mu_1} \frac{\partial^2 \Phi(\mu; \tilde{r})}{\partial \mu \partial \tilde{r}} d\mu d\tilde{r}}_{>0 \text{ by (A-9)}} > 0. \quad (\text{A-10})$$

Note that $\mu_1 \in (\underline{\mu}, \bar{\mu}) \subset (0, 1 - r_2)$, and therefore it is feasible at r_2 . Similarly, μ_2 is feasible at r_1 . By optimality of μ_1 and feasibility of μ_2 at r_1 ,

$$\Phi(\mu_1; r_1) - \Phi(\mu_2; r_1) \geq 0. \quad (\text{A-11})$$

Similarly, by optimality of μ_2 and feasibility of μ_1 at r_2 ,

$$\Phi(\mu_1; r_2) - \Phi(\mu_2; r_2) \leq 0. \quad (\text{A-12})$$

Combining (A-11) and (A-12) yields $(\Phi(\mu_1; r_2) - \Phi(\mu_2; r_2)) - (\Phi(\mu_1; r_1) - \Phi(\mu_2; r_1)) \leq 0$, a contradiction to (A-10). We conclude that $\mu_1 \leq \mu_2$.

To rule out equality, suppose towards a contradiction that $\mu_1 = \mu_2$. Then,

$$\left. \frac{\partial \Phi(\mu; r_2)}{\partial \mu} \right|_{\mu=\mu_2=\mu_1} - \left. \frac{\partial \Phi(\mu; r_1)}{\partial \mu} \right|_{\mu=\mu_1=\mu_2} = \int_{r_1}^{r_2} \underbrace{\left(\left. \frac{\partial^2 \Phi(\mu; \tilde{r})}{\partial \mu \partial \tilde{r}} \right|_{\mu=\mu_1=\mu_2} \right)}_{>0 \text{ by (A-9)}} d\tilde{r} > 0. \quad (\text{A-13})$$

Since $\mu^*(r_1) \subset (0, 1 - r_1)$, μ_1 is an interior optimizer and therefore satisfies the first-order condition:

$$\left. \frac{\partial \Phi(\mu; r_1)}{\partial \mu} \right|_{\mu=\mu_1} = 0. \quad (\text{A-14})$$

Similarly,

$$\left. \frac{\partial \Phi(\mu; r_2)}{\partial \mu} \right|_{\mu=\mu_2} = 0. \quad (\text{A-15})$$

Combining (A-14) and (A-15) yields $\left. \frac{\partial \Phi(\mu; r_2)}{\partial \mu} \right|_{\mu=\mu_2} - \left. \frac{\partial \Phi(\mu; r_1)}{\partial \mu} \right|_{\mu=\mu_1} = 0$, a contradiction to (A-13).

We conclude that $\max \mu^*(r_1) = \mu_1 < \mu_2 = \min \mu^*(r_2)$. Since $\beta^*(r) = \left\{ \beta = \frac{\mu}{1-\mu} \frac{r}{1-r} : \mu \in \mu^*(r) \right\}$ for all $r \in R$, we conclude $\max \beta^*(r_1) < \min \beta^*(r_2)$. $\square$

Proof of Proposition 2. The proof uses the following implication of the likelihood ratio order: If f_1 is larger than f_0 in the likelihood ratio order, by Theorem 1.C.1 of **Shaked and Shanthikumar**

(2007),

$$\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \text{ is increasing in } \mu. \quad (\text{A-16})$$

For $k \in \{0, 1\}$, let $V_k(\beta; r)$ denote the regime's objective function under density f_k . By (3),

$$V_k(\beta; r) = -1 + (\mu_s + (1 - \mu_s)\beta) \cdot \left(1 - F_k\left(\frac{\beta}{\beta + \frac{r}{1-r}}\right)\right).$$

Then, $\beta_k^*(r) = \arg \max_{\beta \in [0,1]} V_k(\beta; r)$. Take $\beta_0 \in \beta_0^*(r)$ and $\beta_1 \in \beta_1^*(r)$. If $\beta_1 \geq \beta_0$, then $\max\{\beta_0, \beta_1\} = \beta_1 \in \beta_1^*(r)$ and $\min\{\beta_0, \beta_1\} = \beta_0 \in \beta_0^*(r)$, and the result follows. For the remainder of the proof, assume $\beta_0 > \beta_1$. Then,

$$\begin{aligned} V_1(\beta_0; r) + 1 &= (\mu_s + (1 - \mu_s)\beta_0) \cdot \left(1 - F_1\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)\right) \\ &= (\mu_s + (1 - \mu_s)\beta_0) \cdot \left(1 - F_0\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)\right) \frac{1 - F_1\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)}{1 - F_0\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)} \\ &= (V_0(\beta_0; r) + 1) \frac{1 - F_1\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)}{1 - F_0\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)} \\ &\geq (V_0(\beta_1; r) + 1) \frac{1 - F_1\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)}{1 - F_0\left(\frac{\beta_0}{\beta_0 + \frac{r}{1-r}}\right)} && (\text{since } \beta_0 \in \beta_0^*(r)) \\ &\geq (V_0(\beta_1; r) + 1) \frac{1 - F_1\left(\frac{\beta_1}{\beta_1 + \frac{r}{1-r}}\right)}{1 - F_0\left(\frac{\beta_1}{\beta_1 + \frac{r}{1-r}}\right)} && (\text{since } \beta_0 > \beta_1 \text{ and by (A-16)}) \\ &= (\mu_s + (1 - \mu_s)\beta_1) \cdot \left(1 - F_0\left(\frac{\beta_1}{\beta_1 + \frac{r}{1-r}}\right)\right) \frac{1 - F_1\left(\frac{\beta_1}{\beta_1 + \frac{r}{1-r}}\right)}{1 - F_0\left(\frac{\beta_1}{\beta_1 + \frac{r}{1-r}}\right)} \\ &= V_1(\beta_1; r) + 1. \end{aligned}$$

Thus, $V_1(\beta_0; r) \geq V_1(\beta_1; r)$. Since $\beta_1 \in \beta_1^*(r)$, this implies $\beta_0 \in \beta_1^*(r)$. We conclude that $\max\{\beta_0, \beta_1\} \in \beta_1^*(r)$.

Moreover, since $\{\beta_0, \beta_1\} \subseteq \beta_1^*(r)$, we have $V_1(\beta_0; r) = V_1(\beta_1; r)$. Therefore, all weak inequalities in the above chain are equalities. This implies $V_0(\beta_0; r) = V_0(\beta_1; r)$. Along with $\beta_0 \in \beta_0^*(r)$, this establishes $\beta_1 = \min\{\beta_0, \beta_1\} \in \beta_0^*(r)$, and the result follows. $\square$

The proof of Proposition 3 uses the following auxiliary Lemma.

Lemma A-5. Consider a random variable X with full support $[0, \bar{x}]$ where $\bar{x} > 0$. Let $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a strictly increasing and convex function with $\alpha(0) = 0$. If

$$\mathbb{E}[\alpha(X)] = \alpha(\mathbb{E}[X])$$

then there exists some $a > 0$ such that

$$\alpha(x) = a \cdot x \quad \text{for all } x \in [0, \bar{x}].$$

Proof of Lemma A-5. Since the support of X is $[0, \bar{x}]$ with $\bar{x} > 0$, we have $\mathbb{E}[X] \in (0, \bar{x})$. By convexity of α , there exists an $a \in \mathbb{R}$ such that

$$\alpha(x) \geq \alpha(\mathbb{E}[X]) + a(x - \mathbb{E}[X]) \quad \text{for all } x \in [0, \bar{x}].$$

Let

$$h(X) \equiv \alpha(X) - \alpha(\mathbb{E}[X]) - a(X - \mathbb{E}[X]) \geq 0.$$

Taking expectations,

$$\mathbb{E}[h(X)] = \mathbb{E}[\alpha(X)] - \alpha(\mathbb{E}[X]) = 0.$$

Since $h(X) \geq 0$, this implies: $h(X) = 0$ almost surely.

Since α is finite and convex with $\alpha(0) = 0$, it is continuous on $[0, \bar{x}]$. Thus, h is continuous on $[0, \bar{x}]$. If $h(x) > 0$ for some $x \in [0, \bar{x}]$, continuity and full support would imply $\Pr(h(X) > 0) > 0$, a contradiction. Therefore,

$$h(x) = 0 \implies \alpha(x) = \alpha(\mathbb{E}[X]) + a(x - \mathbb{E}[X]) \quad \text{for all } x \in [0, \bar{x}].$$

Evaluating at $x = 0$ and using $\alpha(0) = 0$ yields: $\alpha(\mathbb{E}[X]) = a \cdot \mathbb{E}[X]$. Therefore,

$$\alpha(x) = a \cdot x \quad \text{for all } x \in [0, \bar{x}].$$

Finally, strict monotonicity implies $a = \alpha(\bar{x})/\bar{x} > 0$. □

We are now ready to present the proof of Proposition 3.

Proof of Proposition 3. Since f_1 is more concentrated than f_0 ,

$$f_1(x) = \alpha(f_0(x)) \quad \text{for all } x$$

for some strictly increasing and convex function α with $\alpha(0) = 0$. Then, for any $\varepsilon \in [0, 1]$, f_ε is more concentrated than f_0 because,

$$f_\varepsilon(x) = (1 - \varepsilon)f_0(x) + \varepsilon\alpha(f_0(x)) \quad \text{for all } x$$

where $\alpha_\varepsilon(z) = (1 - \varepsilon)z + \varepsilon\alpha(z)$ is a strictly increasing and convex function with $\alpha_\varepsilon(0) = 0$.

Moreover, since $\mu_0^*(r) = \left\{ \mu = \frac{\beta}{\beta + \frac{r}{1-r}} : \beta \in \beta_0^*(r) \right\}$, $\min \beta_0^*(r) \in (0, 1)$ implies $\min \mu_0^*(r) \in (0, 1 - r)$. Let $\mu_0 \equiv \min \mu_0^*(r)$.

As in (5), for any $\varepsilon \in [0, 1]$, let

$$\begin{aligned} \Phi_\varepsilon(\mu; r) &= \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \frac{1 - F_\varepsilon(\mu)}{1 - \mu} \\ &= \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \frac{1 - ((1 - \varepsilon)F_0(\mu) + \varepsilon F_1(\mu))}{1 - \mu} \\ &= \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \left(\frac{1 - F_0(\mu)}{1 - \mu} \left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right) \\ &= \log \left(1 - \mu + \frac{r}{1-r} \frac{1 - \mu_s}{\mu_s} \mu \right) + \log \frac{1 - F_0(\mu)}{1 - \mu} + \log \left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \\ &= \Phi_0(\mu; r) + \log \left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right). \end{aligned} \quad (\text{A-17})$$

Note that $\Phi_\varepsilon(\mu; r)$ is continuous in $(\mu, \varepsilon) \in [0, 1 - r] \times [0, 1]$. Then, by Berge's maximum theorem, the correspondence $\mu_\varepsilon^*(r)$ is upper hemicontinuous in ε at $\varepsilon = 0$. Then,

for every $\delta \in (0, \mu_0)$, there exists $\hat{\varepsilon} > 0$ such that

$$\mu_\varepsilon^*(r) \subset (\mu_0 - \delta, 1 - r] \quad \text{for all } \varepsilon \in [0, \hat{\varepsilon}]. \quad (\text{A-18})$$

Since $\mu_0 \in (0, 1 - r)$, it is an interior optimizer, and hence it satisfies the first-order condition: $\left. \frac{\partial \Phi_0(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0} = 0$. Since, for any $\varepsilon \in (0, 1]$, f_ε is more concentrated than f_0 , Lemma 2 implies:

$$\begin{aligned} \left. \frac{\partial \Phi_\varepsilon(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0} &\geq 0 \quad \underbrace{\iff}_{(\text{A-17})} \underbrace{\left. \frac{\partial \Phi_0(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0}}_{=0} + \frac{\varepsilon \left. \frac{\partial}{\partial \mu} \left(\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right|_{\mu=\mu_0}}{1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu_0)}{1 - F_0(\mu_0)}} \geq 0 \\ &\implies \left. \frac{\partial}{\partial \mu} \left(\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right|_{\mu=\mu_0} \geq 0. \end{aligned}$$

We next consider two exhaustive cases separately.

i. Suppose $\left. \frac{\partial}{\partial \mu} \left(\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right|_{\mu=\mu_0} > 0$. Then, by continuity of $\left. \frac{\partial}{\partial \mu} \left(\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right|_{\mu=\mu_0}$, there exists some $\delta \in (0, \mu_0)$ such that,

$$\left. \frac{\partial}{\partial \mu} \left(\frac{1 - F_1(\mu)}{1 - F_0(\mu)} \right) \right|_{\mu=\hat{\mu}} > 0 \quad \text{for all } \hat{\mu} \in (\mu_0 - \delta, \mu_0). \quad (\text{A-19})$$

By (A-18), there exists some $\bar{\varepsilon} > 0$ such that $\mu_\varepsilon^*(r) \subset (\mu_0 - \delta, 1 - r]$ for all $\varepsilon \in [0, \bar{\varepsilon})$.

Take any $\hat{\mu} \in (\mu_0 - \delta, \mu_0)$. Since $\hat{\mu} < \mu_0 = \min \mu_0^*(r)$, $\hat{\mu} \notin \mu_0^*(r)$. Therefore,

$$\Phi_0(\hat{\mu}) < \Phi_0(\mu_0). \quad (\text{A-20})$$

Moreover, by (A-19), $\frac{1-F_1(\mu)}{1-F_0(\mu)}$ is increasing in μ when $\mu \in (\mu_0 - \delta, \mu_0)$. Therefore,

$$\frac{1 - F_1(\hat{\mu})}{1 - F_0(\hat{\mu})} < \frac{1 - F_1(\mu_0)}{1 - F_0(\mu_0)} \quad (\text{A-21})$$

Combining (A-20) and (A-21) and using (A-17) gives:

$$\begin{aligned} \Phi_\varepsilon(\hat{\mu}) &= \Phi_0(\hat{\mu}; r) + \log \left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\hat{\mu})}{1 - F_0(\hat{\mu})} \right) \\ &< \Phi_0(\mu_0; r) + \log \left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu_0)}{1 - F_0(\mu_0)} \right) = \Phi_\varepsilon(\mu_0), \end{aligned}$$

which implies $\hat{\mu} \notin \mu_\varepsilon^*(r)$. Therefore, $\min \mu_\varepsilon^*(r) \geq \mu_0 = \min \mu_0^*(r)$.

ii. Suppose $\left. \frac{\partial}{\partial \mu} \left(\frac{1-F_1(\mu)}{1-F_0(\mu)} \right) \right|_{\mu=\mu_0} = 0$. Note that, by (A-17) evaluated at $\varepsilon = 1$,

$$\Phi_1(\mu; r) = \Phi_0(\mu; r) + \log \frac{1 - F_1(\mu)}{1 - F_0(\mu)}.$$

Then,

$$\left. \frac{\partial \Phi_1(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0} = \underbrace{\left. \frac{\partial \Phi_0(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0}}_{=0} + \frac{\left. \frac{\partial}{\partial \mu} \left(\frac{1-F_1(\mu)}{1-F_0(\mu)} \right) \right|_{\mu=\mu_0}}{\frac{1-F_1(\mu_0)}{1-F_0(\mu_0)}} = 0.$$

Let $\gamma = \frac{r}{1-r} \frac{1-\mu_s}{\mu_s}$ as in the proof of Lemma 2. By (7), for $k \in \{0, 1\}$,

$$\left. \frac{\partial \Phi_k(\mu; r)}{\partial \mu} \right|_{\mu=\mu_0} = 0 \iff \frac{1 - F_k(\mu_0)}{1 - \mu_0} = f_k(\mu_0) \frac{1 - \mu_0 + \gamma \mu_0}{\gamma}. \quad (\text{A-22})$$

Then, repeating the steps in the proof of Lemma 2, we have

$$\begin{aligned} \frac{1 - F_1(\mu_0)}{1 - \mu_0} &= \frac{\int_{\mu_0}^1 \alpha(f_0(x)) dx}{1 - \mu_0} \geq \alpha \left(\frac{1 - F_0(\mu_0)}{1 - \mu_0} \right) \underbrace{=}_{(\text{A-22}) \text{ with } k=0} \alpha \left(f_0(\mu_0) \frac{1 - \mu_0 + \gamma \mu_0}{\gamma} \right) \\ &\geq \alpha(f_0(\mu_0)) \frac{1 - \mu_0 + \gamma \mu_0}{\gamma} = f_1(\mu_0) \frac{1 - \mu_0 + \gamma \mu_0}{\gamma}. \end{aligned}$$

By (A-22) with $k = 1$, the first and last terms are equal, which implies all weak inequalities in this chain are equalities. Then:

$$\frac{\int_{\mu_0}^1 \alpha(f_0(x)) dx}{1 - \mu_0} = \alpha \left(\frac{1 - F_0(\mu_0)}{1 - \mu_0} \right) = \alpha \left(\frac{\int_{\mu_0}^1 f_0(x) dx}{1 - \mu_0} \right). \quad (\text{A-23})$$

Let $\bar{f}_0 \equiv \max\{f_0(\mu) : \mu \in [\mu_0, 1]\}$. By (A-22) at $k = 0$, $f_0(\mu_0) > 0$. Therefore, $\bar{f}_0 > 0$. Let Y be a random variable with $Y \sim U[\mu_0, 1]$ and define $X = f_0(Y)$. Since f_0 is continuous and $f_0(1) = 0$, the support of X is $[0, \bar{f}_0]$. Moreover,

$$\mathbb{E}[\alpha(X)] = \mathbb{E}[\alpha(f_0(Y))] = \frac{\int_{\mu_0}^1 \alpha(f_0(x)) dx}{1 - \mu_0} \underbrace{=}_{\text{(A-23)}} \alpha\left(\frac{\int_{\mu_0}^1 f_0(x) dx}{1 - \mu_0}\right) = \alpha(\mathbb{E}[f_0(Y)]) = \alpha(\mathbb{E}[X]).$$

By Lemma A-5, we conclude there is some $a > 0$ such that

$$\alpha(z) = a \cdot z \quad \text{for all } z \in [0, \bar{f}_0]. \quad (\text{A-24})$$

We next claim that $f_0(\mu_0) < \bar{f}_0$. This is because

$$\bar{f}_0 = \frac{\int_{\mu_0}^1 \bar{f}_0 dx}{1 - \mu_0} > \frac{\int_{\mu_0}^1 f_0(x) dx}{1 - \mu_0} \underbrace{=}_{\text{(A-22) with } k=0} f_0(\mu_0) \frac{1 - \mu_0 + \gamma \mu_0}{\gamma} \underbrace{\geq}_{r \leq \mu_s \Rightarrow \gamma \leq 1} f_0(\mu_0).$$

We conclude that $f_0(\mu_0) \in (0, \bar{f}_0)$. Then, by continuity of f_0 , there exists some $\delta \in (0, \mu_0)$ such that $f_0(\hat{\mu}) \in [0, \bar{f}_0]$ for all $\hat{\mu} \in (\mu_0 - \delta, 1]$. By (A-24), we conclude that

$$f_1(\hat{\mu}) = \alpha(f_0(\hat{\mu})) = a \cdot f_0(\hat{\mu}) \quad \text{for all } \hat{\mu} \in (\mu_0 - \delta, 1].$$

Integrating both sides, we conclude

$$1 - F_1(\hat{\mu}) = a(1 - F_0(\hat{\mu})) \quad \text{for all } \hat{\mu} \in (\mu_0 - \delta, 1]. \quad (\text{A-25})$$

By (A-18), there exists some $\bar{\varepsilon} > 0$ such that $\mu_\varepsilon^*(r) \subset (\mu_0 - \delta, 1 - r]$ for all $\varepsilon \in [0, \bar{\varepsilon}]$. Take any $\hat{\mu} \in (\mu_0 - \delta, \mu_0)$. Since $\hat{\mu} < \mu_0 = \min \mu_0^*(r)$, $\hat{\mu} \notin \mu_0^*(r)$. Therefore,

$$\Phi_0(\hat{\mu}) < \Phi_0(\mu_0). \quad (\text{A-26})$$

Moreover, by (A-25),

$$\frac{1 - F_1(\hat{\mu})}{1 - F_0(\hat{\mu})} = a = \frac{1 - F_1(\mu_0)}{1 - F_0(\mu_0)}. \quad (\text{A-27})$$

Combining (A-26) and (A-27) and using (A-17) gives:

$$\begin{aligned} \Phi_\varepsilon(\hat{\mu}) &= \Phi_0(\hat{\mu}; r) + \log\left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\hat{\mu})}{1 - F_0(\hat{\mu})}\right) \\ &< \Phi_0(\mu_0; r) + \log\left(1 - \varepsilon + \varepsilon \frac{1 - F_1(\mu_0)}{1 - F_0(\mu_0)}\right) = \Phi_\varepsilon(\mu_0), \end{aligned}$$

which implies $\hat{\mu} \notin \mu_\varepsilon^*(r)$. Therefore, $\min \mu_\varepsilon^*(r) \geq \mu_0 = \min \mu_0^*(r)$.

In either case, $\min \mu_\varepsilon^*(r) \geq \min \mu_0^*(r)$. Since $\beta_\varepsilon^*(r) = \left\{ \beta = \frac{\mu}{1-\mu} \frac{r}{1-r} : \mu \in \mu_\varepsilon^*(r) \right\}$ for all $\varepsilon \in [0, 1]$, we conclude $\min \beta_\varepsilon^*(r) \geq \min \beta_0^*(r)$. $\square$

Proof of Lemma 3. Note that if $\mu > \frac{1}{2} \left(1 - \frac{r}{g} \right) \iff (1 - 2\mu)g - r < 0$, acquiescing is the strictly dominant strategy for citizens. Consequently, in any equilibrium, citizens acquiesce and $A = 0$. Thus, the regime remains in power if and only if $\theta > 0$.

Suppose $\mu \leq \frac{1}{2} \left(1 - \frac{r}{g} \right) \iff (1 - 2\mu)g - r \geq 0$. When citizens use a cutoff strategy with cutoff x^* , the measure of protesters under regime strength θ is

$$A(\theta) \equiv \Pr(x_i \leq x^* \mid \theta).$$

Since $x_i \mid \theta \sim U[\theta - \rho, \theta + \rho]$, $A(\theta)$ is decreasing in θ . The regime remains in power when $\theta > A(\theta)$, where the left hand-side is strictly increasing in θ and the right hand-side is decreasing. We conclude that there is a unique cutoff θ^* such that the regime remains in power if and only if $\theta > \theta^*$. Here, θ^* satisfies $\theta^* = A(\theta^*)$, or,

$$\theta^* = \Pr(x_i \leq x^* \mid \theta = \theta^*). \quad (\text{A-28})$$

From the perspective of a citizen with signal x_i , the probability that the regime remains in power is $\Pr(A(\theta) < \theta \mid x_i) = \Pr(\theta > \theta^* \mid x_i)$. Then, the expected payoff from protesting is

$$\Pr(\theta \leq \theta^* \mid x_i) ((1 - 2\mu)(1 + g)) + \Pr(\theta > \theta^* \mid x_i) (2\mu - 1) - r,$$

and the expected payoff from acquiescing is:

$$\Pr(\theta \leq \theta^* \mid x_i) (1 - 2\mu) + \Pr(\theta > \theta^* \mid x_i) (2\mu - 1).$$

A citizen with signal x_i protests if and only if the expected payoff from protesting is larger than the expected payoff from acquiescing, i.e., if and only if:

$$\Pr(\theta \leq \theta^* \mid x_i) (1 - 2\mu)g - r \geq 0.$$

Note that $\Pr(\theta \leq \theta^* \mid x_i)$ is decreasing in x_i , and so is the left hand-side. For the cutoff x^* to be the equilibrium strategy, the citizen with signal $x_i = x^*$ must be indifferent between the two actions. In other words, x^* satisfies: $\Pr(\theta \leq \theta^* \mid x_i = x^*) (1 - 2\mu)g - r = 0$. Rearranging yields:

$$\Pr(\theta \leq \theta^* \mid x_i = x^*) = \frac{r}{g(1 - 2\mu)} \in \left[\frac{r}{g}, 1 \right]. \quad (\text{A-29})$$

Note that in any equilibrium, $x^* < 1 + \rho$. (For any $x_i \geq 1 + \rho$, $\theta \geq 1$ with probability one. Thus, $A < \theta$ with probability one and acquiescing is a dominant strategy. But then, a citizen with this signal cannot be indifferent.) Thus,

$$x^* < 1 + \rho \quad \underbrace{<}_{\text{by } \bar{\theta} > 1+2\rho} \quad \bar{\theta} - \rho \implies x^* + \rho < \bar{\theta}.$$

Similarly, $x^* - \rho > \underline{\theta}$. We conclude that $\theta|x^* \sim U[x^* - \rho, x^* + \rho]$. Moreover, for (A-29) to hold with $0 < r < g$, we must have: $\theta^* \in [x^* - \rho, x^* + \rho]$. Equivalently, $x^* \in [\theta^* - \rho, \theta^* + \rho]$. This implies

$$\begin{aligned} \Pr(x_i \leq x^* | \theta = \theta^*) & \underbrace{=}_{x_i|\theta^* \sim U[\theta^*-\rho, \theta^*+\rho]} \frac{x^* - (\theta^* - \rho)}{2\rho} = 1 - \frac{\theta^* - (x^* - \rho)}{2\rho} \\ & \underbrace{=}_{\theta|x^* \sim U[x^*-\rho, x^*+\rho]} 1 - \Pr(\theta \leq \theta^* | x_i = x^*). \end{aligned}$$

Overall, we have

$$\Pr(\theta \leq \theta^* | x_i = x^*) = 1 - \Pr(x_i \leq x^* | \theta = \theta^*). \quad (\text{A-30})$$

Combining (A-28), (A-29) and (A-30) gives the desired result. $\square$

Proof of Proposition 4. We first consider the case where

$$\frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r} > 1 \iff \mu_0 > \frac{1}{2} \left(1 - \frac{r}{g} \right).$$

In this case, for any $\beta \in [0, 1]$, $\frac{\mu_0}{\mu_0 + (1 - \mu_0)\beta} \geq \mu_0 > \frac{1}{2} \left(1 - \frac{r}{g} \right)$. Thus, by (8), the optimization problem in (9) becomes:

$$\beta^*(r) = \arg \max_{\beta \in [0, 1]} (1 - \mu_0)(1 - \beta) \cdot \left(-H \left(1 - \frac{r}{g} \right) \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot (-H(0))$$

By assumption, $1 - \frac{r}{g} > 0$, which implies $-H \left(1 - \frac{r}{g} \right) < -H(0)$. Then, the objective function is strictly increasing in β , and $\beta^*(r) = 1$ is the unique optimizer.

We next consider the case where

$$\frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r} \leq 1 \iff \mu_0 \leq \frac{1}{2} \left(1 - \frac{r}{g} \right).$$

In this case,

- Suppose

$$\beta < \frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r} \implies \frac{\mu_0}{\mu_0 + (1 - \mu_0)\beta} > \frac{1}{2} \left(1 - \frac{r}{g} \right).$$

By (8), the objective function in this region is

$$(1 - \mu_0)(1 - \beta) \cdot \left(-H \left(1 - \frac{r}{g} \right) \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot (-H(0))$$

which is strictly increasing in β .

- Next, suppose

$$\beta \geq \frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r} \implies \frac{\mu_0}{\mu_0 + (1 - \mu_0)\beta} \leq \frac{1}{2} \left(1 - \frac{r}{g} \right).$$

The objective function in this region is:

$$\begin{aligned} & (1 - \mu_0)(1 - \beta) \cdot \left(-H \left(1 - \frac{r}{g} \right) \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(-H \left(1 - \frac{r}{g \left(1 - 2 \frac{\mu_0}{\mu_0 + (1 - \mu_0)\beta} \right)} \right) \right) \\ &= (1 - \mu_0)(1 - \beta) \cdot \left(-H \left(1 - \frac{r}{g} \right) \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(-H \left(1 - \frac{r}{g} \frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} \right) \right) \\ &= (1 - \mu_0)(1 - \beta) \cdot \left(-\frac{1 - \frac{r}{g} - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(-\frac{1 - \frac{r}{g} \frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} - \underline{\theta}}{\bar{\theta} - \underline{\theta}} \right) \\ &= -\frac{1}{\bar{\theta} - \underline{\theta}} \left[(1 - \mu_0)(1 - \beta) \cdot \left(1 - \frac{r}{g} - \underline{\theta} \right) + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(1 - \frac{r}{g} \frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} - \underline{\theta} \right) \right] \\ &= -\frac{1}{\bar{\theta} - \underline{\theta}} \left[1 - \frac{r}{g} - \underline{\theta} + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(-\frac{r}{g} \frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} + \frac{r}{g} \right) \right] \\ &= -\frac{1}{\bar{\theta} - \underline{\theta}} \left[1 - \frac{r}{g} - \underline{\theta} + (\mu_0 + (1 - \mu_0)\beta) \cdot \left(-\frac{r}{g} \right) \frac{2\mu_0}{(1 - \mu_0)\beta - \mu_0} \right] \\ &= -\frac{1}{\bar{\theta} - \underline{\theta}} \left[1 - \frac{r}{g} - \underline{\theta} - 2\mu_0 \frac{r}{g} \frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} \right] \end{aligned}$$

Since $\frac{\partial}{\partial \beta} \left(\frac{(1 - \mu_0)\beta + \mu_0}{(1 - \mu_0)\beta - \mu_0} \right) = -\frac{2\mu_0(1 - \mu_0)}{((1 - \mu_0)\beta - \mu_0)^2} < 0$, the objective function is strictly decreasing in β in this region.

Combining both observations, we conclude that $\beta^*(r) = \frac{\mu_0}{1 - \mu_0} \frac{g + r}{g - r}$ is the unique maximizer. The result follows from combining the characterizations in both cases and analytically verifying that $\beta^*(r)$ is increasing in r . $\square$

Proof of Lemma A-1. From (A-1), it can be derived (see, e.g., [Kolotilin, Mylovannov and Zapechelnyuk, 2022](#), p.565) that, for $k \in \{0, 1\}$,

$$\frac{\partial v_k(\omega^*; r)}{\partial \omega^*} = g(\omega^*) [f_k(1 - \mu^* - r)(\mu^* - \omega^*) - (F_k(1 - \omega^* - r) - F_k(1 - \mu^* - r))] . \quad (\text{A-31})$$

Take any $\omega^* \in (0, 1)$ with $\frac{\partial v_0(\omega^*; r)}{\partial \omega^*} \leq 0$. If $g(\omega^*) = 0$, (A-31) immediately implies that $\frac{\partial v_k(\omega^*; r)}{\partial \omega^*} = 0$ for $k \in \{0, 1\}$. Hence, suppose $g(\omega^*) > 0$. By (A-31),

$$\frac{F_0(1 - \omega^* - r) - F_0(1 - \mu^* - r)}{\mu^* - \omega^*} \geq f_0(1 - \mu^* - r) \implies \frac{\int_{\omega^*}^{\mu^*} f_0(1 - \omega - r) d\omega}{\mu^* - \omega^*} \geq f_0(1 - \mu^* - r).$$

Then,

$$\begin{aligned} \frac{F_1(1 - \omega^* - r) - F_1(1 - \mu^* - r)}{\mu^* - \omega^*} &= \frac{\int_{\omega^*}^{\mu^*} f_1(1 - \omega - r) d\omega}{\mu^* - \omega^*} = \frac{\int_{\omega^*}^{\mu^*} \alpha(f_0(1 - \omega - r)) d\omega}{\mu^* - \omega^*} \\ &\geq \underbrace{\alpha}_{\text{Jensen's}} \left(\frac{\int_{\omega^*}^{\mu^*} f_0(1 - \omega - r) d\omega}{\mu^* - \omega^*} \right) \\ &\geq \alpha(f_0(1 - \mu^* - r)) \\ &= f_1(1 - \mu^* - r). \end{aligned}$$

Therefore, by (A-31), $\frac{\partial v_1(\omega^*; r)}{\partial \omega^*} \leq 0$. □

Proof of Lemma A-2. For any $\mu \in [0, 1]$, incentive compatibility requires

$$U_\sigma(\mu, \mu) \geq U_\sigma(\mu, \tilde{\mu}) = \max \{1 - \mu - r, (1 - \mu)(1 - \sigma(\tilde{\mu}))(1 - r)\} \quad \text{for all } \tilde{\mu} \in [0, 1].$$

Taking a sequence of reports whose corresponding censorship intensities converge to $\hat{\beta}$ yields

$$U_\sigma(\mu, \mu) \geq \max \{1 - \mu - r, (1 - \mu)(1 - \hat{\beta})(1 - r)\}. \quad (\text{A-32})$$

Moreover, by definition, $\sigma(\mu) \geq \inf_{\tilde{\mu} \in [0, 1]} \sigma(\tilde{\mu}) = \hat{\beta}$. Then,

$$\begin{aligned} \max \{1 - \mu - r, (1 - \mu)(1 - \hat{\beta})(1 - r)\} &\geq \max \{1 - \mu - r, (1 - \mu)(1 - \sigma(\mu))(1 - r)\} \\ &= U_\sigma(\mu, \mu). \end{aligned} \quad (\text{A-33})$$

Equations (A-32) and (A-33) imply that

$$U_\sigma(\mu, \mu) = \max \{1 - \mu - r, (1 - \mu)(1 - \hat{\beta})(1 - r)\} \quad \text{for all } \mu \in [0, 1]. \quad (\text{A-34})$$

Let

$$\hat{\mu} \equiv \frac{\hat{\beta}}{\hat{\beta} + \frac{r}{1-r}}.$$

We consider two separate cases.

1. If $\mu > \hat{\mu}$,

$$1 - \mu - r < (1 - \mu)(1 - \hat{\beta})(1 - r). \quad (\text{A-35})$$

Suppose $\sigma(\mu) > \hat{\beta}$. Then,

$$(1 - \mu) (1 - \sigma(\mu)) (1 - r) < (1 - \mu) (1 - \hat{\beta}) (1 - r). \quad (\text{A-36})$$

Equations (A-35) and (A-36) imply

$$U_\sigma(\mu, \mu) = \max \{1 - \mu - r, (1 - \mu) (1 - \sigma(\mu)) (1 - r)\} < (1 - \mu) (1 - \hat{\beta}) (1 - r),$$

which contradicts (A-34). We conclude that $\sigma(\mu) = \hat{\beta}$. Then, public censorship with intensity $\hat{\beta}$ generates the same probability of protesting for a citizen with prior μ .

2. If $\mu < \hat{\mu}$, then

$$\mu < \frac{\hat{\beta}}{\hat{\beta} + \frac{r}{1-r}}.$$

Thus, under public censorship with intensity $\hat{\beta}$, the citizen with prior μ protests following $m' = \text{silence}$. Moreover, since $\sigma(\mu) \geq \hat{\beta}$,

$$\mu < \frac{\hat{\beta}}{\hat{\beta} + \frac{r}{1-r}} \leq \frac{\sigma(\mu)}{\sigma(\mu) + \frac{r}{1-r}}.$$

Therefore, under the private censorship strategy, the citizen with prior μ protests following $m'_\mu = \text{silence}$ as well. We conclude that the citizen protests with the same probability under mechanism σ and under public censorship with intensity $\hat{\beta}$.

Under Assumption 1, citizens with prior $\hat{\mu}$ have zero mass. Therefore, private censorship mechanism σ and public censorship with intensity $\hat{\beta}$ induce the same probability of protesting in each state for almost every citizen. $\square$

Proof of Lemma A-3. As f is continuously differentiable by Assumption 1, $v''(\mu; r) = \frac{\partial^2 v(\mu; r)}{\partial \mu^2}$ exists. We will show that $v''(\mu; r)$ satisfies the *strict single-crossing-from-above property*:

$$\text{If } v''(\mu_1; r) \geq 0 \text{ for some } \mu_1 \in [0, 1], \text{ then } v''(\mu_2; r) > 0 \text{ for all } \mu_2 < \mu_1,$$

which implies the result.

By (A-3), $v(\mu; r) = -F(g(\mu))$ where:

$$g(\mu) \equiv \frac{1 - \mu}{1 - \mu + \gamma \mu}, \quad \gamma \equiv \frac{r}{1 - r} \frac{1 - \mu_s}{\mu_s} > 0.$$

Then,

$$v''(\mu; r) = -f'(g(\mu)) \cdot (g'(\mu))^2 - f(g(\mu)) \cdot g''(\mu).$$

Suppose $v''(\mu_1; r) \geq 0$ for some $\mu_1 \in [0, 1]$. This implies

$$\frac{f'(g(\mu_1))}{f(g(\mu_1))} \leq -\frac{g''(\mu_1)}{(g'(\mu_1))^2} = \frac{2}{\gamma} \left((1-\gamma) - (1-\gamma)^2 \mu_1 \right). \quad (\text{A-37})$$

Take any $\mu_2 < \mu_1$. Because $g(\mu)$ is strictly decreasing in μ , $g(\mu_2) > g(\mu_1)$. Because $f(\mu)$ is strictly log-concave,

$$\frac{f'(g(\mu_2))}{f(g(\mu_2))} < \frac{f'(g(\mu_1))}{f(g(\mu_1))}. \quad (\text{A-38})$$

Moreover, $((1-\gamma) - (1-\gamma)^2 \mu)$ is decreasing in μ , so equations (A-37)–(A-38) yield that

$$\frac{f'(g(\mu_2))}{f(g(\mu_2))} < \frac{2}{\gamma} \left((1-\gamma) - (1-\gamma)^2 \mu_2 \right) = -\frac{g''(\mu_2)}{(g'(\mu_2))^2}$$

Therefore, $v''(\mu_2; r) > 0$. □

Proof of Lemma A-4. Take any μ with $\frac{\partial \Phi_0(\mu; r)}{\partial \mu} \geq 0$. By (7),

$$\frac{1 - F_0(\mu)}{1 - \mu} \geq f_0(\mu) \frac{1 - \mu + \gamma \mu}{\gamma}. \quad (\text{A-39})$$

Then,

$$\begin{aligned} \frac{1 - F_1(\mu)}{1 - \mu} (1 - \mu + \gamma \mu) &= \frac{\int_{\mu}^1 f_1(x) dx}{1 - \mu} (1 - \mu + \gamma \mu) \\ &= \frac{\int_{\mu}^1 f_1(x) dx}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} \\ &= \frac{\int_{\mu}^1 \alpha \left(f_0(x) \frac{(1-x+\gamma x)^2}{\gamma} \right) \frac{\gamma}{(1-x+\gamma x)^2} dx}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} && (\text{by (A-4)}) \\ &= \frac{-\int_{\frac{1-\mu}{1-\mu+\gamma\mu}}^0 \alpha \left(f_0 \left(\frac{1-t}{1-t+\gamma t} \right) \frac{\gamma}{(1-t+\gamma t)^2} \right) dt}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} && (\text{change of variables, } t = \frac{1-x}{1-x+\gamma x}) \\ &= \frac{\int_0^{\frac{1-\mu}{1-\mu+\gamma\mu}} \alpha \left(f_0 \left(\frac{1-t}{1-t+\gamma t} \right) \frac{\gamma}{(1-t+\gamma t)^2} \right) dt}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} \\ &\geq \alpha \left(\frac{\int_0^{\frac{1-\mu}{1-\mu+\gamma\mu}} f_0 \left(\frac{1-t}{1-t+\gamma t} \right) \frac{\gamma}{(1-t+\gamma t)^2} dt}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} \right) && (\text{by Jensen's Inequality}) \\ &= \alpha \left(\frac{-\int_1^{\mu} f_0(u) du}{\frac{1 - \mu}{1 - \mu + \gamma \mu}} \right) && (\text{change of variables, } u = \frac{1-t}{1-t+\gamma t}) \end{aligned}$$

$$\begin{aligned}
&= \alpha \left(\frac{\int_{\mu}^1 f_0(u) du}{1 - \mu} (1 - \mu + \gamma\mu) \right) \\
&= \alpha \left(\frac{1 - F_0(\mu)}{1 - \mu} (1 - \mu + \gamma\mu) \right) \\
&\geq \alpha \left(f_0(\mu) \frac{1 - \mu + \gamma\mu}{\gamma} (1 - \mu + \gamma\mu) \right) && \text{(by (A-39) and increasing } \alpha) \\
&= \alpha \left(f_0(\mu) \frac{(1 - \mu + \gamma\mu)^2}{\gamma} \right) \\
&= f_1(\mu) \frac{(1 - \mu + \gamma\mu)^2}{\gamma} && \text{(by (A-4))}
\end{aligned}$$

Therefore,

$$\frac{1 - F_1(\mu)}{1 - \mu} (1 - \mu + \gamma\mu) \geq f_1(\mu) \frac{(1 - \mu + \gamma\mu)^2}{\gamma} \implies \frac{1 - F_1(\mu)}{1 - \mu} \geq f_1(\mu) \frac{1 - \mu + \gamma\mu}{\gamma},$$

which, by (7), implies $\frac{\partial \Phi_1(\mu; r)}{\partial \mu} \geq 0$. □