The Dictator’s Dilemma: A Theory of Propaganda and Repression

A. Arda Gitmez and Konstantin Sonin

MAY 2023
Repression and information manipulation are two main tools of any modern authoritarian regime. Our theoretical model demonstrates how repression and propaganda complement each other: when the regime's opponents are facing stricter punishment, the effect of persuasion is stronger, and propaganda is used by the regime more heavily. Similarly, when repression eliminates those citizens who are relatively more skeptical about the regime, the rest can be more heavily influenced. Finally, we show that when citizens self-select into receiving information from individual sources, the dictator cannot do better than resorting to public messaging.

Keywords: Authoritarian regimes, repression, propaganda, Bayesian persuasion.

JEL Classification: D85, L82.
**Introduction**

Repression and propaganda have always been considered the primary tools of autocratic control (Svolik, 2012). In the 20th century, information manipulation was a central focus in the study of totalitarian dictatorships such as Hitler’s Germany, Stalin’s Russia, and Mao’s China, in which the state tried to control all aspects of subjects’ lives (Arendt, 1951; Friedrich and Brzezinski, 1956; Cassinelli, 1960). With the demise of totalitarian dictatorships, propaganda is no longer considered a means of ideological indoctrination, but rather as a tool used by a leader to maintain his reputation as a strong and competent hand (Guriev and Treisman, 2019). Still, repression remains a critical instrument in a dictator’s arsenal. In the year following the protests of the summer of 2020, 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 (de Vogel, 2022).

The tradition to consider repression and propaganda together goes back centuries. In *The Prince*, Niccolo Machiavelli writes on whether it is better to be feared than loved: “The answer is that one would like to be both the one and the other.” (Machiavelli, 2019). The early formal theory of nondemocratic government (Wintrobe, 1990, 1998) focused on a simple trade-off: the dictator was deciding how to optimally allocate resources between “repression” and “benefits” to population aimed to make the dictator more popular.

Guriev and Treisman (2022) argue that subtle propaganda might be a substitute for brutal repression, persuading citizens in the leader’s competence – unless the dictator’s circumstances change for the worse. Since the start of Russia’s invasion of Ukraine in February 2022, Putin’s regime has doubled down on both tools of authoritarian control. It dramatically increased information manipulation, closing down all independent media outlets, harassing journalists, and jailing ordinary people for sharing information on social media. Simultaneously, it increased repression – with thousands of people opposing war and the country’s course being arrested and opposition politicians getting prison terms of 8-25 years (Stoner, 2022; Treisman, 2022).

In this paper, we study how propaganda naturally complements repression. The basic logic of our model is as follows. Harsher repression make those who were marginally willing to protest inframarginal. The now-marginal citizen’s attitude towards the dictator (the prior about his competence) is higher, which makes the new marginal citizen more amenable to persuasion. This complementarity occurs regardless of whether repression is *ex post*, punishing those who protest, or *ex ante*, targeting the most skeptical citizens. When citizens who are most skeptical about the regime are singled out and repressed, the rest tends to have more favorable attitudes towards the government. With a lower level of repression, they would have been “underpersuaded” or persuaded sub-optimally from the perspective of the dictator. With repression, propaganda directed towards these citizens is more effective and therefore, optimally, more intense. Once the most
disloyal elements of the society are taken out, the rest can be manipulated more. Thus, the dictator
does not face the choice of repression versus propaganda, but rather looks for an optimal bundle
of the two.

In our basic setup, repressions happen *ex post*, targeting those who did not support the dictator.
This does not imply that every non-supporter is punished: the repression works in expectation.
However, further results show that our logic extends to the case, when repression happens *ex ante*,
before propaganda.\(^1\) In this case, when the distribution of attitudes towards the dictator changes,
as a result of repression, the optimal level of information manipulation increases (see discussion
in Subsection 3.2).

Citing dozens of examples, Guriev and Treisman (2022) document the change in the relative
weights of repression and propaganda from the totalitarian dictatorships of the 20\(^{th}\) century to
the “spin dictators” of the 21\(^{st}\)\). Modern dictators do not arrest hundreds of thousands as Stalin,
Hitler, or Mao did or carry public executions as Iraq’s Saddam Hussein, Cambodia’s Pol Pot, or
Equatorial Guinea’s Macias Nguema. They combine repressions with selective censorship, digital
surveillance, and sophisticated propaganda (see Gehlbach et al., 2022, for a model that dis-
tinguishes between different forms of information manipulation, and Shadmehr, 2014, for a model,
in which censorship is used to hide the extent of repression). Our results show that a population
that has more positive priors about the leader’s quality receives a heavier doze of information ma-
nipulation.

To model information manipulation, we use the basic model of Bayesian persuasion (Ka-
menica and Gentzkow, 2011; Bergemann and Morris, 2019) on an audience with heterogeneous
priors (Alonso and Câmara, 2016\(^a\); Laclau and Renou, 2017; Gitmez and Molavi, 2022).\(^2\) Compared
to other communication protocols, the model of Bayesian persuasion assumes fuller commitment
on behalf of the sender.\(^3\) This makes sense in an applied model: dictators do not edit news in
the real time. Instead, they pass laws, establish institutions of censorship, and appoint editors
to control the flow of information. The choice of an institutional bias or an editor of known
ideological preferences corresponds to the choice of the main control parameter in the model.

Yet there are theoretical advantages of using the Bayesian persuasion model as well. First
and most importantly, the model allows one to study the *maximum propaganda*: it provides
the upper limit on the amount of persuasion that can be done via any information exchange
between a sender and a receiver. At the same time, our qualitative results easily translate to other
information-exchange models such as such as cheap talk in Crawford and Sobel (1982), verifiable

---
\(^1\) Arendt (1951) makes a distinction between the *dictatorial terror*, aimed against the well-identified opponents of the regime, from an all-
pervasive *totalitarian terror* of purges, mass executions, and concentration camps. Modern theories of repressions with strategic targeting
and selection include Myerson (2015), Tyson (2018), and Dragu and Przeworski (2019).
\(^2\) Also see Kosterina (2022), who studies Bayesian persuasion with unknown priors and a worst-case-maximizing sender, and Shimoji
(2022), who develops a linear programming approach to Bayesian persuasion with heterogeneous priors.
\(^3\) Though our model assumes full commitment, this assumption is not necessary; the results still hold if the probability that the dictator
can renge on his commitment is not to high.
messaging in Milgrom (1981) (see also Titova, 2022), and signaling in Spence (1973). Though the machinery of a model of propaganda and repression based on other communication protocols would be different, the main intuition is the same.

An important element of our model is that the government organizes information manipulation as public communication: it establishes an institution, which learns the true state of the world, and then makes a public report. A natural question is whether the government could do better if it were possible to target different citizens with different messages. Is it possible to do more persuasion if persuasion is based on private characteristics? In Section 4, we demonstrate that the possibility of private persuasion does not add to the government’s persuasion power. Substantively, this explains why many authoritarian regimes use blank, one-size-fits-all messaging rather than target groups with different attitudes individually. As a technical matter, Proposition 4 justifies our assumption that the government sticks to the public persuasion mechanism. This result mirrors the main result in Kolotilin et al. (2017). The difference is that while in Kolotilin et al. (2017) the receivers have heterogeneous preferences, in our model they have heterogeneous priors.

Repressions have been shown to change citizens’ behavior. Montagnes and Wolton (2019) and Rozenas (2020) use communist purges in Stalin’s Russia and Mao’s China to demonstrate this effect. Physical elimination as in Esteban, Morelli and Rohner (2015) changes the composition of the society; other forms of political disenfranchisement might have the same effect as well. In addition to mass executions, Stalin relocated hundreds of thousands from the 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); in our model, this would correspond to the dictator repressing citizens based on imperfect information about their attitudes. In the realm of democratic politics, Glaeser and Shleifer (2005) show that the incumbent politician might deliberately choose policies that drive voters who oppose him out of the district. Our theory applies to such situations as well. After repression, the rest of the society will be exposed to more information manipulation.

The rest of the paper is organized as follows. Section 2 studies the case when the leader cannot repress, only persuade, the citizens. Section 3 analyzes the main case, when the leader optimally combines repression and persuasion. Finally, Section 4 deals with targeted propaganda.

---

4Wang (2015) and Chan et al. (2019) also compare public and private persuasion, where the receivers’ heterogeneous preferences are known by the sender. Alonso and Câmara (2016b) and Inostroza and Pavan (2022) study public persuasion towards heterogeneous receivers with known preferences, whereas Bardhi and Guo (2018), Arieli and Babichenko (2019), Taneva (2019) and Mathevet, Perego and Taneva (2020) study private persuasion towards heterogeneous receivers with known preferences. On private persuasion towards heterogeneous receivers with private preferences, Guo and Shmaya (2019) characterize the optimal information structure, and Heese and Lauermann (2021) study a voting setting.
2 Propaganda without Repression

We begin with a discussion of our model with information manipulation only, and introduce the possibility of repression in Section 3.

2.1 Setup

There is a sender \( s \) (the leader) and a continuum of receivers \( I = [0, 1] \) (the citizens, where a generic citizen is denoted with \( i \)). A state of the world is denoted by \( \omega \in \{0, 1\} \). Here, \( \omega = 1 \) is the state of the world, in which the citizens’ preferences align with the leader, i.e., the state, in which the leader is good), and \( \omega = 0 \) is the state where there is a misalignment. Citizen \( i \)'s prior on the leader’s being good is

\[
\mu_i \equiv \Pr_i \{\omega = 1\}
\]

Citizens have heterogeneous priors; the distribution of priors among citizens has density \( f(\mu) \).

Throughout the paper, we maintain the following assumption on the distribution of priors.

**Assumption 1.** Probability density function \( f \) has full support on \([0, 1]\), is continuously differentiable with \( f(0) = f(1) = 0 \), and is strictly log-concave, that is, \( \frac{\partial^2}{\partial \mu^2} \log f(\mu) < 0 \) for all \( \mu \in [0, 1] \).

The assumption that the density function \( f \) is log-concave is a mild one; see Bagnoli and Bergstrom (2005) for a list of commonly used log-concave distributions, which include, among others, uniform, (truncated) normal, and beta (with both parameters \( \geq 1 \)) distributions. The leader’s prior is \( \mu_s \in (0, 1) \), and \( F(\mu) \) denotes the cdf associated with \( f(\mu) \). Both \( \mu_s \) and \( f \) are common knowledge.

Given her information about \( \omega \), each citizen \( i \in I \) takes an action \( a_i \in \{0, 1\} \). Here, \( a_i = 1 \) corresponds to supporting the leader, and \( a_i = 0 \) corresponds to protesting. Citizen \( i \)'s payoff is \( u(a_i, \omega) \), which is given in Table 1.

\[
\begin{array}{c|cc}
   & \omega = 0 & \omega = 1 \\
\hline
a_i = 0 & q & 0 \\
a_i = 1 & 0 & 1 - q \\
\end{array}
\]

Table 1. Citizen Payoffs without Repression

Here, \( q \in (0, 1) \) is interpreted as the relative cost of supporting the leader, or, equivalently, \( 1 - q \) is the cost of protesting. The optimal action is to take \( a_i = 1 \) as long as the posterior beliefs remain above \( q \).

The leader’s payoff is the total support he receives from the citizens:

\[
u_s(\{a_i\}_{i \in I}) = \int_{i \in I} a_i \cdot di
\] (1)
Citizens do not have any information about \( \omega \) beyond the information conveyed by the leader. The leader uses a public persuasion mechanism that sends messages from \( M \). That is, the leader 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 message drawn, \( m \in M \), is publicly observable to each citizen.

In an environment where each media source is freely accessible to citizens, restriction to public persuasion is without loss of generality. Indeed, consider a setup where there are multiple information sources \( 1, \ldots, n \) with message spaces \( M_1, \ldots, M_n \). Let source \( j \in \{1, \ldots, n\} \) use information structure \( \{\sigma_j(\cdot|\omega)\}_{\omega \in \{0,1\}} \in \Delta(M_j) \). As long as the citizens can observe messages from various sources, one can define
\[
M \equiv M_1 \times \ldots \times M_n
\]
and, for each \( m = (m_1, \ldots, m_n) \in M \), let
\[
\sigma(m|\omega) = \sigma_1(m_1|\omega) \cdot \ldots \cdot \sigma_n(m_n|\omega) \quad \text{for all } \omega \in \{0,1\},
\]
so that the same outcome can be implemented via a public persuasion mechanism. We assume that \( |M| \) is large enough, so that there is a sufficient number of action recommendations for each receiver. As we will show later, under Assumption 1, the leader uses at most two messages.

The assumption that each citizen can access many information sources is a reasonable starting point. However, we will go further, and demonstrate that the leader cannot do better even if there were a possibility of private persuasion. In Section 4, we focus on the case with each citizen having access to at most one media source. Naturally, she picks the media source that gives the highest expected payoff. This setup imposes an *incentive-compatibility* constraint. We will see that the leader cannot do better with private persuasion than with the public one: the outcome of an incentive-compatible private persuasion mechanism can always be achieved via a public persuasion mechanism.

### 2.2 Persuasion with Heterogeneous Priors

We begin by constructing the value function of the leader as a function of his own posterior belief \( \mu = \Pr_s(\omega = 1|m) \), and then use the concavification approach of Kamenica and Gentzkow (2011).

Suppose the leader’s posterior is \( \mu \in [0,1] \). By Proposition 1 of Alonso and Câmara (2016a), citizen \( i \) has the following posterior:
\[
\mu_i' = \frac{\mu \mu_i}{\mu \mu_i + (1 - \mu) \frac{1-\mu_i}{1-\mu}}, \tag{2}
\]
Citizen $i$ supports the leader as long as $\mu_i' \geq q$. Substituting (2) into this inequality and rearranging terms, any citizen $i$ with prior

$$\mu_i \geq \frac{1 - \mu}{1 - \mu + \mu \frac{1 - q}{q} \frac{1 - \mu_s}{\mu_s}}$$

supports the leader. Given (1), the leader’s value function, as a function of leader’s posterior $\mu$, is

$$v(\mu; q) \equiv 1 - F\left(\frac{1 - \mu}{1 - \mu + \mu \frac{1 - q}{q} \frac{1 - \mu_s}{\mu_s}}\right)$$

(3)

The optimal solution relies on the characterization of the concave closure of $v(\mu; q)$. An inspection of (3) immediately reveals that $v(\mu; q)$ is increasing in $\mu$, with $v(0; q) = 0$ and $v(1; q) = 1$. The next Lemma 1 shows that the value function has a specific shape.

**Lemma 1.** $v(\mu; q)$ is strictly S-shaped, i.e., there is some $\tilde{\mu} \in [0, 1]$ such that $v(\mu; q)$ is strictly convex for $\mu \in [0, \tilde{\mu}]$ and strictly concave for $\mu \in [\tilde{\mu}, 1]$.

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

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

which implies that there exists some $\tilde{\mu}$ such that $v''(\mu; q) > 0$ for $\mu \leq \tilde{\mu}$ and $v''(\mu; q) < 0$ for $\mu \geq \tilde{\mu}$.

By (3), $v(\mu; q) = 1 - F(g(\mu))$ where:

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

Then,

$$v''(\mu, q) = -f' (g(\mu)) \cdot (g'(\mu))^2 - f (g(\mu)) \cdot g''(\mu)$$

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

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

(4)

Take any $\mu_2 < \mu_1$. Because $g(\mu)$ is strictly decreasing in $\mu$, $g(\mu_2) > g(\mu_1)$. Because $f(\mu)$ is strictly log-concave by Assumption 1,

$$\frac{f'(g(\mu_2))}{f(g(\mu_2))} < \frac{f'(g(\mu_1))}{f(g(\mu_1))}$$

(5)

Moreover, $((1 - \gamma) - (1 - \gamma)^2 \mu)$ is decreasing in $\mu$, so equations (4-5) 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, $v''(\mu_2; q) > 0$. \qed
Figure 1 illustrates the leader’s value function in light of Lemma 1. Visual inspection reveals that the leader’s optimal manipulation policy invokes two posteriors. This can be achieved by using two messages about the leader’s quality: \( m \in \{bad, good\} \). Moreover, one of the posteriors in the support is \( \mu = 0 \), i.e., one of the messages perfectly reveals \( \omega = 0 \). This can be achieved by setting \( \sigma(m = good|\omega = 1) = 1 \) in the optimal policy. Therefore, the optimal policy is characterized by a single-dimensional object:

\[
\beta = \sigma(m = good|\omega = 0) \in [0, 1].
\]

The most natural interpretation of \( \beta \) is the level of propaganda: it is the likelihood that the message is “the leader is good” when the leader is, in fact, bad. Let \( \beta^*(q) \) denote the propaganda level chosen by the leader under the optimal policy as a function of \( q \). Let

\[
V^*(q) \equiv (\mu_s + (1 - \mu_s)\beta^*(q)) \cdot v\left(\frac{\mu_s}{\mu_s + (1 - \mu_s)\beta^*(q)}\right)
\]

denote the sender’s (subjective) payoff under the optimal information manipulation policy.

Our first comparative statics results are critical stepping stones in our analysis.

**Proposition 1.** Consider two distributions of priors, \( f_1 \) and \( f_2 \), that both satisfy Assumption 1 and where \( f_1 \) is larger than \( f_2 \) in the likelihood ratio order:

\[
\frac{f_1(\mu)}{f_2(\mu)} \text{ is increasing in } \mu.
\]

Then, the propaganda level chosen by the leader under \( f_1 \) is larger than the propaganda level chosen

---

5Our notation suggests that a receiver takes the sender’s favorite action when indifferent. Since \( f \) is continuously differentiable by Assumption 1, it has no mass points, so the measure of citizens with priors equal to \( q \) is zero. Therefore, this choice of notation is inconsequential for the analysis.

6When \( \mu_s \geq \hat{\mu} \) in Figure 1, the optimal policy does not reveal any information. In this case, any policy with two messages where \( \sigma(m = good|\omega = 0) = \sigma(m = good|\omega = 1) \) is optimal. Among many optimal policies, we choose the one where \( \sigma(m = good|\omega = 0) = 1 \), so that the “bad” message is never sent and the beliefs following \( m = bad \) are free. Consequently, one can choose the posterior following \( m = bad \) to be zero, to remain consistent with the discussion.
two posteriors for the leader:

\[ \beta_1^*(q) \geq \beta_2^*(q). \]

Intuitively, the leader with distribution \( f_1 \) is a “more universally loved” leader than a leader with distribution \( f_2 \), because priors under \( f_1 \) tend to be larger compared to priors under \( f_2 \). In equilibrium, a more popular leader rides his popularity and provides less information to citizens, resulting in more propaganda. In Section 3, we will use Proposition 1 to demonstrate the effect of ex ante repression on the optimal level of propaganda.

**Proof of Proposition 1.** For \( k \in \{1, 2\} \), let \( v_k(\mu; q) \) denote the associated value function under \( f_k(\mu) \). Let \( co v_k(\mu; q) \) denote the concave closure of \( v_k(\mu; q) \). By Lemma 1, \( v_k(\mu; q) \) is strictly S-shaped, and therefore the set where \( v_k(\mu; q) \) coincides with its concave closure takes the following form:

\[ \{ \mu \in [0, 1] : v_k(\mu; q) = co v_k(\mu; q) \} = \{ 0 \} \cup [\hat{\mu}_k, 1] \text{ for some } \hat{\mu}_k \in [0, 1]. \]

By Corollary 2 of Kamenica and Gentzkow (2011), when \( \mu_s < \hat{\mu}_k \), the optimal policy generates two posteriors for the leader: \( \mu \in \{0, \hat{\mu}_k\} \). When \( \mu_s \geq \hat{\mu}_k \), the optimal policy does not reveal any information. Moreover, \( \hat{\mu}_k \) satisfies the following properties:

- \( v'_k(\mu; q) \cdot \mu < v_k(\mu; q) \) for all \( \mu \in (0, 1) \) if and only if \( \hat{\mu}_k = 0 \).
- \( v'_k(\mu; q) \cdot \mu > v_k(\mu; q) \) for all \( \mu \in (0, 1) \) if and only if \( \hat{\mu}_k = 1 \).
- When \( \hat{\mu}_k \in (0, 1) \),

\[ v'_k(\hat{\mu}_k; q) \cdot \hat{\mu}_k = v_k(\hat{\mu}_k; q). \] (6)

Let

\[ y_k(\mu) \equiv v'_k(\mu; q) \cdot \mu - v_k(\mu; q) \quad \text{for all } \mu \in [0, 1] \] (7)

Note that \( y'_k(\mu) = v''_k(\mu; q) \). By Lemma 1, then, \( y'_k(\mu) \) satisfies single-crossing from above. This means \( y_k(\mu) \) is first increasing and then decreasing. Furthermore, \( v'_k(0; q_k) = v'_k(1; q_k) = 0 \) because \( f_k(1) = f_k(0) = 0 \). Then, \( y_k(0) = 0 \) and \( y_k(1) = -1 \).

Define the set \( \mathcal{U}_k \equiv \{ \mu \in [0, 1] : y_k(\mu) \geq 0 \} \). Based on our discussion so far, we conclude that \( \mathcal{U}_k = [0, \hat{\mu}_k] \), for \( k \in \{1, 2\} \).

Next, we show that \( \hat{\mu}_1 \leq \hat{\mu}_2 \). Consider some \( \mu \in [0, \hat{\mu}_1] \), that is, \( y_1(\mu) \geq 0 \). By (7), \( v'_1(\mu; q) \cdot \mu \geq v_1(\mu; q) \). By (3), it implies that

\[
\frac{f_1 \left( \frac{1-\mu}{1-\mu+\frac{q-q_{1-\mu_s}}{\mu_s}} \right)}{1 - F_1 \left( \frac{1-\mu}{1-\mu+\frac{q-q_{1-\mu_s}}{\mu_s}} \right) \left( 1 - \mu + \frac{1-q-q_{1-\mu_s}}{\mu_s} \right)} \geq 1
\]
Because the likelihood ratio order implies the hazard rate order (see, e.g., Theorem 1.C.1 of Shaked and Shanthikumar, 2007), \( \frac{f_2(x)}{1-F_2(x)} \geq \frac{f_1(x)}{1-F_1(x)} \) for all \( x \in [0,1] \), we have
\[
\frac{v'_2(\mu; q)}{v_2(\mu; q)} \mu \geq 1,
\]
which is equivalent to \( y_2(\mu) \geq 0 \). Therefore, \( \mu \in [0, \hat{\mu}_2] \) and \( \hat{\mu}_2 \geq \hat{\mu}_1 \).

To conclude the proof, consider three cases. If \( \mu_s \geq \hat{\mu}_2 \), the optimal policy does not reveal any information in either case. Given that we already set \( \sigma(m=good|\omega=1) = 1 \), the optimal policy includes \( \sigma(m=good|\omega=0) = 1 \). Therefore, \( \beta_1^*(q) = \beta_2^*(q) = 1 \). If \( \hat{\mu}_2 > \mu_s \geq \hat{\mu}_1 \), the optimal policy under \( v_1(\mu; q) \) does not reveal any information. In this case, \( \beta_1^*(q) = 1 \) and \( \beta_2^*(q) < 1 \). Finally, if \( \mu_s > \hat{\mu}_2 \), the optimal propaganda levels of propaganda \( \beta_1^*(q) \) and \( \beta_2^*(q) \) satisfy
\[
\frac{\mu_s}{\mu_s + (1-\mu_s)\beta_k^*(q)} = \hat{\mu}_k, \quad \text{for } k \in \{1,2\}.
\]
Then, \( \hat{\mu}_2 \geq \hat{\mu}_1 \) implies \( \beta_2^*(q) \leq \beta_1^*(q) \).

The next Proposition 2 deals with comparative statics with respect to \( q \), the benefit of opposing the leader. Intuitively, a lower \( q \) means the citizens are more inclined to take \( a_t = 1 \) without any information. Realizing this, the leader sends less informative messages; that, is, chooses a higher level of propaganda. It is vice-versa when \( q \) increases.

**Proposition 2.** The propaganda level chosen by the leader, \( \beta^*(q) \), is decreasing in the payoff of opposing the regime, \( q \).

**Proof of Proposition 2.** Consider \( 0 < q_1 \leq q_2 < 1 \). For \( k \in \{1,2\} \), the value function under \( q_k \) is \( v(\mu; q_k) \). Denoting \( y_k = v'(\mu; q_k)\mu - v(\mu; q_k) \) as in the proof of Proposition 1, let us show that \( \hat{\mu}_1 \leq \hat{\mu}_2 \).

As in Lemma 1,
\[
y_k(\mu) = f(g_k(\mu))(1-g_k(\mu)) \frac{1-g_k(\mu) + g_k(\mu)\gamma_k}{\gamma_k} - (1-F(g_k(\mu)))
\]
where
\[
g_k(\mu) \equiv \frac{1-\mu}{1-\mu+\gamma_k\mu}, \quad \gamma_k \equiv \frac{1-q_k}{q_k} \frac{1-\mu_s}{\mu_s}.
\]
Note that \( g_k(\mu) \) is strictly decreasing in \( \mu \), with \( g_k(0) = 1 \) and \( g_k(1) = 0 \). Therefore, the range of \( g_k(\mu) \) is \([0,1]\). For any \( t \in [0,1] \) define
\[
z_k(t) \equiv f(t)(1-t) \frac{1-t+\gamma_k t}{\gamma_k} - (1-F(t))
\]
Then, \( y_k(\mu) = z_k(g_k(\mu)) \). Because \( y_k(\mu) \) is first increasing and then decreasing in \( \mu \), and \( g_k(\mu) \) is monotonic in \( \mu \), we conclude that \( z_k(t) \) is first increasing and then decreasing in \( t \). Moreover, \( z_k(0) = y_k(1) = -1 \), \( z_k(1) = y_k(0) = 0 \). As a result, the set \( \mathcal{V}_k = \{ t \in [0,1] : z_k(t) \geq 0 \} \) has the form \( \mathcal{V}_k = [\hat{t}_k, 1] \), where \( \hat{t}_k = g_k(\hat{\mu}_k) \).

Based on our discussion so far, characterizing \( \hat{\mu}_k \) is a two-step procedure:
1. First, construct \( V_k \) and find \( \hat{\mu}_k \).

2. Second, calculate 
\[
\hat{\mu}_2 = \frac{1 - \hat{\lambda}_2}{1 - \hat{\lambda}_2 + \gamma_2 \hat{\lambda}_2} \geq \frac{1 - \hat{\lambda}_2}{1 - \hat{\lambda}_2 + \gamma_1 \hat{\lambda}_2} \geq \frac{1 - \hat{\lambda}_1}{1 - \hat{\lambda}_1 + \gamma_1 \hat{\lambda}_1} = \hat{\mu}_1
\]

where the first inequality follows from \( \gamma_1 \geq \gamma_2 \) and the second inequality follows from \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \).

Replicating the argument in the proof of Proposition 1 yields that \( \beta^*(q_1) \geq \beta^*(q_2) \). \( \square \)

### 3 Propaganda with Repression

We now include the possibility of repression in our model. The repression level is determined before the choice of propaganda is made. It takes the following form: repressing the citizens reduces the citizens’ payoffs from taking action \( a_i = 0 \) by some amount \( r \in [0, q] \). However, repression is costly for the ruler. In particular, if the leader chooses a repression level \( r \in [0, q] \), she pays the cost of \( q \cdot c(r) \). Here, \( q > 0 \) parameterizes the cost of repression. We impose the following assumption on the cost of repression:

**Assumption 2.** \( c : [0, q] \rightarrow \mathbb{R} \) is strictly convex, with \( c(0) = c'(0) = 0 \) and \( \lim_{r \to q} c(r) = \infty \).

Assumption 2 ensures that the solution is well-behaved. When repression level is \( r \in [0, q] \), citizens’ payoff functions are given in Table 2.

<table>
<thead>
<tr>
<th>( a_i = 0 )</th>
<th>( \omega = 0 )</th>
<th>( \omega = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q - r )</td>
<td>-r</td>
<td>0</td>
</tr>
<tr>
<td>( 1 - q )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Citizen Payoffs with Repression level \( r \).

In such a framework, any act by the leader that increases the cost of not supporting the leader for citizens is, effectively, *repression* such as harassment, intimidation, imprisonment, torture, or execution. The incidence of repression is determined by the citizens’ choice of actions and, therefore, could be thought of as the expected payoff of not supporting the leader. This is important as even in totalitarian dictatorships only a small share of population is actually punished. (The Great Terror, 1937-1939, the most intensive years of Stalin’s repressions, had about 700,000 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.)

---

\( ^7 \)In line with the logic of our model, Stalin’s Great Terror was accompanied by a massive propaganda campaign (*Conquest*, 2008; Kotkin, 2017). In another example, propaganda has played a critical role ever since the Chinese Communist Party took over in 1949, yet it became even more ferocious during the early years of the Cultural Revolution, 1966-76, which was a combination of elite purges by Mao’s faction and mass terror (*MacFarquhar and Schoenhals, 2006*).
repression, when people are targeted based on imperfect information about their priors, in Subsection 3.2.

The timing of the game with repression is as follows: first, the leader chooses repression level \( r \in [0, q] \). Then, the leader commits to an information structure. Finally, the message is drawn, the citizens take their actions, and payoffs are realized.

### 3.1 Optimal Combination of Repression and Propaganda

Consider a citizen 7 with posterior \( \mu'_7 \). She supports the leader if and only if \( \mu'_7 \cdot (1-\varphi) \geq (1-\mu'_7) \cdot \varphi - r \), which simplifies to \( \mu'_7 \geq q-r \). Therefore, the leader’s value function is \( v(\mu; q-r) \) defined in (3), and the characterization results of Section 2.2 still apply. Consequently, the propaganda level chosen by the leader is \( \beta^*(q-r) \), and the leader’s (subjective) payoff from propaganda is \( V^*(q-r) \). Then, the leader’s optimal level of repression, \( r^* \), is the solution to the following problem:

\[
    r^* \in \arg \max_{r \in [0,q]} \{ V^*(q-r) - \phi \cdot c(r) \}, \tag{8}
\]

and the equilibrium propaganda level is \( \beta(q-r^*) \).

We are interested in how the leader’s chosen repression and propaganda levels are affected by the cost of repression, which is parameterized by \( \phi \). Our next result shows, naturally, that when repression is cheaper, it is used more intensely.

**Lemma 2.** The leader’s chosen repression level, \( r^* \), is decreasing in the cost of repression, \( \phi \).\(^8\)

**Proof.** By Assumption 2, the objective function \( V^*(q-r) - \phi \cdot c(r) \) is strictly submodular in \( \phi \) and \( r \). The result follows from usual supermodularity arguments (e.g., Theorem 5 of Milgrom and Shannon, 1994). \( \square \)

The following Proposition 3, our main result, summarizes the above analysis.

**Proposition 3.** The leader’s optimal propaganda level, \( \beta^* \), is decreasing in the cost of repression, \( \phi \).

**Proof.** By Lemma 2, \( r^* \) is decreasing in \( \phi \). By Proposition 2, \( \beta^*(q-r^*) \) is increasing in \( r^* \), and therefore it is decreasing in \( \phi \). \( \square \)

Proposition 3 shows that, when the cost of repression is lower, a higher intensity of repression is accompanied by a higher level of propaganda. Propaganda and repression are complements.

### 3.2 Ex ante Repressions

One critical feature of repression as defined above is that the punishment applies to those who act, in equilibrium, against the dictator. Let us consider another type of repression – when citizens are

\(^8\)Our treatment does not exclude the possibility that there are multiple optimal levels of \( r^* \). In that case, Lemma 2 and Proposition 3 should be interpreted as decreasing in the sense of strong set order (Milgrom and Shannon, 1994).
targeted *ex ante*, rather than *ex post*. For instance, the dictator might identify, perhaps with some noise, those who are hostile to the regime and eliminate them. In this case, repression alters the distribution of attitudes towards the regime, rather than the incentives for citizens. Below, we use Proposition 1 to demonstrate that the effect of *ex ante* repression is qualitatively the same: the more “pro-regime” is the resulting distribution, the higher is the level of propaganda for those who are not eliminated.\(^9\)

Let \(f_2\) be any density function that satisfies Assumption 1. The leader has access to some informative signal about citizens’ priors. As in the model of identity-based Stalinist repression in Gregory, Schröder and Sonin (2011), suppose that there is an institution, e.g., a secret police, that assigns a label \(\ell_i \in \{\text{Skeptic, Supporter}\}\) to each citizen \(i\) with prior \(\mu_i\). The secret police’s labeling technology is noisy:

\[
\begin{align*}
\Pr\{\ell_i = \text{Supporter} \mid \mu_i\} &= \rho(\mu_i), \\
\Pr\{\ell_i = \text{Skeptic} \mid \mu_i\} &= 1 - \rho(\mu_i),
\end{align*}
\]

where \(\rho(\mu_i) \in [0, 1]\). Suppose \(\rho(\mu)\) is continuously differentiable, increasing and log-concave in \(\mu \in [0, 1]\): 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.

It is straightforward to see that the leader will prefer to purge only those who are labeled *Skeptic*. Suppose for simplicity that the leader purges all citizens labeled \(\ell_i = \text{Skeptic}\) from the society, and only those with label \(\ell_i = \text{Supporter}\) remain. Denote the distribution of priors of the remaining citizens by \(f_1\):

\[
f_1(\mu) = \frac{\rho(\mu)f_2(\mu)}{\int_0^1 \rho(\tilde{\mu})f_2(\tilde{\mu})d\tilde{\mu}}.
\]

As long as \(\rho(\mu)\) is continuously differentiable and log-concave, \(f_1(\mu)\) satisfies Assumption 1.\(^{10}\) Moreover,

\[
\frac{f_1(\mu)}{f_2(\mu)} = \frac{\rho(\mu)}{\int_0^1 \rho(\tilde{\mu})f_2(\tilde{\mu})d\tilde{\mu}}
\]

which is increasing in \(\mu\) because \(\rho(\mu)\) is increasing. Therefore, Proposition 1 applies, and the leader engages in more propaganda following a purge of skeptics. Again, repression and propaganda complement each other: with skeptics purged, the rest accepts more persuasion.

---

\(^9\)In a setup with a continuum of media, heterogeneous preferences of the audience, and common priors, Kolotilin, Mylovanov and Zapechelnyuk (2022) show that the government opts to censor more media when preferences of the audience are more aligned.

\(^{10}\)Most importantly, \(\log f_1(\mu) = \log \rho(\mu) + \log f_2(\mu) - \log \int_0^1 \rho(\tilde{\mu})f_2(\tilde{\mu})d\tilde{\mu}\), which is strictly concave in \(\mu\), and therefore \(f_1(\mu)\) is strictly log-concave.
4 Targeted Propaganda

In the analysis of ex ante repression, the crucial precondition for the complementarity of propaganda and repression was reliance on purges being targeted towards citizens who take a particular action. A natural question to ask is whether the leader also benefits from targeted propaganda. In this section, we argue that the answer is no. In particular, allowing for private persuasion, i.e., the opportunity to design type-specific propaganda, does not actually expand the menu of the leader’s tools. We show that it is never (strictly) optimal for the leader to create different information sources that appeal to different citizens. Therefore, under the optimal policy, the leader sticks with public propaganda.

Our result on targeted persuasion is closely related to the “impossibility of private persuasion” result in Kolotilin et al. (2017); the difference is that our result is in a setup with heterogeneous priors, rather than with heterogeneous preferences. We cannot rely on the garbling result of Blackwell (1953) as Kolotilin et al. (2017) does, as Blackwell’s famed result requires common posteriors following a public message. However, we could modify the proof of Proposition 2 in the Online Appendix to Kolotilin et al. (2017) to obtain a similar result in our setting.

Consider the setup in Section 2, which dealt with persuasion in the absence of repression, but suppose that instead of being restricted to use a public persuasion mechanism, the leader can use private persuasion mechanisms. Each citizen receives a message \( m_i \in \{0, 1\} \) from the news source designed for her. To this end, consider the following setup: each citizen reports her prior \( \mu \in [0, 1] \), and given the reported prior, the mechanism sends an action recommendation \( d \in \{0, 1\} \).

**Definition 1.** A persuasion mechanism \( \sigma \) is \( \sigma = \{\sigma(\mu, 0), \sigma(\mu, 1)\}_{\mu \in [0,1]} \) where

\[
\sigma(\mu, \omega) = \Pr(m = 1|\mu, \omega) \in [0, 1] \quad \text{for } \mu \in [0, 1], \omega \in \{0, 1\}
\]

Consider a receiver \( i \) with prior \( \mu_i = \mu \). Her (subjective) payoff from following reporting a prior \( \tilde{\mu} \), taking action \( d_0 \in \{0, 1\} \) following message \( m = 0 \), and taking action \( d_1 \in \{0, 1\} \) following \( m = 1 \) is

\[
U_\sigma(\mu, \tilde{\mu}, d_0, d_1) \equiv \mu \cdot \sigma(\tilde{\mu}, 1) \cdot d_1 \cdot (1 - q) + (1 - \sigma(\tilde{\mu}, 1)) \cdot d_0 \cdot (1 - q) \\
+ (1 - \mu) \cdot \sigma(\tilde{\mu}, 0) \cdot (1 - d_1) \cdot q + (1 - \sigma(\tilde{\mu}, 0)) \cdot (1 - d_0) \cdot q
\]

The critical element of private persuasion is incentive compatibility. The essence is that if a private persuasion mechanism is in place, then each citizen observes and follows her news source, not others’ news sources. So if \( \sigma \) is an incentive compatible private persuasion mechanism, citizen \( i \) voluntarily chooses to get the messages designed exclusively for her, \( (\sigma(\mu, 0), \sigma(\mu, 1)) \).

The assumption that each citizen follows only one news source requires an explanation of its own. One standard explanation is cognitive constraints; another explanation is a budget con-
straint in terms of opportunity costs. If citizens were able to follow more than one source, then of course they would follow all possible news sources. In this case, any persuasion mechanism will be equivalent to a public persuasion mechanism; we analyzed this case in Section 2.

**Definition 2.** A persuasion mechanism $\sigma$ is **incentive compatible** if, for all $\mu \in [0, 1]$,

$$U_\sigma(\mu, \mu, 0, 1) \geq U_\sigma(\mu, \tilde{\mu}, d_0, d_1)$$

for all $\tilde{\mu} \in [0, 1]$, $d_0, d_1 \in \{0, 1\}$.

Let the subjective payoff of a truthful and obedient citizen (i.e., a citizen who follows the news source designed for her and takes $d_0 = 0, d_1 = 1$) with prior $\mu$ be $U_\sigma(\mu) \equiv U_\sigma(\mu, \mu, 0, 1)$. Standard mechanism design arguments yield the following result:

**Lemma 3.** If $\sigma$ is an incentive compatible persuasion mechanism, then $U_\sigma(\mu)$ is convex, with $U_\sigma(0) = q$ and $U_\sigma(1) = 1 - q$. Moreover, both $\sigma(\mu, 0)$ and $\sigma(\mu, 1)$ are increasing in $\mu$, with

$$U_\sigma'(\mu) = (1 - q)\sigma(\mu, 1) - q(1 - \sigma(\mu, 0)),$$

$$U_\sigma(\mu) - \mu U_\sigma'(\mu) = q(1 - \sigma(\mu, 0))$$

for all $\mu \in [0, 1]$.

We are going to prove that for any incentive compatible private persuasion mechanism, there exists a direct public mechanism that achieves the same outcome. Formally, a public persuasion mechanism $c : \{0, 1\} \rightarrow \Delta(M)$ is a distribution of messages in each state. Let $M = [0, 1]$, and consider direct public mechanisms for the sender: for each $\mu \in M$, $\Pr_s{\omega = 1 | \mu} = \mu$. That is, under a direct public mechanism, the sender’s posterior following a message is the message itself. Given a state $\omega$, let $\pi(\mu | \omega)$ denote the cdf of messages. Consequently, the ex ante distribution of messages (from the sender’s perspective) is $\pi_s(\mu) = \mu_s \pi(\mu | \omega = 1) + (1 - \mu_s) \pi(\mu | \omega = 0)$. For a mechanism to be direct, it needs to satisfy $\mathbb{E}_s[\mu] = \mu_s$.

We say that a direct public mechanism $\pi$ **achieves the same outcome** as a private persuasion mechanism $\sigma$ if it induces the same distribution of actions for each receiver in each state. Note that because of heterogeneous priors, citizens expect different states to realize with different probabilities, so the ex ante distributions of actions still differ.

**Proposition 4.** For any incentive compatible private persuasion mechanism $\sigma$, there exists a direct public mechanism $\pi$ that achieves the same outcome.

Proposition 4 implies that the leader can achieve any outcome via a public persuasion mechanism, where he offers the same information structure for each citizen. This justifies our focus on public mechanisms in Section 2 and 3, and it implies that the optimal policy we characterized yields a higher payoff to the leader than any incentive compatible private persuasion mechanism.

---

12Because we are working with heterogeneous priors, the posterior following a message is different across receivers. It is, therefore, impossible to define a direct public mechanism for everyone. We opt for defining a direct public mechanism for a particular agent, the sender.
Intuitively, this is because the incentive compatibility constraints are extremely binding for the leader, to the extent that a public mechanism (which satisfies incentive compatibility trivially) can yield the same payoff. Substantively, this provides an explanation why many authoritarian regimes prefer a standardized approach to censorship and mass propaganda.

**Proof.** Consider the public persuasion mechanism defined as

\[ \pi(\mu|\omega) = 1 - \sigma(g(\mu), \omega) \quad \text{for all } \mu \in [0, 1], \omega \in \{0, 1\} \]

where again

\[ g(\mu) \equiv \frac{1 - \mu}{1 - \mu + \gamma \mu}, \quad \gamma \equiv \frac{1 - q}{q} \frac{1 - \mu_s}{\mu_s}. \]

Note that \( g(\mu) \) is strictly decreasing in \( \mu \), with \( g(0) = 1 \) and \( g(1) = 0 \). Moreover, \( g(g(\mu)) = \mu \) for all \( \mu \in [0, 1] \).

Take \( \mu, \mu' \in [0, 1] \) with \( \mu \geq \mu' \). Since \( g \) is strictly decreasing, \( g(\mu) \leq g(\mu') \). By Lemma 3, \( \sigma(\mu, \omega) \) is increasing in \( \mu \), which implies: \( \sigma(g(\mu), \omega) \leq \sigma(g(\mu'), \omega) \). Therefore, \( \pi(\mu|\omega) \geq \pi(\mu'|\omega) \). We conclude that \( \pi(\mu|\omega) \) is increasing in \( \mu \) and it is a cdf.

Next, we show that \( \mathbb{E}_s[\mu] = \mu_s \), so that this is indeed a direct mechanism. Using \( (1 - q)\sigma(g(\mu), 1) - q(1 - \sigma(g(\mu), 0)) = U'_\sigma(g(\mu)) \) and \( (1 - \gamma)q(1 - \sigma(g(\mu), 0)) = U_\sigma(g(\mu)) - g(\mu)U'_\sigma(g(\mu)) \), both by Lemma 3, we obtain

\[ \pi_s(\mu) = \mu_s\pi(\mu|\omega = 1) + (1 - \mu_s)\pi(\mu|\omega = 0) \]

\[ = \mu_s(1 - \sigma(g(\mu), 1) + (1 - \mu_s)(1 - \sigma(g(\mu), 0)) \]

\[ = \mu_s - \frac{\mu_s}{1 - q} \left[ (1 - \gamma)U_\sigma(g(\mu)) + (1 - g(\mu) + \gamma g(\mu))U'_\sigma(g(\mu)) \right]. \]

Multiplying the second term by \( \frac{1}{g(\mu) + \gamma g(\mu)} g'(\mu) = 1 \) and re-arranging terms yields

\[ \pi_s(\mu) = \mu_s + \frac{1 - \mu_s}{q} \left[ \frac{1 - \gamma}{(1 - g(\mu) + \gamma g(\mu))^2}U_\sigma(g(\mu)) + \frac{1}{1 - g(\mu) + \gamma g(\mu)}U'_\sigma(g(\mu)) \right] g'(\mu). \]

Taking the integral and changing variables with \( v = g(\mu) \) gives

\[ \int_0^1 \pi_s(\mu) d\mu = \mu_s + \frac{1 - \mu_s}{q} \int_{v=g(0)}^{v=g(1)} \left[ \frac{1 - \gamma}{(1 - v + \gamma v)^2}U_\sigma(v) + \frac{1}{1 - v + \gamma v}U'_\sigma(v) \right] dv. \]

Since \( g(0) = 1 \) and \( g(1) = 0 \), then

\[ \int_0^1 \pi_s(\mu) d\mu = \mu_s - \frac{1 - \mu_s}{q} \int_0^1 \left( \frac{d}{d\mu} \frac{U_\sigma(\mu)}{1 - v + \gamma v} \right) dv \]

\[ = \mu_s - \frac{1 - \mu_s}{q} \left( \frac{U_\sigma(1)}{\gamma} - U_\sigma(0) \right) \]

By Lemma 3, \( U_\sigma(0) = q \) and \( U_\sigma(1) = 1 - q \), so \( \int_0^1 \pi_s(\mu) d\mu = 1 - \mu_s \). Therefore,

\[ \mathbb{E}_s[\mu] = \mu_s, \]

15
and we conclude that the public persuasion mechanism \( \pi \) is direct.

Finally, we show that \( \pi \) induces the same distribution of actions as \( \sigma \) for each receiver in each state. This is true because for each receiver \( i \) with prior \( \mu_i \) and each state \( \omega \), the probability of taking action \( a_i = 1 \) is

\[
\Pr\{a_i = \omega\} = \Pr\left\{ \frac{\mu_i}{\mu_i + (1 - \mu_i) \frac{1 - \mu}{1 - \mu_i}} \geq q \middle| \omega \right\}
\]

\[
= \Pr\left\{ \mu_i \geq \frac{1}{\mu_i + \gamma} \middle| \omega \right\}
\]

\[
= \Pr\{\mu \geq g(\mu_i)\mid \omega\}
\]

\[
= 1 - \pi(g(\mu_i)\mid \omega).
\]

Since \( \pi(\mu\mid \omega) = 1 - \sigma(g(\mu), \omega), \Pr\{a_i = \omega\} = \sigma(g(g(\mu_i)), \omega) \). Moreover, since \( g(g(\mu)) = 1 \) for each \( \mu \in [0, 1] \), \( \Pr\{a_i = \omega\} = \sigma(\mu_i, \omega) \), and the result follows. \( \square \)

5 Conclusion

We offer a model of information manipulation and repression, two main tools in any autocrat's arsenal, considering both public and private persuasion and different types of repression. With a higher level of repression, the leader’s marginal supporter is more disposed towards support and, therefore, can be more heavily manipulated. In George Orwell’s Oceania, people are forced to use the *newspeak*, a special language designed to limit their ability to articulate anti-government concepts, cannot switch off radio that translates propaganda, and are forced to participate in ideological indoctrination meetings. Yet the ultimate message of *1984* is that it is the physical torture, applied to some, that makes citizens believe what the government wants them to believe.
References


