IEB Working Paper 2026/06

SEGMENT AND RULE: MODERN CENSORSHIP IN AUTHORITARIAN REGIMES

Kun Heo, Antoine Zerbini

Version February 2026

Political Economy

IEB Institut d'Economia de Barcelona

IEB Working Paper 2026/06

SEGMENT AND RULE: MODERN CENSORSHIP IN AUTHORITARIAN REGIMES

Kun Heo, Antoine Zerbini

The **Barcelona Institute of Economics (IEB)** is a research centre at the University of Barcelona (UB) which specializes in the field of applied economics. The IEB is a foundation funded by the following institutions: "La Caixa" Foundation, Saba, the Barcelona City Hall, the Barcelona Metropolitan Area, the University of Barcelona, the Autonomous University of Barcelona, the Barcelona Provincial Council, Agbar, Cuatrecasas, Consorci Zona Franca Barcelona, Fundación Miquel y Costas, AENA and Fira de Barcelona.

The IEB Research Programme in **Political Economy** explores the relationship between political and economic systems. Its primary goal is to gain deeper insight into topics such as the determinants of voting behavior and policy-making decisions, the role of institutions in shaping economic policy outcomes, the political economy of immigration, technological change, trade shocks, climate change, and the effects of fiscal and political decentralization. The programme also aims to provide insights into the impact of political institutions and actors on public service and goods provision, ultimately generating valuable knowledge for policy-making and institutional design. The program enjoys the support from the IEB-Foundation.

Postal Address:
Institut d'Economia de Barcelona
Facultat d'Economia i Empresa
Universitat de Barcelona
C/ John M. Keynes, 1-11
(08034) Barcelona, Spain
Tel.: + 34 93 403 46 46
ieb@ub.edu
http://www.ieb.ub.edu

IEB Working Paper 2026/06

SEGMENT AND RULE: MODERN CENSORSHIP IN AUTHORITARIAN REGIMES

Kun Heo, Antoine Zerbini*

ABSTRACT: We analyze the incentives of authoritarian regimes to segment access to censored content through technology. Citizens choose whether to pay to access censored online content at a cost fixed by the regime: the firewall. A low firewall segments access and generates more compliance than full censorship – a high firewall – ever could. Regime opponents self-select into consuming censored content, and comply conditional on positive independent reporting. Regime supporters exclusively consume state propaganda, which secures their compliance. This segment-and-rule strategy can be engineered by making local news outlets uninformative, or by affecting the intrinsic benefit from access.

JEL Codes: D72, D82, D83, O33
Keywords: Censorship, Internet, Rational Attention, Cheap Talk, Information Design, Segmentation

Kun Heo
London School of Economics
E-mail: b.heo@lse.ac.uk

Antoine Zerbini
Universitat Autònoma de Barcelona,
IEB & BSE
E-mail: antoine.zerbini@bse.eu

*We thank Emiel Awad, Benjamin Blumenthal, Roel Bos, Laurent Bouton, Luis Bosshart, Martin Castillo Quintana, Anna Denisenko, Torun Dewan, Hanna Folsz, Catherine Hafer, Gleason Judd, Navin Kartik, Dimitri Landa, Gilat Levy, Elliot Lipnowski, Sergi Martinez, John Marshall, Andrea Mattozzi, Justin Melnick, Giacomo Ponzetto, Carlo Prato, Kris Ramsay, Roxanne Rahnama, Arturas Rozenas, Stephane Wolton, Noam Yuchtman and seminar and conference participants at Columbia, EEPGW, FTCP, MPSA, EPSA, NYU, LSE-NYU 2023, Oxford, PSE, Bruneck 2023, VFTW, Bologna, UAB, Cologne Economics of Media Bias 2024, the 2024 BSE Public Economics Summer Forum, IEB, Universidad de los Andes and SWET Paris 2025 for helpful feedback and comments. Zerbini gratefully acknowledges support from the Spanish Ministry of Science and Innovation (grants CNS2022-135749 and PID2022-137707NB-I00).

1 Introduction

> The internet is uncontrollable. And if the internet is uncontrollable, freedom will win. It's as simple as that.
>
> — Ai Weiwei in 2012 in the Guardian.

The roll-out of the internet across the world was first hailed as a liberating technology for citizens of authoritarian regimes (Diamond, 2010). On top of being able to communicate more easily (Manacorda and Tesei, 2020; Acemoglu, Hassan, and Tahoun, 2018), individual citizens gained the right to decide whether to consume foreign outlets that were previously censored and inaccessible. According to the liberating view, this increased autonomy would then empower citizens in their struggle against authoritarian regimes. This optimism is substantiated by empirical evidence about the difficulties of authoritarian leaders with online censorship. Indeed, millions bypass censorship firewalls everyday.[1]

We contend that authoritarian regimes can benefit from having a positive yet low cost of access – a VPN subscription suffices to bypass it – to ensure that a specific segment of the population self-selects into access to the uncensored internet. On the one hand, the firewall's mild deterrent effect dissuades supporters from seeking out banned content that could otherwise undermine their support for the regime.[2] On the other hand, it does not prevent regime opponents from gaining access. Crucially, the compliance of regime opponents can hardly be garnered without exposure to "positive" reporting from credible sources, such as those censored domestically.[3] For instance, some citizens of authoritarian regimes can become more supportive of their governments after experiencing disillusionment with the outside world. Some Chinese students report increased patriotism

[1] Empirical evidence suggests that at least 5-10% of internet users have at some point used circumvention softwares to bypass the firewall in China (Chen and Yang, 2019; Hobbs and Roberts, 2018; Shen and Zhang, 2018; Mou, Wu, and Atkin, 2016). Similar evidence has been provided in Iran (Dal and Nisbet, 2022), Russia (Fung, 2022; Xue et al., 2022), Egypt (Lutscher, 2023) or Turkmenistan (Nourin et al., 2023). Further, the very observation of online censorship may reinforce the citizens' incentives to bypass firewalls (Hobbs and Roberts, 2018) and organize political movements (Boxell and Steinert-Threlkeld, 2019; Pan and Siegel, 2020).

[2] Experimental evidence shows that VPN access reduces regime support among regime supporters in China (Chen and Yang, 2019) and Rwanda (Bowles, Marshall, and Raffler, 2024). The latter study also documents a null effect among regime opponents.

[3] Positive and credible reporting need not involve independent outlets praising the authoritarian regime. When US media outlets known to be biased against the CCP report negatively about domestic US issues, such as the opioid crisis, pandemic response, or gun violence—this constitutes positive reporting for the CCP by painting it in a favorable light relative to the US. Huang (2015) and Huang and Yeh (2019) provide empirical evidence of such "relative updating" after exposure to foreign news in China, while Chester (2023) documents this "relative argumentation" by CCP-controlled outlets.

following their studies abroad[4] while a few North Korean defectors have returned home after growing disenchanted with capitalist society (Kim, Seo et al., 2022).

Thus, by imposing an intermediary cost of access to independent information, authoritarian regimes can ensure that their citizens self-segment into supporters and opponents, with only the latter accessing independent information. This strategy of "segment-and-rule", by allowing for tailored communication with the population without the need for information on its citizenry, can garner greater compliance than full censorship ever could. While recent work has documented how modern authoritarian regimes exploit technological advances such as AI for surveillance purposes (Tirole, 2021; Dragu and Lupu, 2021; Beraja et al., 2023; Xu, 2023), we present an even bleaker picture: the internet can entrench authoritarian regimes precisely *because* it empowers citizens.

Illustrative example. A leader maximizes compliance from a population of citizens indexed by their political type $\theta_i \in [0, 1]$, which defines their payoff from *not* complying. The payoff from compliance depends on an underlying state of the world $\omega \in \{0, 1\}$ with $Pr(\omega = 1) = 0.5$. Prior to choosing whether to comply ($a_i \in \{0, 1\}$), each citizen decides whether to pay to access an independent media. Access to this outlet (i) delivers a report that perfectly reveals the state of the world and (ii) provides some intrinsic payoff that could depend on a citizen's type $\alpha(\theta_i)$.[5] The cost to access this outlet is chosen by the leader at the beginning of the game.

Without any access to information, a citizen will comply if and only if her expectation over the state of the world is larger than her payoff from non complying: $E[\omega] = Pr(\omega = 1) = 0.5 > \theta_i$. Thus, by setting a sufficiently high cost, the leader obtains a payoff of $F(0.5)$, where F denotes the cdf associated with the distribution of political types. If, instead, all citizens gain access to the report – e.g., if the cost of access is null or negative – the leader obtains a payoff equal to the likelihood of good news: namely $0.5 * F(1) = 0.5$.

Our main point is that the leader can do better than these two extremes, even without additional knowledge about a citizen's political leaning. Suppose that the leader selects an intermediary cost that leads to some form of segmented access to the independent outlet. To understand who would gain access, notice first that the informational value of access to the outlet is given by $0.5 * 1 + 0.5 * \theta_i - 0.5 * 1 = 0.5\theta_i$ for a "supporter" ($\theta_i \leq 0.5$) and for $0.5 * 1 + 0.5 * \theta_i - \theta_i = 0.5(1 - \theta_i)$

[4] See exhibit 6 in Mao (2020) and Fish (2018).

[5] Empirically, citizens of authoritarian regimes have been shown to bypass the firewall to (mostly) access a variety of non-informational content (Chen and Yang, 2019; Hobbs and Roberts, 2018).

for an "opponent" ($\theta_i > 0.5$). That is, the more ex-ante uncertain a citizen is, the more valuable information is. Suppose that the more a citizen is opposed to the regime, the more they intrinsically enjoy access to banned foreign content: $\alpha(\theta_i) = 0.51\theta_i$. Then the total benefit is given by $1.01\theta_i$ for supporters and $0.5 + 0.01\theta_i$ for opponents. The leader could impose a cost of at least 0.51, thus imposing full censorship, which generates, again, a payoff of $F(0.5)$. Instead, he will maximize compliance by imposing a lower cost of 0.505, thus securing the certain compliance of supporters while getting that of opponents sometimes, i.e. $F(0.5) + (1 - F(0.5)) * 0.5$. We refer to such a strategy as one of *segment-and-rule*.

In the paper, we generalize the conditions under which such a strategy benefits the sender. In our baseline model (sections 3 and 4) the leader is endowed with a private signal about the state of the world; he then chooses the cost of access and sends a public cheap talk message. Citizens observe these two actions prior to choosing among a wide range of banned foreign outlets once bypassing the firewall, and finally decide whether to comply. We first show that as long as the willingness to pay for the independent outlet increases in opposition to the leader, a strategy of segment-and-rule is implemented in equilibrium. Further, despite the potential for signaling through a lower cost of access, as long as segment-and-rule is feasible, no information is conveyed in equilibrium through the choice of the cost of access (nor the cheap-talk message): both a "good" and a "bad" leader pool on the segmenting cost.

Next, to understand and fully characterize the optimal censorship and persuasion strategy of the leader, in section 5 (i) the leader controls a *state-media* and can commit to a reporting strategy, (ii) we restrict attention to log-concave distributions of citizens and (iii) we impose that the informativeness of the independent media is common across citizen.[6] Equilibrium play is delineated along one primitive: the strength of the correlation between opposition to the leader and the intrinsic benefit. When there is no (or a small and positive) correlation between the intrinsic benefit from access and opposition to the leader, no citizen gains access in equilibrium: the leader imposes the lowest cost such that no citizen gains access. To see why, observe that the firewall imposes a *common* cost of access. Further, the informational value is highest for "moderates", who are unsure whether to comply after the report from the state media, and, in equilibrium, represent a sizable share of the

[6]This can be interpreted as the leader not being sure about how a citizen's type map into a particular choice of independent outlet, and thus taking a "robust" approach by assuming that all citizens consume from the most or least informative independent outlet.

population. Picking a lower cost would induce some opponents – a "minority" – *and* some moderates into bypassing the firewall. Thus a strategy of segment-and-rule is not feasible, and the leader resorts to full censorship, which characterizes a lower bound on the level of compliance.

At the other extreme, a strong correlation ensures that segment-and-rule is always feasible, as in the baseline model. Relative to the full censorship baseline, compliance is increased through two effects. Fixing the reporting strategy of the state media, the leader can lower the cost of access to obtain – with positive probability – the compliance of opponents not convinced by good news from the state media; a *direct* effect. Further, since opponents now condition their compliance on the report of the independent media (conditional on observing it), the leader can now focus on maximizing compliance among his base, by making the state media less informative; an *indirect* effect. When the correlation is intermediary, the leader *engineers* a strategy of segment-and-rule via his control of the state media. There, segment-and-rule is not feasible while picking the same reporting strategy as in the strong correlation case, yet the leader can do better than full censorship. By making the state media less informative, he increases the informational benefit from access. This dampens the negative effect on the informational benefit, of increasing opposition to the leader (among opponents). In turn, sorting into access occurs more along the intrinsic dimension, allowing for segment-and-rule to be feasible at lower levels of correlation. Nevertheless, this strategy requires making the state media less informative than would otherwise be optimal: the level of compliance is monotonically increasing in the correlation between opposition to the leader and the intrinsic benefit.

So far, we suggest one explanation for the empirical pattern of selective bypassing of the firewall, which rests on a form of endogenous sorting and yields a clear implication: the set of citizens bypassing the firewall includes the strongest opponents, as suggested by the limited empirical evidence.[7] An alternative explanation for this empirical pattern posits that authoritarian regimes impose a low cost of access because of technological or economic motives (e.g., to ensure cheap access to the "free" internet to internationally-oriented businesses). In section 6.2 we provide suggestive evidence against this being the *only* explanation for selective bypassing. Indeed, under this alternative explanation, as long as authoritarian regimes are not too constrained by technological or economic constraints,

[7] E.g., see Mou, Wu, and Atkin (2016) or Shen and Zhang (2018). Also, existing VPN users in Chen and Yang (2019) tend to be less optimistic about the economy, trust the Chinese government less and are more willing to act against the regime.

the strongest opponents of the regime do *not* gain access, because their informational benefit from doing so is too low, which seems at odds with the empirical evidence. More generally, our theoretical framework suggests caution when empirically measuring censorship. We show in section 6.3 that when a strategy of segment-and-rule is implemented in equilibrium, as one (top-down) measure of censorship increases – the reporting slant of the state media, which can be empirically measured Qin, Strömberg, and Wu (2018) – another (bottom-up) measure – the share of citizens with access to independent news, empirically harder to measure – goes down.

Finally, we use our framework to understand the instrumental value of controlling the production and access to entertainment for authoritarian regimes, to tailor information provision to different segments of citizens. We show in section 7.1 that even if entertainment contains absolutely no informational content, authoritarian regimes have an incentive to produce domestic entertainment that mostly appeals to their base and/or to selectively ban entertainment that mostly appeals to regime opponents, to generate the appropriate endogenous sorting along partisan lines that makes a strategy of segment-and-rule feasible.[8]

While our focus is on the effect of the internet on censorship and authoritarian rule, we believe the larger theoretical implications of our framework to be useful for applications in other contexts. Our broader point is that a sender can drastically benefit from the existence of an independent source of information available to the receivers he faces, even if he knows this source to have misaligned interests and to be most likely to reveal "bad" news. In section 6.1 we show that when a strategy of segment-and-rule is implemented in equilibrium, making the independent source more informative need not discipline the sender (Gentzkow and Kamenica, 2017; Galvis, Snyder, and Song, 2016; Kronick and Marshall, 2024). On the extensive margin, making the independent source sufficiently more informative might indeed render segment-and-rule impossible, as sorting then mostly happens along the informational dimension. On the intensive margin, however, a marginal increase in the independent's source informativeness pushes the sender to reveal less information, as opponents become more likely to comply irrespective of how much information he discloses. The optimal

[8]In present-day China, beyond the imposition of a low but positive cost of access via the firewall, the CCP also strategically invests in domestic entertainment such as high profile historical war movies – e.g., *The Battle at Lake Changjin* – or police drama about a corrupt administration – e.g., *The Knockout*; see also Liu and Yao (2023) for empirical evidence of the use of entertainment for propaganda purposes. Further, while high budget and mostly apolitical movies with broad appeal are allowed (e.g., *Jurassic World* or *Transformers*), movies that mostly appeal to opponents are banned (e.g., *Cockroach*, *Eternal Spring* or *Top Gun - Maverick*). Esberg (2020*b*) also documents the censorship of entertainment favored by regime opponents in the Chilean dictatorship.

independent media to weaken such an authoritarian regime is thus either fully informative – to prevent segment-and-rule – or fully uninformative – to render segment-and-rule useless to the regime.

The rest of the paper is organized as follows. We first review the related literature prior to presenting the baseline setup in section 3. We present equilibrium results in sections 4 and 5. In section 6 we turn to comparative statics and conceptual, empirical and policy implications. We then discuss in section 7 how entertainment can be used strategically by authoritarian regimes, and present some extensions prior to concluding.

2 Related Literature

Our main substantive contribution is to the political economy literature on technological change and modern censorship (Chen and Yang, 2019; Zhuravskaya, Petrova, and Enikolopov, 2020; Egorov and Sonin, 2023; Bowles, Marshall, and Raffler, 2024). To explain the variation in censorship levels across autocracies, scholars have highlighted that censorship may backfire when it leads to a loss in valuable entertainment (Kronick and Marshall, 2024) or that it may come at an economic cost by making the monitoring of the state apparatus more difficult (Egorov, Guriev, and Sonin, 2009; Lorentzen, 2014). While Edmond (2013) shows in a global game setting that authoritarian regimes may benefit from the roll-out of the internet if the information technology is easily centralized, we argue that authoritarian regimes purposely engage in a strategy of segmented access to banned content, which implies a low cost of access, irrespective of technological capacity. In that sense we provide one micro-foundation of *how* "informational autocracies" (Guriev and Treisman, 2019, 2020) leverage various sources of information to maximize compliance in heterogeneous citizenries. The strategy of segment-and-rule differs from the age-old strategy of *divide-and-rule* because, instead of disrupting coordination among (groups of) citizens (Acemoglu, Robinson, and Verdier, 2004) or elites (Luo and Rozenas, 2023) through the strategic distribution of resources or information, our leader exploits the heterogeneity of political preferences, in a setting without any collective action problem. As in related political economy papers (Gratton and Lee, 2024; Gitmez and Sonin, 2023; Heo and Zerbini, 2024; Gitmez and Molavi, 2023), beyond the cheap-talk setup, we model the autocrat's censorship problem using an information design approach. Unlike in the cheap-talk setting, when the sender has commitment power and faces a rationally inattentive audience, he can use this commitment power to affect whom listens to what, which in turn further facilitates his persuasion problem.

Theoretically our main contribution lies in the literature on persuasion (Crawford and Sobel, 1982), information design (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2019) and rational inattention (Maćkowiak, Matějka, and Wiederholt, 2023). In the information design literature on price discrimination (Bergemann, Brooks, and Morris, 2015) and private bayesian persuasion (Bardhi and Guo, 2018; Chan et al., 2019; Arieli and Babichenko, 2019) the designer leverages individual level information, to target her communication to the type of the receiver. Similarly in our setting the sender benefits from targeting communication to the type of the receiver. Crucially however, our sender does not know whom he is facing and thus cannot communicate privately: he can only impose a *common* cost of access and design any *public* experiment.[9] Closest to our framework is Matysková and Montes (2023), who consider a game of bayesian persuasion with rational inattention. They show that the sender's payoff is decreasing in the receiver's cost of information acquisition. By modeling a sender that faces a heterogeneous set of receivers, we instead show the sender's payoff is non-monotonic and single-peaked in the receiver's cost of information acquisition.

3 Model

Consider a game between an authoritarian leader A (he) and a $[0,1]$ continuum of citizens (they or she) indexed by the subscript i.

Citizens' actions and preferences. The citizens choose between two actions: $a_i \in \mathcal{A} = \{0,1\}$. $a_i = 1$ represents compliance with the regime and $a_i = 0$ represents non-compliance. The leader maximizes compliance in the citizenry: his payoff is given by $V(s_{\mathcal{S}}, c) = \int_0^1 a_i di$, with $s_{\mathcal{S}}$ and c defined below. A citizen's payoff from compliance depends on the state of the world $\omega \in \{0,1\}$, where $\Pr(\omega = 1) = p \in (0,1)$. In contrast, a citizen's payoff from non-compliance depends on their private type $\theta_i \in [0,1]$. A citizen's utility function is given by $u_i(a_i; \theta_i, \omega) = \mathbf{1}\{a_i = 1\}\omega + \mathbf{1}\{a_i = 0\}\theta_i$. We assume that when indifferent between compliance and non-compliance, a citizen complies. Each

[9] This is not entirely trivial, given that private persuasion through elicitation does not improve on public persuasion in a binary state binary action environment (Kolotilin et al., 2017): our sender leverages the existence of an independent source to avoid the incentive-compatibility constraints imposed by elicitation and targets her communication without any ability to discriminate. An obvious implication is that if the sender could design two public experiments, and choose the cost, he would do so. Empirically this does not seem to be the case in most authoritarian regimes where critical outlets are scarce in the domestic market, but available on the non-censored internet.

citizen privately observes her political type θ_i. The higher θ_i is, the more ex-ante opposed to the regime a citizen is and the harder she is to persuade to comply. The continuum of heterogeneous citizens is distributed according to a cdf F with full support on $[0,1]$.

Regime communication and censorship. At the beginning of the game, the leader privately observes a signal about the state of the world $\hat{\omega} \in \hat{\Omega} = \{0,1\}$ with $Pr(\hat{\omega} = \omega|\omega) = q \in [0.5, 1]$. He then sends a public message $s_{\mathcal{S}} \in \mathcal{M} \subset \mathbb{R}$ and picks the common cost of access to the foreign media $c \in \mathbb{R}^+$. Whenever the leader is indifferent between a non-empty range of costs, we select the lowest such cost: denote $C^* = \arg\max_c V(s_{\mathcal{S}}, c)$ then we select $c^* = min(C^*)$.

Foreign media and sorting into access. Having observed $s_{\mathcal{S}}$, each citizen decides whether to circumvent the firewall to access the banned *foreign media*. The conditional probability that citizens observe $s_{\mathcal{F}} \in \{0,1\}$ from the foreign media is given by:

$$\begin{aligned} &\Pr[s_{\mathcal{F}} = 0|\omega = 0] = 1 && \Pr[s_{\mathcal{F}} = 1|\omega = 0] = 0 \\ &\Pr[s_{\mathcal{F}} = 0|\omega = 1] = \beta(\theta_i) && \Pr[s_{\mathcal{F}} = 1|\omega = 1] = 1 - \beta(\theta_i). \end{aligned}$$

where $\beta(\theta_i) \in [0,1)$. First, the foreign media is banned by the regime, and thus never hides true bad news for the regime. Second, when citizens bypass the firewall, they may access different sources of information. $\beta(\theta_i) \in [0,1]$ flexibly captures potential heterogeneity in information consumption across citizens, e.g. due to differential cognitive constraints.

Sorting into access. Circumventing the firewall benefits the citizens in two ways. First, there is an informational benefit $b_i(\theta_i, s_{\mathcal{S}}, \beta(\theta_i)) \in [0,1]$ which depends on the citizen's type, the message (and equilibrium strategy) of the regime, and the informativeness of the foreign media, which can depend on the citizen's type (more on this below). Second, there is a relative intrinsic non-informational benefit $\alpha(\theta_i) \in \mathbb{R}^+$.[10] The net benefit from gaining access to the foreign media of citizen i is given by:

$$\underbrace{\delta_i(\theta_i, s_{\mathcal{S}}, \beta(\theta_i), c)}_{\text{net benefit}} = \underbrace{\underbrace{b_i(\theta_i, s_{\mathcal{S}}, \beta(\theta_i))}_{\text{informational benefit}} + \underbrace{\alpha(\theta_i)}_{\text{relative intrinsic benefit}}}_{\text{willingness to pay}} - \underbrace{c}_{\text{cost of access}} \tag{1}$$

We denote the observed report from the foreign media by $\hat{s}_{\mathcal{F}} \in \{s_{\mathcal{F}}, \emptyset\}$. We denote the decision to gain access of a citizen i by $z_i \in \mathcal{Z} = \{0,1\}$ with $z_i = 1$ denoting a citizen gaining access to the

[10]The regime faces a censorship problem: if $c = 0$ a non-empty set of citizens gain access.

foreign media. To ensure that the regime faces a censorship problem, we assume that $\alpha(\theta_i) \geq 0$ such that all citizens consume the foreign media absent a positive cost of access.

Timing. The sequence of the game is as follows:

1. Nature determines ω and the leader observes $\hat{\omega}$. The leader sends $s_{\mathcal{S}} \in \mathcal{M}$ and sets $c \in \mathbb{R}^+$, which are publicly observed.
2. Nature privately reveals θ_i to each citizen i, who chooses z_i; if $z_i = 1$, she observes $s_{\mathcal{F}}$.
3. Each citizen chooses whether to comply. Payoffs are realized. Game ends.

The equilibrium concept is weak Perfect Bayesian Equilibrium which requires that all players are best-responding and updating their beliefs according to Bayes' rule whenever possible. Let ΔS denote the set of probability distributions over the set S. Strategies are, for the leader (s, c) : $\hat{\Omega} \times \mathcal{M} \to \Delta(m) \times \mathbb{R}^+$ and for each citizen $(z_i, a_i) : \mathcal{M} \times \Theta \times \mathbb{R}^+ \to \mathcal{Z} \times \mathcal{A}$. We restrict attention to sender-preferred equilibria.

Comments on the Setup

Interpretation of compliance. Compliance captures any action that benefits the leader, relative to non-compliance, such as not joining a foreign movement, not criticizing the regime nor protesting, or actively supporting the regime by joining the ruling party. There are situations (e.g., the leader is more competent than any challenger, or more resilient to large-scale non compliance) where compliance benefits all, such that the leader's and citizens' incentives align. Importantly, conditional on learning for sure that $\omega = 1$, it is optimal for any citizen to comply, irrespective of one's ex-ante alignment with the regime: that is, we focus on the persuadable citizenry.[11]

Interpretation of the intrinsic benefit. The intrinsic benefit $\alpha(\theta_i)$ captures *any* benefit citizens of authoritarian regimes may derive from gaining access to an uncensored internet, *above and beyond* the informational benefit. Among other things, this captures (i) an entertainment benefit, e.g. from gaining access to censored streaming or social media platforms or foreign sports (Chen and Yang, 2019), (ii) an intrinsic benefit from consuming informational content from an independent source,

[11] We provide sufficient conditions under which the incentives to segment extend to a setting where citizens choose their compliance level from a continuum in Lemma A.17.

(iii) an economic benefit, e.g., for business elites operating internationally, (iv) a personal benefit independent of entertainment (e.g., accessing banned travel agencies or tourism related websites).

Bypassing the firewall and banned outlets. In the baseline model, we effectively assume that conditional on bypassing the firewall, there is a rich set of foreign media to choose from: $\beta(\theta_i)$ can be interpreted as the equilibrium chosen source of information beyond the firewall of citizen i. We make little assumption on $\beta(\theta_i)$ to effectively allow each citizen to flexibly choose whichever informational source she benefits the most from. However, $\beta(\theta_i)$ is commonly known and we take the strategies of these banned foreign outlet as given: their reporting slant is chosen either to maximize revenue or some ideological goal outside of the country – e.g., foreign newspapers (the *Liberty Times*, the *Guardian*, or the *New York Times*) or entertainment outlets (e.g., *HBO* or *Netflix*) – or by starch opponents of the regime who have been banned from the country.[12] In the baseline model, we show that conditional on some $\beta(\theta_i)$, the leader benefits from *some* citizens bypassing the firewall. Later, we instead assume that $\beta(\theta_i) = \beta \in [0, 1)$ for all citizens and explain how this can be micro-founded in sections 7.1 and 7.3.[13] Notice that our theoretical argument would also apply if the regime could domestically credibly control both a state and an opposition media and impose a higher cost of access to the later, which seems at odds with the limited domestic tolerance of critical outlets in authoritarian settings.

4 Analysis: Cheap Talk

Given a posterior $\mu(s_{\mathcal{S}})$, compliance is maximized by a strategy of *segment-and-rule* (assuming it is part of an equilibrium) whereby only opponents gain access $z_i = 1 \iff \theta_i > \mu(s_{\mathcal{S}})$ and the compliance level is given by $F(\mu(s_{\mathcal{S}})) + \int_{\mu(s_{\mathcal{S}})}^{1} \mu(s_{\mathcal{S}})(1 - \beta(\theta_i))f(\theta_i)d\theta_i$. Since the leader ignores the citizen's private type, they can only impose a *common* cost of access. In turn, a particular form of endogenous sorting must take place for them to be able to implement a strategy of segment-and-rule: those convinced given $\mu(s_{\mathcal{S}})$ (i.e. $\theta_i \leq \mu(s_{\mathcal{S}})$) must have a lower willingness to pay for the foreign

[12] For a given banned independent outlet, their reporting slant β can be formally micro-founded along the lines of the framework of Gehlbach and Sonin (2014). A foreign outlet targeting a foreign audience chooses β by balancing the two competing goals of (i) garnering advertising revenue by truthfully reporting the state and (ii) impacting the behavior of its audience by slanting the reporting towards the media-preferred citizen behavior.

[13] This can be interpreted as (i) the leader being unsure about the shape of $\beta(\theta_i)$ and considering a worst-case scenario or averaging myopically on the expected β or (ii) there not being a rich enough range of outlets for citizens to choose from beyond the firewall.

media than those not convinced.

To characterize equilibrium play, denote the posterior belief by $\mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}) := Pr(\omega = 1|s_{\mathcal{S}}, \hat{s}_{\mathcal{F}})$ and the willingness to pay for the foreign media by $\psi(s_{\mathcal{S}}, \theta_i, \beta(\theta_i)) := b_i(s_{\mathcal{S}}, \theta_i, \beta(\theta_i)) + \alpha_i(\theta_i)$.

Lemma 1. *In any equilibrium, $a_i^*(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}) = 1$ if and only if $\mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}) \geq \theta_i$.*

Next, denote the (possibly) empty set of segment-and-rule costs, given a posterior $\mu(s_{\mathcal{S}})$ by $\tilde{c}(\mu(s_{\mathcal{S}})) = \{c \in \mathbb{R}^+ : \psi(s_{\mathcal{S}}, \theta_i, \beta(\theta_i)) > c \iff \theta_i > \mu(s_{\mathcal{S}})\}$; recall that by assumption, when indifferent between multiple costs, the leader picks the lowest such cost.

Proposition 1. *Suppose that $\psi(\theta_i)$ is strictly increasing in θ_i. Then, if $q = 1/2$, in the unique equilibrium $c^* = \tilde{c}(p)$.*
If $q > 1/2$, then in the unique sender-preferred equilibrium, $s_{\mathcal{S}}^*(1) = s_{\mathcal{S}}^*(0), c^*(1) = c^*(0) = min(\tilde{c}(p))$.

The first part of the proposition generalizes the main idea of the example in the introduction. Absent informational asymmetries, if incentives to gain access are increasing in opposition to the leader, the leader simply picks the cost that ensures that *only* citizens that would never have complied absent access, do in fact, gain access.

Next, observe that endowing the leader with private information about ω (i.e. $q > 1/2$) could allow for some information through the leader's message $s_{\mathcal{S}}$ and cost of access c. We show that as long as segmentation is feasible, no (semi-)separating exists: the message and cost cannot be used to convey information to citizens. To understand this result, notice that given an interim belief $\mu(s_{\mathcal{S}})$, the incentives of both types of leaders are perfectly aligned: they both want to prevent access to any convinced citizen $\theta_i \leq \mu(s_{\mathcal{S}})$ and to grant access to all others $\theta_i > \mu(s_{\mathcal{S}})$. In turn, both types pool on the segmenting-cost at the prior. We note that this result holds for any $\beta(\theta_i)$, provided that the incentives to gain access are increasing in a citizen's type.[14]

5 Analysis: Information Design

We now make some simplifying assumptions to fully characterize the leader's censorship strategy and examine how the segmentation and persuasion problems of the leader interact when the leader can credibly commit to a reporting strategy.

[14] In the appendix we provide less demanding sufficient conditions for Proposition 1 to hold: see Proposition A.1.

Modified Setup. First, we impose two assumptions on the shape of F: its density $f = F'$ is log-concave and F has a unique fixed point $\theta^\dagger \in (0,1)$ s.t. $F(\theta_i) \geq \theta_i \iff \theta_i \geq \theta^\dagger$.[15] Substantively this requires that there are not too many extreme supporters ($\theta_i \approx 0$) or extreme opponents ($\theta_i \approx 1$).[16]

Second, instead of observing a signal about the state and then sending a message, the leader now commits to an information structure at the beginning of the game. There is a *state-media* whose reporting strategy is chosen by the leader. The leader publicly commits to the pro-regime slant of the state media, $\sigma \in [0,1]$. The whole citizenry observes σ and the realized message $s_{\mathcal{S}} \in \{0,1\}$ from the state media. Given σ, the conditional probability that citizens observe $s_{\mathcal{S}} \in \{0,1\}$ is given by:

$$\Pr[s_{\mathcal{S}} = 0|\omega = 0] = 1 - \sigma \qquad \Pr[s_{\mathcal{S}} = 0|\omega = 1] = 0$$
$$\Pr[s_{\mathcal{S}} = 1|\omega = 0] = \sigma \qquad \Pr[s_{\mathcal{S}} = 1|\omega = 1] = 1.$$

The higher σ is, the more uninformative the state media. Note that the state media does not hide good news: given the unimodality of f this is an equilibrium result. The timing of the game is otherwise unaffected.

Third, we impose structure on the informational and intrinsic benefit. Regarding the informational benefit, we impose that every citizen observes the same information beyond the firewall: $\beta(\theta_i) = \beta \in [0,1)\ \forall \theta_i \in [0,1]$. We micro-found this assumption in section 7. Regarding the intrinsic benefit, for simplicity and without loss we assume that it takes a linear-form with $\alpha_i(\theta_i) = d + \gamma\theta_i$ with $d \in \mathbb{R}^+, \gamma \in \mathbb{R}^+$ so that the leader faces a censorship problem.[17]

[15] This implies that f is unimodal and has a unique interior peak $\hat{\theta} \in (0,1)$. See Lemma 1 in Brändén (2015).

[16] The existence of $\theta^\dagger \in (0,1)$ and log-concavity are sufficient but not necessary. The former rules out a very weak leader ($F(\theta_i) \leq \theta_i\ \forall \theta_i$) and ensures the analysis is smooth. The latter rules out "weird" s-shaped cdfs whose convexity would change many times.

[17] We relax these assumption in section 7.2. More generally, this simple formulation allows for heterogeneous weighting between the informational and intrinsic benefit across citizens, to reflect for instance, different motivations for access. A citizen's net benefit from gaining access can be written as follows $\delta_i(\theta_i, \sigma, \beta, s_{\mathcal{S}}, \alpha(\theta_i)) = \rho(b, \theta_i) \times b_i(\theta_i, \sigma, \beta, s_{\mathcal{S}}) + \rho(\alpha(\theta_i), \theta_i) \times \alpha(\theta_i) - c$ with $\rho(b, \theta_i)$ and $\rho(\alpha(\theta_i), \theta_i)$ capturing the extent to which an individual is more or of an information or intrinsic-content seeker. Since this involves some comparison across individuals of motivations for gaining access, we normalize $\rho(b, \theta_i) = 1$ while $\rho(\alpha(\theta_i), \theta_i) \in \mathbb{R}$. The effective intrinsic benefit can be denoted by $\chi(\theta_i) := \rho(\alpha(\theta_i), \theta_i) \times \alpha(\theta_i)$ and all the results of Proposition 2 through 5 follow.

5.1 Preliminary Intuition

We now solve the game by backward induction. As in the baseline game, a citizen complies if and only if she is sufficiently aligned with the regime, *given* her learning: $a_i^*(s_\mathcal{S}, \hat{s}_\mathcal{F}) = 1 \iff \mu_i(s_\mathcal{S}, \hat{s}_\mathcal{F}) \geq \theta_i$. Thus, since the foreign media does not produce false positives, $a_i^*(s_\mathcal{S}, \hat{s}_\mathcal{F} = 1) = 1$.

Prior to that, a citizen decides whether to gain access to the foreign media's report, given the state-media's slant σ and report $s_\mathcal{S}$ and the cost of access c. Here, we make two observations. First, conditional on bad news from the state media, there is nothing to be learned, because the state media does not produce false negatives: i.e. $b_i(\sigma, s_\mathcal{S} = 0, \cdot) = 0 \forall \theta_i \in [0, 1]$. Second, conditional on good news, the informational benefit is (i) null for sufficiently aligned (and opposed) citizens $\theta_i \in [0, \mu(s_\mathcal{S} = 1, s_\mathcal{F} = 0)] \cup \{1\}$ and single-peaked and maximized at $\theta_i = \mu(s_\mathcal{S} = 1, \hat{s}_\mathcal{F} = \emptyset)$, and (ii) linear in θ_i. (i) formalizes the idea that information has value only if it *can* affect equilibrium behavior. In turn, the informational benefit is maximized for the citizen indifferent between complying and not doing so after the state-media's positive report. (ii) follows from the linearity of the citizen's utility function.

We now characterize equilibrium play. There are three cases to consider, delineated by the level of the correlation between a citizen's type and her intrinsic benefit from consumption, γ.

5.2 Low Correlation: Full Censorship

Hereafter, we refer to the citizen such that $\theta_i = \mu(s_\mathcal{S} = 1, \hat{s}_\mathcal{F} = \emptyset | \sigma)$ as the "target citizen" and denote it by $\theta(\sigma)$.

Proposition 2. *There exists a unique $\underline{\gamma} \in (0, 1 - \beta)$ such that, if the correlation is low ($\gamma \in [0, \underline{\gamma}]$) then,*

- *there exists a unique equilibrium reporting slant $\sigma^* = \sigma^L$ and target citizen $\theta(\sigma^L) \in (max\{p, \hat{\theta}\}, 1)$,*
- *the leader imposes the lowest cost of access such that no citizen bypasses the firewall: $c^* = \overline{c}(\theta(\sigma^L))$ is such that $z_i = 0 \forall \theta_i \in [0, 1]$.*
- *a citizen complies if and only if $s_\mathcal{S} = 1$ and $\theta_i \leq \theta(\sigma^L)$.*

When the correlation between politics and the intrinsic benefit is null or positive but small, the leader maximizes compliance by ensuring that no citizens bypasses the firewall. The reporting slant

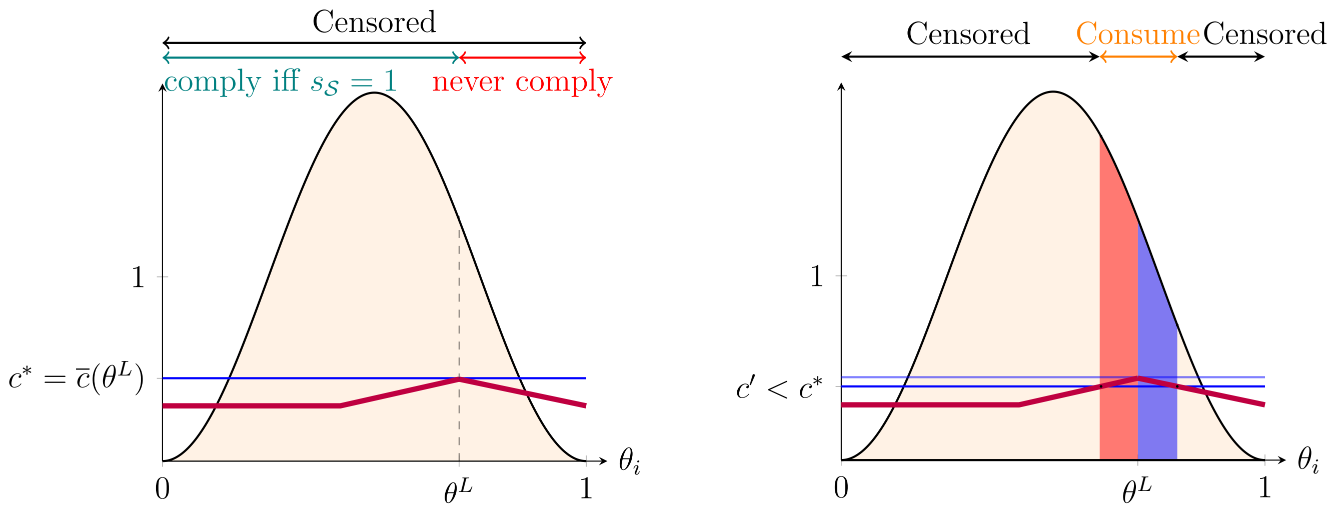

Figure 1: For this illustration, $\beta = 0.31$, $p = .5$ and $f(\theta) = cos(2\pi(\theta - 0.5)) + 1$; $d = 0.6, \gamma = 0, c = 0$. The black lines plots $f(\theta)$. The red line plots the maximum willingness to pay to access the foreign media as a function of θ_i, given $\theta(\sigma) = 0.7$ and $s_{\mathcal{S}} = 1$. The blue line plots the cost of access. The left panel depicts equilibrium behavior under full censorship. The right panel illustrates the asymmetric effects of an off-path reduction in the cost of access.

of the state media ensures that the target citizen is more opposed than the modal one such that only opponents $\theta_i > \theta(\sigma^L)$ *never* comply. This equilibrium strategy is illustrated in the left panel of Figure 1. In principle, the leader could try to censor selectively by imposing a lower cost of access $c' < \bar{c}(\theta(\sigma^L))$ as in the right panel of Figure 1. A reduction in the cost of access would induce citizens in the blue and red areas to bypass the firewall. This would generate two asymmetric effects.

First, reducing the cost of access asymmetrically affects the *share* of citizens who bypass the firewall on either side of the target citizen. There are more people to the left of the target citizen than to the right. This follows from the target citizen being more opposed to the leader than the modal one: $\theta^L > max\{p, \hat{\theta}\}$. Second, reducing the cost of access asymmetrically affects the decision of citizens on either side of the target citizen, conditional on positive reporting from the state media. A citizen to the right of the target citizen changes her decision in favor of the regime only when the state of the world is good – with probability p – *and* the foreign media reports truthfully – with probability $(1 - \beta)$. A citizen to the left of the target citizen changes her decision *against* the regime when the foreign media truthfully reports that the state is bad - with probability $(1 - p)$ – but also when they report untruthfully against the regime – with probability $p\beta$. These two forces work in the same direction to ensure that full censorship is optimal.

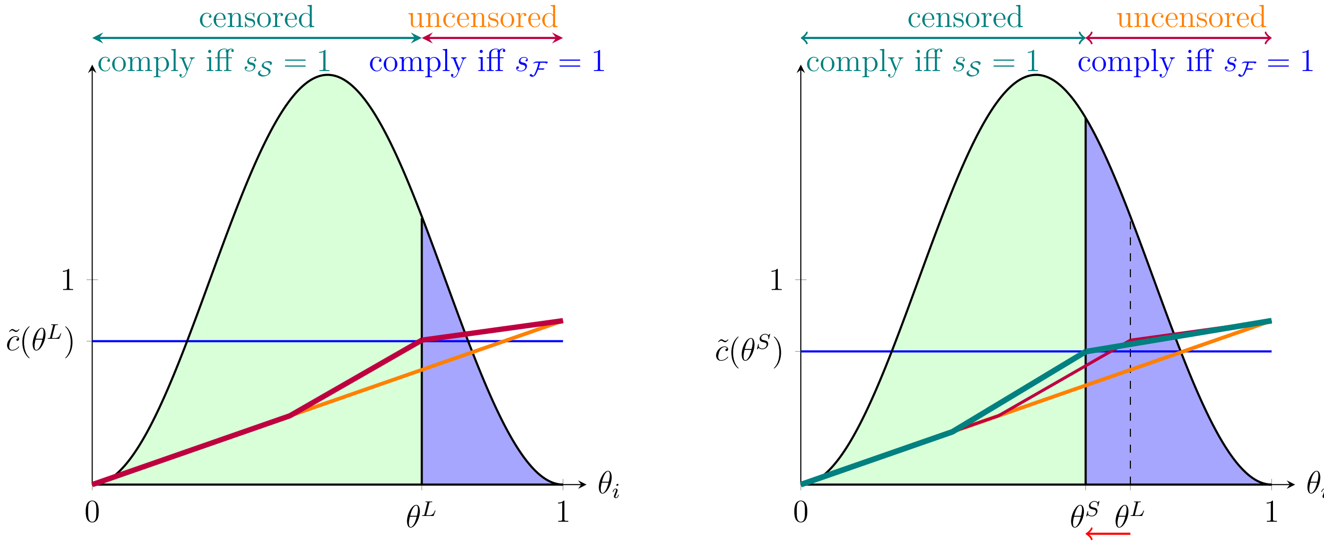

Figure 2: Same parameter values as in Figure 1 with the exception of $d = 0, \gamma = 0.8$. The left panel plots a partial equilibrium picture: the state media reports as if full censorship was implemented. The right panel depicts equilibrium behavior.

5.3 Strong Correlation: Segment-and-Rule

If the intrinsic benefit is sufficiently positively correlated with a citizen's type (high γ), then the intrinsic benefit becomes the main source of heterogeneous sorting into access: regime opponents have the strongest incentives to bypass the firewall and segment-and-rule is feasible.

Proposition 3. *There exists a unique $\overline{\gamma} \in [\underline{\gamma}, 1 - \beta)$ s.t. if $\gamma \geq \overline{\gamma}$ then the leader engages in segment-and-rule in equilibrium:*

- *there exists a unique equilibrium reporting slant $\sigma^* = \sigma^S \in [\sigma^L, 1]$ and target citizen $\theta(\sigma^S) \leq \theta(\sigma^L)$,*
- *only opponents gain access: $c^* = \tilde{c}(\theta(\sigma^S))$ such that $z_i = 1 \iff \theta_i > \theta(\sigma^S)$.*
- *a citizen complies if and only if $s_{\mathcal{S}} = 1$ and either (i) $\theta_i \leq \theta(\sigma^S)$ or (ii) $\theta_i > \theta(\sigma^S)$ and $s_{\mathcal{F}} = 1$,*
- *the level of compliance is maximized and constant in γ.*

When the correlation is strong enough the regime can implement a strategy of segment-and-rule, as in the introductory example and cheap-talk game. The strong correlation (the orange line in Figure 2) ensures that the *total* benefit from bypassing the firewall (the red line) is increasing in a citizen's misalignment with the regime, for any $\sigma \in (0, 1]$. This benefits the leader in two ways, relative to the case of a low-correlation.

First, by lowering the cost – relative to the low correlation case – the leader implements a strategy of segment-and-rule which ensures that only opponents select into bypassing the firewall. Then, they comply with probability $p(1-\beta)$, while the equilibrium best-response of those persuaded by the state-media's report ($\theta_i \leq \theta(\sigma^L)$) is unaffected. This partial equilibrium logic is illustrated on the left panel of Figure 2. Second, segment-and-rule allows the regime to target their communication to those citizens who do condition their behavior on $s_{\mathcal{S}}$: i.e. the regime's "base". Notice that *given* full censorship, the leader faces the following censorship trade-off: increasing the share of compliers ($\uparrow F(\sigma)$) requires making the state media more informative ($\downarrow \sigma$). Formaly, he solves:

$$\max_{\sigma} \quad V(\sigma, c = \bar{c}(\theta(\sigma))) \equiv \underbrace{[p + (1-p)\sigma]}_{Pr(\text{positive reporting})} \; F(\theta(\sigma)) \; + \; \underbrace{(1-p)(1-\sigma)}_{Pr(\text{true bad news})} F(0) \tag{2}$$

In contrast, *given* segment-and-rule, the regime secures the compliance of opponents as long the foreign media reports positively, with probability $p(1-\beta)$. I.e. the leader solves the following problem:

$$\max_{\sigma} V(\sigma, c = \tilde{c}(\theta(\sigma)) \equiv \underbrace{[p + (1-p)\sigma]}_{Pr(\text{positive reporting})} \; F(\theta(\sigma)) \; + \; \underbrace{p(1-\beta)}_{Pr(\text{true good news})} \; [F(1) - F(\theta(\sigma))] \tag{3}$$

Thus, to increase the share of citizens convinced by the state-media's good news $F(\sigma)$ the state media must be more informative either way. However, under segment-and-rule, in so doing the regime also reduces the benefit from segmented access as the share of opponents decreases ($[F(1) - F(\theta(\sigma))] \downarrow$). Thus, when segment-and-rule is possible the state media is less informative than under full censorship. In some cases it fully parrots the party line: $\sigma^S = 1$ and $\theta^S = p$.[18]

5.4 Intermediary Correlation: Engineering Segment-and-Rule

In the previous section, the high correlation ensured that segment-and-rule was always feasible, effectively allowing the leader to pick the optimal information structure *given* that such segmentation was feasible. Instead, at intermediary levels of γ, the leader's two problems – segmenting optimally and selecting the optimal information structure – can no longer be solved in isolation. Since the informational benefit is only decreasing in the citizens' type among opponents ($\theta_i > \theta(\sigma)$), segment-

[18]E.g., high prior p relative to the shape of the distribution or low reporting slant of the foreign outlet β.

and-rule is feasible if and only if:

$$\frac{\partial \overbrace{\delta_i(\theta_i, \sigma, \beta, s_{\mathcal{S}} = 1)}^{\text{total benefit}}}{\partial \theta_i} > 0 \iff \underbrace{\frac{\partial \alpha(\theta_i)}{\partial \theta_i}}_{\text{correlation}} = \gamma > \underbrace{\theta(\sigma)(1-\beta)}_{\text{reduction in } b_i(\cdot) \text{ among opponents}} \tag{4}$$

In the strong correlation case, $\gamma \geq \overline{\gamma}$ ensured that the correlation was sufficiently large relative to the informativeness of the foreign media $(1-\beta)$, such that the regime could pick *any* reporting slant (and thus target citizen $\theta(\sigma)$) and engage in segment-and-rule. Here, instead, the regime must pick an appropriate target citizen – one sufficiently aligned with the regime – to *engineer* segment-and-rule. This requires making the state media less informative ($\sigma \uparrow$). Intuitively, as the state media loses informational content, the value of an additional report increases for *any* opponent. Thus the informational benefit becomes less dependent of a citizen's political type; the slope of the informational benefit becomes flatter. In turn, even a moderate correlation is sufficient for a strategy of segment-and-rule to be implemented in equilibrium.

Proposition 4. *If the correlation is intermediate $\gamma \in [\underline{\gamma}, \overline{\gamma})$ the leader engages in segment-and-rule,*

- *there exists a unique equilibrium reporting slant $\sigma^* = \sigma^I \geq \sigma^S$ and target citizen $\theta(\sigma^I) \leq \theta(\sigma^S)$,*
- *only opponents gain access: $c^* = \tilde{c}(\theta^I)$ such that $z_i = 1 \iff \theta_i \geq \theta(\sigma^I)$,*
- *a citizen complies if and only if $s_{\mathcal{S}} = 1$ and either (i) $\theta_i \leq \theta(\sigma^I)$ or (ii) $\theta_i > \theta(\sigma^I)$ and $s_{\mathcal{F}} = 1$,*
- *compliance $V(\sigma^I, \tilde{c}(\theta(\sigma^I))$ is increasing in γ and $V(\sigma^I, \tilde{c}(\theta(\sigma^I)) \in [V(\sigma^L, \overline{c}(\theta(\sigma^L)), V(\sigma^S, \tilde{c}(\theta(\sigma^S))]$.*

Figure 3 illustrates the engineering logic: if the state media is as informative as in the strong correlation case then segment-and-rule is impossible as the total benefit is decreasing in a citizen's type for opponents. If the state media is sufficiently less informative then sorting occurs along the intrinsic dimension and segment-and-rule is feasible. Importantly this engineering comes at a cost: the leader must pick a persuasion-wise suboptimal reporting slant so that segment-and-rule is possible. Thus the level of compliance is increasing in the correlation between politics and entertainment. At one extreme ($\gamma = \overline{\gamma}$) segment-and-rule requires no compromise and the level of compliance reaches its upper bound. At the other ($\gamma = \underline{\gamma}$) the leader is indifferent between segment-

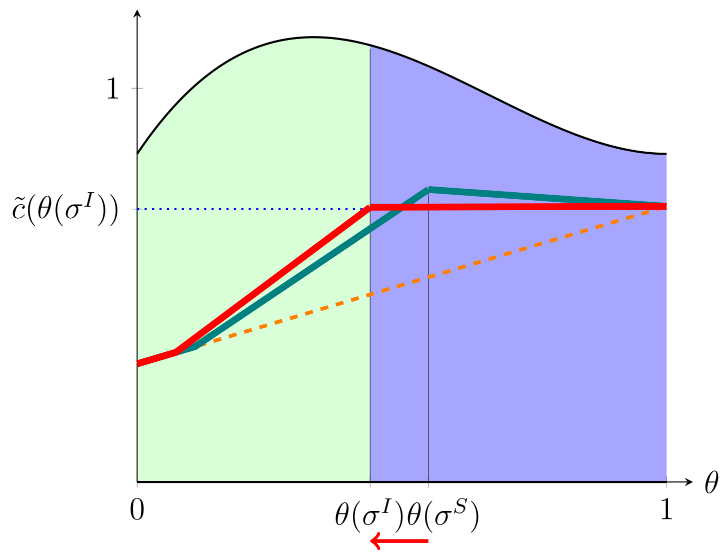

Figure 3: For this illustration, $f(x) = 2x(1-x)^2 + \frac{5}{6}$, $\beta = 0.1$ and $\alpha(\theta_i) = 0.3 + 0.4\theta_i$. The green line plots the willingness to pay conditional on the state media reporting as if there is a strong cleavage ($\gamma \geq \overline{\gamma}$). The red line plots the willingness to pay in equilibrium, given an intermediate correlation ($\gamma \in [\underline{\gamma}, \overline{\gamma}]$) and the equilibrium target citizen $\theta(\sigma^I)$.

and-rule and engaging in full censorship. The "cost" of such engineering pins downs $\underline{\gamma}$. This form of engineering is only observed in some contexts: the $[\underline{\gamma}, \overline{\gamma}]$ interval can be empty.[19]

To recap, modern censorship involves a strategy of segment-and-rule whenever heterogeneous sorting into access occurs sufficiently along the intrinsic dimension ($\gamma \geq \underline{\gamma}$). Then, the state media secures the compliance of the regime's base by parroting the party line while the foreign media inadvertently helps the regime by occasionally persuading opponents to comply.

6 Comparative Statics and Implications

We now consider to what extent the regime can leverage the citizen's agency in accessing foreign content. To do so, we consider how the informativeness of the banned outlets $(1-\beta)$ affects equilibrium compliance and censorship.[20]

Proposition 5. *For any $\gamma \in [0, \theta(\sigma^S))$, there exists a unique $\underline{\beta}$ and $\overline{\beta}$ with $0 < \underline{\beta} \leq \overline{\beta} < 1$ such that*

- *the equilibrium reporting slant σ^* is non-monotonic in β: it is constant for any $\beta < \underline{\beta}$, jumps*

[19]Engineering is never possible when the prior p is high relative to where most of the mass of citizen lie, such that the state media is already very uninformative when the correlation is strong (i.e. $\sigma^S \approx 1$). We provide precise conditions in the formal appendix.

[20]The same comparative statics are derived with respect to γ and presented in Corollary A.1.

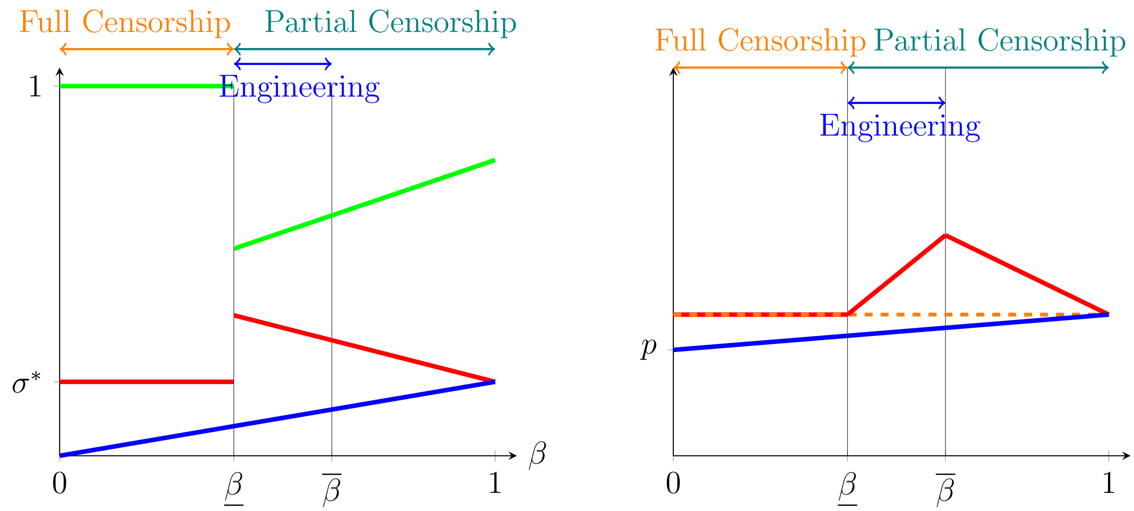

Figure 4: The left panel plots the equilibrium reporting slant (σ^*) absent any censorship (in blue) and given the possibility of manipulating the cost of access (in red), and illustrates the share of citizens who do not bypass the firewall in equilibrium (in green). The right panel illustrates the equilibrium level of compliance absent any censorship (in blue), given full censorship (in dashed orange) and given the possibility of manipulating the cost of access (in red).

at $\beta = \underline{\beta}$ and is decreasing in β otherwise.

- *the equilibrium share of citizens who do not bypass the firewall is 1 for any $\beta < \underline{\beta}$, falls down at $\beta = \underline{\beta}$ and is increasing in β otherwise.*
- *the equilibrium compliance is non-monotonic and single-peaked in the informativeness of the foreign media and maximized at $\beta = \overline{\beta}$.*

If $\gamma \geq \theta(\sigma^S)$ then $\underline{\beta} = \overline{\beta} = 0$.

Fixing some intermediary correlation $\gamma < \overline{\gamma}$, varying the informativeness of the foreign media $(1 - \beta)$ affects whether segment-and-rule is feasible, and if so, at what cost, as illustrated in Figure 4. The more informative the foreign outlet, the more a citizen's type determines her informational benefit (see equation (4)). Thus if the foreign outlet is too informative ($\beta < \underline{\beta}$, possibly an empty interval) then segment-and-rule is impossible, and full censorship takes place. When segment-and-rule is feasible ($\beta \geq \underline{\beta}$) a clear pattern emerges (left panel of Figure 4). As the foreign outlet becomes less informative ($\beta \uparrow$), the share of citizens who bypass the firewall (green line) and the reporting slant (red line) move in opposite directions. This result is striking in three ways.

6.1 Non-Disciplining Effect of Competition

First, it nuances the common wisdom that in models of competition between senders, adding a competing sender usually acts as a disciplining device that induces more information revelation (Gentzkow and Kamenica, 2017). In our context, in a hypothetical scenario *without any* cost of access, the more biased the foreign outlet, the more biased the state outlet, as illustrated by the blue line in the left panel of Figure 4.

Instead, whenever segment-and-rule takes place in equilibrium ($\beta \geq \underline{\beta}$), the state media becomes less informative as the foreign outlet becomes more informative (the red line on the left panel). To understand this argument, notice that as the foreign media becomes more informative ($\beta \downarrow$), the likelihood of positive reporting that ensures the compliance of opponents increases. By making the state media less informative the leader faces a trade-off. First, his supporters are more likely to comply, since the state media parrots the party line more. Second, his base is smaller, and thus more citizens gain access. As the likelihood of positive reporting from the foreign outlet increases, the loss from the smaller size of the leader's base is compensated by the increased propensity of opponents to comply; then, making the foreign media more informative acts as a *non*-disciplining device.[21]

6.2 Limited Censorship Capacity

So far we have intentionally assumed that the leader faces no censorship capacity constraint, to show that an infinite cost of access would not be optimal even if it was feasible (unless if the correlation is low). In reality the capacity to impose a cost of access varies across regimes. In an extension (see Proposition A.2 in the appendix) we formally consider how a binding constraint – formally $c < \overline{C}$ – on the regime's censorship capacity matters if the correlation is low ($\gamma < \underline{\gamma}$). Then, in equilibrium, as in Proposition 2, if the leader wants to ensure that no citizen bypasses the firewall, but cannot, the leader imposes the highest possible cost of access.

When the regime has a limited censorship capacity and the correlation is low ($\gamma \leq \underline{\gamma}$), our framework suggests two observationally equivalent explanations for the empirical pattern of selective bypassing of the firewall. To distinguish between them we first characterize the range of citizens

[21] Li and Norman (2018) already provide conditions for the the disciplining result of Gentzkow and Kamenica (2017) not to hold.

that bypass the firewall when the leader would like to censor all citizens, but cannot.

Lemma 2. *Suppose that the correlation is low ($\gamma < \underline{\gamma}$) and that full censorship is impossible ($\overline{C} < \overline{c}(\theta^L)$). There exists a unique $\overline{\overline{C}} \in (0, \overline{c}(\theta^L))$ such that the strongest opponent of the regime ($\theta_i = 1$) bypasses the firewall if and only if and only if $\overline{C} \leq \overline{\overline{C}}$.*

If an authoritarian regime can impose a non-negligible cost of access ($\overline{C} > \overline{\overline{C}}$) and aims to minimize the share of citizens bypassing the firewall ($\gamma \leq \underline{\gamma}$), then the range of citizens bypassing it does *not* include the most ex-ante opposed to the regime citizens ($\theta_i = 1$).

Consider present-day China or Iran. Empirical evidence suggests that the regime strongest opponents do bypass the firewall (Shen and Zhang, 2018; Mou, Wu, and Atkin, 2016; Dal and Nisbet, 2022). Our theoretical framework suggests that if these regimes possesses the technological capacity to impose a non-trivial cost of access ($\overline{C} \in (\overline{\overline{C}}, \overline{c}(\theta^L))$, then selective bypassing is unlikely to be brought about by technological or economic constraints *alone*.

6.3 Empirical and Policy Implications

Second, we show that in equilibrium a negative association emerges between two empirical measures of censorship – how freely information flows locally among regime controlled outlets, σ^* – and another – the share of citizens who only consume regime controlled outlets. This is illustrated in the left panel of Figure 4. As one measure of censorship – the share not gaining access, plotted in green – goes up, another – the reporting slant of the state-media, plotted in red – goes down.

Lastly, Proposition 5 suggests two potential policy implications for foreign actors interested in weakening an authoritarian leader via strategic investments in their own media landscape.

Implication 1. 1. *Compliance is lowest when banned outlets are most informative ($\beta = 0$) or most uninformative ($\beta = 1$).*

2. *Entertainment content that is polarizing across political lines (high γ) is (i) most likely to be banned and (ii) facilitates segment-and-rule and thus helps the leader achieve higher levels of compliance.*

Informational outlets should avoid being partially biased, in order to attempt to make segment-and-rule impossible ($\beta \leq \underline{\beta}$), or pointless ($\beta = 1$). Further, our framework suggest that strategies of "soft-power" via entertainment could backfire. Outright criticism of the authoritarian regime's culture and norms may backfire by creating content that specifically appeals to regime opponents; we make this point formally in section 7.1.

7 Extensions

7.1 The Instrumental Value of Entertainment Control

In this section, building on the experimental evidence of Chan et al. (2019) we focus on one interpretation of the intrinsic benefit: the entertainment value from by-passing the firewall. We consider how authoritarian regimes can instrumentally control the production and access to entertainment in order to make the control of information flows more efficient.

Investing in domestic entertainment. Rewrite the *relative* intrinsic benefit as follows:

$$\alpha(\theta_i) = \underbrace{d_{\mathcal{F}} + \gamma_{\mathcal{F}} * \theta_i}_{\text{foreign media intrinsic benefit}} - \underbrace{(d_{\mathcal{S}} + \gamma_{\mathcal{S}} * \theta_i)}_{\text{state media intrinsic benefit}}$$

such that $d = d_{\mathcal{F}} - d_{\mathcal{S}}$ and $\gamma = \gamma_{\mathcal{F}} - \gamma_{\mathcal{S}}$. Given some foreign content $(d_{\mathcal{F}}, \gamma_{\mathcal{F}})$, if the regime can manipulate both the quality of local entertainment $(d_{\mathcal{S}})$ and its relative appeal $(\gamma_{\mathcal{S}})$, they can achieve their upper bound compliance payoff. First, they create content which mostly appeals to their base. Formally this requires:

$$\gamma \geq \overline{\gamma} \iff \gamma_{\mathcal{S}} \leq \gamma_{\mathcal{F}} - \underbrace{\theta^S}_{\text{target citizen given strong correlation}} (1-\beta) \equiv \overline{\gamma}_{\mathcal{S}} \tag{5}$$

Next, they ensure that only opponents of the regime do gain access to censored content. This can be done by picking the appropriate cost of access c, *or* by picking the appropriate quality of local entertainment $d_{\mathcal{S}}$; $d_{\mathcal{S}}$ and c are substitutable levers. Formally this requires that:[22]

$$d_{\mathcal{S}} = \underbrace{b_i(\theta^S, \sigma^S, \beta, s_{\mathcal{S}} = 1)}_{\text{informational benefit}} + d_{\mathcal{F}} + \theta^S(\gamma_{\mathcal{F}} - \overline{\gamma}_{\mathcal{S}}) \tag{6}$$

The minimal level of domestic quality to ensure that the regime can segment access optimally is

[22] We assume that the leader pick the lowest $\gamma_{\mathcal{S}}$, in absolute value, whenever indifferent.

increasing in both the quality of foreign content ($d_{\mathcal{F}}$) and the cleavage along political lines ($\gamma_{\mathcal{F}} - \overline{\gamma}_{\mathcal{S}}$).

Strategic bans. Chen and Yang (2019) also document that Chinese citizens *may* be exposed to information-intensive content *after* having bypassed the firewall with the goal of consuming low information and high entertainment content (social media, HBO, Youtube, etc.). Our theoretical framework can be interpreted in a similar manner. Then, the regime can think about affecting what entertainment *is* banned, in order to affect sorting into access and make segment-and-rule possible.

Suppose that there are n foreign outlets, each with endowed with a reporting slant β_j and non-informational content parameters $d^j_{\mathcal{F}}$ and $\gamma^j_{\mathcal{F}}$. Out of these n outlets, $\tilde{n} < n$ have no informational content ($\beta_j = 1 \forall j \in \{1, 2, ..., \tilde{n}\}$). We explicitly assume that all minimally informational foreign outlets (any outlet with $\beta_j < 1$) are banned and can only be accessed by bypassing the firewall. This is done to focus on how a strong correlation between politics and entertainment can be engineered by strategically banning some of the $\tilde{n}$ purely non-informational outlets. Then the regime selects $k \leq \tilde{n}$ of the non-informational outlets to ban. In turn we define

$$d_{\mathcal{F}} \equiv \sum_{i=1}^{k} \frac{d^i_{\mathcal{F}}}{k}, \quad \gamma_{\mathcal{F}} \equiv \sum_{i=1}^{k} \frac{\gamma^i_{\mathcal{F}}}{k}$$

That is, the entertainment value of the "foreign media" is defined as the average entertainment content value of all the k outlets banned by the regime.[23] Notice that the banning decision only affects the entertainment value from bypassing the firewall. The regime then solves for a dual problem: they look for a list of banned outlets that ensures (i) that $\gamma_{\mathcal{F}} = \gamma_{\mathcal{S}} + \theta^S(1 - \beta) \equiv \overline{\gamma}_o$ such that a strong correlation exists and (ii) that $d_{\mathcal{F}} = d_{\mathcal{S}} - b_i(\theta^s, \cdot) - \theta^S(\overline{\gamma}_o - \gamma_{\mathcal{S}})$ such that only opponents ($\theta_i > \theta^S$) bypass the firewall. These two goals may clash. If such a list does not exist then the regime can also make use of the cost of access c to adjust on the second problem and to instead focus on creating the required correlation by banning only the most polarizing outlets (the highest $\gamma^j_{\mathcal{F}}$).

In the case of present-day China, the CCP appears to be using a combination of (i) a positive cost of access – requiring that citizens invest in some VPN to bypass the firewall – and (ii) strategic investments in polarizing local entertainment – such as *The Battle of Lake Changjin* or *The Knockout*[24] – and (iii) strategic bans of entertainment – e.g., banning *Cockroach*, *Eternal Spring*

[23] For simplicity and without loss we assume equal weights for each banned outlet.

[24] In an experimental setting, Yao (2023) finds suggestive evidence that those most ex-ante inclined towards state

or *Top Gun - Maverick* while not banning *Jurassic World Fallen Kingdom* or *Transformers Age of Extinction*.

We suggest an alternative explanation for the empirical pattern of domestic entertainment appealing mostly to regime supporters in authoritarian contexts. Domestic entertainment does not appeal to regime supporters because the regime rewards them (Esberg, 2020*a*); rather, domestic entertainment is unappealing to regime opponents so as to ensure that they self-select into consuming banned entertainment, and in turn information.

7.2 Modeling the Intrinsic Benefit

In this section we derive sufficient conditions on the intrinsic benefit $\alpha(\theta_i)$ for the regime to reach their upper bound payoff by segmenting access and setting $\sigma^* = \sigma^S$. Given such an equilibrium reporting slant, no restrictions need be applied to $\alpha(\theta_i)$ for all unconditional compliers ($\theta_i \leq \mu(s_{\mathcal{F}} = 0, s_{\mathcal{S}} = 1, \sigma^S))$) as they never condition their compliance decision on the report of the foreign outlet. Hereafter we focus on rest of the citizenry.

Intuitively, the regime can reach their upper bound equilibrium payoff from segment-and-rule whenever they can find a cost of access that ensures that (i) all *conditional compliers* ($\theta_i \in [\mu(s_{\mathcal{F}} = 0, s_{\mathcal{S}} = 1, \sigma^S)), \theta(\sigma^S)))$ do not gain access and (ii) *opponents* ($\theta_i \in [\theta(\sigma^S), 1]$) do gain access, *given* the optimal reporting slant σ^S for some distribution of political preferences f.

Proposition 6. If $max(\psi(\theta_i \in [\mu(s_{\mathcal{F}} = 0, s_{\mathcal{S}} = 1, \sigma^S); \sigma^S, \beta)) < min(\psi(\theta_i \in (\theta(\sigma^S), 1]; \sigma^S, \beta))$ then the regime sets $\sigma^* = \sigma^S$ and $c^* = \tilde{c}(\theta^S)$ to achieve their upper bound equilibrium payoff.

The straightforward logic of Proposition 6 is illustrated in Figure 5 with an example where the intrinsic benefit is non-monotonic in the misalignment with the regime. Notice that the distribution of political preferences f determines the equilibrium reporting slant and maximal attainable payoff of the regime, but not directly the feasibility of this equilibrium payoff, other than through the reporting slant σ^S. To recap, a strong (quasi-) linear positive association between misalignment with the regime and the intrinsic benefit is sufficient but not necessary for segment-and-rule; this strategy is also feasible when the intrinsic benefit is most enjoyed by both supporters and opponents and less by moderates.

propaganda in China – possibly regime supporters – are most likely to consume state propaganda. That is, CCP

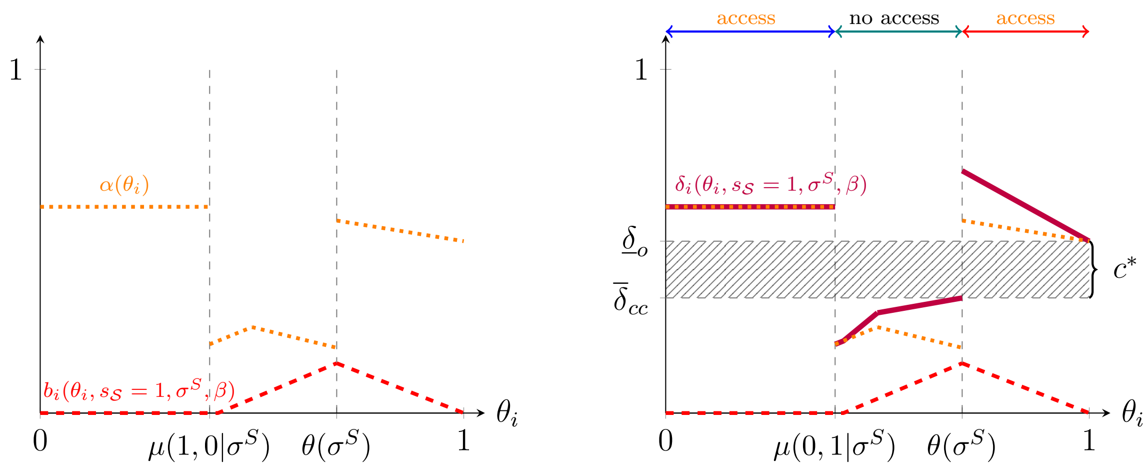

Figure 5: Same parameter values as in Figure 4. Left panel: the dashed red line plots the informational benefit given $s_{\mathcal{S}} = 1$ and $\sigma^* = \sigma^S$ while the dotted orange line plots a given non-monotonic step function intrinsic benefit $\alpha(\theta_i)$ such that unconditional compliers have the highest intrinsic benefit. Right panel: the purple line plots the total willingness to pay for access and the grey dashed area characterizes the set of costs the regime can choose from to secure their upper bound payoff.

7.3 Additional Extensions

Exposure to multiple reports. Once past the firewall citizens could (choose to) observe reports from multiple media outlets who vary in their informativeness. Formally, let us suppose that there are n banned foreign outlets outside of the regime's control and accessible once past the firewall. Each outlet is indexed by $j \in \{1, 2, ...n\}$ and endowed with a reporting slant β_j and intrinsic content parameters $z^j_{\mathcal{F}}$ and $\gamma^j_{\mathcal{F}}$. Each outlet's report is independently drawn. Conditional on a by-passer observing at least one piece of good news from one banned outlet, then this citizen updates that $\omega = 1$, and complies. Then define $\beta \equiv Pr(s_j = 0 \ \forall j \in \{1, ..., n\})$ and notice that the leader's problem is unaffected vis-a-vis the baseline model.

There are at least three intuitive ways to explicitly model exposure to the stream of signals. First, by-passers may observe *all* the signals from each banned outlet. The ex-ante probability that firewall by-passers comply is then given by $p(1 - \prod_{j=1}^n \beta_j)$ and we can here define $\beta \equiv \prod_{j=1}^n \beta_j$. Second, it might be that by-passers consume only one report. For instance they may only consume the most informative report. Then it suffices to define $\beta \equiv min_{j\in\{1,...n\}}\beta_j$. Alternatively, a by-passer may be equally likely to observe a single report from any of the n outlets. Then each citizen complies with probability $p(1 - \frac{\sum_{j=1}^n \beta_j)}{n})$ and it suffices to define $\beta \equiv \frac{\sum_{j=1}^n \beta_j}{n}$. The attention rule could range

propaganda seems to appeal mostly to the regime's base.

anywhere between these two extremes without affecting our central message: if opponents *can* be exposed to at least one report from an independent source that sways them into complying, then the sender benefits from the most skeptical receivers self-selecting into gaining access.

Intertwined levers. In our baseline game, the correlation γ is a primitive and orthogonal to the reporting slant of the state media (σ). One may argue that such a high correlation exists (high γ) *because* the domestic outlets parrot the party line (high σ), which "annoys" opponents and lead to the development of an intrinsic benefit from consuming news from an independent source. We show that if γ is an increasing function of σ then the same qualitative results are derived, as the regime can engineer segmentation mechanically by increasing their reporting slant (Lemma A.15).

Domestic Segmentation. We consider a game with two domestic outlets and without a foreign media to consider under which conditions segmentation could be achieved locally. We show that, as long as the regime cannot credibly commit to reporting negatively on itself, then all domestic outlets have the same reporting slant and the regime achieves its lower bound payoff of full-censorship.[25] The fact that there are no state-controlled outlets known to be outright critics of the regime they operate under, suggests that to segment access to information, authoritarian regimes must rely on banned independent (foreign) outlets.

Non-binary compliance. For tractability reasons we assumed that the citizens' decision to comply is binary. One may be concerned that the attractiveness of a strategy of segment-and-rule relies on this modeling choice: with a continuous compliance level to choose from, as opponents are most of the time exposed to negative news from the foreign outlet, aggregate compliance could be lower under segment-and-rule than full censorship. To address this concern we allow for citizens to choose any compliance level $a_i \in [0, 1]$ and derive sufficient conditions for a strategy of segment-and-rule to dominate full censorship, assuming that some equilibrium compliance profile $a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta))$ exists. If the compliance level (i) decreases in misalignment with the regime, (ii) increases in the belief of a citizen *and* (iii) the usual increasing difference assumption holds – such that the more misaligned a citizen is with the regime, the smaller the marginal effect of a higher belief on ω – then segment-and-rule still dominates full censorship for any reporting slant $\sigma \in [0, 1]$ (Lemma A.17).

[25] Formally, as in Heo and Zerbini (2024), the assumption that the regime cannot take over an opposition outlet and still credibly commit to reporting "against itself" – i.e. $Pr(s_{\mathcal{S}} = 1|\omega = 1) = 1$ by assumption – is crucial; see Lemma A.16.

8 Conclusion

We have presented a model of informational and intrinsic – e.g., entertainment – content control by an authoritarian regime, and consumption by a population of heterogeneous citizens. Using this framework we provide one explanation for the selective bypassing of firewalls: authoritarian leaders purposely impose a low cost of access to tailor information provision to different segments of citizens. This strategy of segment-and-rule leverages both the citizens' agency of information acquisition – gained in the post-internet era – and their heterogeneous political preferences: it keeps the regime's base in the dark and provides opponents with a credible foreign source of information. For this strategy to be feasible, regime opponents must benefit from bypassing the firewall more than regime moderates. Then, we show, to ensure this particular sorting pattern, authoritarian regimes can strategically control access to and production of non-informational content, such as entertainment. This highlights the instrumental value of entertainment: while it does not affect beliefs per se, it helps "subsidise" the consumption of information from local outlets for the regime's base, and from foreign outlets for regime opponents.

We speak to a growing discussion on the use of modern technologies to facilitate authoritarian control. Recent works have argued that developments in AI could entrench autocrats more than they empower citizens. Because digital surveillance is less intrusive than in-person surveillance it can be rolled out with less resistance (Xu, 2023). This surveillance infrastructure can then facilitate pre-emptive suppression of organized dissent (Dragu and Lupu, 2021) and induce compliance via social-scoring rules (Tirole, 2021). In turn, digital surveillance can help authoritarian leaders reduce their provision of public good (Xu, 2021). To make matters worse, there exists a self-reinforcing dynamic between innovation in AI and the entrenchment of authoritarian regimes (Beraja et al., 2023).

We depict a yet grimmer picture by showing that simpler and cheaper technologies that do not involve any surveillance or data collection can be leveraged by authoritarian leaders. In the context of censorship, a naive intuition would suggest that by empowering citizens with more agency over their content consumption, the internet would have made censorship more difficult and helped citizens bring down authoritarian regimes. Unfortunately, it is precisely this agency gain that made possible a segmentation of the citizenry which authoritarian regimes use to improve their grasp on power. In this respect, the internet entrenched authoritarian regimes.

Bibliography

Acemoglu, Daron, James A Robinson, and Thierry Verdier. 2004. "Kleptocracy and Divide and Rule: A Model of Personal Rule." *Journal of the European Economic Association* .

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun. 2018. "The Power of the Street: Evidence from Egypt's Arab Spring." *The Review of Financial Studies* 31(1): 1–42.

Arieli, Itai, and Yakov Babichenko. 2019. "Private Bayesian persuasion." *Journal of Economic Theory* 182: 185–217.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2006. "Monotone comparative statics for models of politics." *American Journal of Political Science* 50(1): 214–231. Publisher: Wiley Online Library.

Bardhi, Arjada, and Yingni Guo. 2018. "Modes of persuasion toward unanimous consent." *Theoretical Economics* 13(3): 1111–1149. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/TE2834.

Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman. 2023. "AI-tocracy*." *The Quarterly Journal of Economics* 138(3): 1349–1402.

Bergemann, Dirk, and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57(1): 44–95.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris. 2015. "The Limits of Price Discrimination." *American Economic Review* 105(3): 921–57.

Bowles, Jeremy, John Marshall, and Pia. Raffler. 2024. "Access to social media and support for elected autocrats: field experimental evidence observational evidence from Uganda.".

Boxell, Levi, and Zachary Steinert-Threlkeld. 2019. "Taxing dissent: The impact of a social media tax in Uganda." arXiv.

Brändén, Petter. 2015. "Unimodality, log-concavity, real-rootedness and beyond." *Handbook of enumerative combinatorics* 87: 437.

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang. 2019. "Pivotal persuasion." *Journal of Economic Theory* 180: 178–202.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109(6): 2294–2332.

Chester, Patrick J. 2023. "Framing Democracy: Characterizing China's Negative Legitimation Propaganda using Word Embeddings." Unpublished paper, . https://patrickjchester.com/publication/chprop1/

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society* , 1431–1451. Publisher: JSTOR.

Dal, Aysenur, and Erik C. Nisbet. 2022. "Walking Through Firewalls: Circumventing Censorship of Social Media and Online Content in a Networked Authoritarian Context." *Social Media + Society* 8(4): 20563051221137738. Publisher: SAGE Publications Ltd.

Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy* .

Dragu, Tiberiu, and Yonatan Lupu. 2021. "Digital Authoritarianism and the Future of Human Rights." *International Organization* 75(4): 991–1017. Publisher: Cambridge University Press.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *The Review of Economic Studies* 80(4 (285)): 1422–1458. Publisher: [Oxford University Press, The Review of Economic Studies, Ltd.].

Egorov, Georgy, and Konstantin Sonin. 2023. "The Political Economics of Non-democracy." *Journal of Economic Literature* .

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *The American Political Science Review* 103(4): 645–668. Publisher: [American Political Science Association, Cambridge University Press].

Esberg, Jane. 2020*a*. "Anticipating Dissent: The Repression of Politicians in Pinochet's Chile.".

Esberg, Jane. 2020*b*. "Censorship as Reward: Evidence from Pop Culture Censorship in Chile." *American Political Science Review* 114(3): 821–836. Publisher: Cambridge University Press.

Fish, Eric. 2018. "Caught in a Crossfire: Chinese Students Abroad and the Battle for Their Hearts.".

Fung, Brian. 2022. “Russian internet users are learning how to beat Putin’s internet crackdown | CNN Business.”.

Galvis, Ángela Fonseca, James M. Snyder, and B. K. Song. 2016. “Newspaper Market Structure and Behavior: Partisan Coverage of Political Scandals in the United States from 1870 to 1910.” *The Journal of Politics* 78(2): 368–381. Publisher: The University of Chicago Press.

Gehlbach, Scott, and Konstantin Sonin. 2014. “Government control of the media.” *Journal of Public Economics* 118: 163–171.

Gentzkow, Matthew, and Emir Kamenica. 2017. “Competition in Persuasion.” *The Review of Economic Studies* 84(1): 300–322.

Gitmez, A. Arda, and Konstantin Sonin. 2023. “The Dictator’s Dilemma: A Theory of Propaganda and Repression.”.

Gitmez, A. Arda, and Pooya Molavi. 2023. “Informational Autocrats, Diverse Societies.” arXiv.

Gratton, Gabriele, and Barton E Lee. 2024. “Liberty, Security, and Accountability: The Rise and Fall of Illiberal Democracies.” *The Review of Economic Studies* 91(1): 340–371.

Guriev, Sergei, and Daniel Treisman. 2019. “Informational autocrats.” *Journal of economic perspectives* 33(4): 100–127. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Guriev, Sergei, and Daniel Treisman. 2020. “A theory of informational autocracy.” *Journal of public economics* 186: 104158. Publisher: Elsevier.

Heo, Kun, and Antoine Zerbini. 2024. “Censorship in Large Societies.” OSF.

Hobbs, William, and Margaret E Roberts. 2018. “How Sudden Censorship Can Increase Access to Information.”.

Huang, Haifeng. 2015. “International knowledge and domestic evaluations in a changing society: The case of China.” *American Political Science Review* 109(3): 613–634. Publisher: Cambridge University Press.

Huang, Haifeng, and Yao-Yuan Yeh. 2019. "Information from abroad: Foreign media, selective exposure and political support in China." *British Journal of Political Science* 49(2): 611–636. Publisher: Cambridge University Press.

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

Kim, Joye, Jae-pyeong Seo et al. 2022. "Why some North Korean defectors return to one of the world's most repressive regimes.".

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. "Persuasion of a privately informed receiver." *Econometrica* 85(6): 1949–1964. Publisher: Wiley Online Library.

Kronick, Dorothy, and John Marshall. 2024. "Collateral Censorship: Theory and Evidence from Venezuela." *Work. Pap., Univ. California, Berkeley, CA. https://dorothykronick.com/RCTV_2024. pdf* .

Li, Fei, and Peter Norman. 2018. "On Bayesian persuasion with multiple senders." *Economics Letters* 170: 66–70.

Liu, Hanzhang, and Linan Yao. 2023. "Entertainment as Trojan Horse: Voluntary Consumption of Propaganda in China." Unpublished paper, . http://linanyao.com/#research

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2): 402–414. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12065.

Luo, Zhaotian, and Arturas Rozenas. 2023. "Ruling the Ruling Coalition: Information Control and Authoritarian Power-Sharing." *Quarterly Journal of Political Science* 18(2): 183–213. Publisher: Now Publishers, Inc.

Lutscher, Philipp M. 2023. "When Censorship Works: Exploring the Resilience of News Websites to Online Censorship." *British Journal of Political Science* , 1–9. Publisher: Cambridge University Press.

Manacorda, Marco, and Andrea Tesei. 2020. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica* 88(2): 533–567. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14392.

Mao, Yishu. 2020. "CONFLICTED HEARTS AND MINDS: A SURVEY OF POLITICAL ATTITUDES OF CHINESE STUDENTS IN GERMANY.".

Matysková, Ludmila, and Alfonso Montes. 2023. "Bayesian persuasion with costly information acquisition." *Journal of Economic Theory* 211: 105678.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature* 61(1): 226–273.

Milgrom, Paul, and Chris Shannon. 1994. "Monotone comparative statics." *Econometrica: Journal of the Econometric Society* , 157–180. Publisher: JSTOR.

Mou, Yi, Kevin Wu, and David Atkin. 2016. "Understanding the use of circumvention tools to bypass online censorship." *New Media & Society* 18(5): 837–856. Publisher: SAGE Publications.

Nourin, Sadia, Van Tran, Xi Jiang, Kevin Bock, Nick Feamster, Nguyen Phong Hoang, and Dave Levin. 2023. "Measuring and Evading Turkmenistan's Internet Censorship: A Case Study in Large-Scale Measurements of a Low-Penetration Country." Paper presented at the Proceedings of the ACM Web Conference 2023,.

Pan, Jennifer, and Alexandra A. Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114(1): 109–125.

Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media Bias in China." *American Economic Review* 108(9): 2442–2476.

Shen, Fei, and Zhi'an Zhang. 2018. "Do circumvention tools promote democratic values? Exploring the correlates of anticensorship technology adoption in China." *Journal of Information Technology & Politics* 15(2): 106–121. Publisher: Routledge _eprint: https://doi.org/10.1080/19331681.2018.1449700.

Tirole, Jean. 2021. "Digital Dystopia." *American Economic Review* 111(6): 2007–2048.

Xu, Xu. 2021. "To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance." *American Journal of Political Science* 65(2): 309–325. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12514.

Xu, Xu. 2023. "The Unintrusive Nature of Digital Surveillance and Its Social Consequences.".

Xue, Diwen, Benjamin Mixon-Baca, ValdikSS, Anna Ablove, Beau Kujath, Jedidiah R. Crandall, and Roya Ensafi. 2022. "TSPU: Russia's decentralized censorship system." Paper presented at the Proceedings of the 22nd ACM Internet Measurement Conference, Nice France.

Yao, Linan. 2023. "From Screen to Mind: An Experimental Examination of the Power of Propagandist Entertainment on Chinese Public Opinion.".

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12(1): 415–438. _eprint: https://doi.org/10.1146/annurev-economics-081919-050239.

For Online Publication: Appendix

Proofs of section 4

Proposition A.1. *Suppose that $\psi(\theta_i)$ is strictly increasing in θ_i. Then, if $q = 1/2$, in the unique equilibrium $c^* = \tilde{c}(p)$.*

If $q > 1/2$, then in the unique sender-preferred equilibrium, $s^*_{\mathcal{S}}(1) = s^*_{\mathcal{S}}(0), c^*(1) = c^*(0) = min(\tilde{c}(p))$.

Proof. **Part 0: $q = 1/2$.** Then there is no asymmetric information and the leader's problem boils down to picking the optimal cost, given that $\mu(s_{\mathcal{S}}) = p$. With $\psi(\theta_i)$ strictly increasing in θ_i, and since the population's "interim" belief is held at p by $q = 1/2$, compliance is maximized by ensuring that $\hat{s}_{\mathcal{F}} = \emptyset$ if $\theta_i \leq p$ and $\hat{s}_{\mathcal{F}} \neq \emptyset$ otherwise, which is achieved by setting $c = \tilde{c}(p)$.

Part 1: there does not exist any (semi-) separating equilibrium. It is useful to note that $\mathbf{1}\{\theta_i \leq \mu\} + \mathbf{1}\{\theta_i > \mu\} = 1$. In a separating equilibrium, $\mu(c_1, s_{\mathcal{F}}(1)) > \mu(c_0, s_{\mathcal{F}}(0))$. We restrict attention to separating equilibria that are sender-optimal given separation: i.e., $c^*_j = \arg\max_{c_j} V(c_j|\mu(c_j, s_{\mathcal{S}}(j)) \neq \mu(c_{j'}, s_{\mathcal{S}}(j')))$ for any $j \in \{0,1\}, j' \neq j$.

No separation via cheap talk. Suppose that there exists a (semi-) separating equilibrium with $c(1) = c(0)$ but $s_{\mathcal{S}}(1) \neq s_{\mathcal{F}}(0)$. Then since $s_{\mathcal{F}}$ is cheap-talk, there exists a profitable deviation for the $\hat{\omega} = 0$-type since $\mu(\cdot, s_{\mathcal{S}}(1)) > \mu(\cdot, s_{\mathcal{S}}(0))$.

Separation via cost. Instead suppose that there exists a (semi-)separating equilibrium with $c(1) \neq c(0)$. Then the cheap-talk message is itself irrelevant and fixing some $(c(1), c(0))$ pair generating a separating equilibrium, any such separating equilibria are outcome-equivalent: abusing notation, in any such separating equilibrium we simply assume that $s_{\mathcal{S}}(1) = s_{\mathcal{S}}(0)$. Then, let $t_i(c_{\hat{\omega}}) := \mathbf{1}\{\psi(\theta_i; \mu(c_{\hat{\omega}})) \geq c_{\hat{\omega}}\}$ and $\mu^\beta_i(c_\omega) := \frac{\beta(\theta_i)\mu(c_{\hat{\omega}})}{\beta(\theta_i)\mu(c_{\hat{\omega}})+1-\mu(c_{\hat{\omega}})}$. Define

$$G(c_{\hat{\omega}}) := \int_0^1 t_i(c_{\hat{\omega}})(1 - \beta(\theta_i))\mathbf{1}\left\{\theta_i > \mu^\beta_i(c_\omega)\right\} dF(\theta_i)$$

as the level of compliance, given $c_{\hat{\omega}}$ among types who (i) gain access to $s_{\mathcal{F}}$ and (ii) with $a^*_i(s_{\mathcal{F}} =$

$1) \neq a_i^*(s_{\mathcal{F}} = 0)$. Next denote

$$H(c_{\hat{\omega}}) := \int_0^1 (1 - t_i(c_{\hat{\omega}}))\mathbf{1}\{\theta_i \leq \mu(c_{\hat{\omega}})\} + t_i(c_{\hat{\omega}})\mathbf{1}\left\{\theta_i \leq \mu_i^{\beta}(c_{\omega})\right\} dF(\theta_i)$$

as the level of compliance, given $c_{\hat{\omega}}$ from citizens such that they either (i) do not gain access to $s_{\mathcal{F}}$ or (ii) with $a_i^*(s_{\mathcal{F}} = 1) = a_i^*(s_{\mathcal{F}} = 0)$.

A type $\hat{\omega}$ leader derives an equilibrium payoff of $\mu(\hat{\omega})G(c_{\hat{\omega}}) + H(c_{\hat{\omega}})$ from playing strategy $c_{\hat{\omega}}$. For such an equilibrium to exists, the two following incentive compatibility conditions must simultaneously hold:

$$\mu(1)G(c_1) + H(c_1) \geq \mu(1)G(c_0) + H(c_0) \iff G(c_1) - G(c_0) \geq \frac{H(c_0) - H(c_1)}{\mu(1)} \qquad (\text{IC}, \hat{\omega} = 1)$$

$$\mu(0)G(c_1) + H(c_1) \leq \mu(0)G(c_0) + H(c_0) \iff G(c_1) - G(c_0) \leq \frac{H(c_0) - H(c_1)}{\mu(0)} \qquad (\text{IC}, \hat{\omega} = 0)$$

If $q > 1/2$ and $H(c_1) \geq H(c_0)$, then the two conditions cannot hold at the same time since

$$(\text{IC}, \hat{\omega} = 1) \cap (\text{IC}, \hat{\omega} = 0) \iff -\frac{H(c_1) - H(c_0)}{\mu(0)} \geq G(c_1) - G(c_0) \geq -\frac{H(c_1) - H(c_0)}{\mu(1)},$$

yet by $q > 1/2$, $\mu(0) < \mu(1)$. In turn, to conclude the proof, it suffices to show that

$$H(c_1) - H(c_0) = \int_0^1 \begin{pmatrix} (1 - t_i(c_1))\mathbf{1}\{\theta_i \leq \mu(c_1)\} + t_i(c_1)\mathbf{1}\left\{\theta_i \leq \mu_i^{\beta}(c_1)\right\} \\ -(1 - t_i(c_0))\mathbf{1}\{\theta_i \leq \mu(c_0)\} - t_i(c_0)\mathbf{1}\left\{\theta_i \leq \mu_i^{\beta}(c_0)\right\} \end{pmatrix} dF(\theta_i) \geq 0.$$

Since $\mu(c_1) \geq \mu(c_0)$ and $\mu(c_{\hat{\omega}}) > \hat{\mu}_{\beta}(c_{\hat{\omega}})$, we have:

$$(1 - t_i(c_1))\mathbf{1}\{\theta_i \leq \mu(c_1)\} + t_i(c_1)\mathbf{1}\left\{\theta_i \leq \mu_i^{\beta}(c_1)\right\} - (1 - t_i(c_0))\mathbf{1}\{\theta_i \leq \mu(c_0)\} - t_i(c_0)\mathbf{1}\left\{\theta_i \leq \mu_i^{\beta}(c_0)\right\}$$

$$= \begin{cases} 1 - t_i(c_1) & \text{if } \theta_i \in (\max\{\mu_i^{\beta}(c_1), \mu(c_0)\}, \mu(c_1)) \\ 1 & \text{if } \mu_i^{\beta}(c_1) > \mu(c_0) \text{ and } \theta_i \in (\mu(c_0), \mu_i^{\beta}(c_1)] \\ t_i(c_0) - t_i(c_1) & \text{if } \mu_i^{\beta}(c_1) < \mu(c_0) \text{ and } \theta_i \in (\mu_i^{\beta}(c_1), \mu(c_0)] \\ 1 - t_i(c_0) & \text{if } \theta_i \in (\mu_i^{\beta}(c_0), \min\{\mu_i^{\beta}(c_1), \mu(c_0)\}] \\ 0 & \text{otherwise.} \end{cases}$$

In turn, if $t_i(c_0) - t_i(c_1)$ for $\mu_i^\beta(c_1) < \mu(c_0)$ and $\theta_i \in (\mu_i^\beta(c_1), \mu(c_0)]$ the proof is concluded. Notice that if $\psi(\theta_i)$ is strictly increasing in θ_i then $t_i(c_0) = 1$ over this interval which concludes the proof. Alternatively, if $\psi(\theta_i)$ is strictly increasing in $\theta_i \forall \theta_i \geq \mu(c_1)$ and $min(\psi(\theta_i; \theta_i \geq \mu(c_1))) > max(\psi(\theta_i; \theta_i < \mu(c_1)))$ then $t_i(c_1) = 0$ over this interval which is another sufficient condition.

Part 2: Pooling Equilibrium with strictly increasing $\psi(\theta_i)$. There exists many different pooling equilibria: we focus on the sender-preferred one: $c^*(\hat{\omega}) = \tilde{c}(p) \forall \hat{\omega} \in \{0, 1\}$ and $s_{\mathcal{S}}(1) =_{\mathcal{S}} (0)$, which is sustained with an off-path belief $\mu(c' \neq \tilde{c}(p)) \leq p$, and survives the Intuitive Criterion. No profitable deviation from such an equilibrium exists as long as

$$\begin{aligned} V(c'; \hat{\omega}) &= \int_0^1 \mathbf{1}\{\theta_i \leq \mu(c')\} + \mu(\hat{\omega})(1 - \beta(\theta_i))\mathbf{1}\{\theta_i > \mu(c')\} dF(\theta_i) \\ &= \int_0^1 \mu(\hat{\omega})(1 - \beta(\theta_i)) + (1 - \mu(\hat{\omega})(1 - \beta(\theta_i)))\mathbf{1}\{\theta_i \leq \mu(c')\} dF(\theta_i) \\ &\leq \int_0^1 \mu(\hat{\omega})(1 - \beta(\theta_i)) + (1 - \mu(\hat{\omega})(1 - \beta(\theta_i)))\mathbf{1}\{\theta_i \leq p\} dF(\theta_i) = V(\tilde{c}(p); \hat{\omega}) \end{aligned}$$

Notice that this holds for any $\hat{\omega}$ as long as $\mu(c') \leq p$.

Consider now a pooling equilibrium with an off-path belief $\mu(c') > p$. To rule out any equilibrium with such an off-path belief, we impose the following restriction. Receivers are Bayesian off-path whenever possible: $\mu(c') = \frac{Pr(c'|\hat{\omega})Pr(\hat{\omega})}{Pr(c')}$ with $Pr(c'|\hat{\omega}) = 1$ if $V(c'; \hat{\omega}) \geq V(\tilde{c}(p); \hat{\omega})$ and $Pr(c'|\hat{\omega}) = 0$ otherwise. Finally, if $Pr(c') = 0$ then $\mu(c') = p$. That is, an off-path belief $\mu(c') > p$ can be rationalized if and only if the $\hat{\omega} = 1$-type has a profitable deviation, given such an off-path belief, but the $\hat{\omega} = 0$-type does not. If both types have a profitable deviation, only $\mu(c') = p$ can be rationalized.

Notice that as both types of leaders have the same incentives to increase the interim belief $\mu(c')$, then $V(c'; 1) \geq V(\tilde{c}(p); 1) \iff V(c'; 0) \geq V(\tilde{c}(p); 0) \iff \mathbf{1}\{\mu(c')\} > \mathbf{1}\{\mu(\tilde{c}(p)\}$, which rules out any $\mu(c') > p$ off-path belief.

Part 2.1: Weakening the assumption on $\psi(\theta_i)$. Notice that ψ being strictly increasing in θ_i is sufficient but not necessary. First, observe that if $\psi(\theta_i)$ is strictly increasing for any $\theta_i \in [p, 1]$ and $max(\psi(\theta_i); \theta_i < p) < min(\psi(\theta_i); \theta_i \geq p)$ then (i) a pooling equilibrium with $c^*(\hat{\omega}) = \tilde{c}(p)$ is sustained with an off-path belief of $\mu(c') \leq p$ and (ii) any pooling equilibria with an off-path belief of $\mu(c') > p$ cannot be rationalized, according to the same argument.

Part 2.2: further weakening. Hereafter we only assume that $min(\psi(\theta_i; \theta_i > p)) > max(\psi(\theta_i; \theta_i \leq p))$ and introduce another notion of reasonable off-path belief.

Let $\hat{\mu}_i(\hat{s}_{\mathcal{F}})$ denote a citizen i's arbitrary belief that $\omega = 1$, following some deviation. Throughout, we impose that $\hat{\mu}_i(\hat{s}_{\mathcal{F}} = 1) = 1$ under any circumstances, as it constitutes a fully revealing signal about the state of the world regardless of the leader's strategy. If $\{\hat{\mu}_i(\hat{s}_{\mathcal{F}})\}$ has the following properties, it constitutes a *reasonable* off-path beliefs. We denote the payoff from some deviation by $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); \hat{\omega})$.

Definition 1 (**Reasonable Off-Path Beliefs**). If $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 1) - V(\tilde{c}(p); 1) > V(c'; 0) - V(\tilde{c}(p); 0)$ then $\min_i\{\hat{\mu}_i\} > p$. If $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 1) - V(\tilde{c}(p); 1) < V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 0) - V(\tilde{c}(p); 0)$ then $\max_i\{\hat{\mu}_i\} < p$.

This belief is reasonable in the following sense: when the high type has greater incentives to deviate, citizens should update their beliefs optimistically upon observing deviation (their worst-case belief should be better than the prior). Conversely, when the low type has greater incentive to deviate, citizens should update pessimistically upon observing deviation (their best-case belief should be worse than the prior).

The leader with $\hat{\omega}$ can profitably deviates with arbitrary separation if and only if $V(c', \hat{\mu}_i; \hat{\omega}) - V(\tilde{c}(p); \hat{\omega}) \geq 0$. Notice that it suffices to show that there does not exist any profitable deviation, for any type of leader, in the best case scenario deviation: i.e. it suffices to show that

$$
\begin{aligned}
V(\{\hat{\mu}_i\}; \hat{\omega}) - V(\tilde{c}(p); \hat{\omega}) &= \int_0^1 \begin{pmatrix} \mathbf{1}\{\theta_i \leq \hat{\mu}_i\} + \mu(\hat{\omega})(1 - \beta(\theta_i))\mathbf{1}\{\theta_i > \hat{\mu}_i\} \\ -\mathbf{1}\{\theta_i \leq p\} - \mu(\hat{\omega})(1 - \beta(\theta_i))\mathbf{1}\{\theta_i > p\} \end{pmatrix} \\
&= \int_0^1 \begin{pmatrix} \mathbf{1}\{\theta_i \leq \hat{\mu}_i\} + \mu(\hat{\omega})(1 - \beta(\theta_i))(1 - \mathbf{1}\{\theta_i \leq \hat{\mu}_i\}) \\ -\mathbf{1}\{\theta_i \leq p\} - \mu(\hat{\omega})(1 - \beta(\theta_i))(1 - \mathbf{1}\{\theta_i \leq p\}) \end{pmatrix} \\
&= \int_0^1 (1 - \mu(\hat{\omega})(1 - \beta(\theta_i)))\Big(\mathbf{1}\{\theta_i \leq \hat{\mu}_i\} - \mathbf{1}\{\theta_i \leq p\}\Big) dF(\theta_i) \leq 0
\end{aligned}
$$

Suppose that $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 1) - V(\tilde{c}(p); 1) < V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 0) - V(\tilde{c}(p); 0)$: then $\max_i\{\hat{\mu}_i\} \leq p$ by the reasonable off-path belief requirement, which in turn implies $\mathbf{1}\{\theta_i \leq \hat{\mu}_i\} - \mathbf{1}\{\theta_i \leq p\} \leq 0$ which implies that no profitable deviation exists for either type: a contradiction.

Suppose instead that $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 1) - V(\tilde{c}(p); 1) > V(c'; 0) - V(\tilde{c}(p); 0)$ then $\min_i\{\hat{\mu}_i\} > p$ so that $\min_i\{\hat{\mu}_i\} \geq p$ by the reasonable off-path beliefs. Then, $\mathbf{1}\{\theta_i \leq \hat{\mu}_i\} - \mathbf{1}\{\theta_i \leq p\} \geq 0$ and

$1-(1-\beta(\theta_i))\mu(\hat{\omega}=0) > 1-(1-\beta(\theta_i))\mu(\hat{\omega}=1) \geq 0$ which implies that $V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 1) - V(\tilde{c}(p); 1) < V(c', \hat{\mu}_i(\hat{s}_{\mathcal{F}}); 0) - V(\tilde{c}(p); 0)$, a contradiction.

Therefore, with such reasonable off-path beliefs, to sustain the pooling equilibrium on $c^*(\hat{\omega}) = \tilde{c}(p)$, the condition that $\psi(\theta_i)$ be strictly increasing can be reduced to $min(\psi(\theta_i; \theta_i > p)) > max(\psi(\theta_i; \theta_i \leq p))$.

□

General Propositions of Section 5

In this section we provide complete and more general formal statements of Propositions 2 through 4. Conditional on full censorship occurring in equilibrium (Proposition A.2), we also assume that there exists a possibly binding upper bound on the cost of access that the regime can access, which captures the censorship technological capacity of the regime, so as to derive the main insight of section 6.2. Formally, $c \in [0, \overline{C}]$ with $\overline{C} \in \mathbb{R}_+$. That is, Proposition A.2 nests Proposition 2.

- $\hat{\theta}$ denotes the mode of $f(\theta)$
- $\theta^\dagger$ is the unique solution to $F(\theta^\dagger) = \theta^\dagger$ which exists by assumption and must be unique when it does (by the unimodality of $f(\theta)$)
- $\theta^L := \arg\max_\theta \frac{F(\theta)}{\theta}$ for a unimodal F with some $\theta^\dagger \in (0,1)$.
- $\theta_\beta := \frac{p\beta}{p\beta+(1-p)} = Pr(\omega = 1 | \sigma = 1, s_{\mathcal{S}} = 1, s_{\mathcal{F}} = 0)$
- $\sigma^L := argmax_\sigma V(\sigma, \beta, \text{no citizen gains access to } s_{\mathcal{F}})$ denotes the optimal reporting slant conditional on no citizen gaining access.
- $\sigma_{nc} := argmax_\sigma V(\sigma, \beta, \text{all citizen gain access to } s_{\mathcal{F}})$ denotes the optimal reporting slant conditional on all citizens gaining access.
- $\pi := Pr(s_{\mathcal{S}} = 1 | \omega = 1)$

Further, for Proposition A.2, since the set of individuals gaining access is closed, we introduce the following definition:

Definition 2. The range of citizens gaining access to the foreign outlet is given by $[\underline{\theta}(c,\sigma), \overline{\theta}(c,\sigma)]$ with $0 \leq \underline{\theta}(c,\sigma) \leq \theta(\emptyset,\sigma) \leq \overline{\theta}(c,\sigma) \leq 1$. Equilibrium thresholds are denoted by $\underline{\theta}(c^*,\sigma^*)$ and $\overline{\theta}(c^*,\sigma^*)$.

Target citizen notation. In the core of the paper we needed only to define the target citizen as follows $\theta(\sigma) := \mu(s_{\mathcal{S}} = 1|\sigma)$. In the appendix we instead more generally write: $\theta(\hat{s}_{\mathcal{F}} = \emptyset, \sigma) := Pr(\omega = 1|\hat{s}_{\mathcal{F}} = \emptyset, s_{\mathcal{S}} = 1, \sigma)$.

Every citizen consumes the state media, $s_{\mathcal{S}} \in \{0,1\}$. However, not all citizens necessarily consume the foreign media: abusing notation, $\hat{s}_{\mathcal{F}} \in \{0,1,\emptyset\}$.

Proposition A.2. *There exists a unique $\underline{\gamma} \in (0, 1-\beta)$ and a unique $\underline{C}(\sigma_{nc}) \in (0, \overline{c}(\theta^L))$ such that if $\gamma \in [0, \underline{\gamma}]$ then in the unique equilibrium,*

- *if $\overline{C} \leq \underline{C}(\sigma_{nc})$ then $c^* = 0$, $\sigma^* = \sigma_{nc}$, $\theta(0,\sigma^*) = \theta^L$ and $\overline{\theta}(c^*,\sigma^*) = 1$. If $z \geq 0$ then $\underline{\theta}(c^*,\sigma^*) = 0$, otherwise $\underline{\theta}(c^*,\sigma^*) = \theta(0,\sigma^*)$. A citizen complies if and only if (i) $s_{\mathcal{F}} = 1$ or (ii) $s_{\mathcal{F}} = 0, s_{\mathcal{S}} = 1, \theta_i \leq \theta(0,\sigma^*)$.*

- *if $\overline{C} > \underline{C}(\sigma_{nc})$*

  - *full censorship takes place whenever possible: $c^* = min\{\overline{c}(\theta^L), \overline{C}\}$. If $\overline{C} < \overline{c}(\theta^L)$ then citizens in the range $[\underline{\theta}(c^*,\sigma^*), \overline{\theta}(c^*,\sigma^*)]$ gain access, with $\underline{\theta}(c^*,\sigma^*) = \theta^L$ and $\overline{\theta}(c^*,\sigma^*) \in (\theta(\emptyset,\sigma^*), 1]$. The share of citizens bypassing the firewall is decreasing in $\overline{C}$.*
  - *$\sigma^*(\overline{C}) \in [\sigma_{nc}, \sigma^L]$ and σ^* increases in $\overline{C}$. $\theta(\emptyset,\sigma^*)$ decreases in $\overline{C}$ with $\lim_{\overline{C}\to\overline{c}(\theta^L)} = \theta^L$.*
  - *A citizen complies if and only if (i) $\hat{s}_{\mathcal{F}} = 1$ or (ii) $\hat{s}_{\mathcal{F}} = \emptyset, s_{\mathcal{S}} = 1, \theta_i \leq \underline{\theta}(c^*,\sigma^*)$.*
  - *compliance (weakly) increases in the censorship capacity of the regime $\overline{C}$.*

Discussion: the importance of the binding constraint on the cost of access in low correlation citizenries. When the leader's ability to impose a cost of access is too limited ($\overline{C} \leq \underline{C}(\sigma_{nc})$) then it is optimal not to impose any cost of access, simply because increasing the cost of access reduces the share of opponents with access ($\downarrow \overline{\theta}(c^*,\sigma^*)$) while not affecting the share of unconditional compliers: formally $\underline{\theta}(\alpha_n,\sigma) < \theta(0,\sigma)$ when the upper bound on the cost of access is too low.

Otherwise, when the leader's ability to impose a cost of access is limited (but not too limited, i.e. $\overline{C} > \underline{C}(\sigma_{nc})$) the regime minimizes the share of citizens bypassing the firewall. This is done via the two levers at their disposal. First, they sets the cost of access as high possible ($c^* = \overline{C}$). Second, they make the state media more informative to reduce the informational benefit from bypassing the firewall. As the regime's ability to impose a cost of access decreases ($\overline{C} \downarrow$) the state media becomes more informative. Formally, among firewall by-passers, the one most aligned with the regime is the same as the one that would just comply after contradictory reporting in the case of no censorship ($\underline{\theta}(c^*, \sigma^*) = \theta^L$). As the regime's ability to impose a cost of access increases ($\overline{C} \uparrow$), $\sigma^* \uparrow$ and $\overline{\theta}(c^*, \sigma^*)$ converges towards θ^L. Compliance is thus maximized under full censorship. Put differently, if partial censorship does occur when $\gamma \leq \underline{\gamma}$ in equilibrium it is because it is unavoidable, not because it is actively pursued by the regime.

Proposition A.3. *There exists a unique $\overline{\gamma} \in (\underline{\gamma}, 1 - \beta)$ s.t. if $\gamma \geq \overline{\gamma}$ then in the unique equilibrium:*

- $\sigma^* = \sigma^S \in (\sigma^L, 1]$ *such that* $\theta(\emptyset, \sigma^S) = \theta^S < \theta^L$.
- *censorship is partial even when full censorship is possible.* $c^* = \tilde{c}(\theta^S)$ *ensures that only citizens more misaligned than* θ^S *gain access;* $\underline{\theta}(c^*, \sigma^*) = \theta^S$ *and* $\overline{\theta}(c^*, \sigma^*) = 1$.
- *A citizen complies if and only if (i)* $\hat{s}_{\mathcal{F}} = 1$ *or (ii)* $\hat{s}_{\mathcal{F}} = \emptyset, s_{\mathcal{S}} = 1, \theta_i \leq \theta^S$.
- *the level of compliance is maximized and constant in* γ *and for any* $\gamma \geq \overline{\gamma}$.

Proposition A.4. *If $\gamma \in [\underline{\gamma}, \overline{\gamma})$ then in the unique equilibrium,*

- $\sigma^* = \sigma^I \in [\sigma^S, 1]$ *such that* $\theta(\emptyset, \sigma^*) = \theta^I \in [p, \theta^S]$.
- $c^* = \tilde{c}(\theta^I)$ *ensures that only citizens* $\theta_i \geq \theta^I$ *gain access;* $\underline{\theta}(c^*, \sigma^*) = \theta^I$ *and* $\overline{\theta}(c^*, \sigma^*) = 1$.
- *A citizen complies if and only if (i)* $\hat{s}_{\mathcal{F}} = 1$ *or (ii)* $\hat{s}_{\mathcal{F}} = \emptyset, s_{\mathcal{S}} = 1, \theta_i \leq \theta^I$.
- *compliance is strictly increasing in* γ *for any* $[\underline{\gamma}, \overline{\gamma}]$ *and bounded between*
  - *a lower bound: the full-censorship payoff* ($\gamma \leq \underline{\gamma}$)*, and*
  - *an upper bound: the segment-and-rule payoff from the strong correlation case* ($\gamma \geq \overline{\gamma}$)*.*

Finally, this corollary is close in spirit to Proposition 5 and not included in the main text.

Corollary A.1. *For any $\beta \in (0, 1)$, the equilibrium*

1. *reporting slant (σ^*) is non-monotonic in γ: it is constant for any $\gamma < \underline{\gamma}$, jumps at $\gamma = \underline{\gamma}$ and is decreasing in γ otherwise with $\lim_{\gamma\to\infty} \sigma^* > max\{\sigma^* \mid \gamma < \underline{\gamma}\}$*

2. *share of citizens who are censored – do not bypass the firewall – is 1 for any $\gamma < \underline{\gamma}$, falls at $\gamma = \underline{\gamma}$, is increasing in γ (but always strictly below 1) otherwise*

3. *compliance is weakly increasing in γ, strictly so for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.*

We now provides proofs for each full proposition and corrolary provided above, as well as any other results presented in the core of the paper.

Proofs of sections 5 and 6

First, we characterize optimal propaganda σ^* given no censorship and full censorship. It is useful to introduce the concept of an upper bound $\overline{C} \in \mathbb{R}_+$ s.t. $c \in [0, \overline{C}]$.

Lemma A.1. *If $\overline{C} = 0$ then there exists a unique $\theta^L \in [max\{p, \hat{\theta}, \theta^\dagger\}, 1)$ and $\sigma_{nc} = min\{\beta \frac{p}{1-p} \frac{1-\theta^L}{\theta^L}, 1\}$ which maximize compliance. Further, compliance increases in β.*
If $\overline{C} > \overline{c}(\theta^L)$ and $c \geq \overline{c}(\theta^L)$ s.t. no citizen gains access then there exists a unique $\theta^L \in [max\{p, \hat{\theta}, \theta^\dagger\}, 1)$ and $\sigma^L = min\{\frac{p}{1-p} \frac{1-\theta^L}{\theta^L}, 1\}$ which maximize compliance.

Proof. Suppose that $\overline{C} = 0$ s.t. $c = 0$ by constraint. The leader's problem is given by

$$\begin{aligned}
&\max_{\sigma} \quad [p + (1-p)\sigma]F(\theta(0,\sigma)) + p(1-\beta)(1 - F(\theta(0,\sigma))) \\
&\max_{\theta(0,\sigma)} \quad p\beta \frac{F(\theta(0,\sigma))}{\theta(0,\sigma)} + p(1-\beta) \\
&(FOC) \quad f(\theta(0,\sigma))\theta(0,\sigma)) - F(\theta(0,\sigma))) = 0
\end{aligned}$$

First, notice that $f(0) \times 0 = F(0) = 0$ and $f(1) \times 1 < F(1) = 1$. Next, notice that, abusing notation, $\frac{\partial}{\partial\theta} f(\theta)\theta = f(\theta) + f'(\theta) \geq \frac{\partial F(\theta)}{\partial\theta} = f(\theta)$ if and only if $\theta \leq \hat{\theta}$. Thus, $f(\theta)\theta > F(\theta)$ for $\theta \leq \hat{\theta}$ and there exists $\theta^L > \hat{\theta}$ such that $f(\theta^L)\theta^L = F(\theta^L)$ by the Intermediate Value Theorem. Further, at θ^L the SOC wrt to θ yields $f'(\theta^L)\theta^L < 0$ since $\theta^L > \hat{\theta}$. Given θ^L, there exists a unique $\sigma^* = \beta \frac{p}{1-p} \frac{1-\theta^L}{\theta^L}$ associated with it. Importantly, it must be the case that the target citizen θ^L can be reached with

$\sigma^* \in [0,1]$. If $\theta_\beta > \theta^L$ then $\sigma^* = 1$. Formally, $\sigma^* := \arg\max_\sigma V(\sigma, c = 0)$ is given by

$$\sigma^* = \begin{cases} 1 & \text{if} \quad \theta_\beta \in [\theta^L, 1] \\ \beta \dfrac{p}{1-p} \dfrac{1-\theta^L}{\theta^L} & \text{if} \quad \theta_\beta \in [0, \theta^L). \end{cases}$$

Notice that σ^* is increasing in β. Notice that $\sigma^*(c > \overline{c}(\theta^L)) = \sigma^*(c = 0, \beta) = 1$. It follows that compliance increases in β.

Finally, suppose that we allow for $\pi := Pr(s_{\mathcal{S}} = 1|\omega = 1) < 1$. We provide this proof for $c = 0$ which implies the proof holds for $\beta = 1$ which is equivalent to $c > \overline{c}(\theta^L)$. For any $\pi < 1$, write

$$\theta_1^\pi := \theta(1, 0, \sigma, \pi) = \frac{p\beta\pi}{p\beta\pi + (1-p)\sigma}$$
$$\theta_0^\pi := \theta(0, 0, \sigma, \pi) = \frac{p\beta(1-\pi)}{p\beta(1-\pi) + (1-p)(1-\sigma)}$$

Note that $\theta(1, \sigma) \geq \theta_1^\pi$ for any $\pi < 1$. Further, $\theta_1^\pi \geq \theta_0^\pi \iff \pi \geq \sigma$. The leader's payoff can be written as

$$V(\sigma, \beta, \pi) = p\Big[\beta\Big[\frac{F(\theta_1^\pi)}{\theta_1^\pi}\pi + \frac{F(\theta_0^\pi)}{\theta_0^\pi}(1-\pi)\Big] + (1-\beta)\Big] \tag{7}$$

We first note that the payoff from true good news from the foreign media is not affected by π nor σ and can thus be ignored. Observe the following:

$$\begin{aligned} V(\sigma, \beta, \pi) &= \pi\Big[p\beta\frac{F(\theta_1^\pi)}{\theta_1^\pi} + (1-\beta)p\Big] + (1-\pi)\Big[p\beta\frac{F(\theta_0^\pi)}{\theta_0^\pi} + (1-\beta)p\Big] \\ &= \pi V(\sigma; \beta, \pi) + (1-\pi)V(1-\sigma; \beta, 1-\pi). \end{aligned}$$

Intuition for the formal argument below: one can maximise individually $V(\sigma; \beta, \pi)$ and $V(1-\sigma; \beta, 1-\pi)$ by solving for the optimal target citizen θ^L (problem already solved in the baseline game). But then $V(\sigma; \beta, \pi)$ and $V(1-\sigma; \beta, 1-\pi)$ cannot be maximised jointly with any $\pi \in (0,1)$.

Formally, we know that $V(\sigma; \beta, \pi)$ attains its unique maximum when $\theta_1^\pi(\sigma) = \theta^L$. Thus, for σ such that $\theta_1^\pi(\sigma) \neq \theta^L$, $V(\sigma; \beta, \pi) < \max_\sigma V(\sigma; \beta, \pi) = V(\sigma^*; \beta, c = 0)$. Similarly, $V(1-\sigma; \beta, 1-\pi) < \max_\sigma V(1-\sigma; \beta, 1-\pi) = V(\sigma^*; \beta, c = 0)$ for σ such that $\theta_0^\pi(\sigma) \neq \theta^L$. Therefore, $V(\sigma, \beta, \pi) \leq V(\sigma^*; \beta, c = 0)$. □

Denote the maximal payoff from $c = 0$ by $V(\theta^L, c = 0) = p[\beta\frac{F(\theta^L)}{\theta^L} + (1-\beta)] \geq p$ where the last inequality follows from the fact that the leader can always ensure a payoff of p by setting $\sigma = 0$. Given some $\theta(\emptyset, \sigma)$, denote the cost of access – assuming it exists – that ensures that only citizens above the target citizen (i.e. $\theta_i \geq \theta(\emptyset, \sigma)$) gain access, by $\tilde{c}(\theta(\emptyset, \sigma))$. Similarly denote the payoff from segment-and-rule (hereafter SAR) assuming it is possible and given some σ by $V(\theta(\emptyset, \sigma), \tilde{c}(\theta(\emptyset, \sigma))) = [p + (1-p)\sigma]F(\theta(\emptyset, \sigma)) + p(1-\beta)[1 - F(\theta(\emptyset, \sigma)]$.

Lemma A.2. *Fixing some target citizen $\theta(\emptyset, \sigma)$, whenever SAR is possible, then $c^* = \tilde{c}(\theta(\emptyset, \sigma))$ maximizes compliance. SAR is possible iff $\frac{\partial \delta_i(\theta_i, \sigma, s_{\mathcal{S}})}{\partial \theta_i} \geq 0 \forall \theta_i \in [\theta^{min}, 1] \iff \gamma \geq \theta(\emptyset, \sigma)(1-\beta)$.*

Proof. Given some $\theta(\emptyset, \sigma)$, all $\theta_i \leq \theta(\emptyset, \sigma)$ comply conditional on only observing good news from the state media. Since δ_i is increasing in θ_i, denote by $\tilde{c}$ the unique solution to $\delta_i(\theta_i = \theta(\emptyset, \sigma), \sigma, s_{\mathcal{S}} = 1, \tilde{c}) = 0$.

Note first that any $c' > \tilde{c}$ is dominated by $\tilde{c}$: it only reduces compliance among types $\theta_i > \theta(\emptyset, \sigma)$ and not affect the decision-making of $\theta_i \leq \theta(\emptyset, \sigma)$. Further, any $c' < \tilde{c}$ is dominated by $\tilde{c}$. Given some c', citizens gain access iff $\theta_i > \theta'$ (with $\theta' < \theta(\emptyset, \sigma)$). Then the leader is better off under $\tilde{c}$ iff

$$(p + (1-p)\sigma - p(1-\beta))[F(\theta(\emptyset, \sigma)) - F(\theta')] > 0$$

which trivially holds. Thus, *fixing* σ, $c^* = \tilde{c}(\sigma)$.

□

Lemma A.3. *Suppose that SAR is feasible ($\frac{\partial \delta_i(\theta_i, \sigma, s_{\mathcal{S}})}{\partial \theta_i} \geq 0 \forall \theta_i \in [max\{p, \hat{\theta}, \theta^\dagger\}, 1]$) such that $c^* = \tilde{c}(\theta(\emptyset, \sigma))$. There exists a unique $\theta^S \in [p, \theta^L)$ and $\sigma^S \in (\sigma^L, 1]$ which maximizes compliance. Further, θ^S increases in β and σ^S decreases in β.*

Proof. We need only consider $c = \tilde{c}(\theta)$. In turn the leader's problem boils down to

$$\max_{\theta(\emptyset, \sigma) \in (p,1)} [p + (1-p)\sigma]F(\theta(\emptyset, \sigma)) + p(1-\beta)(1 - F(\theta(\emptyset, \sigma))) = p\frac{F(\theta(\emptyset, \sigma))}{\theta(\emptyset, \sigma)} + p(1-\beta)(1 - F(\theta(\emptyset, \sigma))) \tag{8}$$

$$(FOC) \quad f(\theta(\emptyset, \sigma))[1 - \theta(\emptyset, \sigma)(1-\beta)] = \frac{F(\theta(\emptyset, \sigma))}{\theta(\emptyset, \sigma)} \tag{9}$$

$$(SOC) \quad \propto f'(\theta(\emptyset, \sigma))(1 - \theta(\emptyset, \sigma)(1-\beta)) = 2f(\theta(\emptyset, \sigma))(1-\beta) \tag{10}$$

First, recall that we assume that $f(1) < 1$ in order to have $\theta^\dagger \in (0,1)$. Then, notice that $f(\theta(\emptyset,\sigma))[1-\theta(\emptyset,\sigma)(1-\beta)]$ is decreasing in $\theta(\emptyset,\sigma)$ forall $\theta(\emptyset,\sigma) > \hat{\theta}$. Also, $\frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ is single peaked, and maximized at θ^L: $\frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ increases in θ iff $\theta(\emptyset,\sigma) \leq \theta^L$. Also $f(\theta(\emptyset,\sigma)) \geq \frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ iff $\theta(\emptyset,\sigma) \leq \theta^L$.
Denote the interior solution to (FOC), if any, by θ^S. Note that $f(\theta^L) = \frac{F(\theta^L)}{\theta^L}$ thus $\theta^S < \theta^L$ and $\sigma^S > \sigma^L$. It must be that $\theta^S < \theta^L$ since $p(1-\beta)[1-F(\theta(\emptyset,\sigma))]$ is maximized at $\theta(\emptyset,\sigma) = p$ for any $\beta < 1$. There are 2 cases to consider

1. Case 1: $f(\theta(\emptyset,\sigma))[1-\theta(\emptyset,\sigma)(1-\beta)] < \frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)} \forall \theta(\emptyset,\sigma) \in [p,1]$. In this case $\theta^S = p$ and $\sigma^S = 1$. Note that this necessitates $f(p)[1-p(1-\beta)] < \frac{F(p)}{p}$.

2. Case 2: $f(\theta(\emptyset,\sigma))[1-\theta(\emptyset,\sigma)(1-\beta)] = \frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ has 1 or more local argmax on the interval $[p,\theta^L]$.
   Step 1: consider some local argmax $\theta^S \geq \hat{\theta}$. For any $\theta \in (\theta^S,\theta^L)$, $\frac{F(\theta)}{\theta} - f(\theta)(1-\theta(1-\beta)) > 0$. $\theta^S \geq \hat{\theta}$ implies that $f'(\theta) < 0 \forall \theta > \theta^S$ which proves the claim.
   Step 2: consider some local argmax $\theta^S < \hat{\theta}$. For any $\theta \in (\theta^S,\theta^L)$, $\frac{F(\theta)}{\theta} - f(\theta)(1-\theta(1-\beta)) > 0$. Suppose not. Then there must exist at least one pair of θ^1 and θ^2 s.t. $\theta^S < \theta^1 < \theta^2 < \hat{\theta}$ where θ^1 is a local argmin and θ^2 a local argmax. Rewrite the SOC evaluated at θ^S as follows:

$$(SOC) \qquad \frac{f'(\theta^S)}{f(\theta^S)} \frac{(1-\theta^S(1-\beta))}{(1-\beta)} < 2$$

   (where the inequality follows from θ^S being a local argmax). Notice that $1-\theta^S(1-\beta)$ decreases in θ^S and so does $\frac{f'(\theta^S)}{f(\theta^S)}$ (weakly) for any log concave f thus no such θ^1 can exist. A contradiction.
   Step 3: if some $\theta^S \in (p,\theta^L)$ exists, then it is unique. By step 1, if the smallest θ^S is above the mode of f, then it is unique. By step 2, if the smallest θ^S is below the mode of f, then it is unique.
   Less stringent assumptions on f. Notice that log-concavity is sufficient but not necessary. It would suffice to require that $\frac{f'(\theta^S)}{f(\theta^S)} \frac{(1-\theta^S(1-\beta))}{(1-\beta)}$ is non-increasing for $\theta \in (p,\hat{\theta})$.

To recap, either a unique interior θ^S exists and is unique, or no such interior θ^S exists and then $\theta^S = p$.
Note that the leader could always pick $\theta = \theta^L$ and $c = \overline{c}(\theta^L)$ in order to attain her ex-ante censorship payoff; since she does not, compliance is strictly higher than under ex-ante censorship.

Finally, observe that as β increases, the LHS of (FOC) increases: $\lim_{\beta\to 1}\theta^S = \theta^L$, and thus σ^S decreases in β.

Finally, notice that it is always optimal to set $\pi = 1$. Suppose not. Then building on the same proof as in Lemma A.1 notice that the compliance level is given by

$$\begin{aligned} V^S(\sigma,\beta,\pi) &= p\Big[\frac{F(\theta_1^\pi)}{\theta_1^\pi}\pi + \frac{F(\theta_0^\pi)}{\theta_0^\pi}(1-\pi) + (1-\beta)(1-F(\theta_1^\pi))\Big] \\ &= \underbrace{p\Big[\frac{F(\theta_1^\pi)}{\theta_1^\pi} + (1-\beta)(1-F(\theta_1^\pi))\Big]}_{V^S(\sigma,\tilde{c}(\theta(\emptyset,\sigma)),\pi=1)} + \underbrace{(1-\pi)\frac{F(\theta_0^\pi)}{\theta_0^\pi} - (1-\pi)\frac{F(\theta_1^\pi)}{\theta_1^\pi}}_{:=Z} \end{aligned}$$

Observe first that $Z > 0 \iff \theta^L < \theta_0^\pi < \theta_1^\pi$ (recall that θ^L is the unique argmax of $\frac{F(\theta)}{\theta}$ and $\frac{F(\theta)}{\theta}$ decreases in θ for any $\theta > \theta^L$) and thus we need only consider this case. Then recall from Lemma A.1 that $\frac{F(\theta_1^\pi)}{\theta_1^\pi}\pi + \frac{F(\theta_0^\pi)}{\theta_0^\pi}(1-\pi) < \frac{F(\theta^L)}{\theta^L}$ and thus

$$\begin{aligned} &p\Big[\frac{F(\theta_1^\pi)}{\theta_1^\pi}\pi + \frac{F(\theta_0^\pi)}{\theta_0^\pi}(1-\pi) + (1-\beta)(1-F(\theta_1^\pi))\Big] < p\Big[\frac{F(\theta^L)}{\theta^L} + (1-\beta)(1-F(\theta_1^\pi))\Big] \\ &< p\Big[\frac{F(\theta^L)}{\theta^L} + (1-\beta)(1-F(\theta^L))\Big] < p\Big[\frac{F(\theta^S)}{\theta^S} + (1-\beta)(1-F(\theta^S)\Big] \end{aligned}$$

where the penultimate inequality follows from $\theta_1^\pi > \theta^L$. Thus notice that, for any $\pi \in (\sigma, 1)$ the regime would be better off with $\pi = 1$ and $\sigma = \sigma^L$. The last inequality follows directly from the core part of proof of this very lemma. □

For the next three lemmas we assume that full censorship is feasible ($\overline{C} > \overline{c}(\theta^L)$) and compare the payoffs from segmentation to that of full censorship. Later on show that when segmentation is dominated by full censorship, the regime maximizes compliance by censoring as much as possible ($c^* = \overline{C}$), or not at all ($c^* = 0$).

Lemma A.4. *Assume $\overline{C} > \overline{c}(\theta^L)$. There exists a unique $\overline{\gamma} \in (0, 1-\beta)$ s.t. if $\gamma \geq \overline{\gamma} := \theta^S(1-\beta)$ then $c^* = \tilde{c}(\theta^S)$, only citizens of type $\theta_i \geq \theta^S$ gain access, $\sigma^* = \sigma^S$ and $\theta(\emptyset, \sigma^*) = \theta^S$.*

Proof. Given some target citizen $\theta(\emptyset,\sigma)$, $\frac{\partial \delta_i(\theta_i,\sigma,s_S)}{\partial\theta_i} \geq 0 \iff \gamma \geq \theta(\emptyset,\sigma)(1-\beta)$.

Recall that, given some β and p, compliance is maximized when the leader can pick $\theta(\emptyset,\sigma^*) = \theta^S$. If $\gamma \geq \theta^S(1-\beta)$ the leader can pick any target citizen $\theta(\emptyset,\sigma) \in [\theta^S, 1]$ and ensure that the net benefit from gaining access is (weakly) increasing in θ_i in equilibrium. In turn, by Lemma A.3, it follows that $c^* = \tilde{c}(\theta^S)$ and $\theta(\emptyset, \sigma^*) = \theta^S$.

□

Lemma A.5. *Assume $\overline{C} > \bar{c}(\theta^L)$ and $\gamma \in (p(1-\beta), \overline{\gamma})$. If $\theta^S > p$ then there exists a unique $\theta^I \in (p, \theta^S)$ s.t. $\theta(\emptyset, \sigma)(1-\beta) \geq \gamma \iff \theta(\emptyset, \sigma) \leq \theta^I$. Otherwise no such θ^I exists.*

Proof. Denote by θ^I the largest $\theta(\emptyset, \sigma) \in [p, \theta^S]$ s.t. $\gamma \geq \theta(\emptyset, \sigma)(1-\beta)$. Denote the reporting slant associated with θ^I by σ^I. Notice that $\theta(\emptyset, \sigma)(1-\beta)$ monotonically increases in $\theta(\emptyset, \sigma)$: by the intermediate value theorem such a θ^I must exist and be unique. □

Lemma A.6. *Suppose θ^I exists and assume $\overline{C} > \bar{c}(\theta^L)$ and $\gamma \in (p(1-\beta), \overline{\gamma})$; further, suppose that $p\frac{F(\theta^L)}{\theta^L} > p\frac{F(p)}{p} + p(1-\beta)[1-F(p)]$. Then there exists a unique $\theta_w^\dagger$ s.t. $p\frac{F(\theta^L)}{\theta^L} \geq p\frac{F(\theta^I)}{\theta^I} + p(1-\beta)[1-F(\theta^I)] \iff \theta^I < \theta_w^\dagger$.*

Proof. By assumption $\theta^I < \theta^S$ and by inspection of (FOC) and (SOC), θ^I dominates any $\theta(\emptyset, \sigma) \in (p, \theta^I)$. Further, any $\theta(\emptyset, \sigma) \in (\theta^I, \theta^S)$ is dominated by θ^I and/or θ^L since it does not allow for SAR *and* leads to suboptimal propaganda.

Consider under which conditions is the leader better off generating SAR (choosing θ^I) rather than ensuring his full ex-ante censorship payoff (choosing θ^L). This is the case iff

$$
\begin{aligned}
&[p + (1-p)\sigma^L]F(\theta^L) < [p + (1-p)\sigma^I]F(\theta^I) + p(1-\beta)[1-F(\theta^I)] \\
\iff &\frac{F(\theta^L)}{\theta^L} < \frac{F(\theta^I)}{\theta^I} + (1-\beta)[1-F(\theta^I)]
\end{aligned}
$$

where the second line follows from (i) $\theta^L = \frac{p}{p+(1-p)\sigma^L}$ and (ii) $\theta^I = \frac{p}{p+(1-p)\sigma^I}$. Note that by Lemma A.3 $\frac{F(\theta^I)}{\theta^I} + (1-\beta)[1-F(\theta^I)]$ is maximized and single-peaked at $\theta^I = \theta^S$. Thus there are two cases to consider:

1. If $\frac{F(\theta^L)}{\theta^L} \leq \frac{F(p)}{p} + (1-\beta)[1-F(p)]$ then $\theta(\emptyset, \sigma^*) = \theta^I$.

2. If not, then by Lemma A.1 and the intermediate value theorem there exists a unique $\theta_w^\dagger \in (p, \theta^S)$ s.t. if $\theta^I \geq \theta_w^\dagger$ then $\theta(\emptyset, \sigma^*) = \theta^I$, and otherwise $\theta(\emptyset, \sigma^*) = \theta^L$.

Finally note that if $\gamma < p(1-\beta) < \theta_w^\dagger(1-\beta)$, then irrespective of the choice of the target citizen, $\frac{\partial \delta_i}{\partial \theta_i} < 0 \ \forall \theta_i > \theta(\emptyset, \sigma)$.

$\square$

In turn, define $\underline{\gamma} := \theta^{\dagger}_{w}(1-\beta)$ as the minimal correlation such that SAR dominates full censorship. Observe that $\underline{\gamma} \in (0, \overline{\gamma})$.

In Lemmas A.7 to A.13 we characterize equilibrium play when the leader's maximal payoff from perfect segmentation is dominated by full censorship: that is, for all such lemmas we assume that $\gamma < \underline{\gamma}$. Within that range, we consider both the cases of a non binding and a binding upper bound on the cost of access: i.e. $\overline{C} < \overline{c}(\theta^L) = z + \gamma\theta^L + b_i(\theta_i = \theta^L | \theta(\emptyset, \sigma) = \theta^L)$ where $b_i(\cdot|\theta(\emptyset, \sigma))$ denotes the informational benefit from consuming $s_{\mathcal{F}}$ for type θ_i, given some target citizen $\theta(\emptyset, \sigma)$.
For the purpose of the following lemmas we write the net common benefit from consuming the foreign media as $\delta_i(\theta_i, \sigma, s_{\mathcal{S}}) = b_i(\theta_i, \sigma, s_{\mathcal{S}}) + \gamma\theta_i + \alpha_n$ where $\alpha_n = z - c$ denotes the net of cost common non informational benefit from consuming the foreign media. Observe that if $\alpha_n \geq 0$, then the entire citizenry consumes the foreign media. We treat α_n as an exogenous parameter, and, in some lemmas, switch from the cost c to the net common benefit α_n.
Hereafter we sometimes use the notation $(\theta(\emptyset, \sigma), c)$ or $(\theta(\emptyset, \sigma), \alpha_n)$ to denote a strategy - a target citizen and a cost of access - for the leader.

Lemma A.7. *Suppose $0 \leq \overline{C} \leq \overline{c}(\theta^L)$. Given some $\theta(\emptyset, \sigma) \in (\theta^{\dagger}, 1)$ there exists two unique cutoffs $\underline{\theta} \in [0, \theta(\emptyset, \sigma)] = \frac{\beta\theta(\emptyset,\sigma) - \alpha_n}{1-(1-\beta)\theta(\emptyset,\sigma)+\gamma}$ and $\overline{\theta} \in [\theta(\emptyset, \sigma), 1] = \frac{\alpha_n + (1-\beta)\theta(\emptyset,\sigma)}{(1-\beta)\theta(\emptyset,\sigma)-\gamma}$ s.t. only citizens of type $\theta_i \in [\underline{\theta}, \overline{\theta}]$ circumvent the firewall after observing $s_{\mathcal{S}} = 1$.*

Proof. θ_i consumes $s_{\mathcal{F}}$ if and only if it increases her expected payoff:

$$\max_{a_i \in \{0,1\}} E[u_i(a_i|\theta_i, s_{\mathcal{S}}, s_{\mathcal{F}}, \sigma, \beta)] + \alpha_n + \theta_i\gamma \geq \max_{a_i \in \{0,1\}} E[u_i(a_i|\theta_i, s_{\mathcal{S}}, \emptyset)]$$

Suppose that $s_{\mathcal{S}} = 0$. Then there is no informational benefit of consuming $s_{\mathcal{F}}$, so the entire citizenry never complies and consumes $s_{\mathcal{F}}$ iff $\alpha_n \geq 0$.
Recall that $\mu(\cdot, 1) := \frac{p}{p+(1-p)\sigma} = \theta(\emptyset, \sigma)$, $\mu(0, 1) := \frac{p\beta}{p\beta+(1-p)\sigma} = \theta(0, \sigma)$. Note that we need only consider $\alpha_n < 0$. For any $\alpha_n \geq 0$ all citizens consume $s_{\mathcal{F}}$ and thus the leader is indifferent between any $c > 0$ s.t. $\alpha_n > 0$ and, by assumption, the leader always picks the lowest cost of access between any cost of access that yields the same level of compliance.

Claim 1: There exists a unique $\underline{\theta}(\alpha_n, \sigma) \in [0, \theta(\emptyset, \sigma)]$, such that a citizen of type $\theta_i \in [0, \theta(\emptyset, \sigma)]$

gains access iff $\theta_i \geq \underline{\theta}(\alpha_n, \sigma)$. Following $s_{\mathcal{S}} = 1$, a citizen gains access iff

$$\underbrace{[\beta\theta(\emptyset,\sigma) + (1-\theta(\emptyset,\sigma))]}_{Pr(s_{\mathcal{F}}=0|s_{\mathcal{S}}=1)}\theta_i + \underbrace{(1-\beta)\theta(\emptyset,\sigma)}_{Pr(s_{\mathcal{F}}=1|s_{\mathcal{S}}=1)} + \underbrace{\alpha_n + \gamma * \theta_i}_{\text{intrinsic benefit}} \geq \underbrace{\theta(\emptyset,\sigma)}_{\text{payoff from not gaining access and } s_{\mathcal{S}}=1}$$

$$\iff \theta_i \geq min\{\frac{\beta\theta(\emptyset,\sigma) - \alpha_n}{1-(1-\beta)\theta(\emptyset,\sigma)+\gamma}, \theta(\emptyset,\sigma)\} := \underline{\theta}(\alpha_n,\sigma)$$

We note that for any $\theta_i \in [0, \theta(\emptyset,\sigma)]$, the informational benefit is given by $b_i(\theta_i, \cdot) = (\beta\theta(\emptyset,\sigma) + 1 - \theta(\emptyset,\sigma))\theta_i - \beta\theta(\emptyset,\sigma)$ and $\frac{\partial^2 b_i(\theta_i,\cdot)}{\partial\theta_i \partial\theta(\emptyset,\sigma)} < 0$: the higher $\theta(\emptyset,\sigma)$, the flatter the slope of $b_i(\theta_i,\cdot)$. Note too that $\underline{\theta}(\alpha_n,\sigma) = \theta(\emptyset,\sigma)$ if $\alpha_n \leq \underline{\alpha}_n := -(1-\beta)\theta(\emptyset,\sigma)(1-\theta(\emptyset,\sigma))$.

Claim 2: There exists a unique $\overline{\theta}(\alpha_n,\sigma) \in [\theta(\emptyset,\sigma), 1]$*, such that* $\theta_i \in [\theta(\emptyset,\sigma),1]$ *gain access iff* $\theta_i \leq \overline{\theta}(\alpha_n,\sigma)$. Citizens with $\theta_i > \theta(\emptyset,\sigma)$ gain access following $s_{\mathcal{S}} = 1$ iff

$$\underbrace{(1-\beta)\theta(\emptyset,\sigma)}_{Pr(s_{\mathcal{F}}=1|s_{\mathcal{S}}=1)} + \underbrace{[\beta\theta(\emptyset,\sigma) + (1-\theta(\emptyset,\sigma))]}_{Pr(s_{\mathcal{F}}=0|s_{\mathcal{S}}=1)}\theta_i + \underbrace{\alpha_n + \gamma * \theta_i}_{\text{intrinsic benefit}} \geq \underbrace{\theta_i}_{\text{payoff from not gaining access}}$$

$$\iff \theta_i \leq max\{\theta(\emptyset,\sigma), min\{\frac{\alpha_n + (1-\beta)\theta(\emptyset,\sigma)}{(1-\beta)\theta(\emptyset,\sigma)-\gamma}, 1\}\} := \overline{\theta}(\alpha_n,\sigma)$$

We note that for any $\theta_i \in (\theta(\emptyset,\sigma), 1]$ the informational benefit is given by $b_i(\theta_i,\cdot) = (1-\beta)(1-\theta_i)\theta(\emptyset,\sigma)$ and $\frac{\partial^2 b_i(\theta_i,\cdot)}{\partial\theta_i \partial\theta(\emptyset,\sigma)} < 0$: the higher $\theta(\emptyset,\sigma)$, the steeper the slope of $b_i(\theta_i,\cdot)$. □

Lemma A.8. *Given some* $\theta(\emptyset,\sigma)$ *(i) there exists a unique* $\theta_F^M = min\{\theta_i \in [0,1] : a_1^*(\theta_i, s_{\mathcal{S}} = 1, \hat{s}_{\mathcal{F}}, \sigma) = 1\} = max\{\theta(0,\sigma), \underline{\theta}(\alpha_n,\sigma)\}$ *such that all* $\theta_i \leq \theta_F^M$ *comply conditional on* $s_{\mathcal{S}} = 1$*, and (ii) there exists a unique* $\underline{\underline{\alpha}}_n \in [\underline{\alpha}_n, 0)$ *s.t.* $\underline{\theta}(\alpha_n,\sigma) \leq \theta(0,\sigma)$ *iff* $\alpha_n \geq \underline{\underline{\alpha}}_n$.

Proof. Claim 1: there exists a unique $\theta_F^M = min\{\theta_i \in [0,1] : a_1^*(\theta_i, s_{\mathcal{S}} = 1, \hat{s}_{\mathcal{F}}, \sigma) = 1\} = max\{\theta(0,\sigma), \underline{\theta}(\alpha_n,\sigma)\}$ *such that all* $\theta_i \leq \theta_F^M$ *comply conditional on* $s_{\mathcal{S}} = 1$. There are two cases to consider. Case 1: $\theta(0,\sigma) \leq \underline{\theta}(\alpha_n,\sigma)$. By definition all $\theta_i \leq \underline{\theta}(\alpha_n,\sigma)$ are exposed to $\hat{s}_{\mathcal{F}} = \emptyset$ and thus comply following $s_{\mathcal{S}} = 1$. Case 2: $\theta(0,\sigma) > \underline{\theta}(\alpha_n,\sigma)$. Here all types $\theta_i \in [\underline{\theta}(\alpha_n,\sigma), \theta(0,\sigma)]$ are exposed to $\hat{s}_{\mathcal{F}} \in \{0,1\}$ but always comply following $s_{\mathcal{S}} = 1$.

Claim 2: there exists a unique $\underline{\underline{\alpha}}_n \in [\underline{\alpha}_n, 0)$ *s.t.* $\underline{\theta}(\alpha_n,\sigma) \leq \theta(0,\sigma)$ *iff* $\alpha_n \geq \underline{\underline{\alpha}}_n$. This follows directly from the facts that (i) $\underline{\theta}(\alpha_n,\sigma)$ decreases in α_n and (ii) for any $\alpha_n \geq 0$ then all citizens gain access ($\underline{\theta}(\alpha_n,\sigma) = 0$) and (iii) for any $\alpha_n \leq \underline{\alpha}_n$ no citizen gains access ($\underline{\theta}(\alpha_n,\sigma) = \theta(\emptyset,\sigma)$) and (iii) $\theta(0,\sigma) \geq \theta_\beta > 0$. We note that this implies the existence of a unique $\underline{C} \in [0, \overline{C}]$ s.t. $\forall c \in [0, \underline{C}]$ then

$\underline{\theta}(\alpha_n, \sigma) \leq \theta(0, \sigma)$. □

Lemma A.9. $\theta_F^M = \theta^L > \hat{\theta}$ for any $\alpha_n \in \mathbb{R}$.

Proof. Claim 1: given some $\theta(\emptyset, \sigma)$ and $\alpha_n < \underline{\underline{\alpha}}_n$ then $\theta_F^M = \underline{\theta}(\alpha_n, \sigma) = \theta^L$. By $\alpha_n < \underline{\underline{\alpha}}_n$ we know that $\underline{\theta}(\alpha_n, \sigma) > \theta(0, \sigma)$. Observe that $\frac{\partial \underline{\theta}(\alpha_n, \sigma)}{\partial \alpha_n} < 0$ and thus $\underline{\theta}(\alpha_n, \sigma)$ reaches its minimum at $\alpha_n = \underline{\underline{\alpha}}_n$. Further, recall that we already pinned down the equilibrium reporting strategy of the state media under no censorship ($\alpha_n \geq \underline{\underline{\alpha}}_n$) and full censorship ($\alpha_n \leq \underline{\alpha}_n$) in Lemma A.1. Thus

$$lim_{\alpha_n \to \underline{\underline{\alpha}}_n} \theta(0, \sigma) = \theta^L = lim_{\alpha_n \to \underline{\underline{\alpha}}_n} \theta(\emptyset, \sigma)$$

This follows from the fact that both with full and no censorship, given some cdf F, the share of compliers conditional on $s_{\mathcal{S}} = 1$ is held constant; only σ^* changes (see Lemma A.1). Thus, by the squeeze theorem, $lim_{\alpha_n \to \underline{\underline{\alpha}}_n} \underline{\theta}(\alpha_n, \sigma) = \theta^L$ and thus $\underline{\theta}(\alpha_n, \sigma) \geq \theta^L$. Importantly, we also know that $lim_{\alpha_n \to \underline{\alpha}_n} \theta(\emptyset, \sigma) = \theta^L$; since $\underline{\theta}(\alpha_n, \sigma)$ is maximized at $\alpha_n = \underline{\alpha}_n$ and since $\underline{\theta}(\alpha_n, \sigma) \leq \theta(\emptyset, \sigma)$, this implies that $\underline{\theta}(\alpha_n, \sigma) \leq \theta^L$. Thus $\underline{\theta}(\alpha_n, \sigma) = \theta^L$. Further we recall that θ^L is on the concave side of F; thus $\underline{\theta}(\alpha_n, \sigma) > \hat{\theta}$.

Claim 2: given some $\theta(\emptyset, \sigma)$ and $\alpha_n \geq \underline{\underline{\alpha}}_n$ then $\theta_F^M = \theta(0, \sigma) = \theta^L$. If in equilibrium $\alpha_n \geq \underline{\underline{\alpha}}_n$ then it is optimal to set $c^* = 0 \implies \alpha_n \geq 0$. By Lemma A.1 this implies that compliance is maximized with $\sigma = \sigma_{nc}$ s.t. $\theta(0, \sigma_{nc}) = \theta^L$. □

Lemma A.10. If $\alpha_n > \underline{\underline{\alpha}}_n \iff \underline{\theta}(\alpha_n, \sigma) < \theta(0, \sigma)$ compliance increases in α_n and is maximized at $\alpha_n \geq 0$. Further, $\sigma(\alpha_n) = \arg\max V_w(\sigma; \alpha_n > \underline{\underline{\alpha}}_n, \beta) = \sigma_{nc}$.

Proof. Given $\alpha_n > \underline{\underline{\alpha}}_n \iff \underline{\theta}(\alpha_n, \sigma) < \theta(0, \sigma)$ then compliance is maximized by imposing no cost of access ($\alpha_n \geq 0$) because $\frac{\partial \bar{\theta}(\alpha_n, \sigma)}{\partial \alpha_n} > 0$ while $\frac{\partial \theta(0, \sigma)}{\partial \alpha_n} = 0$. Further, given that it is optimal not to impose any cost of access, we know from Lemma A.1 that compliance is maximized by setting $\sigma(\alpha_n) = \sigma_{nc}$ s.t. $\theta(0, \sigma_{nc}) = \theta^L$. □

Lemma A.11. If $\alpha_n \leq \underline{\underline{\alpha}}_n \iff \underline{\theta}(\alpha_n, \sigma) \geq \theta(0, \sigma)$ compliance decreases in α_n and is maximized at $\alpha_n = \underline{\alpha}_n$. Further, $\sigma(\alpha_n) = \arg\max V_w(\sigma; \alpha_n \leq \underline{\underline{\alpha}}_n, \beta)$ is unique and weakly decreasing in α_n with $\sigma(\alpha_n = \underline{\underline{\alpha}}_n) = \sigma_{nc}$ and $\sigma(\alpha_n = \underline{\alpha}_n) = \sigma_c$.

Proof. Claim 1: Given some $\theta(\emptyset, \sigma)$ and $\alpha_n \leq \underline{\underline{\alpha}}_n$ then compliance weakly decreases in α_n. Suppose that $\gamma = 0$. Observe the following (either algebraically or from Bayes-plausibility)

$$E[\{\underline{\theta}(\alpha_n,\sigma),\overline{\theta}(\alpha_n,\sigma)\}|s_{\mathcal{S}}=1] = \underbrace{[\theta(\emptyset,\sigma)\beta + 1-\theta(\emptyset,\sigma)]}_{Pr(\underline{\theta}(\alpha_n,\sigma))}\underline{\theta}(\alpha_n,\sigma) + \underbrace{\theta(\emptyset,\sigma)(1-\beta)}_{Pr(\overline{\theta}(\alpha_n,\sigma))}\overline{\theta}(\alpha_n,\sigma) = \theta(\emptyset,\sigma)$$

Recall that $\hat{\theta} < \underline{\theta}(\alpha_n,\sigma) < \overline{\theta}(\alpha_n,\sigma)$. For any $\alpha_n \in (\underline{\alpha}_n, \underline{\underline{\alpha}}_n)$, conditional on $s_{\mathcal{S}} = 1$, the leader's payoff is given by

$$\begin{aligned} V_w(\theta(\emptyset,\sigma);\alpha_n,\beta) =& [\theta(\emptyset,\sigma) + 1-\theta(\emptyset,\sigma)]F(\underline{\theta}(\alpha_n,\sigma)) + \theta(\emptyset,\sigma)(1-\beta)[F(\overline{\theta}(\alpha_n,\sigma)) - F(\underline{\theta}(\alpha_n,\sigma))] \\ =& [\theta(\emptyset,\sigma)\beta + 1-\theta(\emptyset,\sigma)]F(\underline{\theta}(\alpha_n,\sigma)) + \theta(\emptyset,\sigma)(1-\beta)F(\overline{\theta}(\alpha_n,\sigma)) < F(\theta(\emptyset,\sigma)) \end{aligned}$$

Where the inequality follows directly from the facts that (i) $\hat{\theta} < \underline{\theta}(\alpha_n,\sigma) < \overline{\theta}(\alpha_n,\sigma)$ and f is unimodal. Further, since $\frac{\partial\underline{\theta}(\alpha_n,\sigma)}{\partial\alpha_n} < 0$, $\frac{\partial\overline{\theta}(\alpha_n,\sigma)}{\partial\alpha_n} > 0$ and since $lim_{\alpha_n \to \underline{\alpha}_n}\underline{\theta}(\alpha_n,\sigma) = \theta(\emptyset,\sigma)$, $lim_{\alpha_n\to\underline{\alpha}_n}\overline{\theta}(\alpha_n,\sigma) = \theta(\emptyset,\sigma)$ thus $lim_{\alpha_n\to\underline{\alpha}_n}V_w(\theta(\emptyset,\sigma);\alpha_n,\beta) = V_w(\theta(\emptyset,\sigma);\underline{\alpha}_n,\beta)$. Thus, as α_n decreases, $V_w(\theta(\emptyset,\sigma);\alpha_n,\beta)$ increases and converges to the full censorship payoff.[26]

Suppose now that $\gamma \in (0,\underline{\gamma})$: recall that $\underline{\theta}(\alpha_n,\sigma) = \frac{\beta\theta(\emptyset,\sigma)-\alpha_n}{1-(1-\beta)\theta(\emptyset,\sigma)+\gamma}$ and $\overline{\theta}(\alpha_n,\sigma) = \frac{(1-\beta)\theta(\emptyset,\sigma)+\alpha_n}{(1-\beta)\theta(\emptyset,\sigma)-\gamma}$. Recall that $\underline{\theta}(\alpha_n,\sigma)$ is the unique solution to

$$\begin{aligned} \theta(\emptyset,\sigma) = & \underbrace{(\beta\theta(\emptyset,\sigma)+1-\theta(\emptyset,\sigma))}_{Pr(s_{\mathcal{F}}=0|s_{\mathcal{S}}=1)}\underline{\theta}(\alpha_n,\sigma) + \underbrace{(1-\beta)\theta(\emptyset,\sigma)}_{Pr(s_{\mathcal{F}}=1|s_{\mathcal{S}}=1)}\overline{\theta}(\alpha_n,\sigma) \\ & + (1-\beta)\theta(\emptyset,\sigma)(1-\overline{\theta}(\alpha_n,\sigma)) + \gamma\underline{\theta}(\alpha_n,\sigma) + \alpha_n \end{aligned}$$

Observe that $1-\overline{\theta}(\alpha_n,\sigma) = \frac{-\gamma-\alpha_n}{(1-\beta)\theta(\emptyset,\sigma)-\gamma}$ and thus

$$\begin{aligned} (1-\beta)\theta(\emptyset,\sigma)(1-\overline{\theta}(\alpha_n,\sigma)) + \gamma\underline{\theta}(\alpha_n,\sigma) + \alpha_n =& \gamma\underline{\theta}(\alpha_n,\sigma) + \alpha_n + \gamma(1-\overline{\theta}(\alpha_n,\sigma)) \\ & + [(1-\beta)\theta(\emptyset,\sigma)-\gamma](1-\overline{\theta}(\alpha_n,\sigma)) \\ =& \gamma(1+\underline{\theta}(\alpha_n,\sigma)-\overline{\theta}(\alpha_n,\sigma)) + \alpha_n - \gamma - \alpha_n \\ =& \gamma(\underline{\theta}(\alpha_n,\sigma)-\overline{\theta}(\alpha_n,\sigma)) < 0 \end{aligned}$$

[26] Put differently: We can construct a linear function $l_w(\cdot)$ such that $l_w'(\cdot) = \frac{F(\overline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n)) - F(\underline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n))}{\overline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n) - \underline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n)}$, $l_w(\underline{\theta}(\alpha_n,\sigma)) = F(\underline{\theta}(\alpha_n,\sigma))$, $l_w(\overline{\theta}(\alpha_n,\sigma)) = F(\overline{\theta}(\alpha_n,\sigma))$, and $l_w(\theta(\emptyset,\sigma)) = V_w(\theta(\emptyset,\sigma);\alpha_n,\beta)$ because the expectation of a linear function is a linear function of expectation. Notice that holding $\theta(\emptyset,\sigma)$ fixed, as α_n increases, $l_w(\theta(\emptyset,\sigma))$ decreases as its distance to $F(\theta(\emptyset,\sigma))$ increases. In other words, $V_w(\sigma)$ is decreasing in α_n.

Then for any $\gamma \in (0, \underline{\gamma})$,

$$\begin{aligned}E[\underline{\theta}(\alpha_n,\sigma),\overline{\theta}(\alpha_n,\sigma)|s_{\mathcal{S}}=1] =&(\beta\theta(\emptyset,\sigma)+1-\theta(\emptyset,\sigma))\underline{\theta}(\alpha_n,\sigma)+[(1-\beta)\theta(\emptyset,\sigma)]\overline{\theta}(\alpha_n,\sigma)\\ =&(\beta\theta(\emptyset,\sigma)+1-\theta(\emptyset,\sigma)-\gamma)\underline{\theta}(\alpha_n,\sigma)+[(1-\beta)\theta(\emptyset,\sigma)+\gamma]\overline{\theta}(\alpha_n,\sigma)\\ &+\gamma(\underline{\theta}(\alpha_n,\sigma)-\overline{\theta}(\alpha_n,\sigma))=\theta(\emptyset,\sigma)\end{aligned}$$

In turn the same proof as in the case of $\gamma = 0$ applies. Lastly notice that for any $\alpha_n \geq \underline{\underline{\alpha}}_n$ the leader's payoff is constant in α_n. Thus the leader's payoff is continuously (weakly) decreasing in α_n.

Claim 2: $\sigma(\alpha_n) = \arg\max V_w(\sigma; \alpha_n \leq \underline{\underline{\alpha}}_n, \beta)$ is weakly decreasing in α_n. Observe first that given $\alpha_n \geq \underline{\underline{\alpha}}_n$ then $\sigma(\alpha_n) = \sigma^*(c = 0) = \sigma^L$ and similarly, given $\alpha_n \leq \underline{\alpha}$ then $\sigma(\alpha_n) = \sigma^L$.
Consider $\sigma' > \sigma$ and $\alpha_n < \alpha_n^{'}$ s.t. $(\alpha_n, \alpha_n^{'}) \in [\underline{\alpha}_n, 0]$. We want to show that $V_w(\sigma', \alpha_n'; \beta) - V_w(\sigma, \alpha_n'; \beta) \geq 0$ implies $V_w(\sigma', \alpha_n; \beta) - V_w(\sigma, \alpha_n; \beta) \geq 0$, so that σ^* is weakly increasing in $-\alpha_n$, or it is weakly decreasing in α_n by the single-crossing property (Milgrom and Shannon, 1994; Ashworth and Bueno de Mesquita, 2006).

Notice that $V_w(\sigma', \alpha_n'; \beta) - V_w(\sigma, \alpha_n'; \beta) \geq 0$ implies $V_w(\sigma', \alpha_n; \beta) - V_w(\sigma, \alpha_n; \beta) \geq 0$ if

$$\begin{aligned}&V_w(\sigma',\alpha_n';\beta)-V_w(\sigma,\alpha_n';\beta)\leq V_w(\sigma',\alpha_n;\beta)-V_w(\sigma,\alpha_n;\beta)\\ &\iff V_w(\sigma,\alpha_n;\beta)-V_w(\sigma,\alpha_n';\beta)\leq V_w(\sigma',\alpha_n;\beta)-V_w(\sigma',\alpha_n';\beta)\\ &\iff [p\beta+(1-p)\sigma][F(\underline{\theta}(\sigma,\alpha_n))-F(\underline{\theta}(\sigma,\alpha_n'))]+p(1-\beta)\Big[F(\overline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n))-F(\overline{\theta}(\alpha_n,\sigma)(\sigma,\alpha_n'))\Big]\\ &\leq [p\beta+(1-p)\sigma'][F(\underline{\theta}(\sigma',\alpha_n))-F(\underline{\theta}(\sigma',\alpha_n'))]+p(1-\beta)\Big[F(\overline{\theta}(\alpha_n,\sigma)(\sigma',\alpha_n))-F(\overline{\theta}(\alpha_n,\sigma)(\sigma',\alpha_n'))\Big].\end{aligned}$$

Note that $F(\underline{\theta}(\sigma', \alpha_n)) \leq F(\underline{\theta}(\sigma, \alpha_n))$ for any $\alpha_n < 0$ and

$$\begin{aligned}\frac{\partial\underline{\theta}(\sigma,\alpha)}{\partial\theta_c} =&\frac{\beta(1+\gamma)-(1-\beta)\alpha_n}{[1-(1-\beta)\theta_c]^2}\geq 0\\ \frac{\partial^2\underline{\theta}(\sigma,\alpha)}{\partial\theta_c\partial\alpha_n} =&\frac{-(1-\beta)\alpha_n}{[1-(1-\beta)\theta_c]^2}\leq 0 \iff \frac{\partial^2\underline{\theta}(\sigma,\alpha)}{\partial\sigma\partial\alpha_n}\geq 0\end{aligned}$$

Thus, if it was the case that f is uniform such $F(\theta) = \theta \forall \theta \in [0, 1]$, the proof would be complete. Since F is concave at any $\underline{\theta}(\sigma, \alpha)$ and since $\frac{\partial^2\underline{\theta}(\sigma,\alpha)}{\partial\sigma\partial\alpha_n} \geq 0$, then $F(\underline{\theta}(\sigma', \cdot))$ is on a "stiffer" side of F than $F(\underline{\theta}(\sigma, \cdot))$: that is, the fact that the proof is complete for a uniform F implies that the same result necessarily holds for a concave F (at these points): i.e. $F(\underline{\theta}(\sigma', \alpha_n^{'})) - F(\underline{\theta}(\sigma, \alpha_n^{'})) >$

$F(\underline{\theta}(\sigma', \alpha_n)) - F(\underline{\theta}(\sigma, \alpha_n))$.

Note that $F(\overline{\theta}(\alpha_n, \sigma)(\sigma', \alpha_n)) < F(\overline{\theta}(\sigma, \alpha_n))$ for any $\alpha_n < 0$ and

$$\frac{\partial \overline{\theta}(\sigma, \alpha)}{\partial \theta_c} = \frac{(1-\beta)[-\alpha_n - \gamma]}{[(1-\beta)\theta(\emptyset, \sigma)]^2} \geq 0$$
$$\frac{\partial^2 \overline{\theta}(\sigma, \alpha)}{\partial \theta(\emptyset, \sigma)\partial \alpha_n} = \frac{-(1-\beta)}{[(1-\beta)\theta(\emptyset, \sigma)]^2} < 0$$

where the first inequality follows from the fact that $-\alpha_n \geq \gamma$ otherwise $\overline{\theta} > 1$.
The second inequality implies that $\frac{\partial^2 \overline{\theta}(\sigma, \alpha)}{\partial \sigma \partial \alpha_n} > 0$. In turn, the same proof as for $\underline{\theta}$ applies and thus $F(\overline{\theta}(\sigma', \alpha_n^{'})) - F(\overline{\theta}(\sigma, \alpha_n^{'})) > F(\underline{\theta}(\sigma', \alpha_n)) - F(\overline{\theta}(\sigma, \alpha_n))$. □

Lemma A.12. *For any $c \leq \overline{C} \in [z, \overline{c}(\theta^L)] \iff \alpha_n \in [\underline{\alpha}_n, 0]$ compliance increases in c (decreases in α_n): $c^* = min\{\overline{C}, \overline{c}(\theta^L)\}$.*

Proof. Step 1: compliance increases in c with a non-binding $\overline{C}$. From Claim 5 of Lemma A.7, we know that for a given $\theta(\emptyset, \sigma)$, a leader would pick the highest c, so his partial equilibrium payoff is $p\frac{F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ given $\theta(\emptyset, \sigma)$ and the leader can only chooses c. Also, by Lemma A.11, $\sigma(\alpha_n)$ is weakly increasing in c. We know the following two facts. First, $F(\theta^L)/\theta$ is decreasing in θ for $\theta \geq \theta^L$. Second, we know that $\theta(\emptyset, \sigma) \geq \theta^L$. Therefore, the leader is better off picking the highest c.
Step 2: compliance increases in c with a binding $\overline{C}$. By the law of iterated expectation, we know that the leader's payoff is given by $[p\beta + (1-p)\sigma] \underbrace{\theta^L}_{=\underline{\theta}(\alpha_n,\sigma)} + p(1-\beta)\overline{\theta}(\alpha_n, \sigma) + (1-p)(1-\sigma) \cdot 0 = p$.
Furthermore, $[p\beta + (1-p)\sigma]\theta^L + p(1-\beta)\overline{\theta}(\alpha_n, \sigma) = \theta(\emptyset, \sigma)$; the leader's equilibrium payoff can be rewritten as

$$[p + (1-p)\sigma]l_F(\theta(\emptyset, \sigma)) = \frac{l_F(\theta(\emptyset, \sigma))}{\theta(\emptyset, \sigma)}p$$

where $l_F(\theta(\emptyset, \sigma)) = \beta F(\theta^L) + (1-\beta)F(\overline{\theta}(\alpha_n, \sigma)) = EU_l[\cdot|s_{\mathcal{S}} = 1]$ which is evaluated at $\theta(\emptyset, \sigma)$ since $E[\underline{\theta}, \overline{\theta}(\alpha_n, \sigma)|s_{\mathcal{S}} = 1] = \theta(\emptyset, \sigma)$.
We know that $l_F(\theta(\emptyset, \sigma)) < F(\theta(\emptyset, \sigma))$ (by concavity of F on $\theta \in [\theta^L, \overline{\theta}(\alpha_n, \sigma)]$).
Also, $\frac{l_F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$ is decreasing in $\theta(\emptyset, \sigma)$. This follows from two facts. First, $\overline{\theta}(\alpha_n, \sigma) \geq \theta^L$. Second, $F(\cdot)$ is concave on that range. Therefore, $\frac{l_F(\theta(\emptyset,\sigma))}{\theta(\emptyset,\sigma)}$, which is the value of a line segment evaluated at p, between the point $(0, 0)$ and a point in another line segment between the points $(\theta^L, F(\theta^L))$ and $(\overline{\theta}(\alpha_n, \sigma), F(\overline{\theta}(\alpha_n, \sigma)))$. By the concavity of F, this value decreases as $\overline{\theta}(\alpha_n, \sigma)$ increases and

is maximized at $\theta(\emptyset, \sigma) = \theta^L$, $l_F(\theta(\emptyset, \sigma)) = F(\theta(\emptyset, \sigma))$.[27] Finally, from Lemma A.7, we know that $\theta(\emptyset, \sigma)$ (and thus $dxdx\overline{\theta}(\alpha_n, \sigma)))$ is decreasing in c; therefore, the leader's equilibrium payoff is increasing in c. □

So far we have shown the existence of a unique $\underline{\underline{\alpha}}_n$ *given* some $\theta(\emptyset, \sigma)$. We now show that in equilibrium there exists a unique $\underline{\underline{\alpha}}_n(\sigma_{nc})$ which is the unique solution to $\theta(\emptyset, \sigma_{nc}) = \underline{\theta}(\alpha_n, \sigma_{nc})$.

Lemma A.13. There exists a unique $\underline{\underline{\alpha}}_n(\sigma_{nc}) \in (\underline{\alpha}_n, 0)$ which is the unique solution to $\theta(\emptyset, \sigma_{nc}) = \underline{\theta}(\alpha_n, \sigma_{nc})$. In the unique equilibrium,

- if $\alpha_n > \underline{\underline{\alpha}}_n(\sigma_{nc})$ then $c^* = 0, \sigma^* = \sigma_{nc}$.
- if $\alpha_n \leq \underline{\underline{\alpha}}_n(\sigma_{nc})$ then $c^* = \overline{C}, \sigma^* \in [\sigma_{nc}, \sigma^L]$ and σ^* increases in $\overline{C}$.

Proof. By Lemma A.7 through A.12 we have already characterized equilibrium play for any α_n, given some $\theta(\emptyset, \sigma)$ and $\underline{\underline{\alpha}}_n$. It thus suffices to show that $\underline{\underline{\alpha}}_n(\sigma_{nc})$ exists and is unique.

To show existence notice that, $\underline{\underline{\alpha}}_n(\sigma_{nc})$ is the unique solution to $\theta(\emptyset, \sigma_{nc}) = \underline{\theta}(\alpha_n, \sigma_{nc})$ (same intermediate value theorem proof as in Lemma A.8).

To show uniqueness, suppose that there exists some $\alpha' \neq \underline{\underline{\alpha}}_n(\sigma_{nc})$. For any $\alpha_n > \alpha'$ we know that $c^* = 0$ and $\sigma^* = \sigma_{nc}$. For any $\alpha_n \leq \alpha'$ we know that $c^* = \overline{C}$ and $\sigma^* \geq \sigma_{nc}$. At $\alpha_n = \alpha'$ the same compliance level can be reached with maximal censorship ($\sigma^* = \sigma_{nc}, c^* = \overline{C}$) and without censorship ($\sigma^* = \sigma_{nc}, c^* = 0$) since $\underline{\theta}(\alpha', \sigma_{nc}) = \theta(0, \sigma_{nc})$; yet the only solution to $\underline{\theta}(\alpha_n, \sigma_{nc}) = \theta(0, \sigma_{nc})$ is $\underline{\underline{\alpha}}_n(\sigma_{nc})$; a contradiction.

Finally, to go back to the notation from Proposition A.2, notice that $\underline{C}(\sigma_{nc}) = z - \underline{\underline{\alpha}}_n(\sigma_{nc})$. □

Proof of Proposition 5.

Proof. Notice the following: $\underline{\gamma} = \theta^I(1 - \beta)$ thus denote $\underline{\beta} := max\{0, 1 - \frac{\gamma}{\theta^I}\}$ and similarly since $\overline{\gamma} = \theta^S(1 - \beta)$ denote $\overline{\beta} := max\{0, 1 - \frac{\gamma}{\theta^S}\}$.

If $\beta \geq \overline{\beta}$ then $\sigma^* = \sigma^S$ and recall that σ^S solves equation (9) and since $\theta^S > \theta^L$ and since $F(\theta/\theta)$ is maximized at $\theta = \theta_L$, an increase in β is associated with a weak increase in θ^S, i.e. a weak decrease in σ^S.

[27] To provide some intuition, notice that the slope of $l_F(\theta(\emptyset, \sigma))$ decreases as $\overline{\theta}(\alpha_n, \sigma)$ increases, since $\underline{\theta} = \theta^L$ for any $c < \overline{c}(\theta)^*$, by the concavity of F. Recall too that the leader's ex-ante payoff is evaluated at the prior p.

If $\beta \in [\underline{\beta}, \overline{\beta}]$ then recall that σ^I is pinned down by $\theta^I(1-\beta) = \gamma$. In turn, as β increases θ^I increases and σ^I decreases. To verify that σ^* jumps at $\beta = \underline{\beta}$ notice that $\sigma^* = \sigma^L \forall \beta < \underline{\beta}$ while $\sigma^I \geq \sigma^S > \sigma^L \forall \beta \geq \underline{\beta}$.

It is very straightforward to verify that compliance decreases in β for any $\beta \geq \underline{\beta}$.

If $\gamma \geq \theta^S$ then $\underline{\beta} = \overline{\beta} = 0$. □

Proofs of Section 7

Extension 1: Intrinsic Benefit and Feasibility of Segment-and-Rule

So far $\alpha(\theta_i) = z + \gamma \times \theta_i$. Hereafter we derive necessary conditions on $\alpha(\theta_i)$ for the regime to reach their upper bound payoff by segmenting and setting $\sigma^* = \sigma^S$.

Given some $\alpha(\theta_i)$ and $\sigma^* = \sigma^S$ two important cutoffs exists: $\theta(0, \sigma^S)$ and $\theta(\emptyset, \sigma^S)$, and there exists a unique equilibrium informational benefit $b_i(\theta_i, \sigma^S, s_{\mathcal{S}} = 1, \beta)$.

Forall $\theta_i \leq \theta(0, \sigma^S)$, the regime need not prevent access to the foreign media since exposure to $s_{\mathcal{F}}$ does not affect the citizens' compliance; no restriction need be applied to $\alpha(\theta_i)$ for any $\theta_i \leq \theta(0, \sigma^S)$. Next, denote

$$\begin{aligned}
\overline{\delta}^{cc} :=& \quad max_{\theta_i} \; \delta_i\Big(\theta_i \in [\theta(0,\sigma^S), \theta(\emptyset, \sigma^S)], \sigma^S, s_{\mathcal{S}} = 1, \beta\Big) \\
\underline{\delta}^{o} :=& \quad min_{\theta_i} \; \delta_i\Big(\theta_i \in (\theta(\emptyset,\sigma^S), 1], \sigma^S, s_{\mathcal{S}} = 1, \beta\Big)
\end{aligned}$$

and similarly denote $\overline{\theta}^{cc}$ as the (largest) solution to $\delta_i(\theta_i \in [\theta(0,\sigma^S), \theta(\emptyset,\sigma^S)], \sigma^S, s_{\mathcal{S}} = 1, \beta) = \overline{\delta}^{cc}$ and $\underline{\theta}^o$ as the (smallest) solution to $\delta_i(\theta_i \in (\theta(\emptyset,\sigma^S), 1], \sigma^S, s_{\mathcal{S}} = 1, \beta) = \underline{\delta}^o$.

Lemma A.14. *For any $\alpha(\theta_i)$ s.t. $\alpha(\theta_i)$ s.t. $\overline{\delta}^{cc} \leq \underline{\delta}^o$ the regime sets $\sigma^* = \sigma^S$ and $c^* \in [\tilde{c}(\overline{\theta}^{cc}), \tilde{c}(\underline{\theta}^o)]$.*

Proof. Given some $\alpha(\theta_i)$ s.t. $\overline{\delta}^{cc} \leq \underline{\delta}^o$ there exists a range of costs $c \in [\tilde{c}(\overline{\theta}^{cc}), \tilde{c}(\underline{\theta}^o)]$ that ensure that $\theta_i \in [\theta(0,\sigma^S), \theta(\emptyset,\sigma^S)]$ do not gain access while $\theta_i \in (\theta(\emptyset,\sigma^S), 1]$, *given* $\sigma^* = \sigma^S$. The regime is indifferent between any of these costs, and reaches their upper bound payoff. □

Extension 2: Propaganda *Creates* Sorting

In this extension, we entertain the possibility that the regime may *directly* make the consumption of foreign content political by making the state-media parrot the party line. That is, γ is no longer a primitive but rather an increasing function of σ. Formally, the relative entertainment benefit is now given by

$$\alpha_i(\theta_i, \sigma) = z + \theta_i \underbrace{(\eta + \rho\sigma)}_{:=\gamma(\sigma)}$$

For simplicity we assume that $\eta = 0$ and focus on the effect of making the state media less informative, that is, ρ. We assume that $\rho > 0$.

Lemma A.15. *There exists a unique $\overline{\rho} \in (0, 1)$ and $\underline{\rho} \in (0, \overline{\rho})$ s.t. in the unique equilibrium*

- *if $\rho > \overline{\rho}$ then $\sigma^* = \sigma^S$ and $c^* = \tilde{c}(\theta^S)$ and $\gamma(\sigma^*) > \overline{\gamma}$*
- *if $\rho \in [\underline{\rho}, \overline{\rho}]$ then $\sigma^* = \sigma^C \in [\sigma^S, max\{\sigma^I\}]$ and $c^* = \tilde{c}(\theta^C)$ and $\gamma(\sigma^*) \in [\underline{\gamma}, \overline{\gamma}]$*
- *if $\rho < \underline{\rho}$ then $\sigma^* = \sigma^L$ and $c^* = \tilde{c}(\theta^L)$ and $\gamma(\sigma^*) < \underline{\gamma}$*

Proof. Recall that $\theta^S = \frac{p}{p+(1-p)\sigma^S}$. Then suppose that

$$\rho\sigma^S \geq \theta^S(1-\beta) \iff \rho \geq \frac{p}{p + (1-p)\sigma^S}\frac{(1-\beta)}{\sigma^S} := \overline{\rho}$$

Then the regime picks $\sigma^* = \sigma^S$ and $c^* = \tilde{c}(\theta^S)$ and achieve its maximal payoff.

Next, suppose that $\rho < \overline{\rho}$. Recall that $max\{\sigma^I\}$ is such that $V(max\{\sigma^I\}, \tilde{c}(\theta^I)) = V(\sigma^L, \overline{c}(\theta^L))$.

Thus the regime may be able to pick $\sigma \in (\sigma^S, max\{\sigma^I\}]$ such that $\rho\sigma \geq \theta(1-\beta)$. Denote by σ^C the unique solution to $\rho\sigma^C = \theta^C(1-\beta)$. Then, if $\sigma^C \in (\sigma^S, max\{\sigma^I\}]$ the regime picks $\sigma^* = \sigma^C$. Denote the target citizen associated with σ^C by θ^C. In turn, segment-and-rule takes place iff

$$\rho \geq \frac{\theta^C}{\sigma^C}(1-\beta) \iff \rho \geq \frac{p}{p + (1-p)\sigma^C}\frac{(1-\beta)}{\sigma^S} := \underline{\rho}$$

If $\rho < \underline{\rho}$ then $\sigma^* = \sigma^L$ and $c^* = \overline{c}(\theta^L)$.

□

Extension 3: Domestic Segmentation

Suppose that the leader can control the reporting slant of two domestic outlets and suppose that there is no foreign outlet. Each citizen must consume one outlet and gains access to its signal. An outlet $i \in \{1,2\}$ reporting strategy is given by a pair (π_i, σ_i) with $\pi_i = Pr(s_i = 1|\omega = 1)$.

Lemma A.16.

- *If the regime can pick any experiment for both outlets ($(\pi_i, \sigma_i) \in [0,1]^2$ $\forall i \in \{1,2\}$) then the regime can replicate his segment-and-rule maximal payoff through domestic segmentation with $\sigma_1^* = \sigma^S, \pi_1^* = 1, \sigma_2^* = 0, \pi_2^* = 1 - (\sigma^S)^2$.*
- *If the regime is constrained to "regime-credible" experiments ($\pi_i = 1, \sigma_i \in [0,1]$ $\forall i\{1,2\}]$) then the regime can do no better than his minimal payoff of full censorship and no domestic segmentation is observed.*

Proof. We assume that when indifferent between two outlets a citizen consumes the most informative outlet.

Case 1: suppose that $\pi_1 = \pi_2 = 1$ *and* $0 \leq \sigma_2 < \sigma_1 \leq 1$. There exists a unique $\theta_1 = \mu(s_1 = 1) < \theta_2 = \mu(s_2 = 1)$.

- $\theta_i < \theta_2$: payoff from consuming the second outlet is $p + (1-p)(1-\sigma_2)\theta_i$ and from consuming the first outlet it is $p + (1-p)(1-\sigma_1)\theta_i$ (for $\theta_i < \theta_1$) or θ_i (for $\theta_i \in [\theta_1, \theta_2]$); i.e. all consume the most informative outlet.
- $\theta_i > \theta_2$: either way the citizen *never* complies. Thus she is indifferent between either outlet and consumes the most informative outlet.

Hence, all citizens (weakly) go for the most informative outlet and it is without loss to restrict attention to a single domestic outlet.

Case 2: suppose that $\pi_1 = 1, \pi_2 \in [0,1)$ *and* $\sigma_1 \in (0,1), \sigma_2 = 0$. There exists a unique $\theta_1 = \mu(s_1 = 1) > \theta_2 = \mu(s_2 = 0)$.

- $\theta_i < \theta_2$ get p from consuming the second outlet and $p + (1-p)(1-\sigma_1)\theta_i$ from consuming the first outlet; i.e. all consume the first outlet. Note that this interval is empty if $\pi_2 = 1$.

- $\theta_i \in [\theta_2, \theta_1]$: there exists a unique $\theta^* = \frac{p(1-\pi)}{p(1-\pi)+(1-p)\sigma} \in (\theta_2, \theta_1)$ such that a citizen consumes the first outlet iff $\theta_i \leq \theta^*$.

- $\theta_i > \theta_1$ get θ_i from the first outlet and $p\pi + [1-p\pi]\theta_i > \theta_i$; i.e. all consume the second outlet.

Then notice that if the regime can pick any $\sigma_1 \in (0,1)$ and $\pi_2 \in (0,1)$ they can set $\sigma_1^* = \sigma^S$ and $\pi_2^* = 1-(\sigma^S)^2$ s.t. $\theta^* = \theta^S$ so that they retrieve their payoff from segment-and-rule in the strong association case ($\gamma \geq \overline{\gamma}$) of the baseline game. Crucially however, this requires the regime to be able to credibly commit to one of its own outlets reporting negatively on the regime: $\sigma_2^* = 0, \pi_2^* < 1$. $\square$

Extension 4: Non-Binary Compliance

Let us now assume that $a_i \in [0,1]$. Instead of solving explicitly for a modified version of the game, we fix a reporting slant σ and target citizen θ, and aim to derive under which conditions segment-and-rule dominates full censorship, given some equilibrium compliance profile $a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta))$. We assume such an equilibrium compliance profile exists and make the following assumptions about it:

1. $a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta)) = \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta)$ if $\mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta) = 0$ or 1.
   All (respectively no) citizens provides full compliance (respectively zero compliance) conditional on perfectly knowing that $\omega = 1$ (respectively $\omega = 0$).

2. $\frac{\partial a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta))}{\partial \theta_i} \leq 0$.
   The more ex-ante misaligned with the regime a citizen, the lower her equilibrium compliance level.

3. $\frac{\partial a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta))}{\partial \mu} \geq 0$.
   The higher a citizen's belief about ω, the higher her equilibrium compliance level.

4. $\frac{\partial^2 a_i^*(\theta_i, \mu(s_{\mathcal{S}}, \hat{s}_{\mathcal{F}}|\sigma, \beta))}{\partial \mu \partial \theta_i} \leq 0$ and 0 at $\theta_i = 1$.
   The more misaligned a citizen is with the regime, the smaller the marginal effect of a higher posterior on ω.

We focus on the behavior of regime opponents ($\theta_i > \theta$) given some reporting slant σ, and conditional on $s_{\mathcal{S}} = 1$ since otherwise no citizens provide any positive level of compliance. We also assume that segment-and-rule is feasible, e.g., $\gamma > 1-\beta$.

Lemma A.17. *Suppose that assumptions 1. through 4 hold and that* $\gamma > 1-\beta$. *Then, given some* $\sigma \in [0,1]$ *and* $\theta \in [p,1]$, *there exists a unique* $\tilde{\theta} \in [\theta, 1)$ *s.t.* $V(\tilde{c}(\tilde{\theta}), \sigma) > max\{V(\bar{c}(\theta), \sigma), V(\tilde{c}(\theta), \sigma)\}$.

Proof. Then, define the gain from a strategy of segment-and-rule at the individual level by

$$\begin{aligned}\Delta_s(\theta_i) =& p(1-\beta)*1 + [1-p(1-\beta)]a_i^*(\theta_i, \mu(1,0|\sigma,\beta)) - a_i^*(\theta_i, \mu(1,\emptyset|\sigma,\beta)) \\ =& p(1-\beta)(1-a_i^*(\theta_i,\mu(1,\emptyset|\sigma,\beta))) - [1-p(1-\beta)](a_i^*(\theta_i,\mu(1,\emptyset|\sigma,\beta)) - a_i^*(\theta_i,\mu(1,0|\sigma,\beta))).\end{aligned}$$

Note that $\frac{\partial^2 a_i^*(\theta_i,\mu(s_{\mathcal{S}},\hat{s}_{\mathcal{F}}|\sigma,\beta))}{\partial\theta_i\partial\mu} \leq 0$ is sufficient (but not necessary) to ensure that $\frac{\partial\Delta_s(\theta_i)}{\partial\theta_i} > 0$.

Then, if $\Delta_s(\theta) > 0$, then there exists a unique $\tilde{\theta} \in [\theta, 1)$ with the above property. Suppose not. Notice that $\Delta_s(1) \geq 0$ by $\frac{\partial^2 a_i^*(\theta_i,\mu(s_{\mathcal{S}},\hat{s}_{\mathcal{F}}|\sigma,\beta))}{\partial\mu\partial\theta_i}(\theta_i = 1) = 0$.[28] Then, by the intermediate value theorem, there exists a unique $\tilde{\theta} \in [\theta, 1)$ s.t. $\Delta_s(\theta_i) \geq 0 \iff \theta_i \geq \tilde{\theta}$. Then, fixing some σ, $V(\tilde{c}(\tilde{\theta}), \sigma) > max\{V(\bar{c}(\theta), \sigma), V(\tilde{c}(\theta), \sigma)\}$. That is, as long as $\Delta_s(\theta) > 0$ – which is always the case when assumptions 1 through 4 hold – the regime can always, fixing a reporting slant σ and target citizen θ, adjusts the cost of access such that a strategy of segment-and-rule dominates a strategy of full censorship. □

[28]Since this assumption implies that $a^*(\theta_i = 1)$ is independent of $\mu(\cdot)$.

IEB Institut d'Economia de Barcelona