

Segment-and-Rule: Modern Censorship in Authoritarian Regimes

Kun Heo (NYU), Antoine Zerbini (LSE)

EEA Barcelona 2023

August 28, 2023

A Timeless Authoritarian Concern



“Politicians must run the newspapers.”

Mao Zedong (1957), Xi Jinping (2016).

Changing Times

1994: Internet roll-out in China

→ citizens gain agency: possibility to choose whether to gain access to foreign outlets

→ internet is a *liberating technology* (Diamond, 2010)

“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.

Authoritarian Response: The Firewall

Firewall: make online access to foreign outlets costly (download a vpn)

But millions bypass the firewall everyday

China: estimated that 5% percent of urban residents actively circumvent internet censorship (Roberts 2017, Chen and Yang 2019), similar evidence in Iran (Dal and Nisbet 2022)

→ Weakened ability to control information flows?

Authoritarian Response: The Firewall

Firewall: make online access to foreign outlets costly (download a vpn)

But millions bypass the firewall everyday

China: estimated that 5% percent of urban residents actively circumvent internet censorship (Roberts 2017, Chen and Yang 2019), similar evidence in Iran (Dal and Nisbet 2022)

→ Weakened ability to control information flows?

This paper: **selective resilience to censorship is a feature of modern censorship, not a bug!**

Modern Censorship

Modern censorship is **selective**

- ▶ **regime supporters** consume local outlets: convinced by government propaganda to comply
- ▶ **regime opponents** bypass the firewall: occasionally comply after positive reporting from a banned foreign outlet

→ a strategy of **segment-and-rule**

Segment-and-Rule

Segment-and-rule requires **entertainment to be consumed along political lines**

→ this can be **engineered** by the regime, by

- ▶ making the domestic outlets parrot the party line
- ▶ investing in domestic entertainment
- ▶ strategically banning foreign entertainment

Related Literature

Propaganda and Censorship

- ▶ Trade-offs inherent to censorship: Egorov et al (2009), Lorentzen (2014), Kronick and Marshall (2022), Edmond (2013)
- ▶ which information (states of the world) to censor/disclose. Kamenica and Gentzkow (2011), Adena et al (2013), Bernhardt and Shadmerh (2015), Gehlbach and Sonin (2014), Little (2018), Heo and Zerbini (2023) Gitmez and Molavi (2022, Kolotilin (2018,2022), Li and Zhou (2022), Gitmez, Sonin (2022)

Other levers of top-down accountability

- ▶ Divide-and-rule: Acemoglu et al (2004), Luo and Rozenas (2023)
- ▶ violence: Montagnes and Wolton (2019), Rozenas (2020)
- ▶ economic incentives: Magaloni (2006), Treisman (2011)
- ▶ selective incentives: De Mesquita et al (2005), Gandhi (2008), Blaydes (2010).

Persuasion of voting-bodies Caillaud and Tirole (2007), Alonso and Camara (2016), Awad (2020)

Setup

(in)Formal Framework

An **authoritarian leader** maximizes compliance in a heterogeneous citizenry and

1. controls **local information flows**

→ *government* media

2. can dissuade access to **non-controlled information flows**:
the *firewall*

→ *foreign* media

(in)Formal Framework

Citizens choose

1. whether to **access non-state controlled content**, which
 - ▶ provides some information and
 - ▶ a non-informational - e.g., entertainment - benefit
 - ▶ may be **associated** with a citizen's political type
 - ▶ is costly: the firewall

(in)Formal Framework

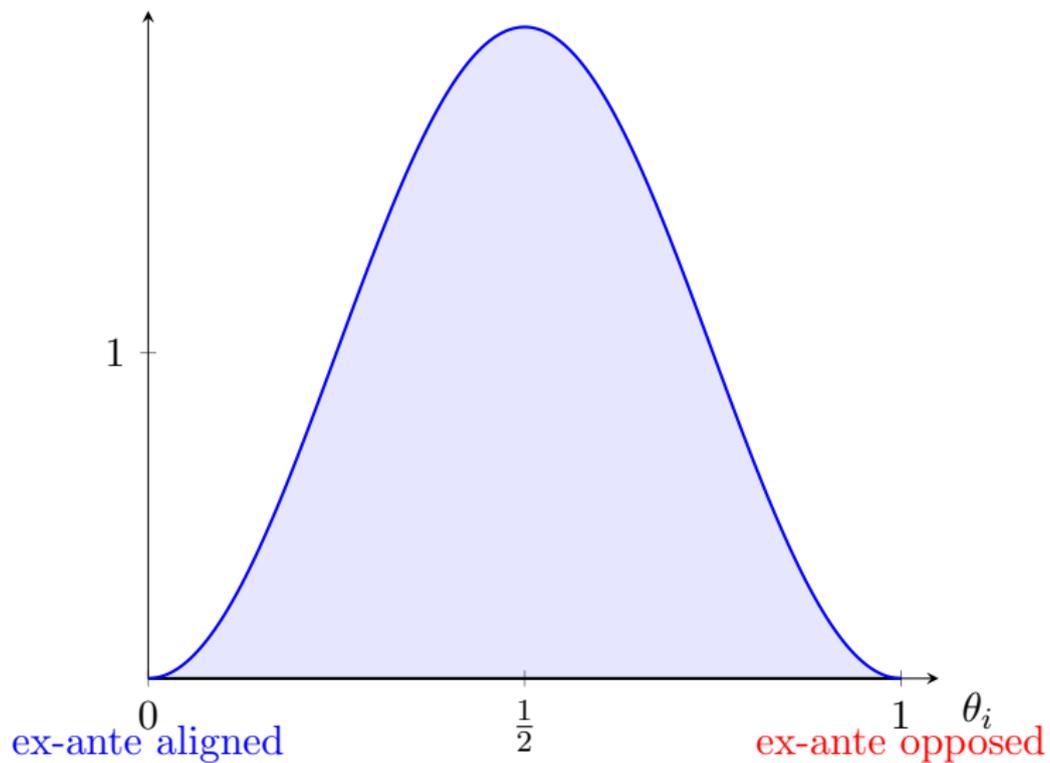
Citizens choose

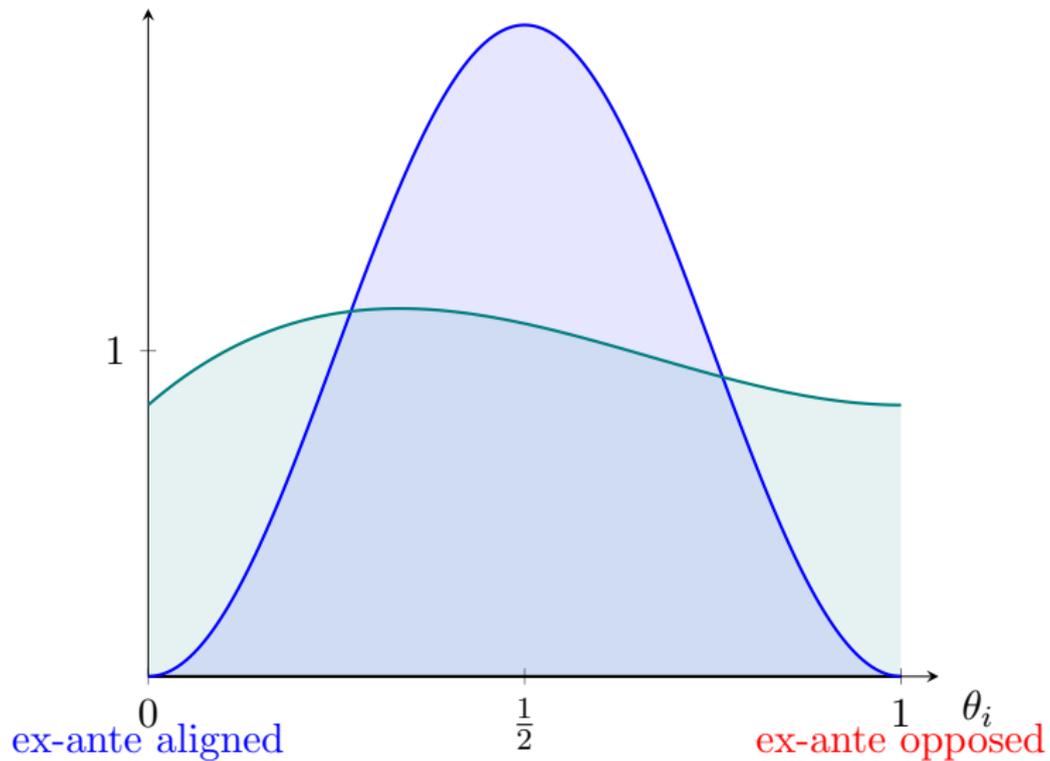
1. whether to **access non-state controlled content**, which
 - ▶ provides some information and
 - ▶ a non-informational - e.g., entertainment - benefit
 - ▶ may be **associated** with a citizen's political type
 - ▶ is costly: the firewall
2. whether to **comply** with the regime
 - ▶ **complying** ($a = 1$) yields
 - ▶ 1 if $\omega = 1$
 - ▶ 0 if $\omega = 0$

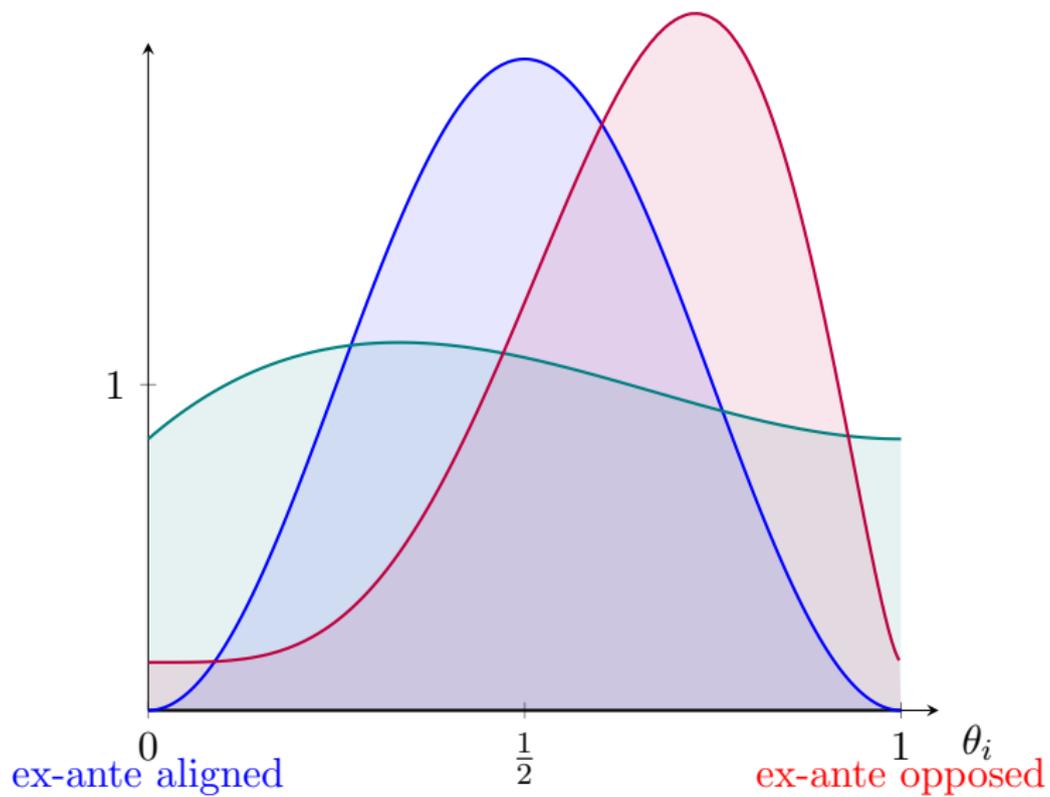
(in)Formal Framework

Citizens choose

1. whether to **access non-state controlled content**, which
 - ▶ provides some information and
 - ▶ a non-informational - e.g., entertainment - benefit
 - ▶ may be **associated** with a citizen's political type
 - ▶ is costly: the firewall
2. whether to **comply** with the regime
 - ▶ **complying** ($a = 1$) yields
 - ▶ 1 if $\omega = 1$
 - ▶ 0 if $\omega = 0$
 - ▶ **not complying** ($a = 0$) yields **her political type**, $\theta_i \in [0, 1]$.
 θ_i is privately observed and distributed according to unimodal and log-concave f .







Timing

1. The leader
 - 1.1 publicly commits to the reporting slant of the government media σ .
 - 1.2 picks the cost of access c
2. Nature determines ω and privately reveals θ_i to citizen i .
3. Nature generates the government media's report s_G as well as the foreign media's report s_F .

Each citizen, having observed s_G decides whether to consume the foreign media; if they do, they observe s_F .
4. The citizens choose whether to comply with the regime. Game ends. Payoffs are realized.

Solution concept: wPBE.

Analysis: Preliminary

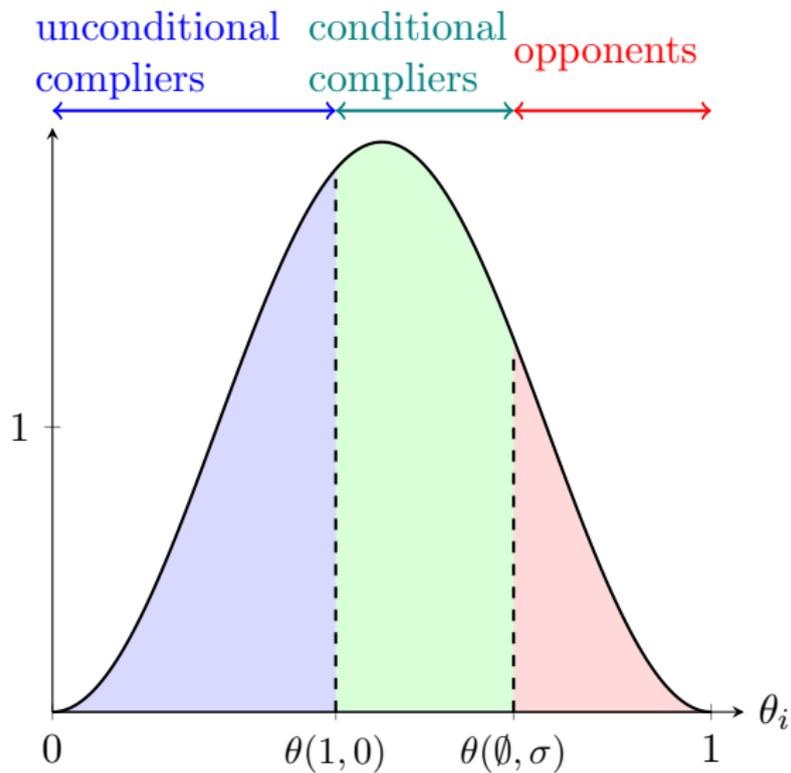
Preliminary

Bad news from the government media: no one complies.

Good news from the foreign media: everyone complies.

Otherwise: only citizens sufficiently aligned with the regime comply

Given some reporting slant σ ,



Censorship trade-off

Fully prevent access to foreign media \implies

- ▶ conditional compliers comply
- ▶ opponents never comply

→ ideally segment access to the foreign media

Three cases to consider, delineated by the **strength of the correlation** between **politics** and **entertainment**: $\gamma \geq 0$

Analysis: Low correlation

Case 1: Low correlation

Proposition 1

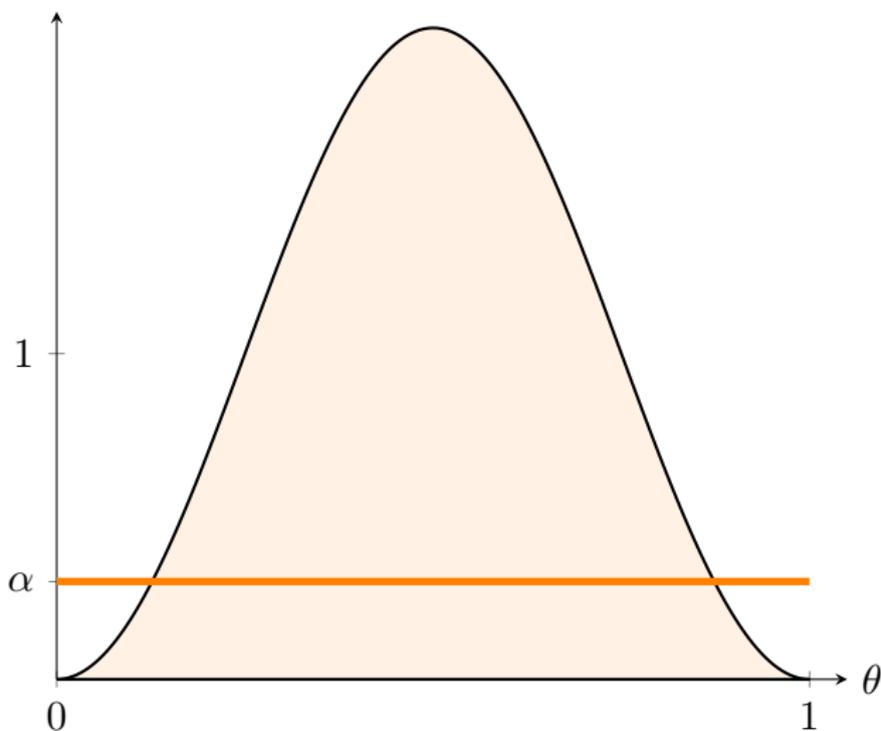
Entertainment is not consumed along political lines ($\gamma < \underline{\gamma}$)

\implies

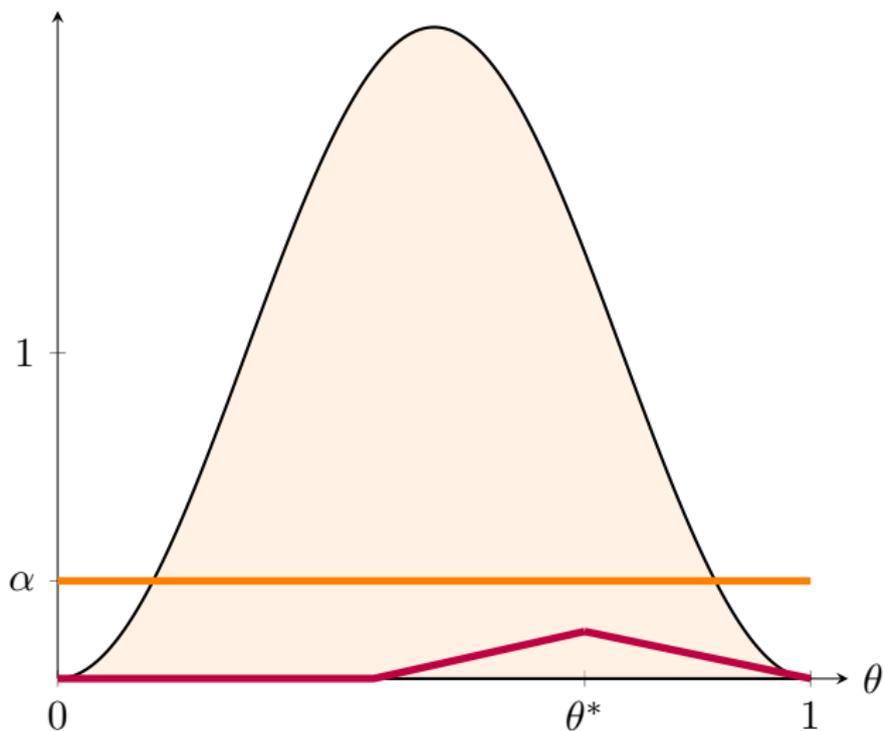
1. **full censorship**: no citizen bypasses the firewall
($c^* = \bar{c}(\theta^*)$)
2. there exists a unique reporting strategy σ^* and target citizen θ^* .

Suppose $\gamma = 0$ for illustration.

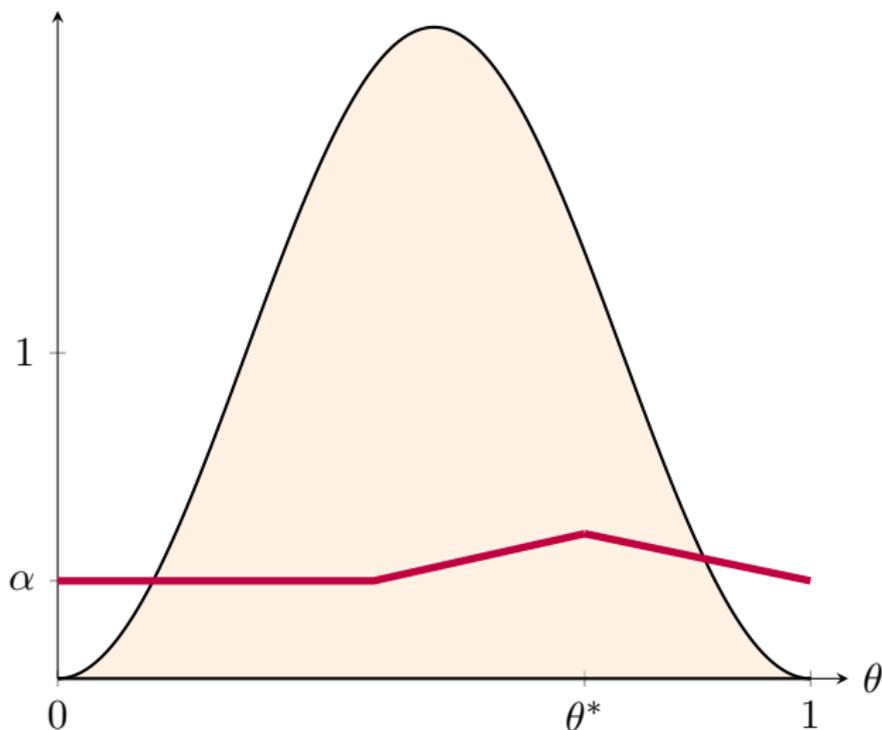
Net benefit: $\delta_i(\theta_i, \sigma, s_{\mathcal{G}}, \beta) =$ $\underbrace{\alpha}_{\text{non-info benefit}}$



$$\text{Net benefit: } \delta_i(\theta_i, \sigma, s_G, \beta) = \underbrace{\alpha}_{\text{non-info benefit}} + \underbrace{b_i(\theta_i, \sigma, s_G, \beta)}_{\text{info benefit}}$$

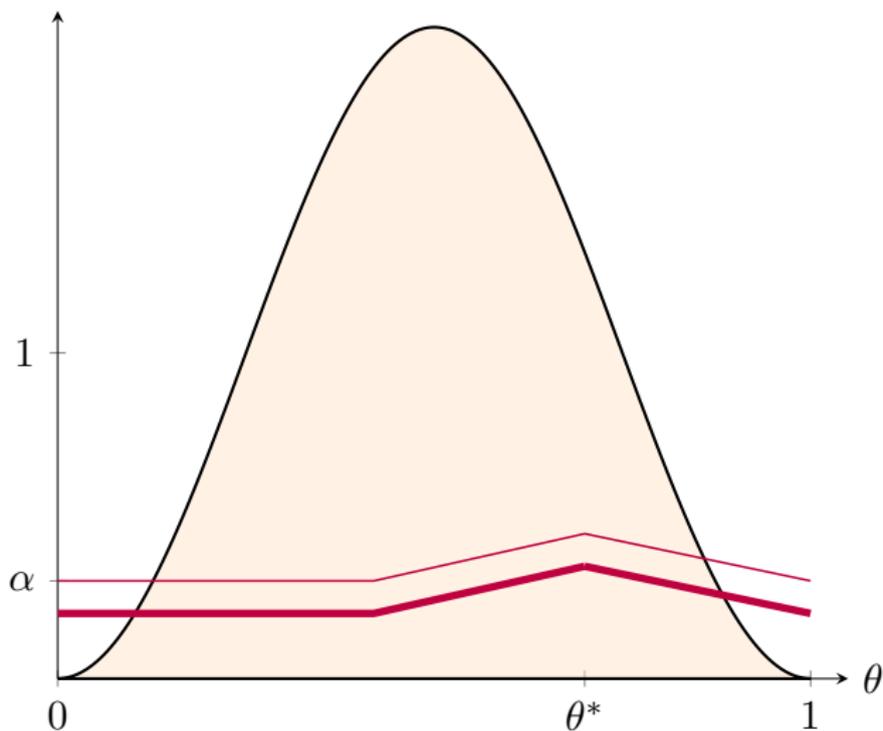


$$\text{Net benefit: } \delta_i(\theta_i, \sigma, s_G, \beta) = \underbrace{\alpha}_{\text{non-info benefit}} + \underbrace{b_i(\theta_i, \sigma), s_G, \beta)}_{\text{info-benefit}}$$

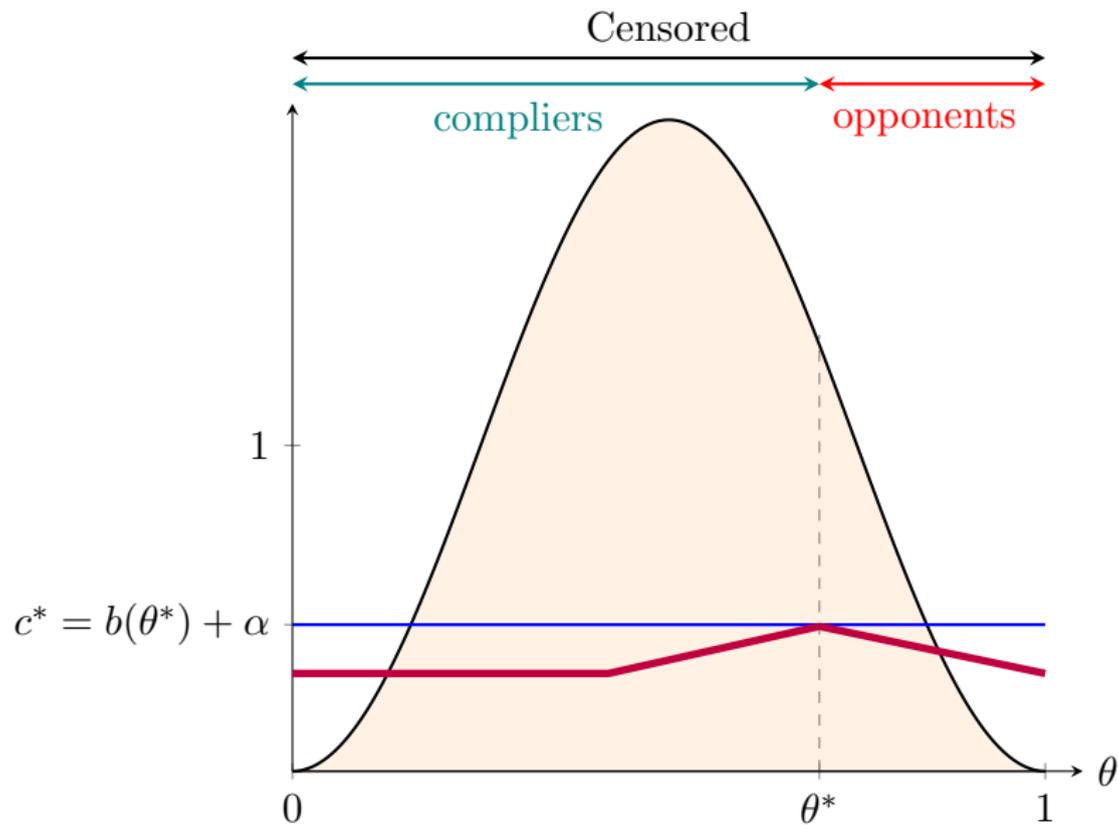


\implies equilibrium target citizen has the highest WTP; δ_i is **single-peaked**

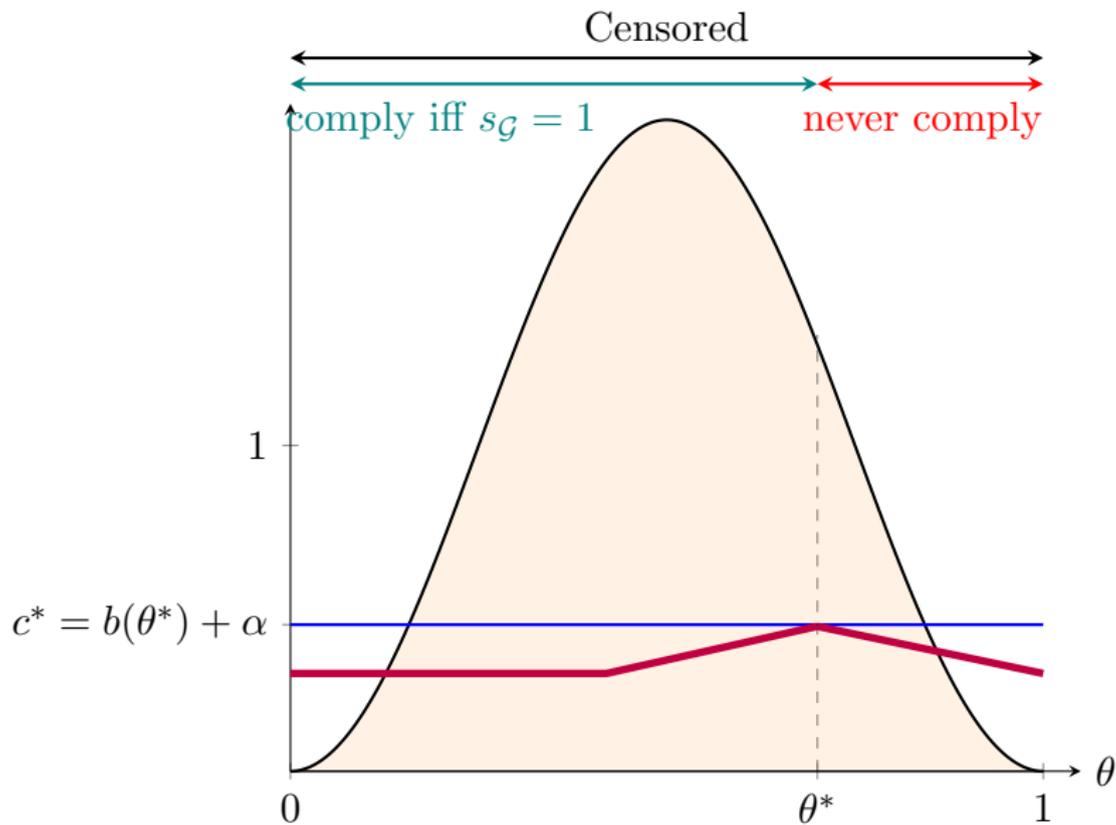
$$\delta_i(\theta_i, \sigma, s_G, \beta) = \underbrace{\alpha}_{\text{non-info benefit}} + \underbrace{b_i(\theta_i, \sigma, s_G, \beta)}_{\text{info-benefit}} - \underbrace{c}_{\text{cost of access}}$$



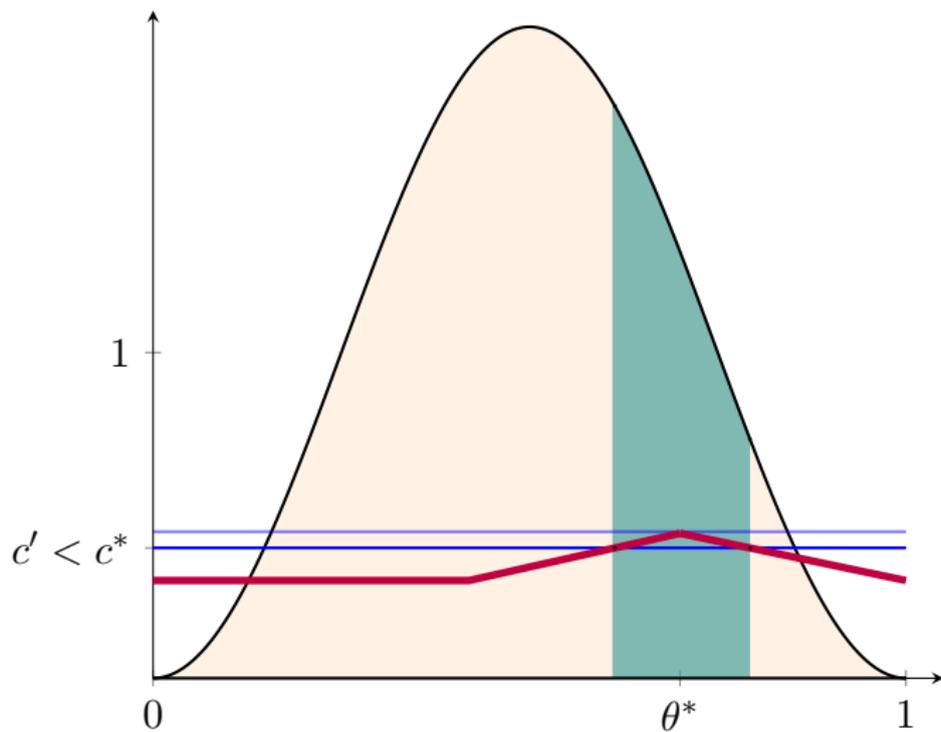
In equilibrium: low correlation \implies full censorship.



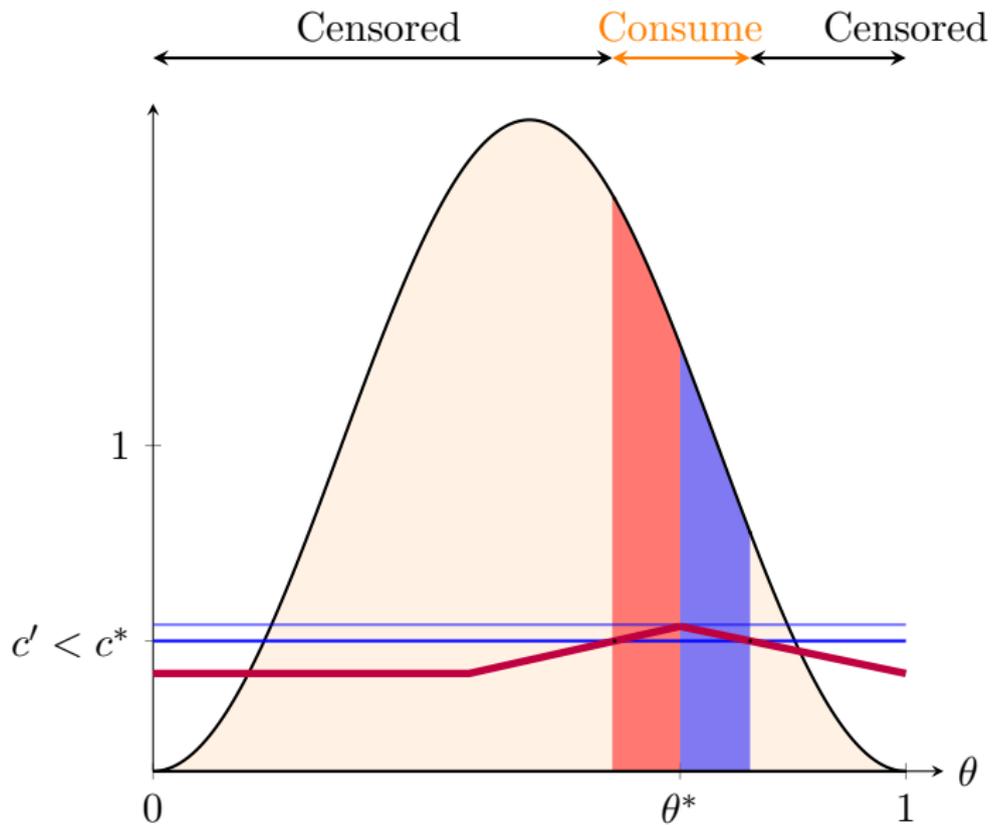
In equilibrium: low correlation \implies full censorship.



Suppose profitable deviation to c' , such that segmentation takes place



some opponents consume, but more compliers do too ...



Analysis: Strong correlation

Entertainment is Political

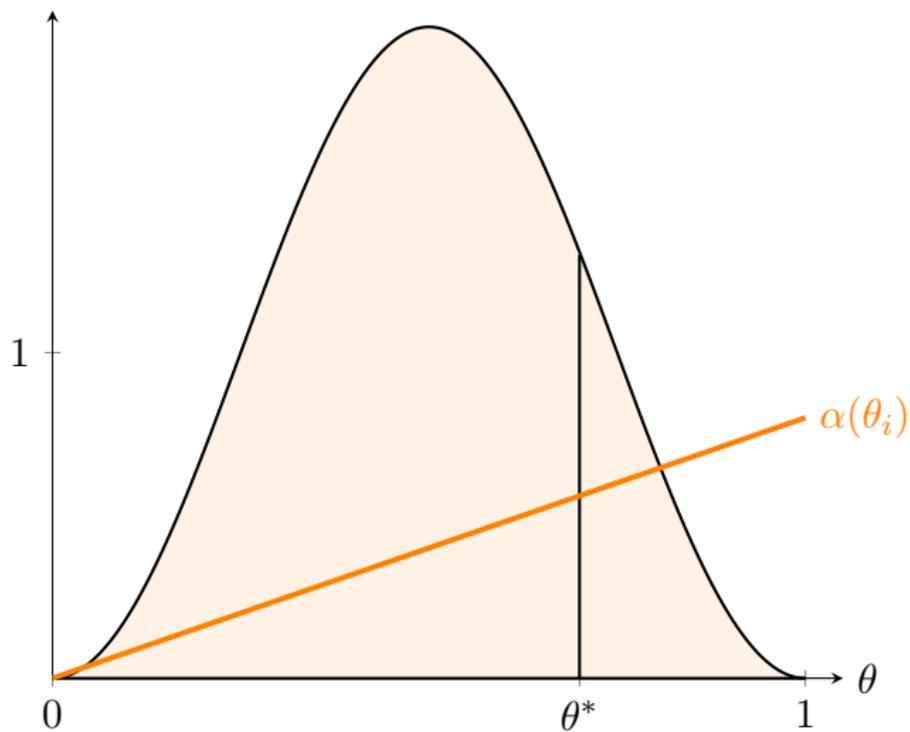
Proposition 2

Entertainment is political \implies strategy of **segment-and-rule**

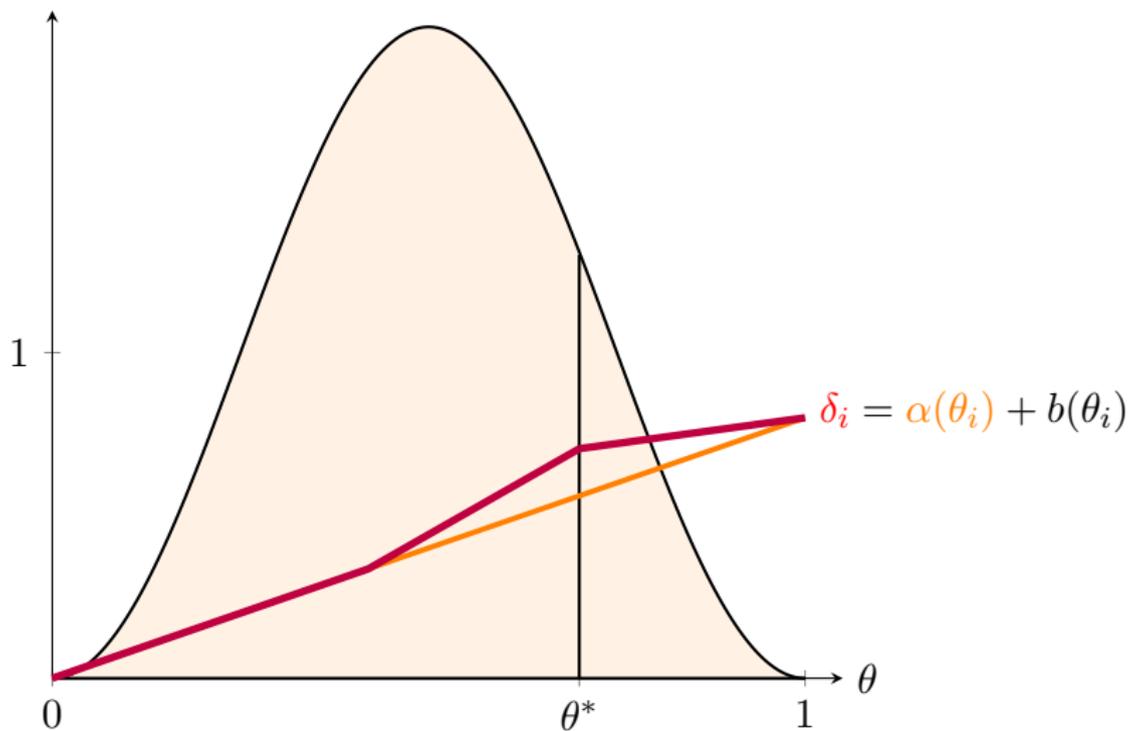
1. **selective censorship**: only opponents ($\theta_i > \theta_w^*$) bypass the firewall ($c^* = \tilde{c}(\theta_w^*)$)
2. the government media is less informative (wrt weak correlation and full censorship)

$$\underbrace{\theta_w^* = \theta_w^s < \theta^*}_{\text{lower target citizen}}, \quad \underbrace{\sigma_w^* = \sigma_w^s > \sigma^*}_{\text{less informative gov media}}$$

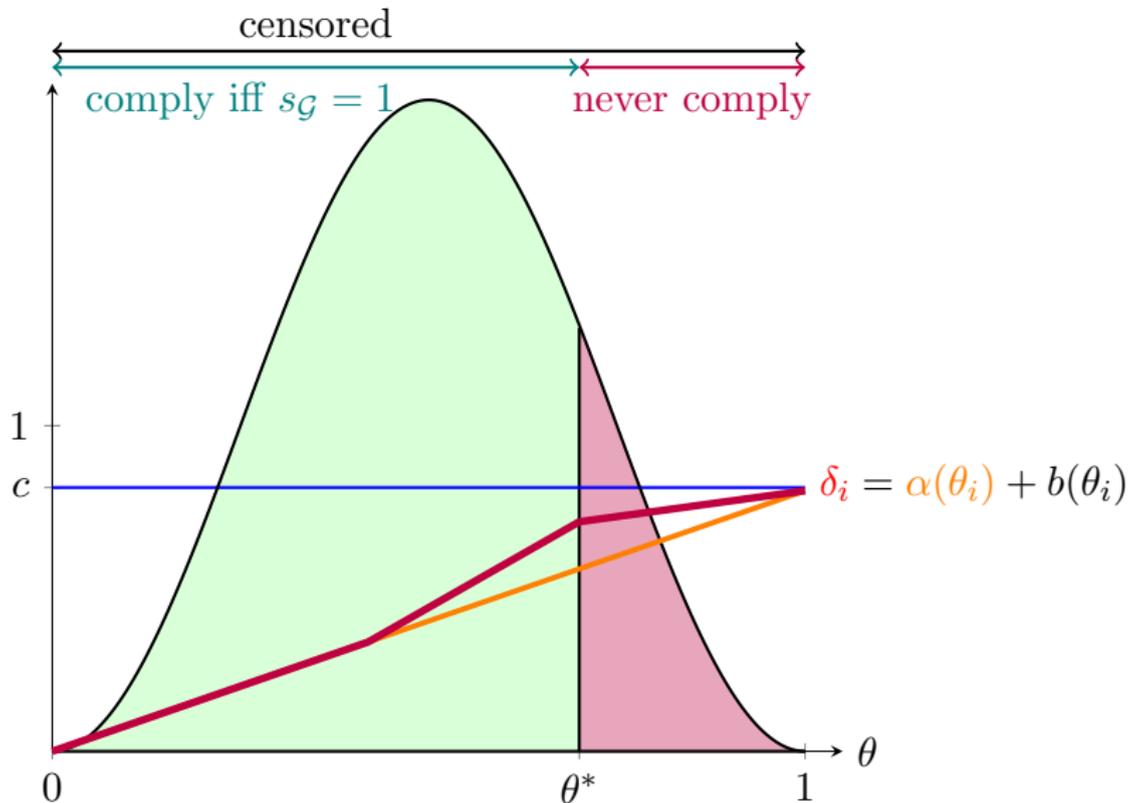
3. **compliance is higher** than under low correlation (and full censorship)



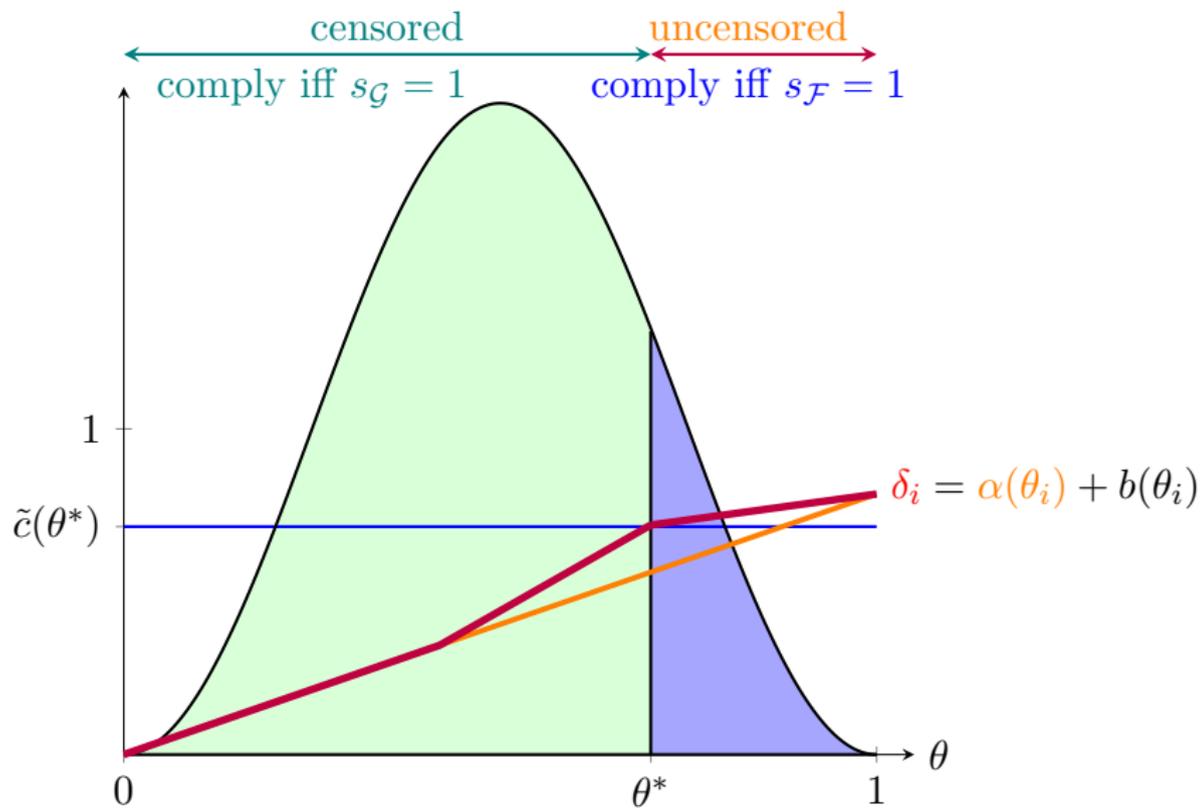
Entertainment benefit **increases** in a citizen's type



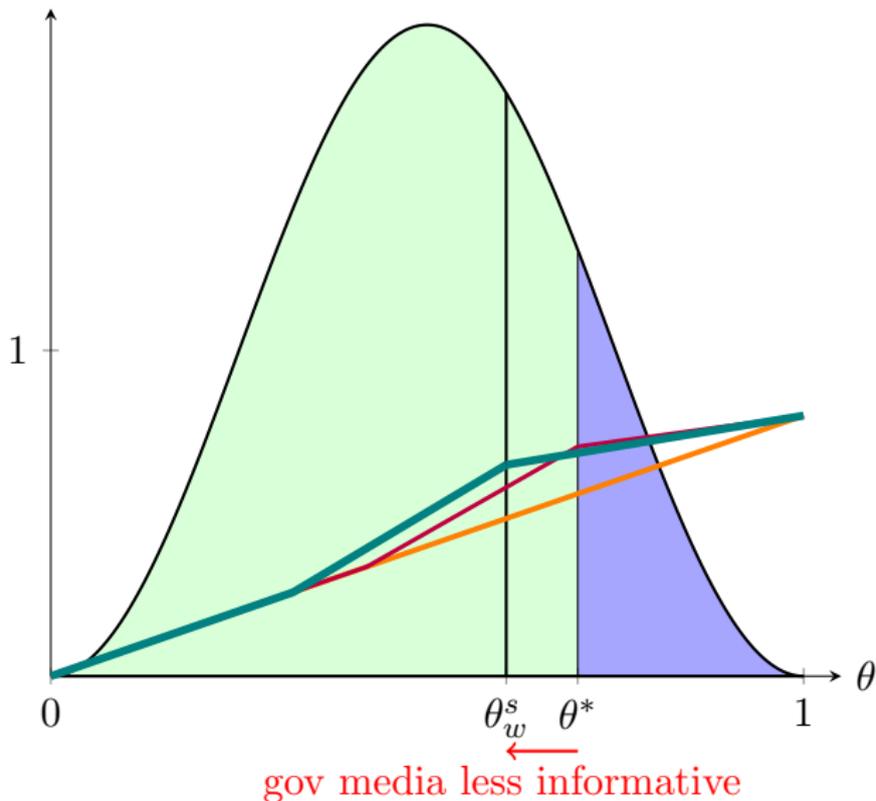
and the **total** benefit also increases in the citizen's type



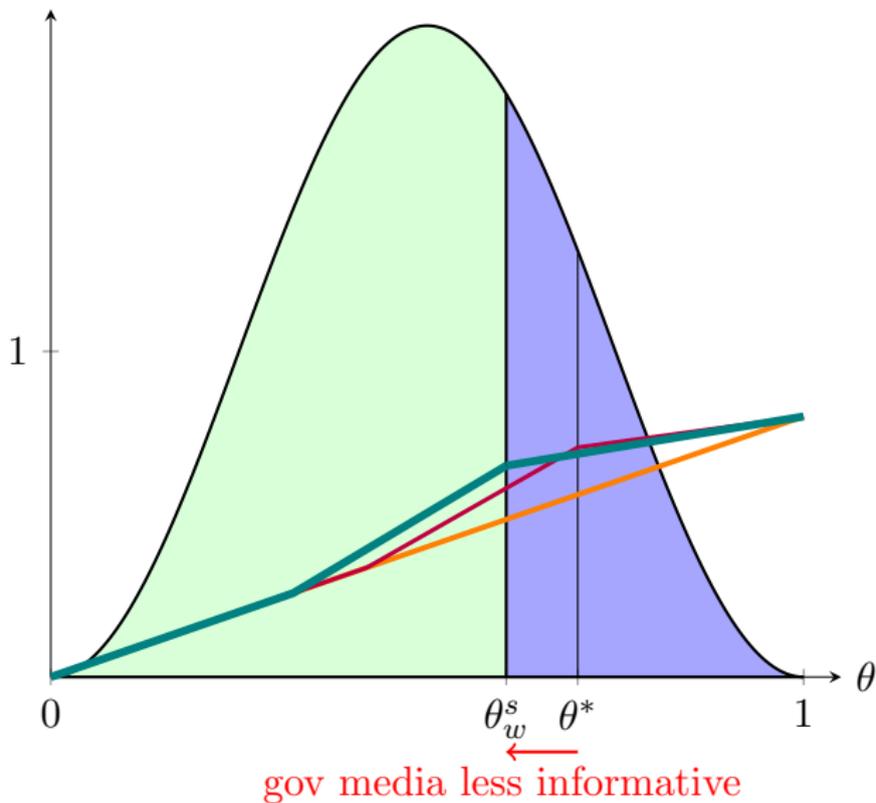
Full censorship compliance level *can* be replicated but...



Fixing propaganda, can improve on full censorship with partial censorship!



Partial censorship \implies opponents are never fully lost \implies different trade-off vis-a-vis full censorship



Speak differently to the base and push opponents towards consumption.

Analysis: Intermediate Correlation

Case 3: Engineering Segment-and-Rule

If entertainment is only somewhat political

⇒ regime depletes government media of information content

⇒ incentivize opponents to consume foreign content

⇒ *engineer* segmentation to make segment-and-rule possible

Proposition 3

If $\gamma \in (\underline{\gamma}, \bar{\gamma})$ then

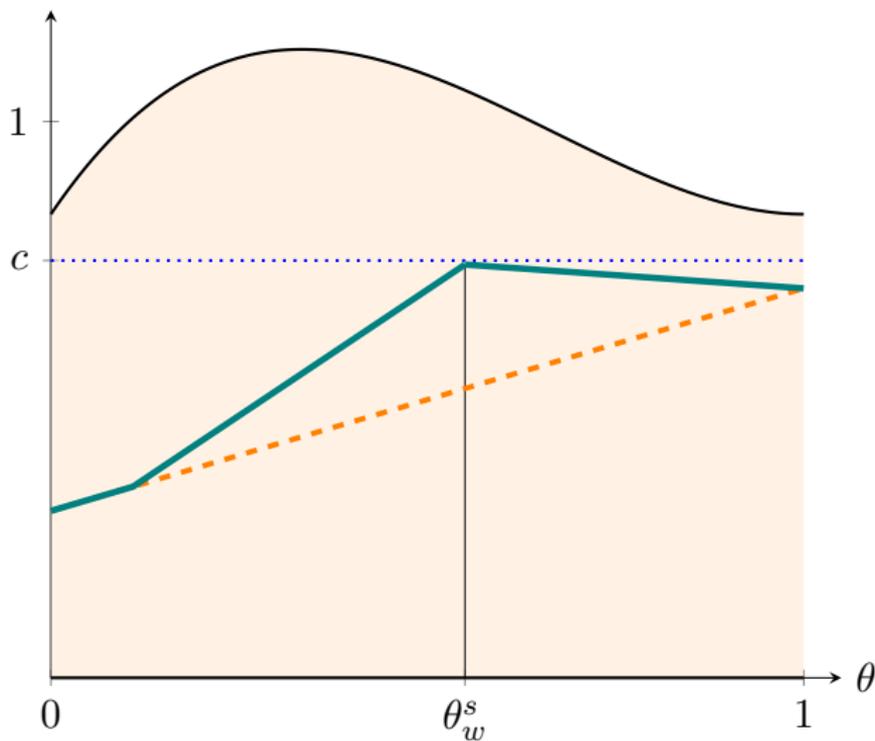
1. the **gov media is least informative** (wrt weak and strong correlation)

$$\theta_w^* = \theta_w(c) < \theta_w^s < \theta^*, \quad \underbrace{\sigma_w^* > \sigma_w^s > \sigma^*}_{\text{least informative reporting}}$$

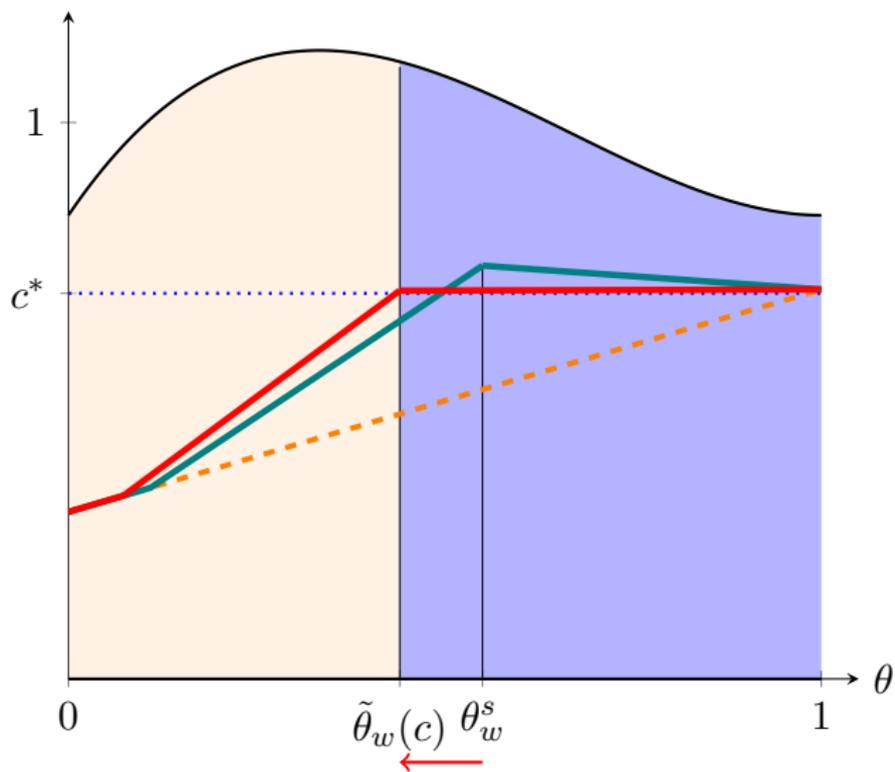
2. the regime **engineers partial censorship** by setting $c^* = \tilde{c}(\theta_w(c))$
3. compliance is bounded between
 - ▶ **lower bound**: payoff from full censorship (low correlation)
 - ▶ **upper bound**: payoff from perfect segmentation (strong correlation)

→ must compromise on propaganda to allow for segmentation

Fixing the optimal informativeness of the gov media: impossible to segment \rightarrow optimal to fully censor



but with a less informative gov media \rightarrow segmentation is possible!



How to Segment

The Segmentation Handbook

1. so far: **Make access costly** → increase the cost of access

A form of “*friction*” (Roberts 2017).

The Segmentation Handbook

1. so far: **Make access costly** → increase the cost of access

A form of “*friction*” (Roberts 2017).

2. **Invest in domestic entertainment** (flooding (Roberts 2017))
 - ▶ improve (common) quality of entertainment: substitute for c (Liu, Yao 2023)
 - ▶ generate content enjoyed heterogenously along political lines
→ e.g. historical (war) movies

The Segmentation Handbook

1. so far: **Make access costly** → increase the cost of access

A form of “*friction*” (Roberts 2017).

2. **Invest in domestic entertainment** (flooding (Roberts 2017))
 - ▶ improve (common) quality of entertainment: substitute for c (Liu, Yao 2023)
 - ▶ generate content enjoyed heterogenously along political lines
→ e.g. historical (war) movies
3. **Strategic ban**: make banned entertainment political → strategic *friction*

High Profile Polarizing Propaganda



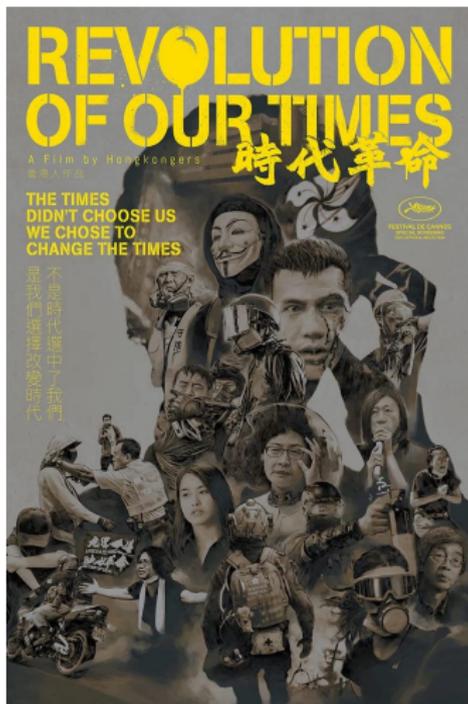
The Knockout

[Details](#)



The Battle of Lake Changjin.

Banned Movies



Political or politicized movie → high correlation → banned

Non-Banned



Non-political (?) → low correlation → approved (and high revenue) [▶ Details](#)

Strategic Bans

(one) interpretation of the “foreign media”: average

- ▶ informational $(1 - \beta)$ and
- ▶ non-informational (z, γ) content

of all banned outlets.

Strategic Bans

(one) interpretation of the “foreign media”: average

- ▶ informational ($1 - \beta$) and
- ▶ non-informational (z, γ) content

of all banned outlets. Consider the set of non-informational outlets ($\beta = 1$) (e.g., cartoons, some movies, sports)

Suppose: if ban all non-state outlets \implies low correlation ($\gamma < \underline{\gamma}$) \iff lower bound compliance payoff

Easy fix: “un-ban” low correlation (low γ) outlets to make circumvention political

Conceptualizing Censorship

We have modelled two layers of censorship

- ▶ how freely information flows locally σ^*

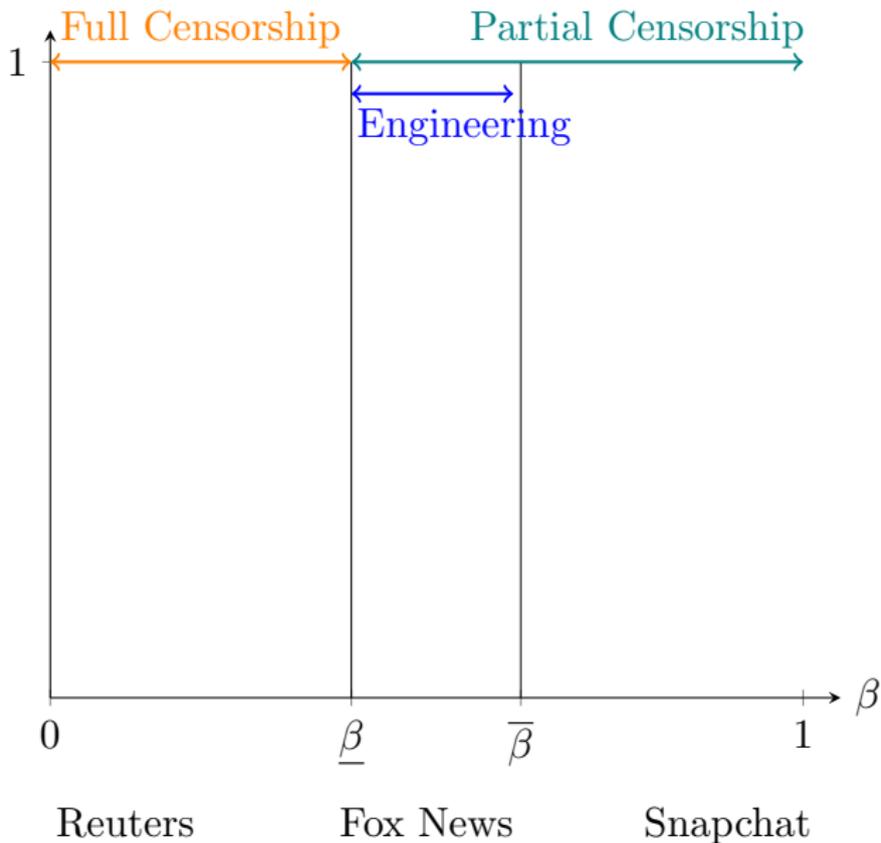
Shadmerh and Bernhard 2011, Gehlbach and Sonin 2014, Kolotilin et al 2022 etc.

- ▶ the share of citizens who gain access to a foreign outlet:
studied empirically

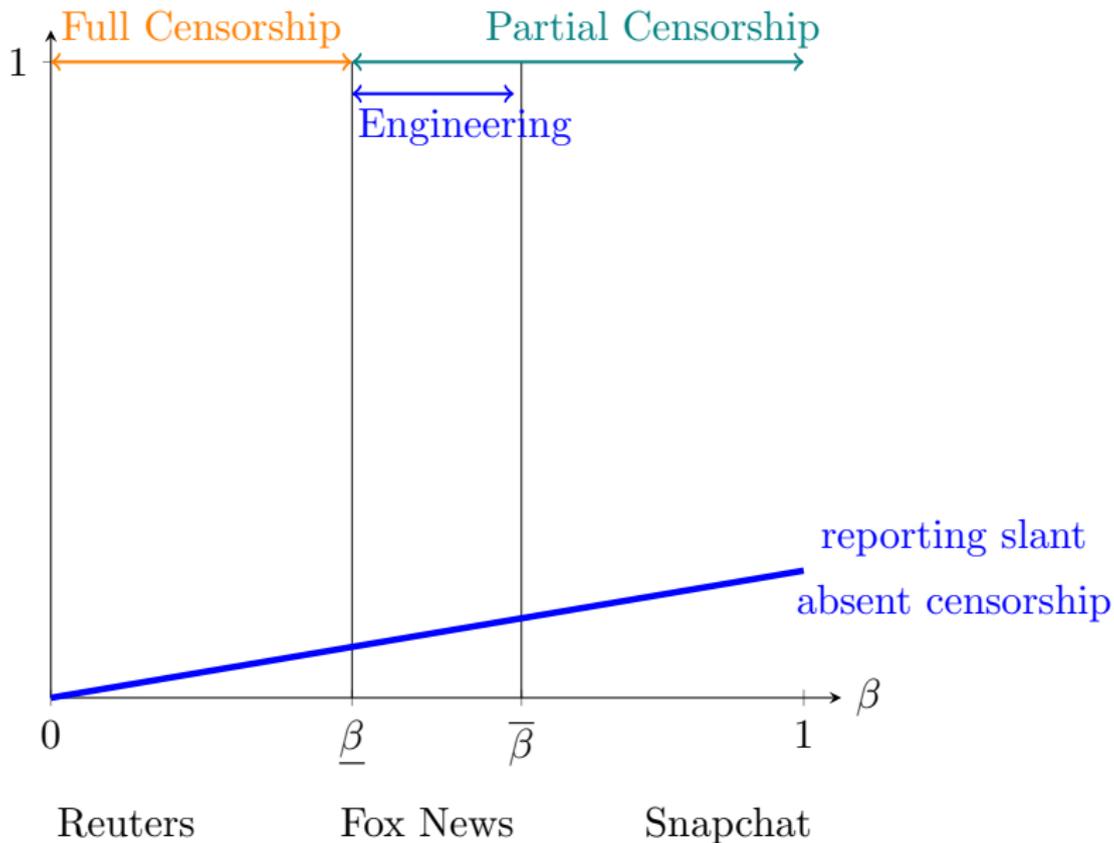
In equilibrium, these **quantities move in opposite directions**

⇒ important asymmetries across citizens in consumption patterns and thus beliefs

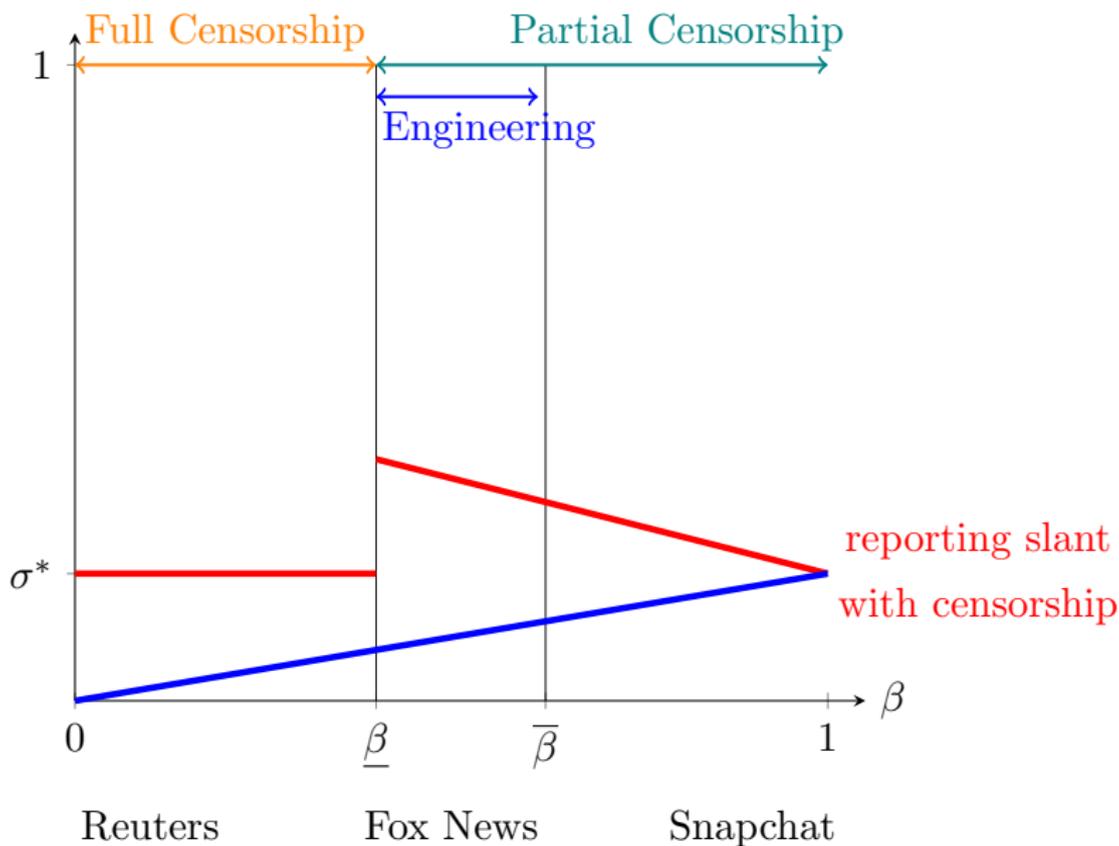
Government Media Informativeness



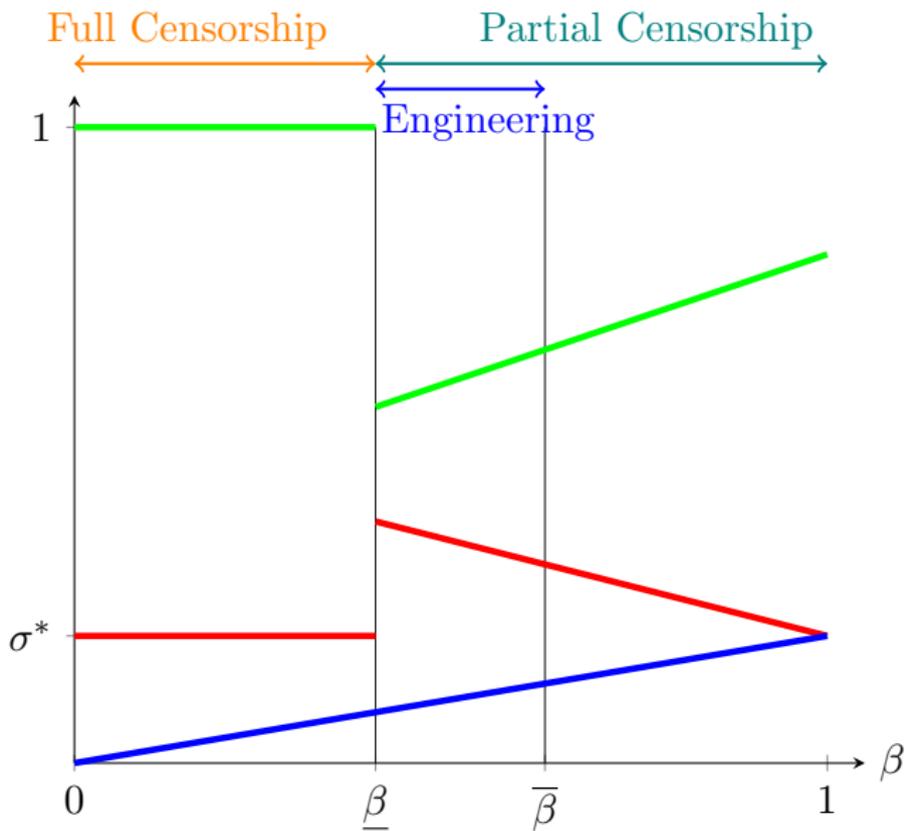
Government Media Informativeness



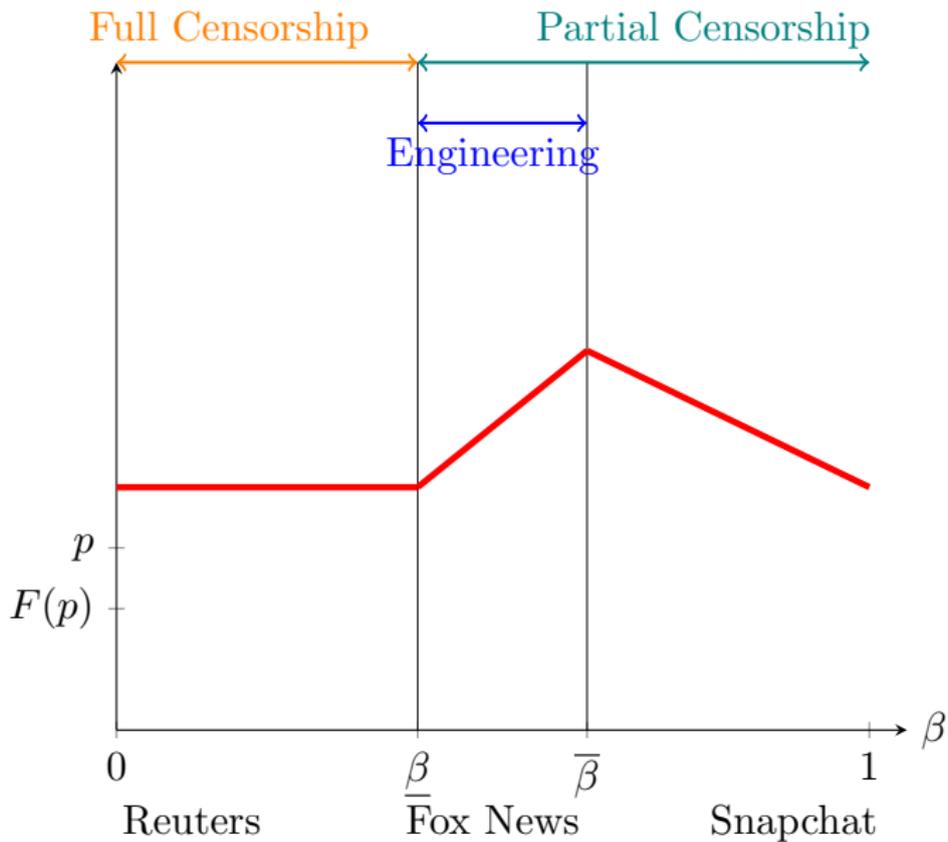
Government Media Informativeness



Government Media Informativeness



Equilibrium Compliance



Conclusion

Modern censorship is selective by design

Segment-and-rule exploits

1. citizen's agency in information acquisition
2. heterogeneity of political preferences

Segment-and-rule requires entertainment to be (made) political:
→ incentives to **generate a cleavage** in entertainment consumption patterns

How to weaken the autocrat?

1. **make dissuasion and segmentation harder**: soft-power propaganda, but only if it appeals to the whole population
→ not polarizing soft-power (e.g., Top Gun, The Interview)
2. **make outlet *extremely* (un)-informative**

Thank you!

bh1875@nyu.edu

a.r.zerbini@lse.ac.uk

The Framework

An **authoritarian leader** maximizes compliance in a heterogeneous citizenry and

1. controls **local information flows**
2. can dissuade access to **non-controlled information flows:**
the *firewall*

The Framework

An **authoritarian leader** maximizes compliance in a heterogeneous citizenry and

1. controls **local information flows**
2. can dissuade access to **non-controlled information flows**:
the *firewall*

Citizens choose

- ▶ whether to **access non-state controlled content**, which
 - ▶ provides some information and
 - ▶ a non-informational - e.g., entertainment - benefit
 - ▶ is costly: the firewall
- ▶ whether to **comply** with the regime

Model

Players. Leader, Mass of Citizens.

Citizens' actions. They choose whether to comply with the regime ($a = 1$) or not ($a = 0$).

The Leader maximises the share of citizens complying

$$U_L = \int_{i \in I} a_i$$

State of the world. $\omega \in \{0, 1\}$. $Pr(\omega = 1) = p \in (0, 1)$.
Symmetrically unknown.

Citizens' Preferences

▶ **complying** ($a = 1$) yields

▶ 1 if $\omega = 1$

▶ 0 if $\omega = 0$

▶ **not complying** ($a = 0$) yields her political type, $\theta_i \in [0, 1]$

θ_i is privately observed.

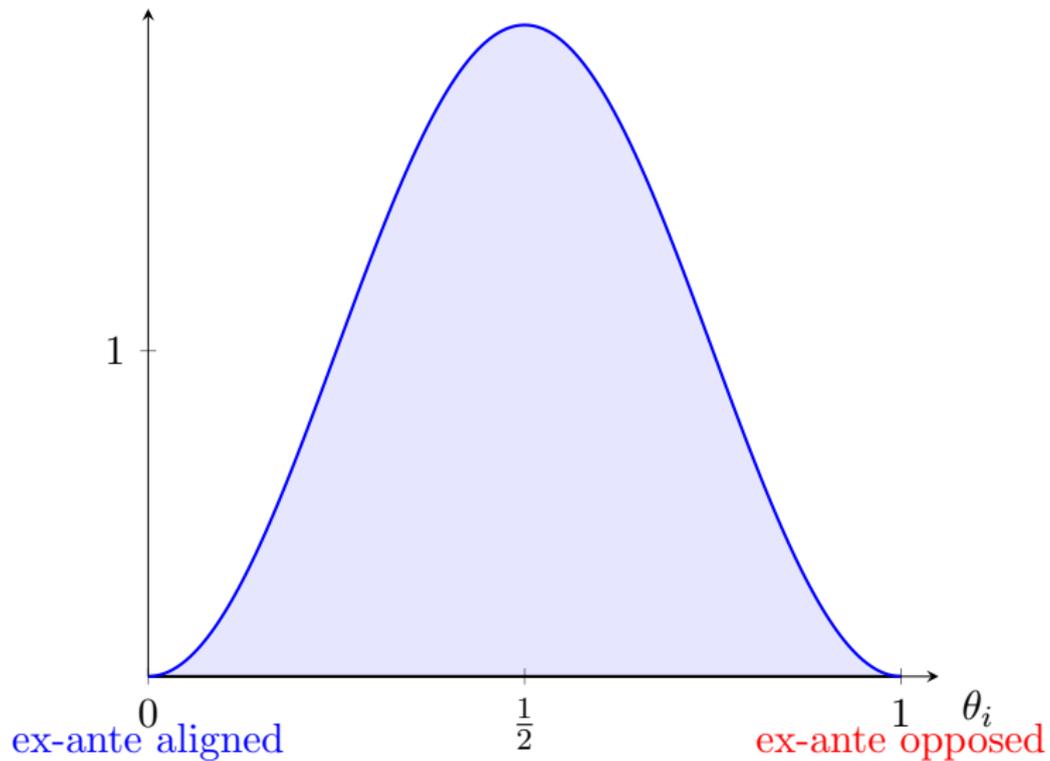
Citizens' Preferences

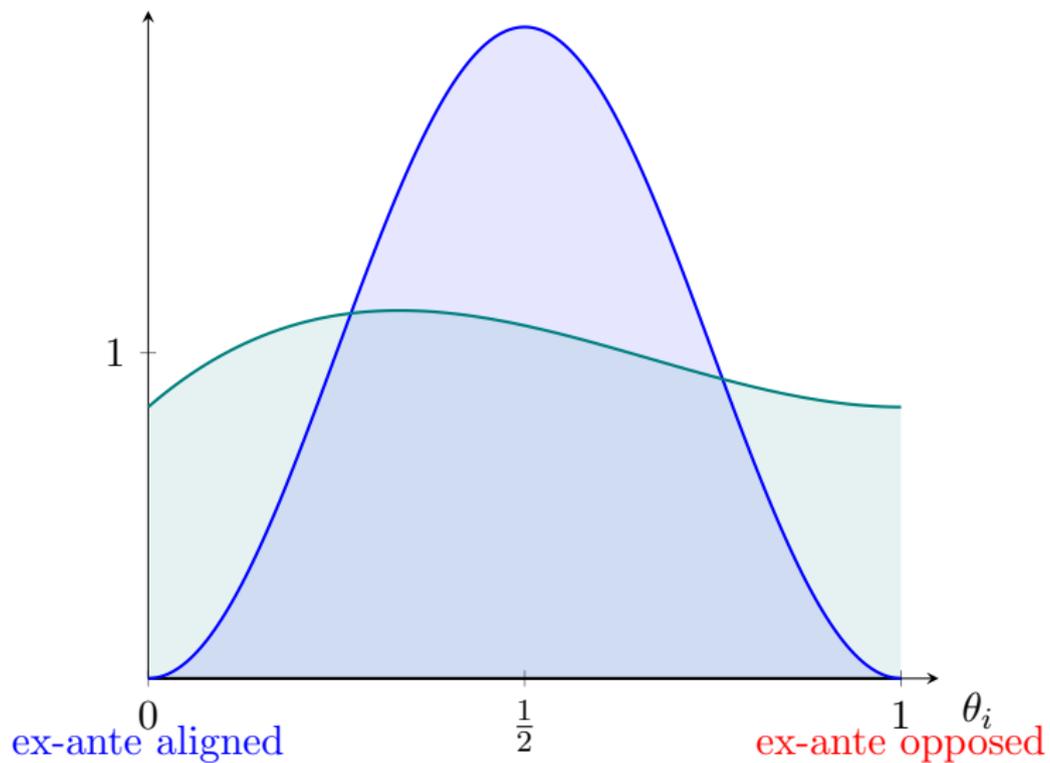
- ▶ **complying** ($a = 1$) yields
 - ▶ 1 if $\omega = 1$
 - ▶ 0 if $\omega = 0$
- ▶ **not complying** ($a = 0$) yields her political type, $\theta_i \in [0, 1]$

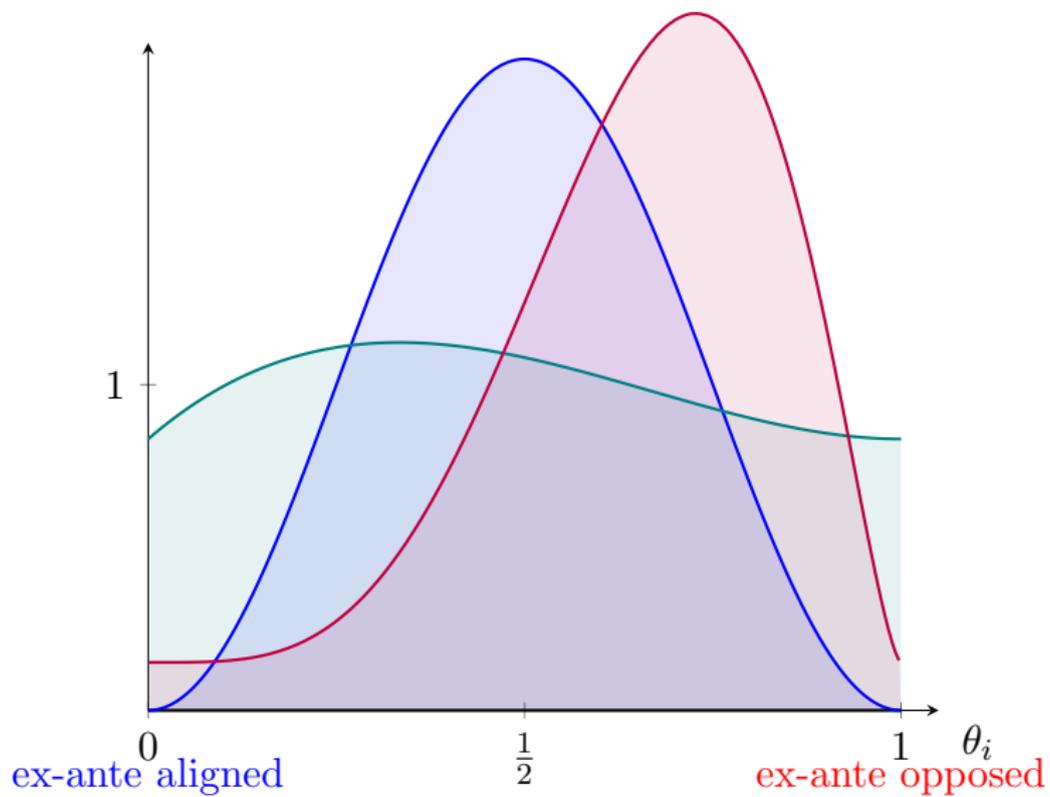
θ_i is privately observed.

Political preferences are distributed according to some pdf f with full support on $[0, 1]$ and cdf F s.t.

- ▶ f is unimodal and log-concave
- ▶ not too many extreme types: $\exists \theta^\dagger \in (0, 1)$ s.t. $F(\theta^\dagger) = \theta^\dagger$.







Information Flows Dichotomy

Differentiate between two sets of information flows

- ▶ those the regime **can control** \approx local outlets

→ “**government media**”

Citizens have **free access** to it.

- ▶ those the regime **can only control through (indirect) censorship** \approx foreign-based outlets

→ “**foreign media**”

Citizen **choose whether to obtain access** to it.

Government Media

Government media commits to an information structure that generates a report s_G s.t.

- ▶ **always** report true good news for the regime

$$Pr(s_G = 1 | \omega = 1) = 1$$

⇒ Bad news must be true

- ▶ chooses reporting slant $\sigma \in [0, 1]$

$$Pr(s_G = 1 | \omega = 0) = \sigma$$

High σ : uninformative reporting.

Foreign Media

Foreign media produces a report $s_{\mathcal{F}}$ on the state of the world, s.t.

- ▶ always report true bad news for the regime

$$Pr(s_{\mathcal{F}} = 0 | \omega = 0) = 1$$

\implies Good news must be true

- ▶ has reporting slant $\beta \in [0, 1]$

$$Pr(s_{\mathcal{F}} = 0 | \omega = 1) = \beta$$

High β : uninformative reporting.

Circumvention Decision

All citizens observe s_G and then decide whether to gain access to the foreign outlet.

Circumvention Decision

All citizens observe s_G and then decide whether to gain access to the foreign outlet.

Net benefit from consuming the foreign media

$$\underbrace{\delta_i(\theta_i, \sigma, s_G, \beta)}_{\text{Net Benefit}} = \underbrace{b_i(\theta_i, \sigma, s_G, \beta)}_{\text{info benefit}} + \underbrace{\alpha(\theta_i)}_{\text{relative non-info benefit}} - \underbrace{c}_{\text{cost of access}}$$

WTP

Circumvention Decision

All citizens observe s_G and then decide whether to gain access to the foreign outlet.

Net benefit from consuming the foreign media

$$\underbrace{\delta_i(\theta_i, \sigma, s_G, \beta)}_{\text{Net Benefit}} = \underbrace{b_i(\theta_i, \sigma, s_G, \beta)}_{\text{info benefit}} + \underbrace{\alpha(\theta_i)}_{\text{relative non-info benefit}} - \underbrace{c}_{\text{cost of access}}$$

WTP

$$\alpha(\theta_i) = z + \gamma * \theta_i$$

- ▶ $\gamma \geq 0$: correlation between political type and non-info benefit for foreign media
- ▶ $z \in \mathbb{R}$: non-info advantage of foreign media

Timing

1. The leader
 - 1.1 publicly commits to the reporting slant of the government media σ .
 - 1.2 picks the cost of access $c \in \mathbb{R}_+$
2. Nature determines ω and privately reveals θ_i to citizen i .
3. Nature generates the government media's report s_G as well as the foreign media's report s_F .

Each citizen, having observed s_G decides whether to consume the foreign outlet; if they do, they observe s_F .
4. The citizens choose whether to comply with the regime. Game ends. Payoffs are realized.

Solution concept: wPBE.

Dual Communication

Ideally, the leader can communicate

- ▶ with his **base** via state-controlled outlets and
- ▶ with the **opponents** via selective access to foreign media

→ citizens that would *never* comply if they had no access to a foreign credible source

Dual Communication

Ideally, the leader can communicate

- ▶ with his **base** via state-controlled outlets and
 - ▶ with the **opponents** via selective access to foreign media
- citizens that would *never* comply if they had no access to a foreign credible source

But the firewall is a *coarse* tool ...

→ need to create a cleavage in the citizenry vis-a-vis foreign content

→ whether **politics drives entertainment consumption** is crucial

Key Results: How to Segment

1. cost of access and propaganda: segmentation can be *engineered* when correlation is intermediate
2. invest in local entertainment
3. strategically (not) ban foreign content

Related Literature

Trade-offs inherent to censorship: monitoring, backlash, etc.

Egorov et al (2009), Lorentzen (2014), Kronick, Marshall (2022), Edmond (2013)

→ *how* to censor outlets that cannot be shut down

Propaganda and Information Design: which information (states of the world) to censor/disclose.

Kamenica and Gentzkow (2011), Adena et al (2013), Bernhardt and Shadmerh (2015), Gehlbach, Sonin (2014), Little (2018), Yu (2021), Gitmez and Molavi (2022), **Kolotilin (2018,2022)**

→ designer can improve on full commitment public persuasion payoff by allowing selective access - through common cost - to an exogenous signal

Linked to persuasion via intermediaries in voting bodies Schnakenberg 2017, Awad 2020

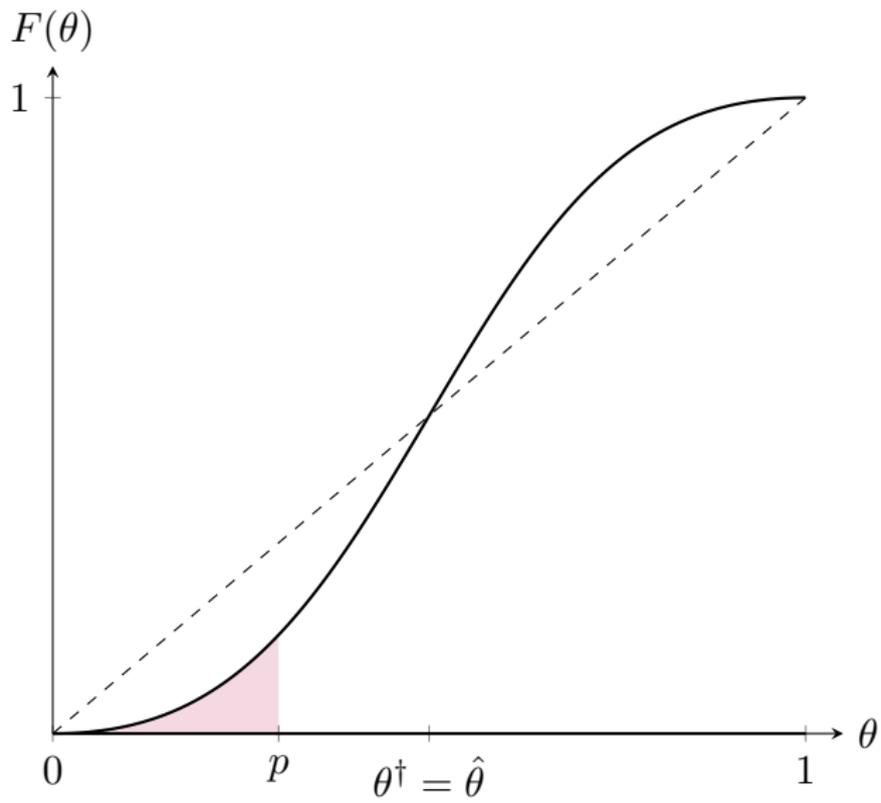
Optimal Propaganda W/o Censorship

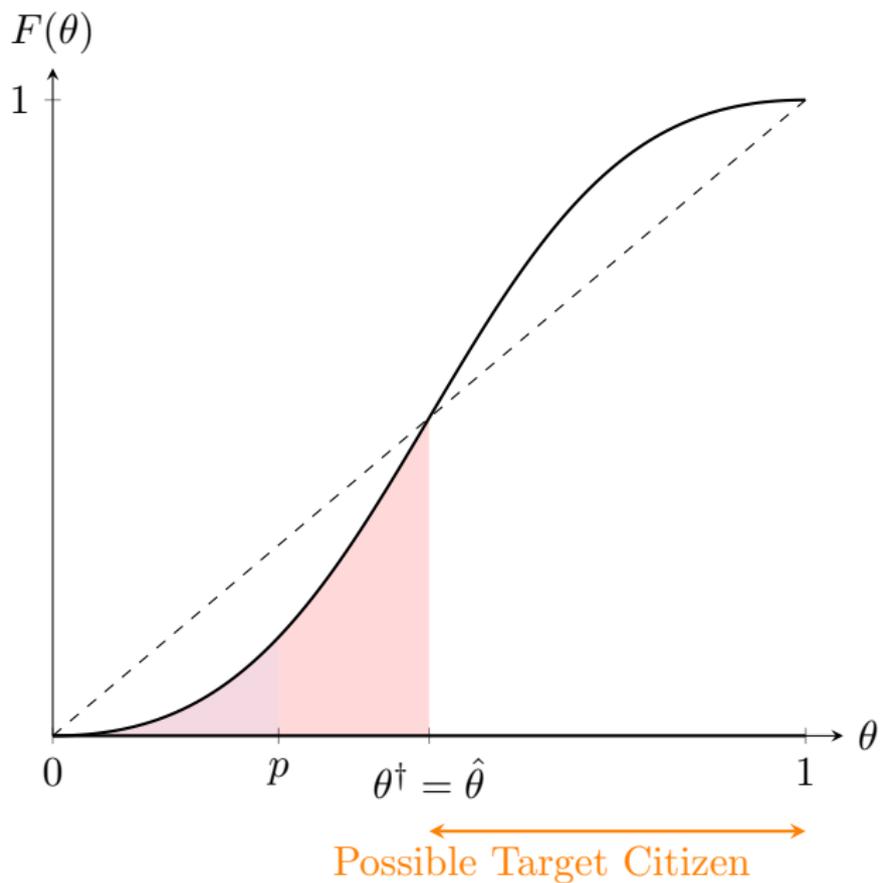
Suppose no firewall and entertainment is positive \implies all citizens consume the foreign outlet

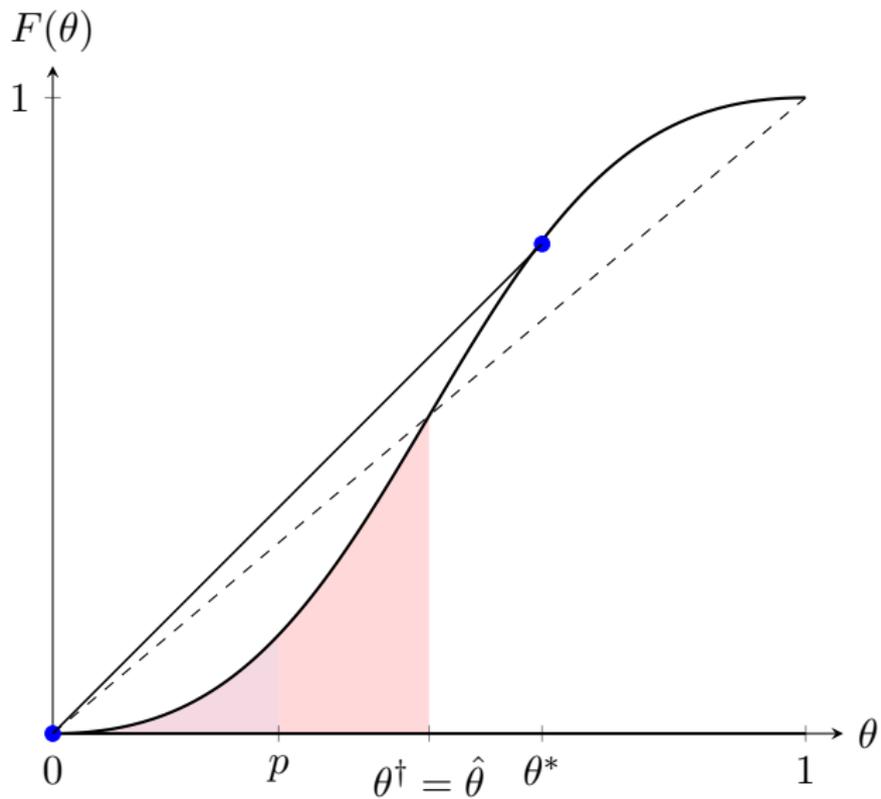
Lemma 1

For any unimodal f , the **equilibrium target citizen** θ^* is always **more misaligned** with the regime than

1. prior citizen
2. modal citizen
3. fixed point citizen







The Informativeness of the foreign Outlet

Lemma 2

The more informative the foreign outlet is ($\uparrow (1 - \beta)$)

1. the less informative the government media is in its reporting ($\sigma^* \uparrow$).

Recall: absent any censorship, informativeness of gov media increases in informativeness of foreign media (credibility constraint).

The Informativeness of the foreign Outlet

Lemma 2

The more informative the foreign outlet is ($\uparrow (1 - \beta)$)

1. the less informative the government media is in its reporting ($\sigma^* \uparrow$).

Recall: absent any censorship, informativeness of gov media increases in informativeness of foreign media (credibility constraint).

→ possibility of partial censorship reverses this logic.

2. $\underline{\gamma}$ increases: segmentation is less likely (in the sense of set inclusion)

⇒ compliance is single-peaked in the foreign media informativeness

Outside of China

Iran

- ▶ Media Freedom Ranking: 178/180
- ▶ Blocked foreign outlets: CBS, NBC, Netflix, Facebook, etc.

Turkmenistan

- ▶ Media Freedom Ranking: 176/180
- ▶ Blocked foreign outlets: Facebook, Instagram, Youtube etc.

Similar dynamics in Russia, Qatar, the UAE, etc.

Related Literature

Propaganda: which information (states of the world) to censor/disclose.

Kamenica, Gentzkow (2011), Adena et al (2013),Bernhardt, Shadmerh (2015), Gehlbach, Sonin (2014), Little (2018), Yu (2021), Gitmez and Molavi (2022, Kolotilin (2018,2022)

Trade-offs inherent to censorship: monitoring, backlash, etc.

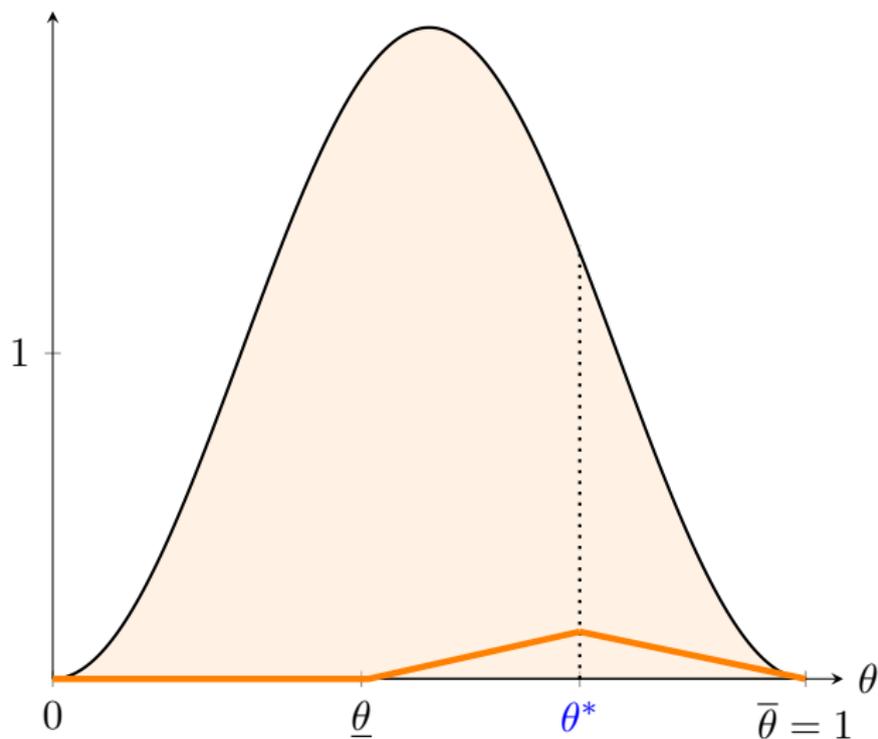
Egorov et al (2009), Lorentzen (2014), Kronick, Marshall (2022), Edmond (2013)

Intersection of propaganda and censorship / other levers Guriev, Treisman (2020), Gehlbach et al (2022), Li, Zhou (2022), Gitmez, Sonin (2022)

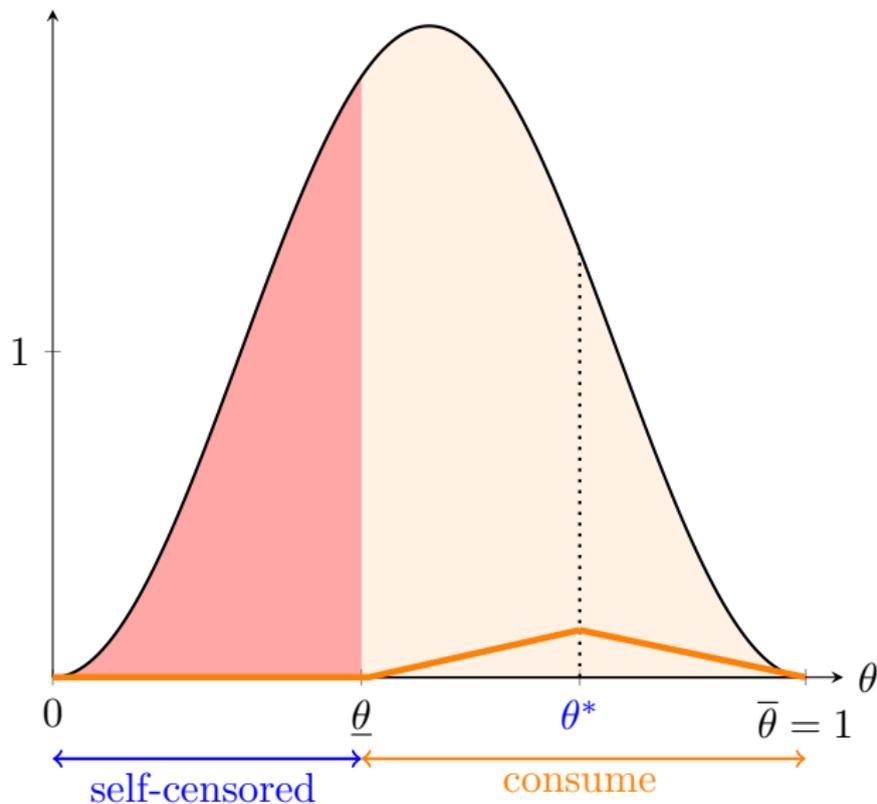
Reviews Stromberg (2015), Gehlbach, Sonin, Svolik (2016), Zhuravskaya et al (2020), Roberts (2020), Egorov, Sonin (2022)

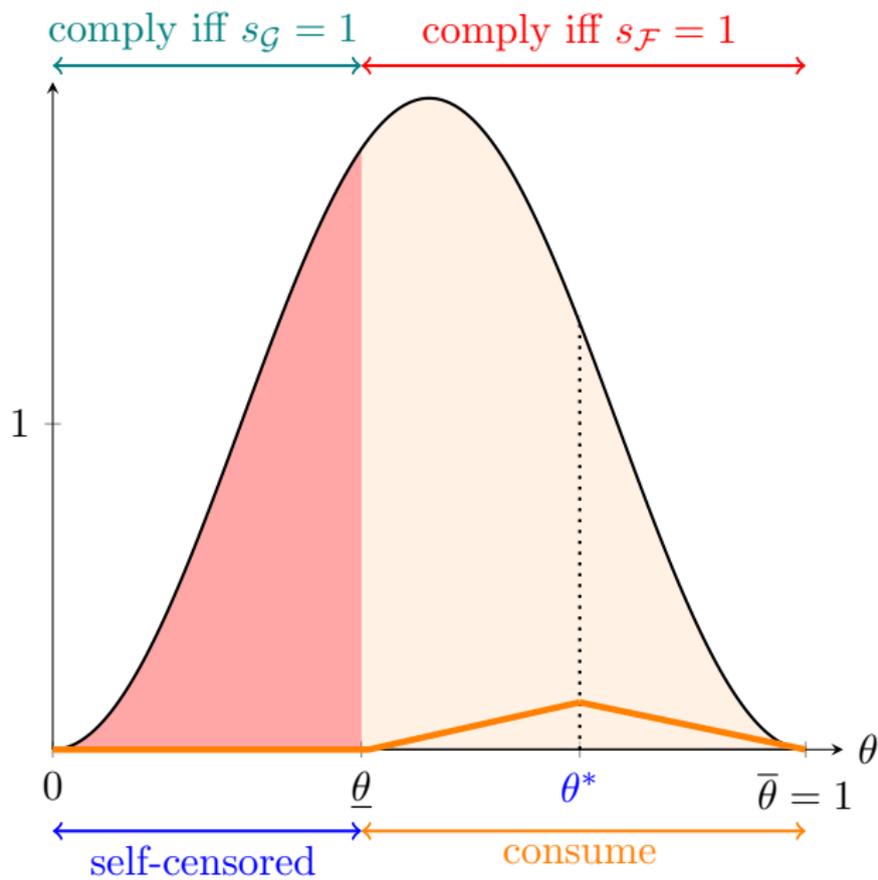
[▶ Back](#)

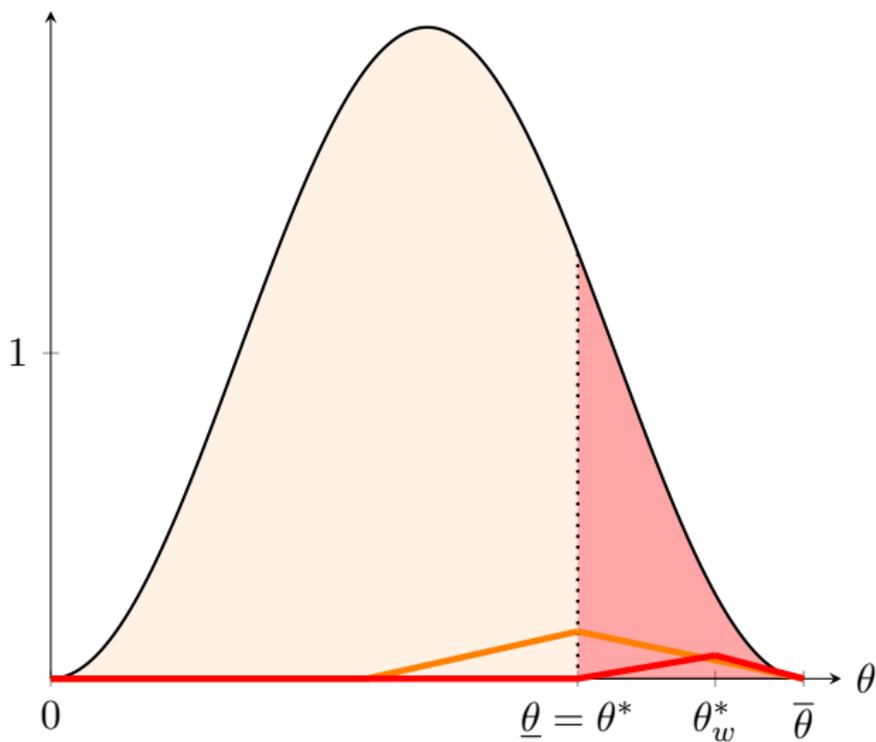
Suppose $\alpha = 0 = \bar{c}$ and **fix** the reporting slant σ^*



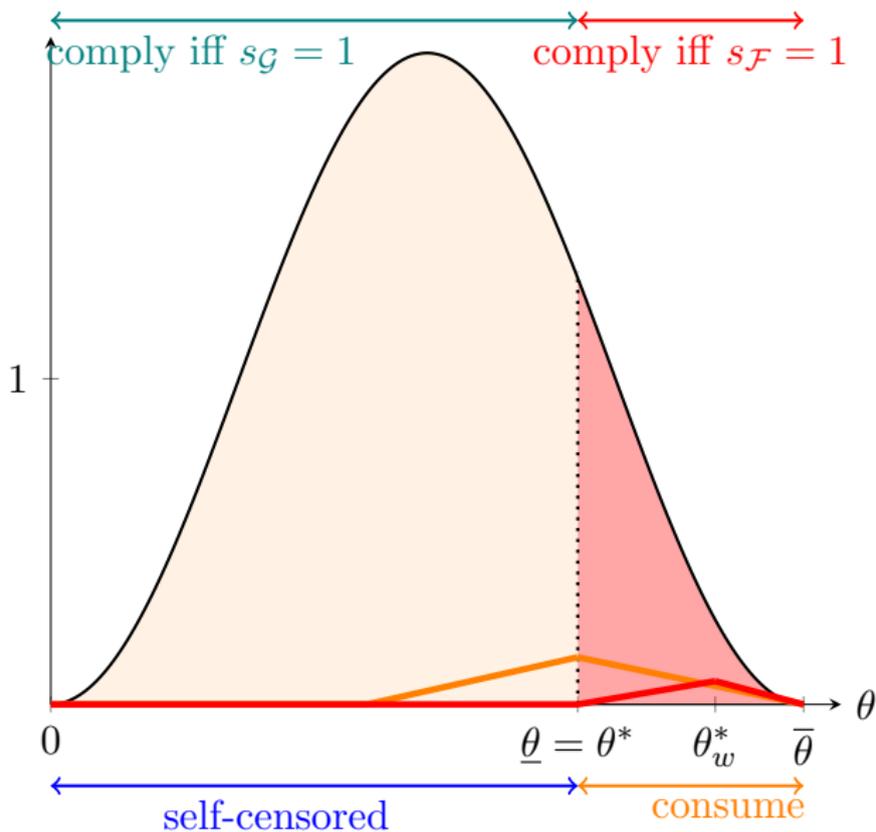
If regime cannot dissuade access, large part of the population consumes $s_{\mathcal{F}}$



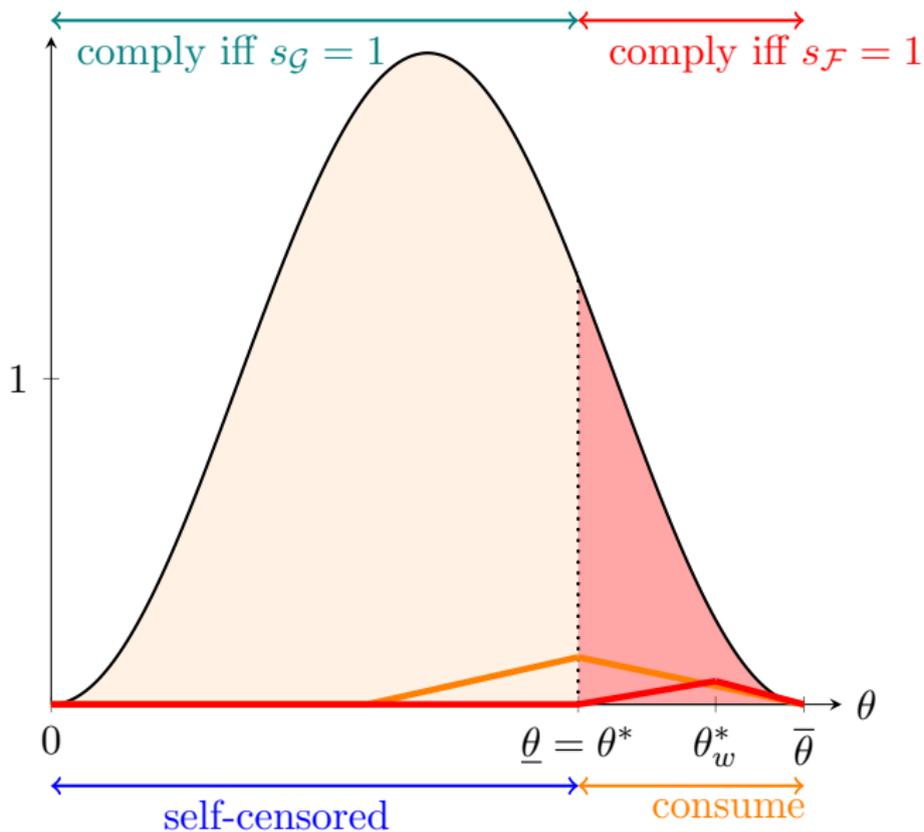




Costly-fix: make the gov media more informative to reduce the share of consumers.

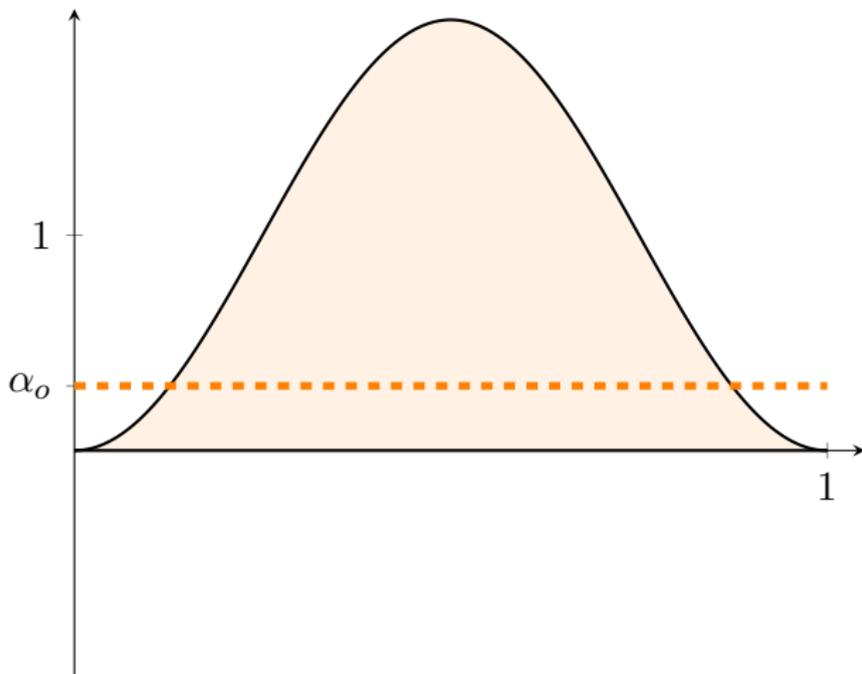


Ensure compliance of same share of citizens ($\theta_i < \theta^*$) but less often.

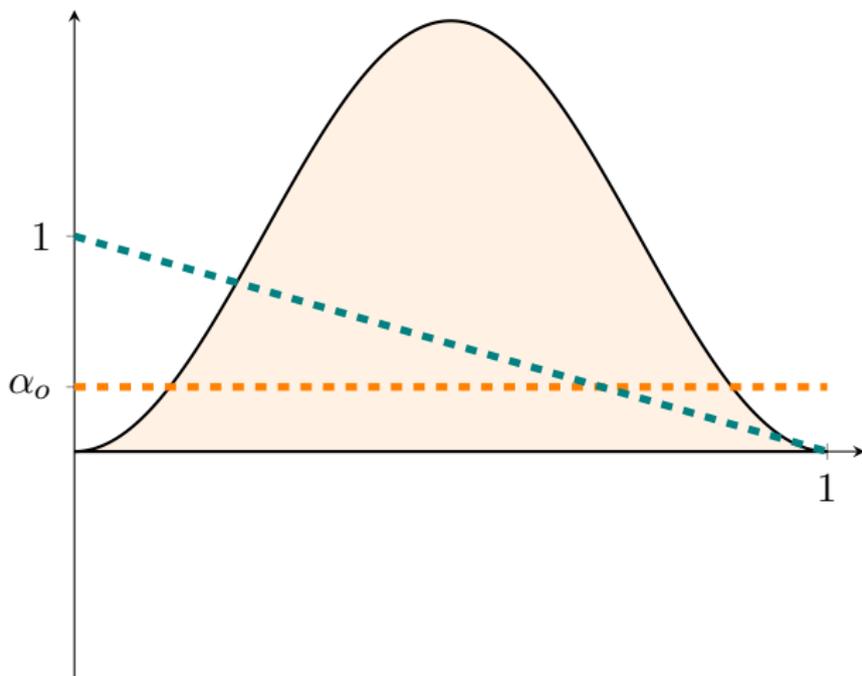


→ illustration of **complementarity between propaganda and censorship.**

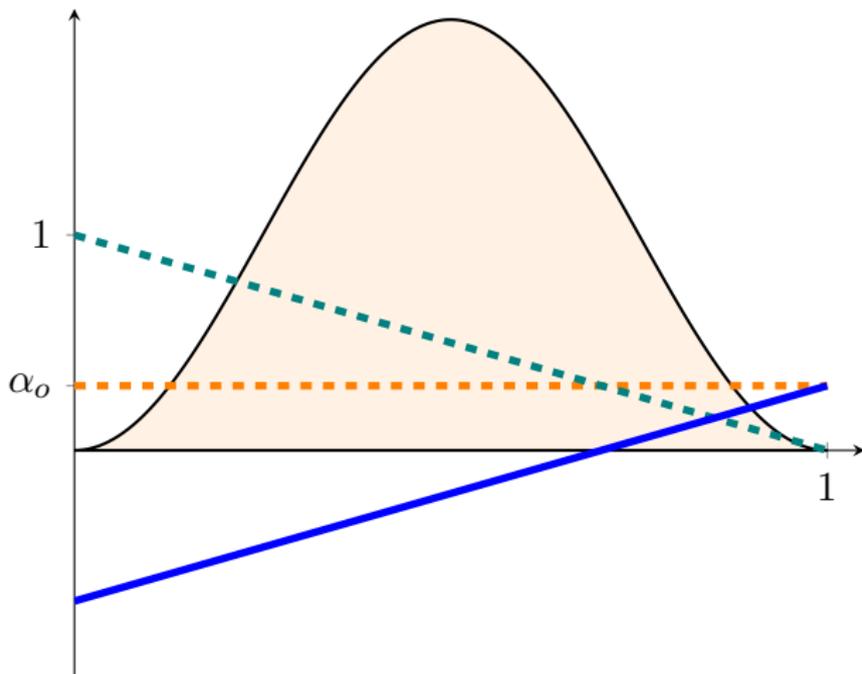
▶ Back Suppose Hollywood movies provide a common entertainment benefit $\alpha_o > 0$.



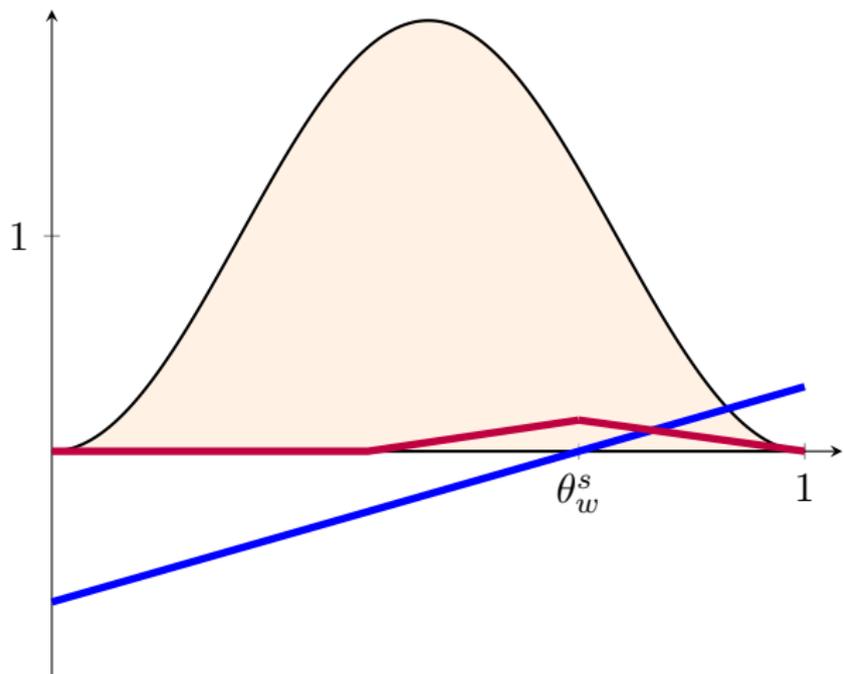
Historical Propaganda movie: $\alpha_g(\theta_i) = 1 - \theta_i$



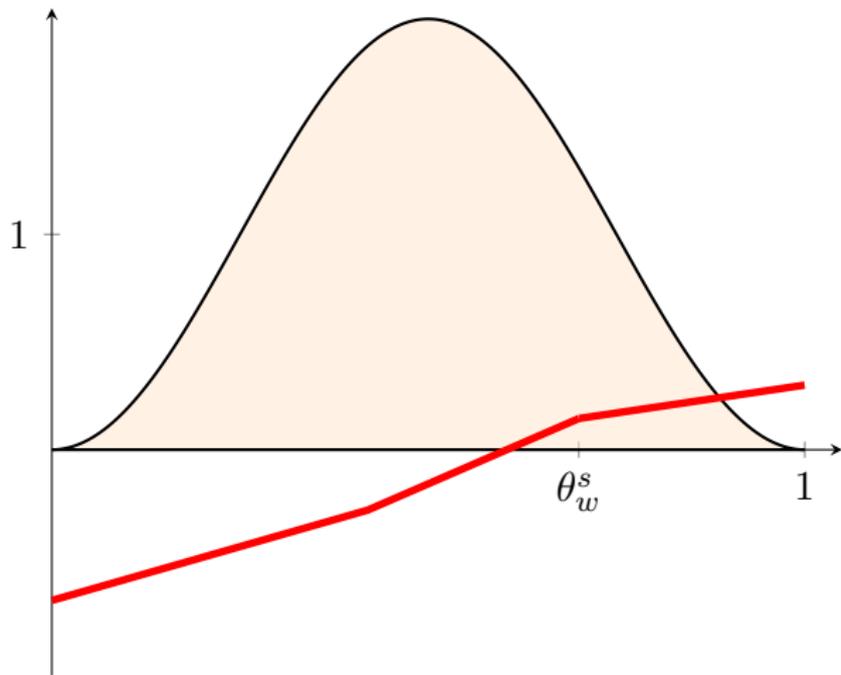
⇒ relative entertainment benefit $\alpha(\theta_i)$ is now strongly associated with a citizen's political views θ_i .



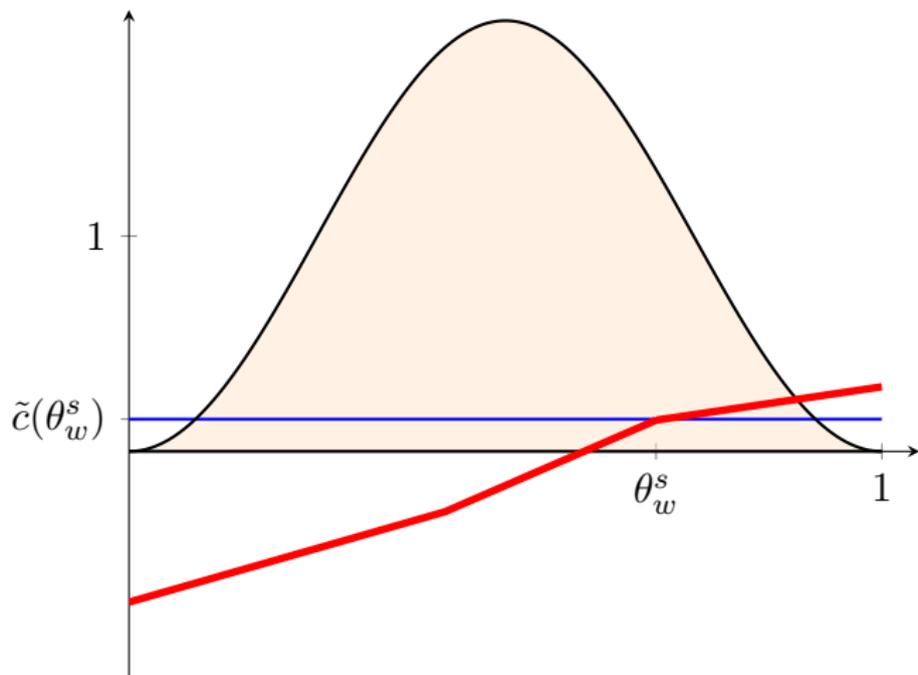
▶ Back



⇒ **net** benefit is also increases in θ_i



⇒ segmentation and partial censorship is optimal



Formally: suppose

$$\alpha = \alpha_o + \alpha_g(\theta_i)$$

and

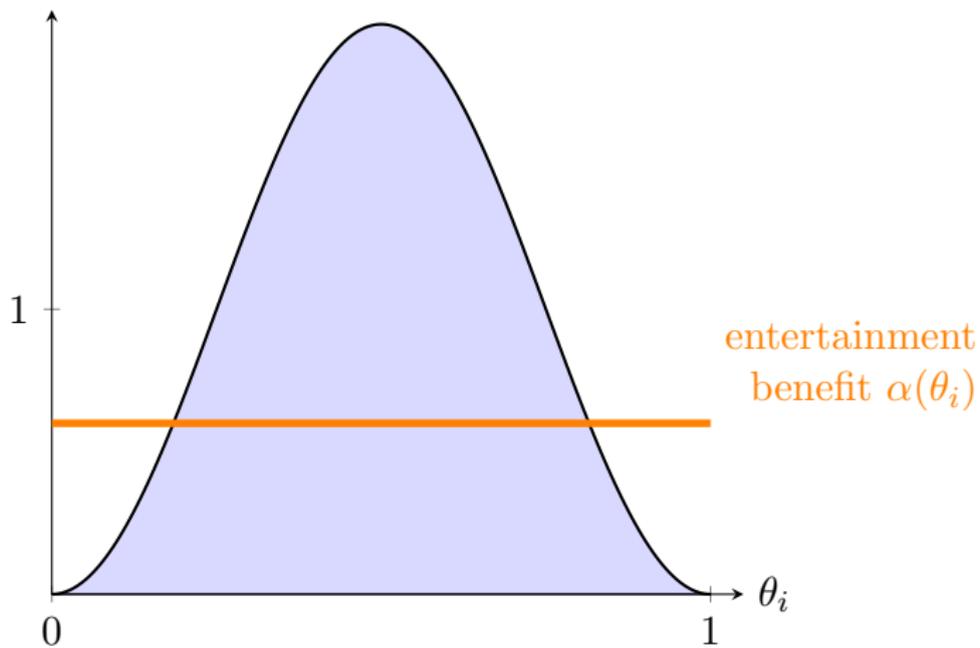
$$\alpha_g(\theta_i) = z - \gamma\theta_i$$

$\gamma < 0 \implies$: disproportionately enjoyed by CCP supporters

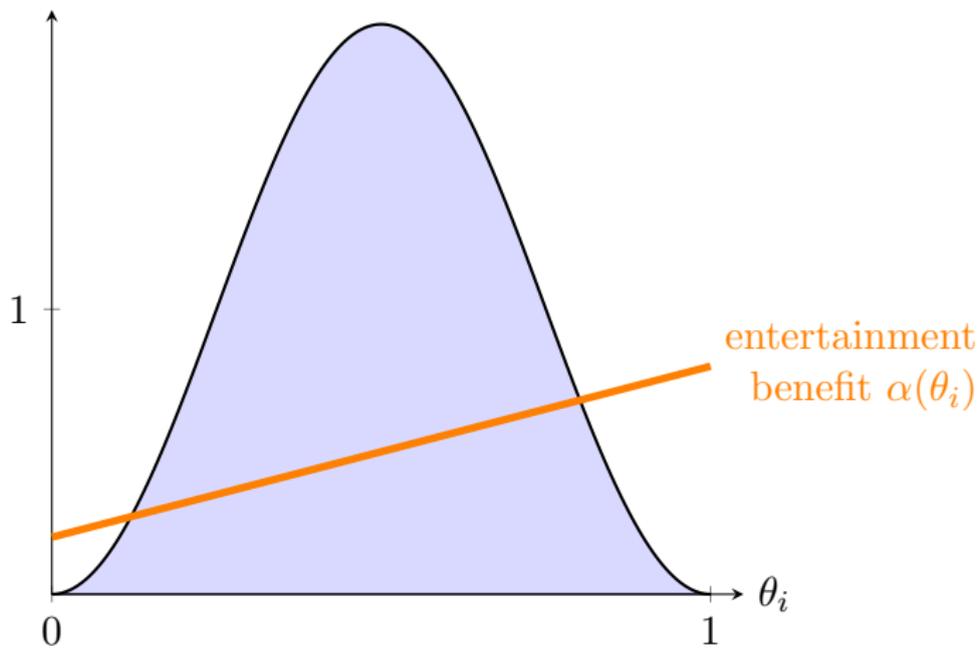
\implies create correlation, even if there's none vis-a-vis foreign entertainment

[▶ Back](#)

No Sorting on Entertainment

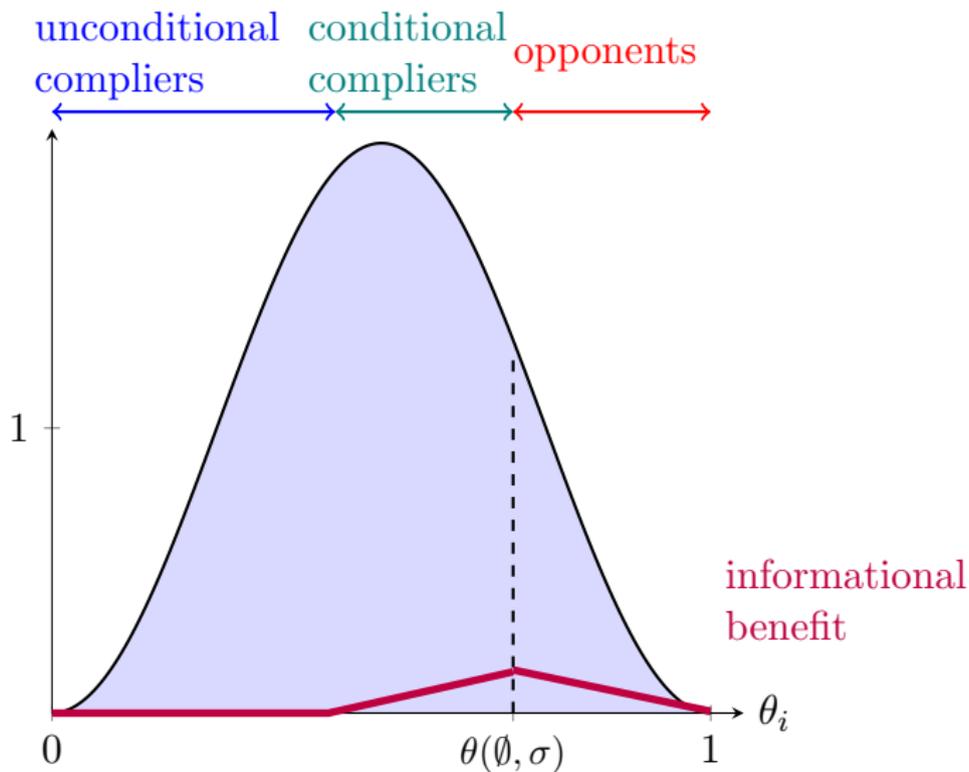


Sorting on Entertainment



Informational Sorting

Given some reporting slant σ and positive reporting



Explaining Observed Segmentation

Proposition 1b

If the correlation is weak ($\gamma < \underline{\gamma}$) and the **cost of access is bounded** ($c < \bar{c}$) then

1. the **regime censors as much as possible** ($c^* = \bar{c}$)
2. as the constraint disappears ($\uparrow \bar{c}$)

2.1 the government media becomes less informative: $\frac{\partial \sigma^*}{\partial \bar{c}} \geq 0$

2.2 the target citizen becomes more ex-ante aligned: $\frac{\partial \theta^*}{\partial \bar{c}} \leq 0$

2.3 compliance increases

→ **compromise on propaganda to induce less consumption**

⇒ segmentation and partial censorship only occur **by constraint**

▶ Graphs

Extension: Learning from segmentation

Use of vpn can reveal a citizen's type

→ can be used to target repression if learn who are the most opposed citizens

Learning is most efficient under intermediate/strong correlation

⇔ most extreme citizens consume

⇒ ramp up incentives to segment

Robustness of Segmentation

Consumption of $s_{\mathcal{F}}$ can reveal θ_i : w.p. $\rho \in (0, 1)$ the regime can observe which individuals consumed $s_{\mathcal{F}}$.

The leader's payoff can then be written as

$$V_l = V(\sigma, c) + \rho \int_{\theta(l)}^1 c_i dF(\theta_i) \quad (1)$$

A citizen's net benefit from consuming $s_{\mathcal{F}}$ is then given by

$$\delta_i(\theta_i, \cdot) = b_i + \alpha(\theta_i) - c - c_l * \rho \quad (2)$$

Denote $C = c + c_l * \rho$ and denote the total equilibrium cost of access by C^* .

Then in equilibrium, simply set $C^* = c^*$ to derive all the previous results. Further, the leader's payoff is strictly higher than previously.

Signalling Through Censorship: Robustness

Chavez shuts down RCTV in 2006 \rightarrow signal $\omega = 0$?

Substantive justification. Censorship signals commitment to freedom of press.

Not capacity to provide public good, or (military) strength of the regime.

[▶ Back](#)

Robustness Exercise

Suppose that the leader has a privately know type, $\tau \in \{h, l\}$, with

▶ $\Pr(\omega = 1|h) = p$

▶ $\Pr(\omega = 1|l) = \lambda \times p$

for $\lambda \in [0, 1)$ and $\rho \equiv \Pr[\tau = h]$

Assume that

- ▶ citizens believe that the dictator is a low type if he censors the foreign media (irrespective of equilibrium conjunctures),
- ▶ and that no signalling can take place through the choice of σ (otherwise always censor)

Then, abusing notation

$$\mu_c \equiv Pr(\tau = h|c, s_G = 1) = \frac{p\lambda}{p\lambda + (1 - p\lambda)\sigma}$$

$$\mu_{nc} \equiv Pr(\tau = h|nc, s_G = 1) = \frac{p\beta}{p\beta + (1 - p)\sigma} = \theta_{nc}$$

$$\hat{V}_c(\sigma; \lambda) = [p + (1 - p)\sigma]F(\mu_c)$$

$$\hat{V}_{nc}(\sigma; \beta) = [p\beta + (1 - p)\sigma]F(\mu_{nc}) + p(1 - \beta) = V_{nc}(\sigma; \beta).$$

Lemma 3

There exists $\lambda^ < 1$ such that $\max_{\sigma} \hat{V}_c(\sigma; \lambda) \geq V_{nc}(\sigma_{nc}^*)$ for $\lambda \geq \lambda^*$.*

As long as the high and low-types are sufficiently different, even when censorship leads to the worst possible updating from the citizenry, a high-type is better off censoring the foreign media.

Gain through censorship $>$ loss from signalling.

Under F , we show that censorship and propaganda are complementary.

Given pooling on censorship, if signalling could occur through propaganda, it would be much better to signal via propaganda.

If propaganda and censorship were substitutes the censorship incentives would be reduced.

Incentives to engage in censorship are strongest given complementarity, even when we introduce the possibility of signalling.

The Censorship Filter

Autocrat can try to “filter” bad news (King et al 2013, Gelbach et al 2022).

The leader can **commit ex-ante** to filtering a share $q \in [0, 1]$ of bad news ($s_{\mathcal{F}} = 0$).

(for exposition) Re-interpret good news from the foreign media as “no news”.

Denote the (possibly filtered) foreign media signal by $\hat{s}_o \in \{0, \emptyset\}$.

$$Pr(\text{no news}) = \underbrace{p(1 - \beta)}_{Pr(s_{\mathcal{F}}=\emptyset)} + \underbrace{q}_{Pr(\text{non-filtered})} \underbrace{[p\beta + (1 - p)]}_{Pr(s_{\mathcal{F}}=0)}$$

Propaganda and Censorship as Substitutes

Lemma 2

For any $q > 0$, q and σ are perfect substitutes iff the foreign media is perfectly informative ($\beta = 0$).

Intuition: absent fake bad news ($\beta = 0$), q and σ both reduce the value of good/no news, while increasing the frequency of good/no news.

► Details

The Filter is a Second-Best Instrument

Proposition 3

In equilibrium, if $\beta > 0$, $q_F^* = 1$.

The optimal filter is “not a (partial) filter”.

It makes the foreign media observable report (\hat{s}_o) completely uninformative.

No Censorship Motives

1. Optimal to report truthfully on ω .

\implies foreign media loses any informational content

\implies no censorship because pointless

When most citizens are ex-ante opposed to the regime.

$F(\theta) < \theta \forall \theta \in (0, 1)$. [▶ Graphs](#)

2. Optimal to “report negatively” on ω .

\implies Approximate optimal reporting through no censorship.

When many citizens are ex-ante aligned with the regime (and few are around the prior) [▶ Graphs](#)

When to Censor?

Any other distribution → **optimal to censor** the foreign media.

Two effects through which censorship is beneficial

1. (distributional) **direct** effect
2. indirect effect through the **complementarity of propaganda and censorship**

▶ Graphs

No Censorship: Motive 1

Optimal to report truthfully on ω .

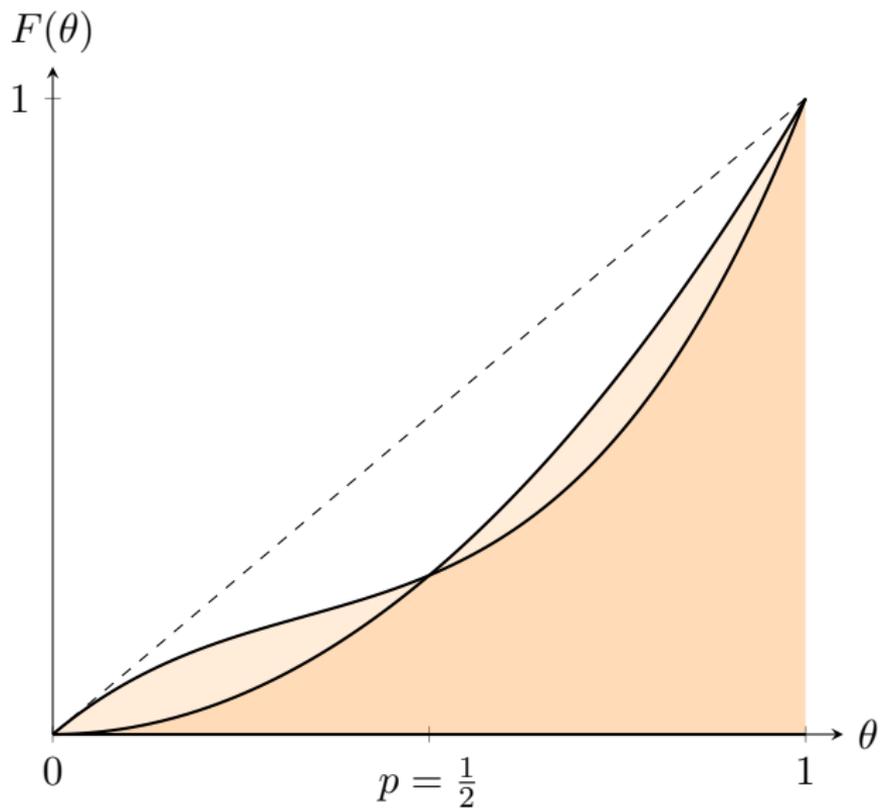
\implies foreign media loses any informational content

\implies no censorship because pointless

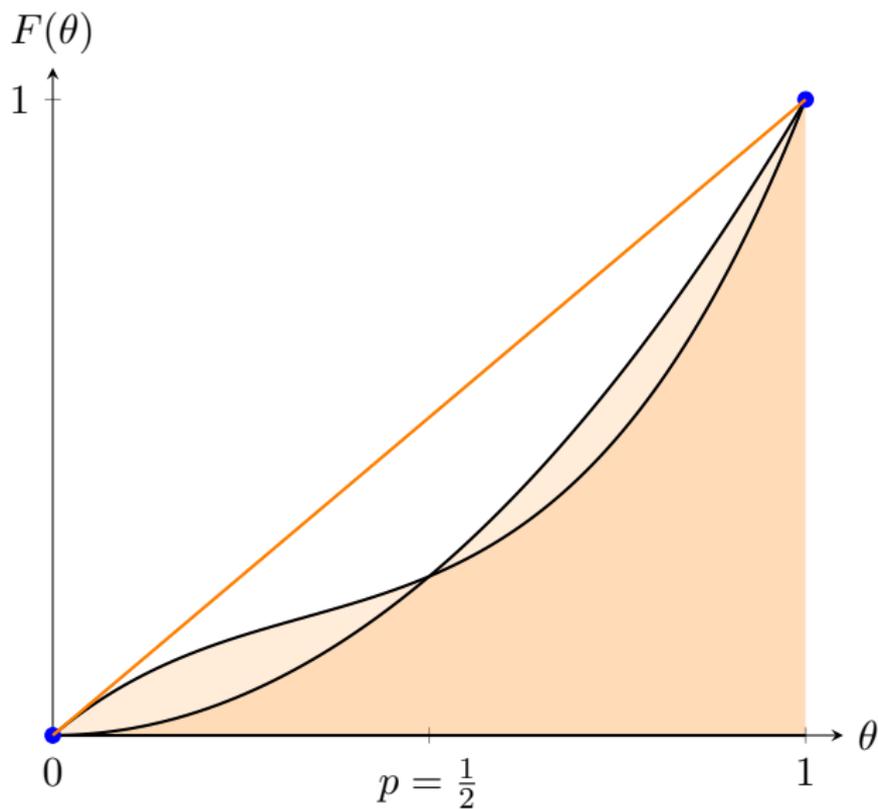
When most citizens are ex-ante opposed to the regime.

$$F(\theta) < \theta \quad \forall \theta \in (0, 1).$$

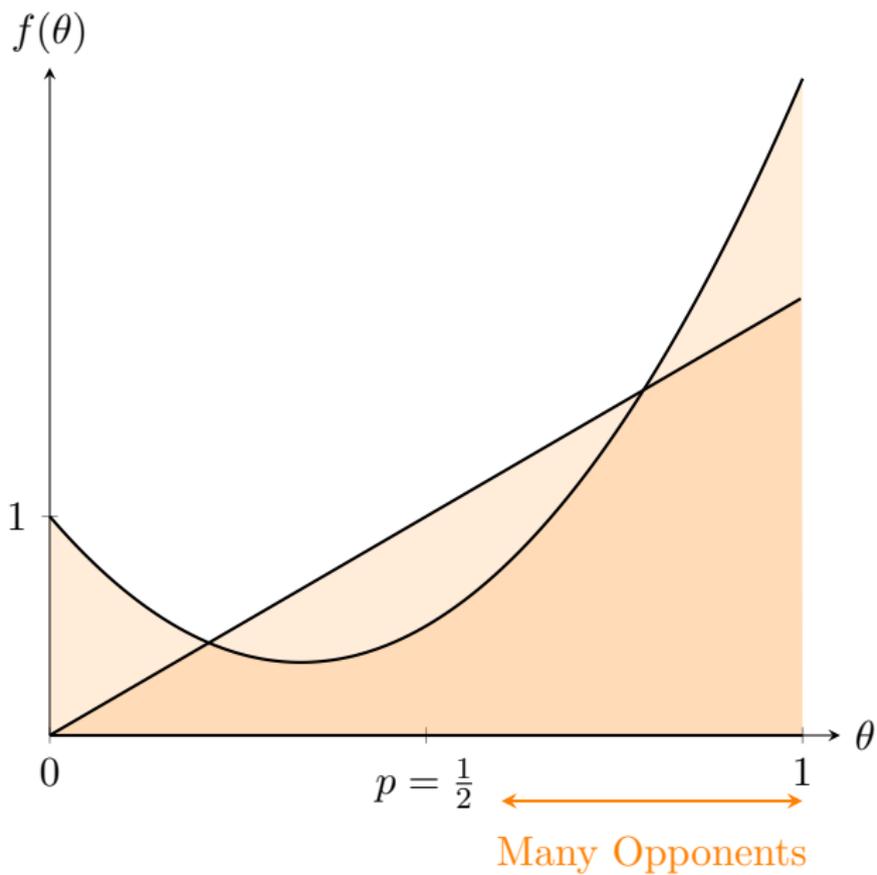
Motive 1: Illustration



Motive 1: Illustration



Motive 1: Illustration



No Censorship: Motive 2

Optimal to “report negatively” on ω .

\implies impossible through the government media because of credibility constraint ($\tau = 1$)

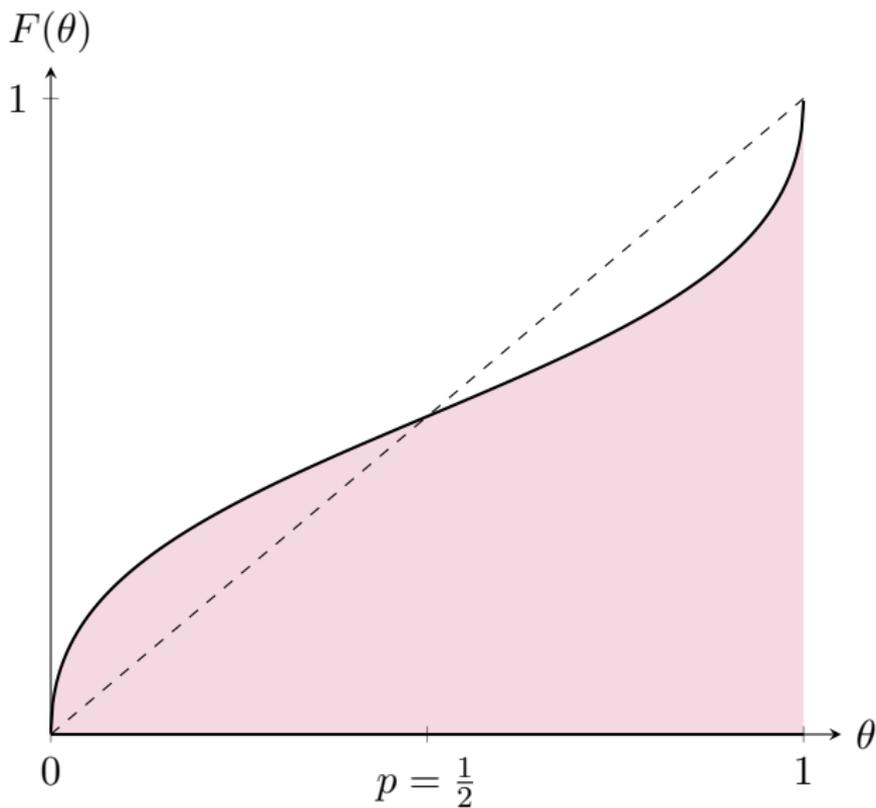
\implies foreign media is the sole source of information

\implies no censorship to approximate the unattainable optimal reporting strategy

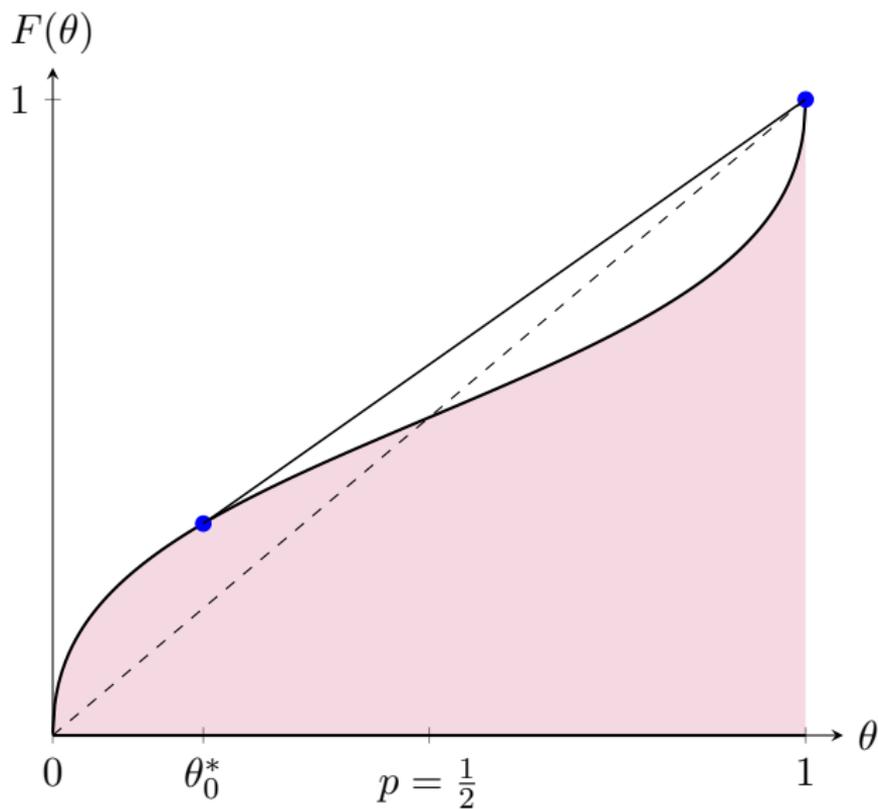
When many citizens are ex-ante aligned with the regime.

Tangent to $D(p)$ lies strictly above the graph of F

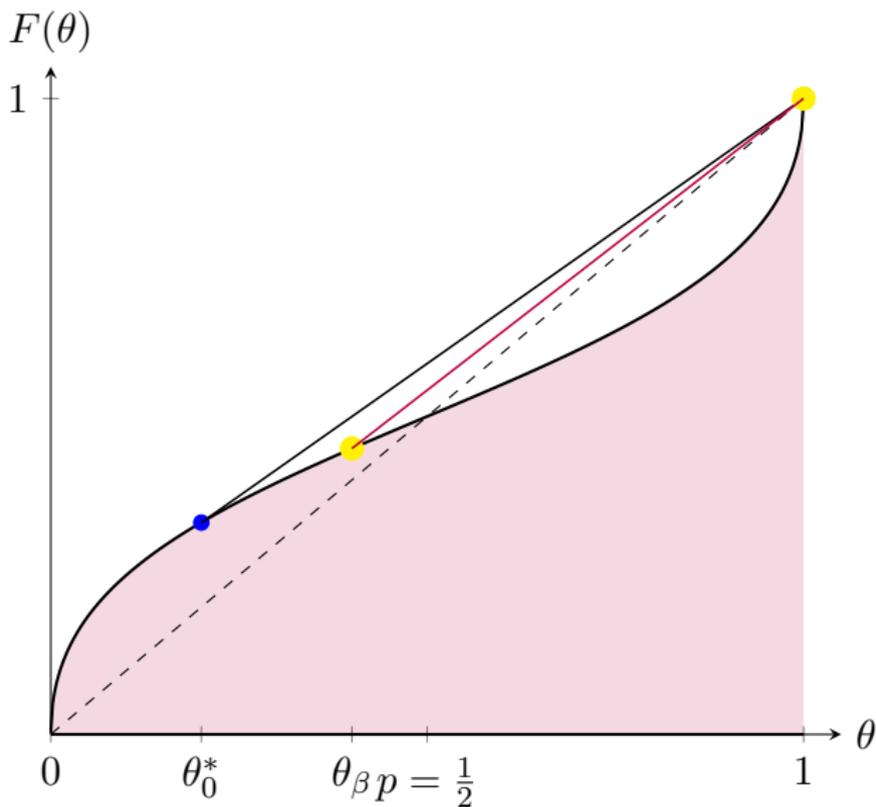
Motive 2: Illustration



Motive 2: Illustration



Motive 2: Illustration



Approximate optimal reporting with (i) self-censorship and (ii) no