Contestability and Optimal Regulation of Social Media Platforms

Martino Banchio, Francesco Decarolis, Carl-Christian Groh, Rafael Jiménez-Durán and Miguel Risco

September 21, 2025

Introduction

Introduction & Motivation

- What? Competition between **social media platforms.**
  - Optimal regulation to improve social welfare?
  - Two keys: **harmful but engaging content** (platform side); **naive users** (user side).
- **Why?**
  - Internal/external evidence of social media platforms harmful effects (Horwitz et al., 2021, Braghieri et al., 2022).
  - Regulatory attention (EU Digital Markets Act).
  - Presence of naive users on social media platforms (Allcott et al., 2022).

$\rightarrow$ This project: theoretical model + experiment

Theoretical model

- Building blocks:
  - Platform competition between an entrant and an incumbent (with **competitive advantage**).
  - Platforms choose share of **harmful but engaging** content to display. Harmful but engaging content maximizes user engagement.
  - Some consumers are **naive**: they neglect the negative effects of harmful content.
- Key trade-offs:
  - Platforms want to **maximize engagement** of users $\Longrightarrow$ incentives to raise share of harmful content.
  - However: if too much harmful content $\Longrightarrow$ rational consumers leave.

Preview of theoretical results

- **Roadmap:**
  - How are the equilibrium outcomes affected by the competitive advantage and the share of rational users?
  - Regulation: we move along the two dimensions (grid).

Main take-aways

1. If all users visit the incumbent, user welfare is larger than in any equilibrium in which some users visit the entrant.
2. If the share of rational users is low, improving contestability has no effect on user welfare.
3. If there is no competitive advantage, the user-optimal outcome can emerge if the share of rational users is large enough.

Insights

- The **share of rational users interacts critically** with the incumbent's competitive advantage.
- Measures that decrease the share of naives are a **necessary condition** for improving social welfare.
- Different social media platforms' markets need **different regulation**.

Experiment and estimation

- We conduct an experiment to **estimate the parameters** of the model.
- We run the experiment in **two different markets**: Instagram vs TikTok and $\mathbb{X}$ vs Meta Threads.
- For each market, we estimate the **share of rational users** and the **competitive advantage**, which allows us to provide policy recommendations.

Related literature

- **What's new?** Model of social media platform competition with naive users (who neglect the costs of harmful content).
- Platform competition: Armstrong (2006), Biglaiser et al. (2022), Beknazar-Yuzbashev et al. (2024), Ichihashi and Kim (2023), Acemoglu et al. (2024).
- Contestability in digital markets: Kades and Scott Morton (2020), Bourreau and Krämer (2023), Dhakar and Yan (2024).
- Empirical evidence on preference inconsistency regarding social media platforms: Hoong (2021), Allcott et al. (2022).

Model

A model of platform competition

- **Players:** Unit mass of users and two platforms $\{I, E\}$.
- **User choices:**
  - Which platform to visit: $j_i \in \{\emptyset, I, E\}$.
  - The time to spend on the platform (engagement): $e_i \in \mathbb{R}$.
- **Platform choices:**
  - Share of harmful (yet engaging) content displayed to any user: h_p.
- **User preferences and heterogeneity**:
  - User i in platform p with engagement e_i obtains $U_p(h_p, e_i)$.
  - Rational users **internalize the costs of harmful content**. Naive users do not. Share $\rho \in (0, 1)$ is rational.
  - Rational users choose platform I instead of k iff $U_I(h_I, e_i^*(h_I))) \geq U_k(h_k, e_i^*(h_k)))$. Naive users join the platform which yields higher perceived utility.

Model assumptions

- **Assumption 1.** The incumbent has a competitive advantage: $U_I^r(h) > U_E^r(h)$ and $U_I^n(h) > U_E^n(h)$ hold for all $h \in [0, 1]$.
- **Assumption 2.** Exposure to harmful content decreases true utility: $U_p^r(h)$ is strictly decreasing in h and $U_p^r(1) < 0 < U_p^r(0)$ holds for both $p \in \{I, E\}$.
- **Assumption 3.** Naive users are drawn to harmful content: $U_p^n(h)$ is strictly increasing in h for both $p \in \{I, E\}$.
- **Assumption 4.** Harmful content is engaging: $e_p^*(h)$ is strictly increasing in h for both $p \in \{I, E\}$.

Platform's preferences

- Platform revenues are **proportional to the engagement** of users who join.

$$\Pi_p = \int_0^1 \mathbb{1}[j_i = p](\mathbb{1}[t_i = r]\pi_p^r(e_p^*(h_p)) + \mathbb{1}[t_i = n]\pi_p^n(e_p^*(h_p)))di.$$

For every $p \in \{I, E\}$ and every $t \in \{r, n\}$, $\pi_p^t(x)$ is an **increasing** function.

Timing & equilibrium

- **Timing:**
  1. Platforms simultaneously choose h_p.
  2. Users choose which platform to visit and their engagement.
- **Equilibrium:** subgame perfect equilibrium accounting for naives' behavior.
  - Given (h_I, h_E), rationals maximize their utility.
  - Given (h_I, h_E), naives maximize their perceived utility.
  - Each platform maximizes revenues given the others' strategies.

Example: a preference framework + specific naiveté

- User i's (true) utility is:

  $$U_p(h_p, e_i) = (\eta_p h_p + \theta_p(1 - h_p))e_i + (1 - h_p) - \delta h_p - \gamma(e_i)^2,$$

  where δ is the cost of consuming harmful content and γ the opportunity cost of time.
- **Naives neglect** the component $-\delta h_p$.
- The **platforms' technology** is characterized by η_p and θ_p:
- Assumptions:
  - $\eta_p > \theta_p \ \ \forall p \in \{I, E\}$.
  - $\eta_I > \eta_E$ and $\theta_I > \theta_E$.
  - δ large enough so $U_p^r(h_p)$ decreases in h_p and $U_p^n(h_p)$ increases in h_p.

Analysis

User-optimal outcome

Lemma

User welfare is maximal if all users join the incumbent and $h_I = 0$.

Interpretation:

- On any platform p, a user's utility is maximal if $h_p = 0$ (given the assumptions).
- Incumbent has technological advantage $\implies$ it's optimal for all users to join the incumbent.

Migration & user welfare

Proposition (Harmful effects of differentiation)

In any pure-strategy equilibrium in which all users visit the incumbent, the utility of all users is strictly larger than in any other pure-strategy equilibrium.

Interpretation and intuition

- User migration is usually viewed as positive and a sign of healthy platform competition. Here, it is more nuanced.
- Intuition: If users split, platforms **differentiate their content** $\Longrightarrow$ intensity of competition $\downarrow$.

Mixed-strategy equilibrium

Sketch of the proof

- In any PSE in which **all users visit the incumbent**, all users obtain **strictly positive utility**.
  - Suppose rational users attain zero utility at the incumbent $\implies$ the entrant would set $h_E = 0$ and attract them.
  - True utility of naives at incumbent = utility of rationals.
- In any equilibrium in which the **entrant is visited by some users**, all users obtain **weakly negative utility.**
  - The platform j which naive users visit must set $h_j = 1$, since this maximizes the engagement and perceived utility of naive users.
  - Implication: Rational consumers would attain negative utility on platform j.
  - If rational consumers attain positive utility on platform p, the platform would marginally raise h_p to boost engagement.
  - Thus: rational users attain zero utility and naives get negative utility.

Equilibrium candidates

We define:

- $\tilde{h}_p$ harmful content share that leaves rationals at zero utility in platform p: $U_p^r(\tilde{h}_p) = 0$.
- $\check{h}_I$ harmful content share at I that leaves rationals indifferent between visiting I and visiting E with no harmful content: $U_I^r(\check{h}_p) = U_E^r(0)$.

Proposition (Equilibrium candidates)

There are three candidates for a PSE:

- *An equilibrium in which* $h_E^* = \tilde{h}_E^r$ *and* $h_I^* = 1$ *(naivety-focused).*
- *An equilibrium in which* $h_E^* = 1$ *and* $h_I^* = \tilde{h}_I^r$.
- *An equilibrium in which* $h_E^* = 0$, $h_I^* = \check{h}_I$ *(market dominance).*

Equilibrium existence

Proposition (PSE: existence)

The existence regions are as follows:

- *An equilibrium with* $h_E^* = \tilde{h}_E$ *and* $h_I^* = 1$ *exists if and only if* $\rho\pi_I^r(e_I^*(\tilde{h}_I)) + (1-\rho)\pi_I^n(e_I^*(\tilde{h}_I)) \leq (1-\rho)\pi_I^n(e_I^*(1))$.
- *An equilibrium in which* $h_E^* = 1$ *and* $h_I^* = \tilde{h}_I$ *exists iff (i)* $U_E^n(1) \geq U_I^n(\tilde{h}_I)$ *and (ii)* $\frac{\rho}{1-\rho} \in \left[\frac{\pi_I^n(e_I^*(1))}{\pi_I^r(e_I^*(\tilde{h}_I^)}, \frac{\pi_E^n(e_E^*(1))}{\pi_E^r(e_E^*(\tilde{h}_E))}\right]$ *jointly hold.*
- *An equilibrium in which* $h_E^* = 0$ *and* $h_I^* = \check{h}_I$ *exists iff the conditions (i)* $U_E^n(1) \leq U_I^n(\check{h}_I)$ *and (ii)* $(1-\rho)(\pi_I^n(e_I^*(1)) - \pi_I^n(e_I^*(\check{h}_I))) \leq \rho\pi_I^r(e_I^*(\check{h}_I))$ *jointly hold.*

The importance of awareness

Corollary (The importance of awareness)

There is a $\underline{\rho} > 0$ such that, if $\rho < \underline{\rho}$, there exists a unique equilibrium in which $h_I^* = 1$ and $h_E^* = \tilde{h}_E$.

Intuition:

- If the share of naives is large enough, the incumbent focuses on them and sets $h_I^* = 1$ (*naivety-focused equilibrium*).
- The entrant focuses on the rationals and offers $h_E^* = \tilde{h}_E$.

Leveling the playing field

Proposition (Leveling the playing field)

Suppose the incumbent has no competitive advantage. There is a $\rho^* \in (0,1)$ *such that, if* $\rho > \rho^*$*, there exists a unique PSE in which* $h_I^* = h_E^* = 0$*.*

Interpretation & intuition:

- The outcome in which $h_E^* = h_I^* = 0$ maximizes user welfare (under our assumptions).
- In this equilibrium, the most profitable deviation for any platform is to set $h = 1 \rightarrow$ this is not optimal if ρ is large enough.

Regulation: Awareness

- Our results suggest **complementarities** between regulation that closes the competitive gap and initiatives that **promote digital literacy.**
- If $\rho \approx 0$, diminishing the competitive gap will **not affect user welfare.**
- Even if platforms are symmetric, the user-optimal outcome only emerges if ρ is large enough.

Conclusion

Conclusion

- Main take-aways:
  - Regulation which **decreases the incumbent's competitive advantage** on social media platform markets **may have non-monotonic effects.**
  - There are important complementarities between competitive advantages and **user sophistication**.
- **Extensions:**
  - Multi-homing.
  - Differences in engagement.
  - Network effects.
  - Captive users.
  - Personalization of content.

Mixed-strategy equilibrium

Lemma (MSE with market dominance)

There exist no MSE in which all users join the entrant with probability 1. In any MSE in which all users join the incumbent with probability 1, the incumbent sets the harmful content share $\check{h}_I$ with probability 1.

Proposition (MSE: user welfare)

User welfare would be strictly smaller in any MSE in which some users join the entrant with positive probability than in an equilibrium in which all users join the incumbent with probability 1.

Back