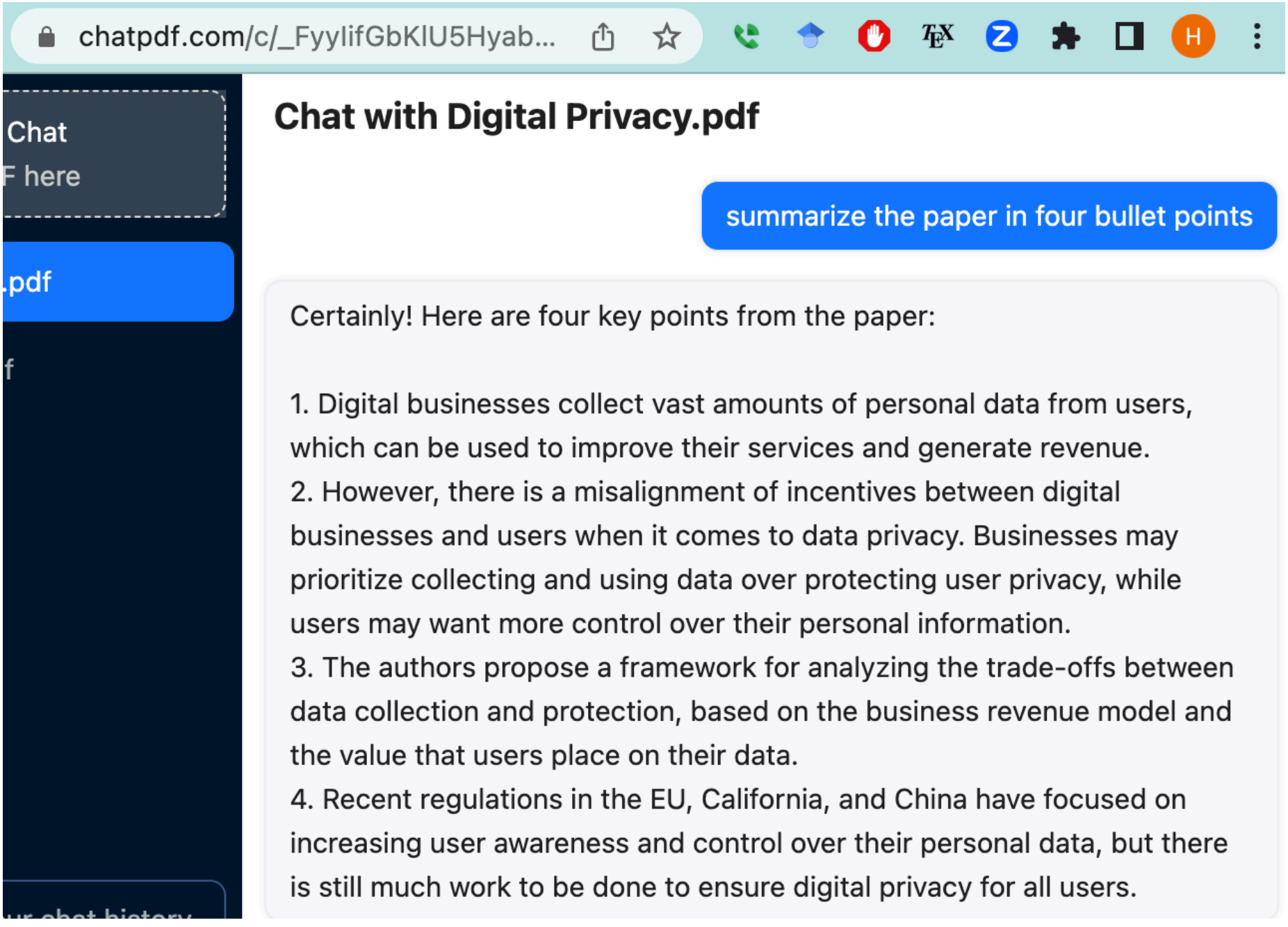
Digital Privacy
by Itay Fainmesser (JHU), Andrea Galeotti (LBS), Ruslan Momot (HEC/Kellogg/Michigan Ross)

Hanzhe Zhang
Michigan State University

Thursday, April 6, 2023

1. Introduction



Questions

- (Users) What are the trade-offs that people face when using online services? What are the determinants of harmful use of users' data?
- (Platform) What are the incentives of digital businesses to collect and protect users' data?
- (Government) What actions should be taken by regulatory authorities in order to protect consumer privacy and maximize social welfare?

Model: Timeline

1. Platform chooses **data strategy** (ξ, C) ("terms and agreements")
   - **data collection** $\xi \in [0, 1]$, proportion of user' activity to record and store
   - **data protection** $C \in [0, \infty)$, a costly investment
2. Users $I \ni i \in [0, 1]$ and $M \to \infty$ adversaries $J \ni j$ simultaneously decide
   - User i chooses **activity** a_i, and $\overline{a} = \int a_i di$ is aggregate activity.
   - Adversary j with characteristic $\gamma_j \sim U[0, M]$ chooses to **attack** or not.

Model: Payoffs

1. Platform chooses (ξ, C):

$$\Pi(\xi, C) = (1 - P) \cdot \bar{a} + P \cdot \xi \bar{a} - \psi C$$

   - $P = 1$: Purely usage-driven
   - $P = 0$: Purely data-driven
   - $P \in [0, 1]$: e.g., ad-driven Google Facebook, user activity + data-based ad

2. User i chooses a_i:

$$U(a_i, \bar{a}) = \underbrace{a_i - {a_i}^2/2}_{\text{usage payoff}} + \underbrace{\beta a_i \bar{a}}_{\text{network effect}} + \underbrace{(\rho - \omega) a_i \xi}_{\text{positive and negative info externalities}}$$

   where ω is the expected number of attacking adversaries.

3. Adversary j with characteristic $\gamma_j \sim U[0, M]$ attacks:

$$\pi(\gamma_j) = \xi \bar{a} - \gamma_j C$$

Equilibrium interactions of users and adversaries

Proposition 1

Fix Platform's data strategy (ξ, C). The ensuing subgame has a unique equilibrium in which users' and adversaries' activities are

$$\overline{a}^*(\xi, C) = \frac{C \cdot (1 + \rho\xi)}{C \cdot (1 - \beta) + \xi^2}, \text{ and}$$

$$\omega^*(\xi, C) = \frac{\xi}{C}\overline{a}^*(\xi, C) = \frac{\xi + \rho \cdot \xi^2}{C \cdot (1 - \beta) + \xi^2}.$$

Proof Sketch

$$\text{adversaries:} \quad \pi(\gamma_j) = \xi\overline{a} - C \geq 0 \Rightarrow \gamma_j \geq \frac{\xi}{C}\overline{a} \Rightarrow \omega^* = \frac{\xi}{C}\overline{a}$$

$$\text{users:} \quad 1 - a_i + \beta\overline{a} + (\rho - \frac{\xi}{C}\overline{a})\xi = 0.$$

Comparative statics of ξ and C

Users' activity $\bar{a}^*(\xi, C)$ and data collected $\xi\bar{a}^*(\xi, C)$ both

- increase in data protection C and
- first increase and then decrease in data collection ξ

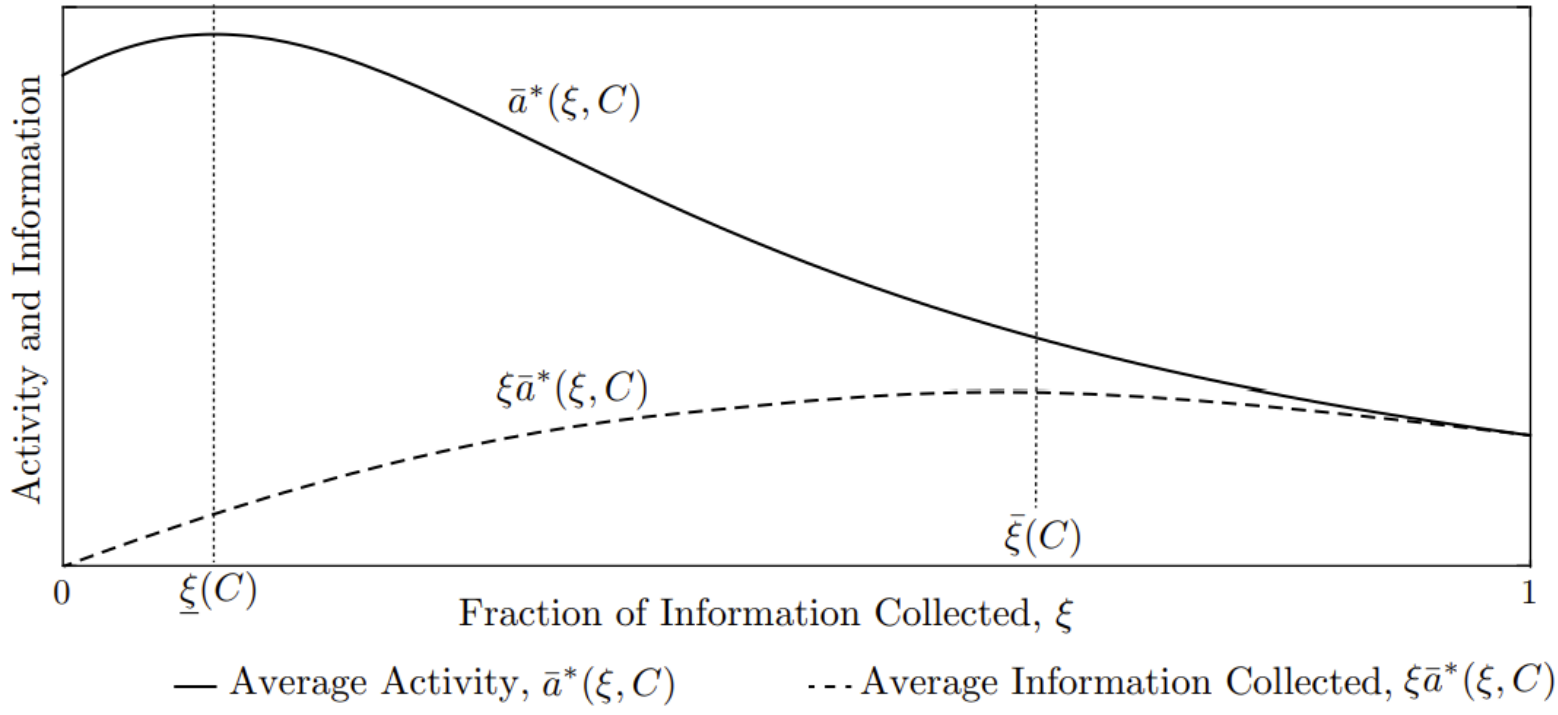


Figure 1: Users' activity $\bar{a}^*(\xi, C)$ and data collected $\xi\bar{a}^*(\xi, C)$ as a function of data collection ξ.

Explanation

- When $\xi \sim 0$ increases, positive externalities $\rho \cdot \xi$ increase from e.g., tailored services to users. An increased demand for platform service.
- When ξ continues to increases, negative externalities become relevant, as more data are available and more adversaries are incentivized to attack. At some point, the cost outweighs the benefit.
- Three regions:
  - Small $\xi < \underline{\xi}(C)$: users' activity and data collected both increase
  - Intermediate $\xi \in (\underline{\xi}(C), \overline{\xi}(C))$: users' activity decreases but data collected increases
  - Large $\xi > \overline{\xi}(C)$: users' activity and data collected both decrease
- Consumer surplus

$$\text{CS}(\xi, C) = \frac{1}{2}\overline{a}^*(\xi, C)^2$$

Platform's optimal collection and protection

$$\Pi(\xi, C) \quad = \quad (1 - P + P \cdot \xi) \cdot \frac{C \cdot (1 + \rho\xi)}{C \cdot (1 - \beta) + \xi^2} - \psi \cdot C$$

$$\xi^* = \max\{0, \min\{1, \frac{1}{2P\rho}(-(1-P)\rho - P + |P - (1-P)\rho| \cdot \sqrt{\frac{\psi}{\max\{0, \psi - P\rho\}}})\}\}$$

$$C^* = \frac{\xi^*}{1 - \beta}\left(-\xi^* + \sqrt{\frac{1}{\psi}(1 + \rho\xi^*)(1 - P(1 - \xi^*))}\right)$$

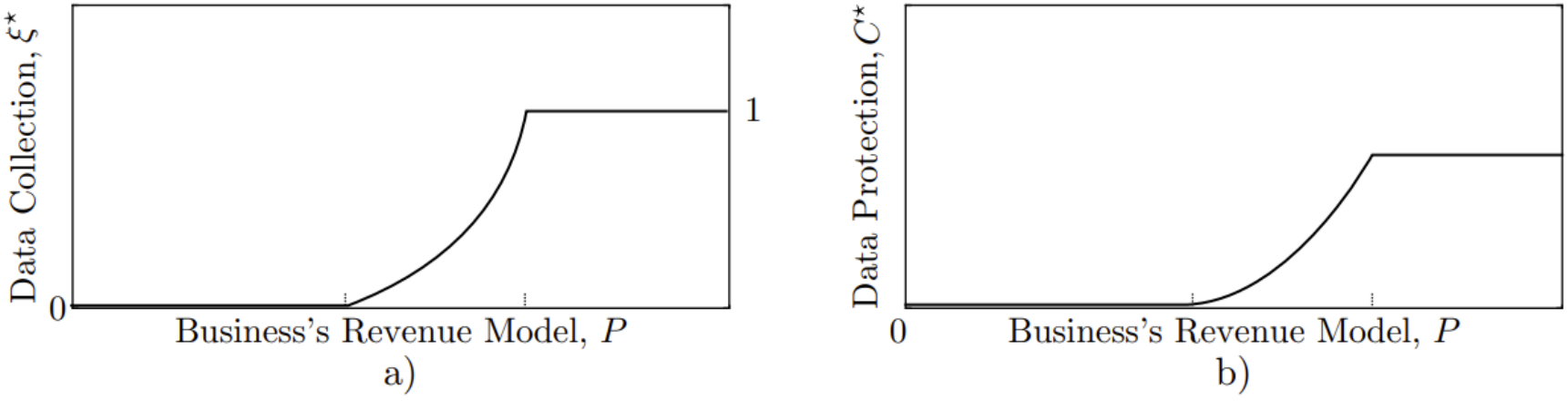
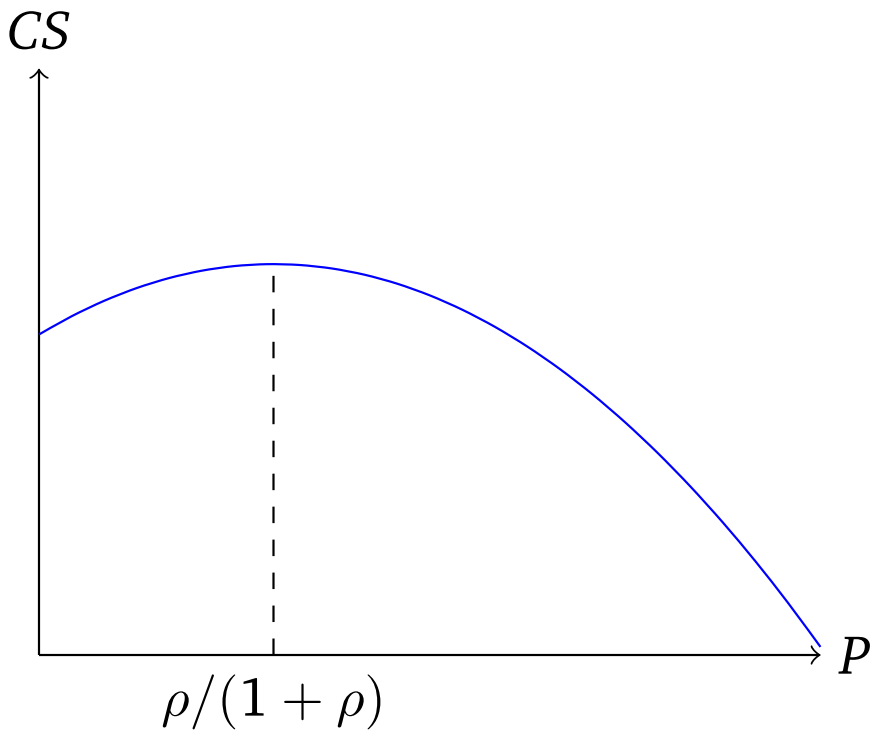


Figure 2: Equilibrium data collection and data protection strategies as a function of the business's revenue model, P (higher P implies a more data-driven business model).

Consumer surplus

Proposition 2

If $P < \rho/(1+\rho)$ [equivalently, $\rho > P/(1-P)$], then a more data-driven (higher P) platform generates weakly higher CS. Otherwise, CS weakly decreases with P.



Explanation

As P increases, two conflicting effects arise

1. Increase in data collection decreases users' activity. [stronger when ρ is smaller]
2. Increase in data protection increases users' activity. [stronger when ρ is larger]

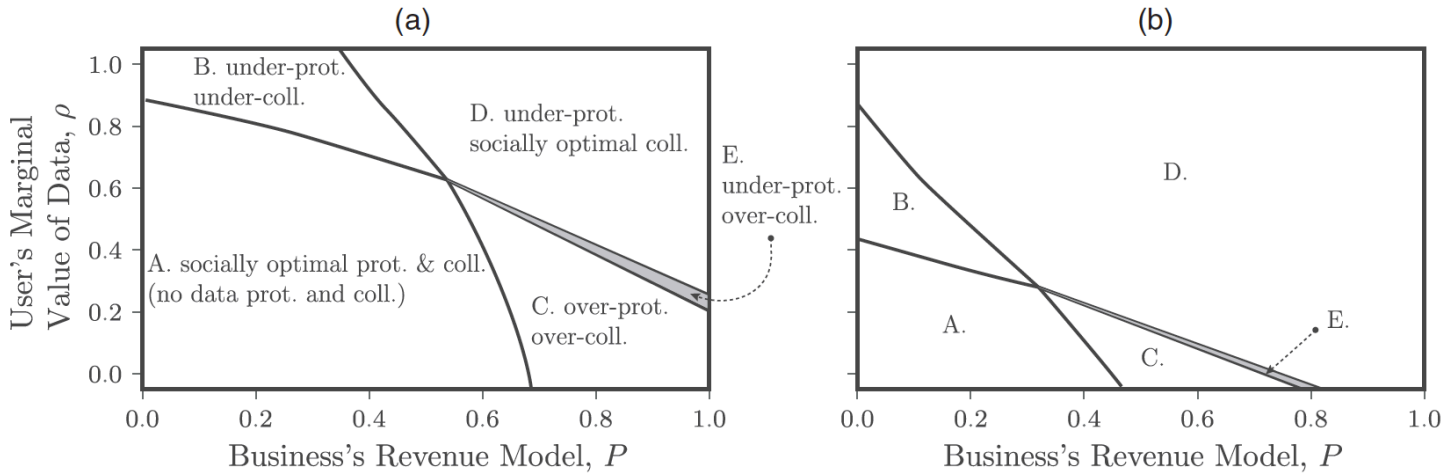
Government: no intervention comparison

Welfare-maximizing government:

$$(\xi^W, C^W) \in \text{argmax}_{\xi,C}\ \alpha \cdot \text{CS}(\xi, C) + (1-\alpha) \cdot \Pi(\xi, C).$$

- never undercollection $\xi^* \geq \xi^W$ and overprotection $C^* \geq C^W$.
- Usage-driven business $P = 0$: undercollection $\xi^* \leq \xi^W$ and underproctection $C^* \leq C^W$.

Figure 3. Regions of Inefficiencies in the Business's Data Strategy Relative to the Socially Optimal Strategy as a Function of the Business's Revenue Model, P, and Users' Marginal Value of Data, ρ



Note. The numerical example is generated with the following parameters: $\alpha = 0.5, \beta = 0$, and (a) $\psi = 0.35$ or (b) $\psi = 0.1$.

Government: optimal interventions

Proposition 4

The following two-pronged policy induces an efficient equilibrium.

- a required minimum data protection level $C_{\min} = C^W$.
  - (Proposition 3: only imposing this requirement, e.g., FTC may reduce CS)
- a liability fine or a data collection tax
  - a liability fine proportional to the expected damage from adversaries

$$\ell^* = \frac{\alpha}{1-\alpha} \frac{C^W}{2r(\xi^W, C^W)(\xi^W)^2}$$

  - a data collection tax rate

$$t^* = \frac{\alpha}{1-\alpha} \frac{\overline{a}^*(\xi^W, C^W)}{r(\xi^W, C^W)\xi^W}$$

Discussions

- Possibility of joint data collection (Appendix C.7)
  - Joint data collection and revenue decision may hurt consumers.
  - Merging dataset but keeping data collection decision (and revenue decision) separate does not affect consumers
- Pay for data (Appendix C.8)
  - Suppose users are paid. There may exist a regime in which users' incentives for activity are so strong that the platform can gain from an imposed payment by increasing data collection.

Contribution

- Different components of a business' data strategy.
- Compare data-driven and usage-driven revenue models.
- Positive and normative implications of data policy design.

Related literature

- "An active interdisciplinary area of research that studies the consequences for market outcomes of the ability of digital institutions to amass large data sets..."
- Related work
  - How user info is disclosed, traded, or inferred (Ali+, Bergemann+, Bonatti+, Fudenberg+, Ichihashi+): user action=user info.
  - Mechanisms to extract info: Fainmesser, Galeotti, Jullien, Ali Benabou, Elliott, Smolin, Bonatti
  - More recent work investigates the impact of privacy regulation on platforms and users: Argenziano Bonatti (2021), Bimpikis Ozdaglar Yildiz (2021 OR), Markovich Yehezkel (2021), Chen (2022)
  - Glaeser Scheinkman (2000), Non-market interactions. (users-adversaries game)

THANK YOU!