DIGITAL PRIVACY

Itay P. Fainmesser (Johns Hopkins University)
Andrea Galeotti (London Business School)
Ruslan Momot (HEC Paris)

User Data

Better products and services:

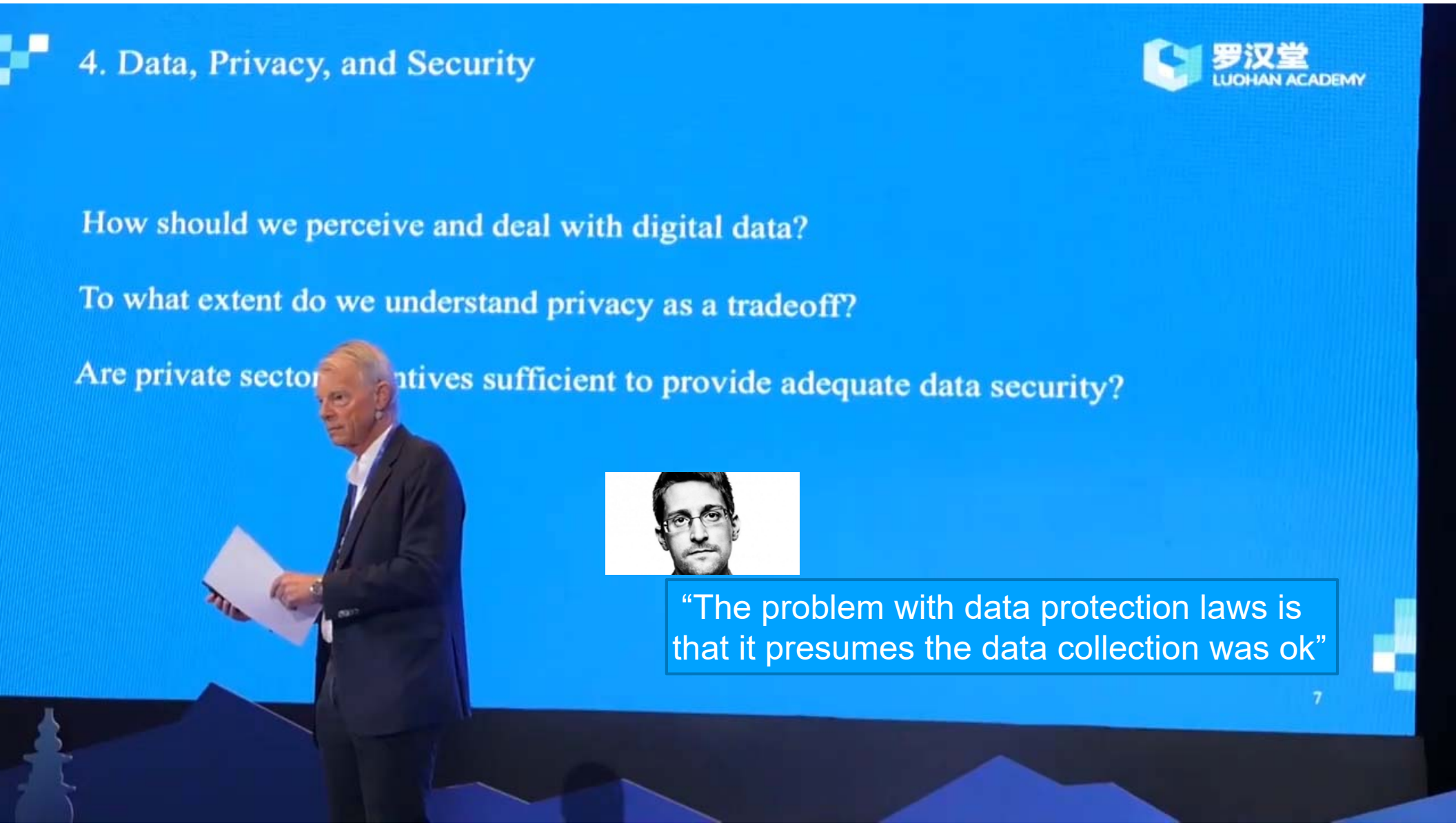
- better matching
- better ads and targeting
- safety
- …

Risk of data being misused

- identity theft
- manipulating public opinion
- government overreach
- …

RESEARCH QUESTIONS

RESEARCH QUESTIONS



RESEARCH QUESTIONS

1. Do digital businesses have the right incentive to collect and protect users' information?

   How do the incentives to collect and protect user data, and their welfare implications depend on the platform's revenue model?

2. What is the scope for regulation of data collection and protection?

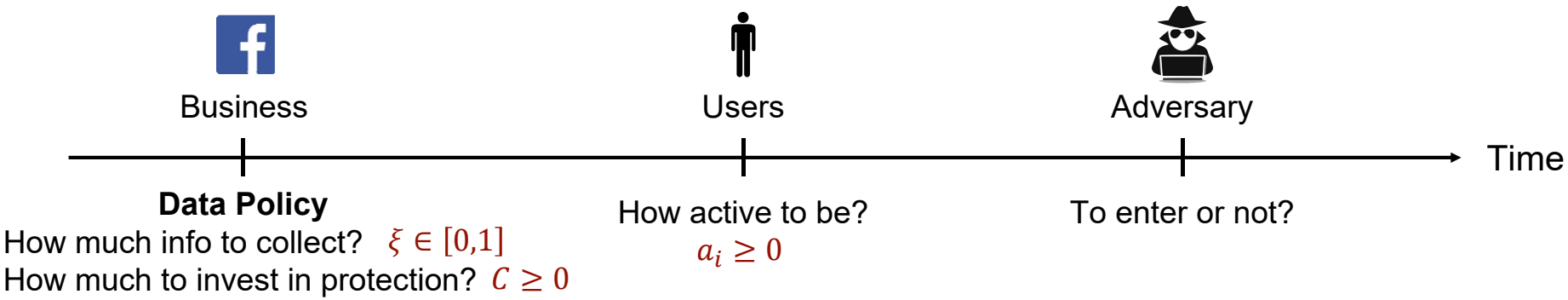
3. How much does data abuse actually hurts consumers?

RELATED LITERATURE

Differential Privacy — design of algorithmic mechanisms for anonymizing individual-level data	Dwork and Roth 2014; Cummings et al. 2015; Ghosh and Roth 2015; Abowd and Schmutte 2019.
How users' information is revealed through users' actions and how firms acquire that data.	Acquisti et al. 2016; Conitzer et al. 2012; Acemoglu et al. 2017; Bergemann, Bonatti, and Gan 2019; Ichihashi 2019a
How revealed information can be used to design personalized pricing, targeted advertisement, shape norms and, in turn affects individuals' behavior	Candogan et al. 2012; Fainmesser and Galeotti 2015; Miguel Villas-Boas 2017; Goldfarb and Tucker 2011; Koh et al. 2015; Jann and Schottmüller 2016; Gradwohl 2019; Ichihashi 2019; Ali and Bénabou 2016; Ichihashi 2019b; Liu, Sockin, and Xiong 2020; Anderson, Baik, Larson 2019; Tirole 2020
The Data Economy	Farboodi and Veldkamp 2019…

MODEL OVERVIEW

ACTIONS AND TIMELINE



USERS

A continuum of mass 1 of users

$$U_i(a_i) = a_i b_i - \frac{1}{2} a_i^2 + \beta a_i \bar{a} + \rho \cdot a_i \xi - \omega_i \cdot a_i \xi$$

ω_i	— expected number of adversarial against user i
$a_i \xi$	— data collected on user i

USERS

A continuum of mass 1 of users

Direct benefit from usage (e.g., messaging, photos, etc.)

Benefit from the business using user's information (e.g., advertisements, better matches)

$$U_i(a_i) = a_i b_i - \frac{1}{2} a_i^2 + \beta a_i \bar{a} + \rho \cdot a_i \xi - \omega_i \cdot a_i \xi$$

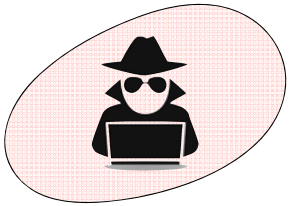
Positive network effects

Expected loss from adversarial activity

ω_i — expected number of adversarial against user i

$a_i \xi$ — data collected on user i

ADVERSARIES



— mass K of adversaries (large K)

Adversaries are heterogeneous in their ability to hack security systems $\gamma_j \sim U[0, K]$

If adversary j chooses to enter, he pays a fixed cost $\gamma_j C$
and takes action against one user i chosen u.a.r.

$$\pi_{j \to i} = a_i \xi - \gamma_j C$$

Benefit to adversary j from action against user i

Cost of hacking the system

$\Rightarrow$ Adversary participates if: $\gamma_j \leq \frac{\overline{a}\xi}{C}$

(where $\overline{a} = average\{a_i\}_i$)

DIGITAL BUSINESS

$$\Pi = \Phi(\overline{a}, \overline{a}\xi) - \psi C$$

User Activity

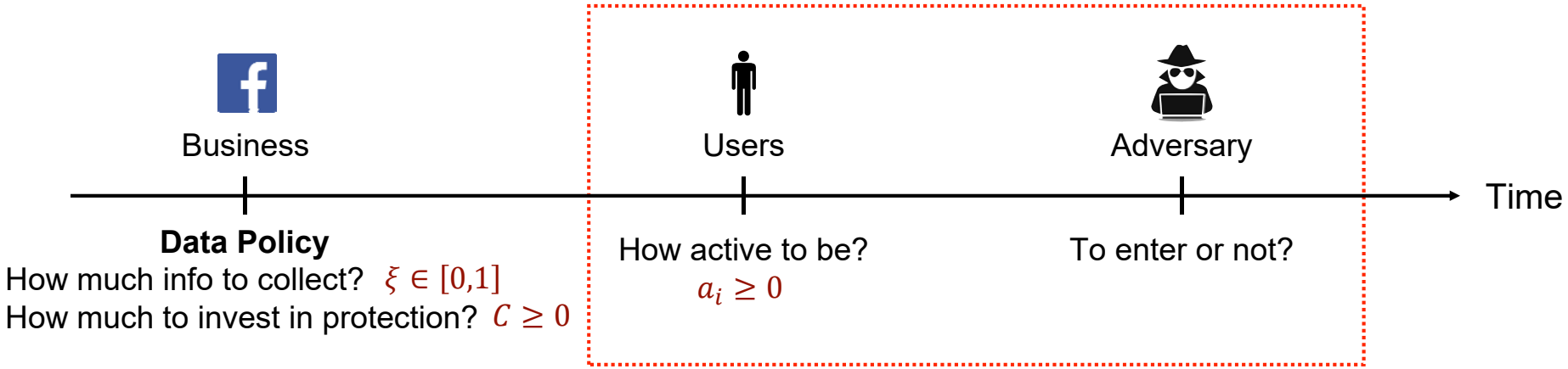
Information collected

We assume that Φ:
1. Increases in both elements
2. Concave in both elements

Transaction-driven business $\Phi'_y = 0$

Advertisement-driven business $\Phi'_x = 0$

TIMELINE


Business
Users
Adversary
Time
Data Policy
How much info to collect? $\xi \in [0,1]$
How much to invest in protection? $C \geq 0$
How active to be?
$a_i \geq 0$
To enter or not?

EQUILIBRIUM

Proposition: There exists a unique equilibrium. In equilibrium, average user activity is $\bar{a}^* = \frac{C(\bar{b}+\rho\xi)}{C(1-\beta)+\xi^2}$ and $\omega_i^* = \frac{\bar{a}^*\xi}{C}$ adversaries enter.

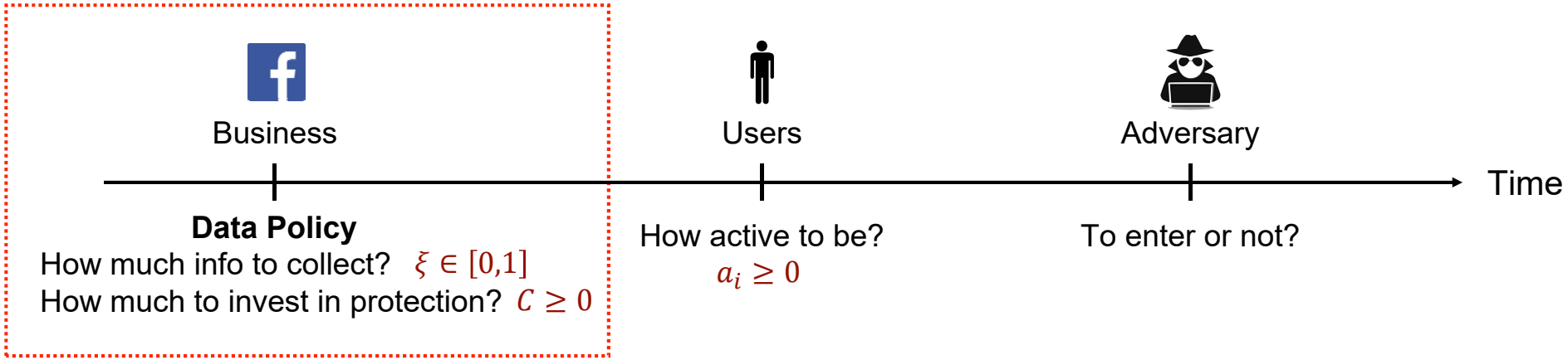
Negative network effects

$$U_i(a_i) = a_i b_i - \frac{1}{2}a_i^2 + \beta a_i \bar{a} + \rho \cdot a_i \xi - \omega^* \cdot a_i \xi$$

Observation: Consumer surplus is proportional to $\bar{a}^{*2}$

Proof: $CS = \frac{1}{2}\int a_i^{*2} = \frac{1}{2}\left(\sigma_b^2 + \bar{a}^{*2}\right)$ where σ_b^2 is exogenous.

TIMELINE


Business
Users
Adversary
Time
Data Policy
How much info to collect? $\xi \in [0,1]$
How much to invest in protection? $C \geq 0$
How active to be?
$a_i \geq 0$
To enter or not?

DATA COLLECTION (EXOGENOUS C)

Let ξ_W and ξ_{CS} be the welfare and consumer surplus maximizing levels of ξ respectively.

Proposition: Let C be fixed.

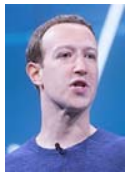
1. A business collects more data than the level that maximizes welfare and consumer surplus. That is $\xi^* \geq \xi_W, \xi_{CS}$.
2. A transaction-driven business collects the efficient amount of data given the level of data protection. That is, $\xi^* = \xi_W, \xi_{CS}$.

DATA COLLECTION (EXOGENOUS C)

Let ξ_W and ξ_{CS} be the welfare and consumer surplus maximizing levels of ξ respectively.

Proposition: Let C be fixed.

1. A business collects more data than the level that maximizes welfare and consumer surplus. That is $\xi^* \geq \xi_W, \xi_{CS}$.
2. A transaction-driven business collects the efficient amount of data given the level of data protection. That is, $\xi^* = \xi_W, \xi_{CS}$.



*"there's the important question of whether the **advertising model** encourages companies like ours [Facebook] to **use and store more information** that we otherwise would."*

DATA COLLECTION (EXOGENOUS C) – REGULATION

Currently in the US

Regulation: minimum level of data protection $(\underline{C})$.

Enforcement:

1. Common: fine for data breaches (liability).
2. Rare: minimum level of data protection $(\underline{C})$.

DATA COLLECTION (EXOGENOUS C) – REGULATION

Currently in the US

Regulation: minimum level of data protection $(\underline{C})$.

Enforcement:

1. Common: fine for data breaches (liability).
2. Rare: minimum level of data protection $(\underline{C})$.

$$\Pi = \Phi(\overline{a}, \overline{a}\xi) - \psi C - \boldsymbol{L \cdot \omega \bar{a} \xi}$$

Proposition: Let C be fixed.

Welfare can be maximized by fines on data breaches, equal to

$$\mathrm{L} = -\frac{\partial CS/\partial \xi}{\partial(\omega \bar{a}\xi)/\partial \xi} \Big|_{\xi_W}$$

Consumer surplus can be maximized by fines on data breaches, equal to

$$\mathrm{L} = \frac{2}{\rho}\Phi'_y \Big|_{\xi_{CS}}$$

DATA COLLECTION AND PROTECTION

		Data collection (ξ)	Data collection and protection (C, ξ)
General	Direction of inefficiency	$(+)$	
	Effective regulation	Liability	
Transaction-driven business	Direction of inefficiency	No inefficiency	
	Effective regulation	N\A	

DATA COLLECTION AND PROTECTION

		Data collection (ξ)	Data collection and protection (C, ξ)
General	Direction of inefficiency	$(+)$	$(-,-)\,;(-,+)\,;(+,+)$
	Effective regulation	Liability	$\underline{C}$ + liability
Transaction-driven business	Direction of inefficiency	No inefficiency	$(-,-)$
	Effective regulation	N\A	$\underline{C}$

Observations:

1. In general, digital businesses may collect and protect too much or too little.
2. In general, a lower bound on protection is insufficient and could have perverse effects if used by itself.
3. FTC's enforcement may be efficiency enhancing.*

THE LOSS MULTIPLIER OF ADVERSARIAL ACTIVITY

Not that much - right? (for entire UK)

$$U_i(a_i) = a_i b_i - \frac{1}{2} a_i^2 + \beta a_i \bar{a} + \rho \cdot a_i \xi - \omega \cdot a_i \xi$$

Direct loss

The direct loss ignores:

- The change in activity when users are fearful of adversaries
- A respective change in businesses' data policies

Taking into account all these effects:

$$\text{Loss Multiplier (LM)} = \frac{CS_{\textbf{without}}\ \text{adversaries} - CS_{\textbf{with}}\ \text{adversaries}}{\text{Direct loss}} \geq \frac{2}{1-\beta}$$

The change in consumer surplus due to the presence of adversaries:

- is **at least 2 times** direct loss;
- quickly **explodes** with strength of network effects β

RESEARCH QUESTIONS

1. Do digital businesses have the right incentive to collect and protect users' information?

   How do the incentives to collect and protect user data, and their welfare implications depend on the platform's revenue model?

2. What is the scope for the regulation of data collection and protection?

3. How much does data abuse actually hurts consumers?

DISCUSSION

Data protections policies (e.g., GDPR)

- In equilibrium data protection and data collection are complements

Potential regulatory / firm actions

- Splitting up Facebook
- Pay for data
- Restricting data collection (also GDPR)

Other questions

- Competition
- Digital equity and data policies
  - an increase in ξ can lead to an increase in $a_i\xi$ for some (high b_i) users and a decrease for others.

THANK YOU!