Competitive Markets for Personal Data

Simone Galperti   Jacopo Perego

UCSD   Columbia

January 2023

Personal data fuels multi-billion dollars industries — essential input in digital economy

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates

- What market arrangements could correct them

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates

- What market arrangements could correct them

Preview of Our Results:

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates

- What market arrangements could correct them

Preview of Our Results:

1. In status quo, many inefficiencies lead to mkt failure

Overview

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates

- What market arrangements could correct them

Preview of Our Results:

1. In status quo, many inefficiencies lead to mkt failure

2. Giving people control over their data does not solve the problem, and in fact can backfire

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates

- What market arrangements could correct them

Preview of Our Results:

1. In status quo, many inefficiencies lead to mkt failure

2. Giving people control over their data does not solve the problem, and in fact can backfire

Overview

Personal data fuels multi-billion dollars industries — essential input in digital economy

Status quo: Consumers are imperfectly compensated for their data, and have limited control over their use

We study **competitive markets** for personal data to understand

- What inefficiencies this status quo generates
- What market arrangements could correct them

Preview of Our Results:

1. In status quo, many inefficiencies lead to mkt failure
2. Giving people control over their data does not solve the problem, and in fact can backfire
3. Efficiency can be achieved by opening more markets   ⤳ how practical?

model

for talk, simplify model a bit

for talk, simplify model a bit

There are I competing e-commerce **platforms**

for talk, simplify model a bit

There are I competing e-commerce **platforms**

There is a third-party **vendor** who is active on platform i and sells a product and sets its price a

for talk, simplify model a bit

There are I competing e-commerce **platforms**

There is a third-party **vendor** who is active on platform i and sells a product and sets its price a

There is a continuum of heterogeneous **consumers**: unit demand and WTP ω

for talk, simplify model a bit

There are I competing e-commerce **platforms**

There is a third-party **vendor** who is active on platform i and sells a product and sets its price a

There is a continuum of heterogeneous **consumers**: unit demand and WTP ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

for talk, simplify model a bit

There are I competing e-commerce **platforms**

There is a third-party **vendor** who is active on platform i and sells a product and sets its price a

There is a continuum of heterogeneous **consumers**: unit demand and WTP ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

When a ω-consumer transacts with platform i's vendor, payoffs realize

for talk, simplify model a bit

There are I competing e-commerce **platforms**

There is a third-party **vendor** who is active on platform i and sells a product and sets its price a

There is a continuum of heterogeneous **consumers**: unit demand and WTP ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

When a ω-consumer transacts with platform i's vendor, payoffs realize

$$
\begin{array}{rll}
\text{consumer:} & g_i(a,\omega) & \text{e.g. surplus} \\
\text{vendor:} & \pi_i(a,\omega) & \text{e.g. profit} \\
\text{platform:} & u_i(a,\omega) & \text{e.g. intermediation fee}
\end{array}
$$

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. info about this consumer's WTP ω

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. info about this consumer's WTP ω (e.g., age, gender, edu, etc.)

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. info about this consumer's WTP ω (e.g., age, gender, edu, etc.)

2. access to this consumer

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. info about this consumer's WTP ω (e.g., age, gender, edu, etc.)

2. access to this consumer (e.g., email, IP address, etc.)

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. info about this consumer's WTP ω

2. access to this consumer

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. <u>Complete</u> info about this consumer's WTP ω

2. access to this consumer

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. Complete info about this consumer's WTP ω

2. access to this consumer

From 1. Type of a record $==$ Type of a consumer   (simplifies notation)

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. <u>Complete</u> info about this consumer's WTP ω

2. <u>Exclusive</u> access to this consumer

From 1. Type of a record $==$ Type of a consumer   (simplifies notation)

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. <u>Complete</u> info about this consumer's WTP ω

2. <u>Exclusive</u> access to this consumer

From 1. Type of a record $==$ Type of a consumer   (simplifies notation)

From 2. Data record is rival good

Each consumer is uniquely identified by a **data record**, which gives its owner:

1. Complete info about this consumer's WTP ω

2. Exclusive access to this consumer

From 1. Type of a record $==$ Type of a consumer   (simplifies notation)

From 2. Data record is rival good

A collection of data records is called a **database**: denoted $q_i \in \mathbb{R}_+^{\Omega}$

Suppose platform i has acquired database q_i

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Suppose platform i has acquired database q_i

  i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

$$\text{such that:} \quad \sum_{\omega} \big(\pi_i(a, \omega) - \pi_i(\hat{a}, \omega) \big) x_i(a, \omega) \geq 0 \qquad \forall\, a, \hat{a} \in A$$

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad\qquad \forall\, \omega \in \Omega$$

Suppose platform i has acquired database q_i

   i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

$$\text{such that:} \quad \sum_{\omega} \big(\pi_i(a, \omega) - \pi_i(\hat{a}, \omega) \big) x_i(a, \omega) \geq 0 \qquad \forall\, a, \hat{a} \in A$$

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad\qquad \forall\, \omega \in \Omega$$

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

platform's
exptd payoff

$$\overbrace{U_i(q_i)} \quad = \max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

$$\text{such that:} \quad \sum_{\omega} \big(\pi_i(a, \omega) - \pi_i(\hat{a}, \omega) \big) x_i(a, \omega) \geq 0 \qquad \forall \, a, \hat{a} \in A$$

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad\qquad \forall \, \omega \in \Omega$$

Suppose platform i has acquired database q_i

    i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\overbrace{U_i(q_i)}^{\substack{\text{platform's}\\\text{exptd payoff}}} = \max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

$$\text{such that:} \quad \sum_{\omega} \big(\pi_i(a, \omega) - \pi_i(\hat{a}, \omega) \big) x_i(a, \omega) \geq 0 \qquad \forall \, a, \hat{a} \in A$$

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad\qquad\qquad \forall \, \omega \in \Omega$$

Denote by $x_{q_i}^*$ a solution (note: it depends on the entire database)

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

platform's
exptd payoff

$$\overbrace{U_i(q_i)} = \max_{x_i: A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

$$\text{such that:} \quad \sum_\omega \big(\pi_i(a, \omega) - \pi_i(\hat{a}, \omega)\big) x_i(a, \omega) \geq 0 \qquad \forall \, a, \hat{a} \in A$$

$$\sum_a x_i(a, \omega) = q_i(\omega) \qquad\qquad \forall \, \omega \in \Omega$$

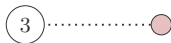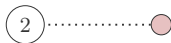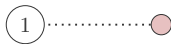Denote by $x_{q_i}^*$ a solution (note: it depends on the entire database)

platforms

1

2

3

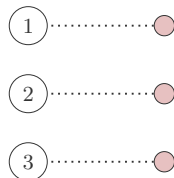platforms · vendors

1 .... ○

2 .... ○

3 .... ○

consumers

ω

ω'

ω''

platforms   vendors

1 .... ○

2 .... ○

3 .... ○

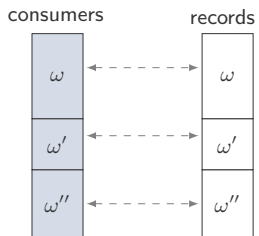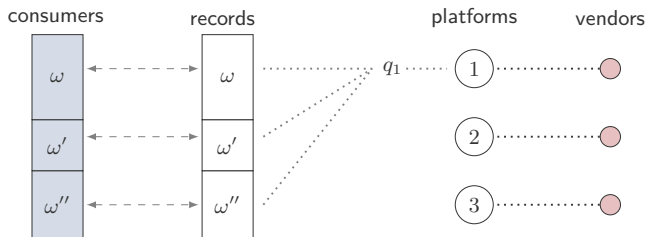consumers   records   platforms   vendors
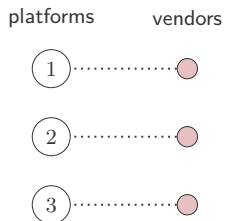
Timing:
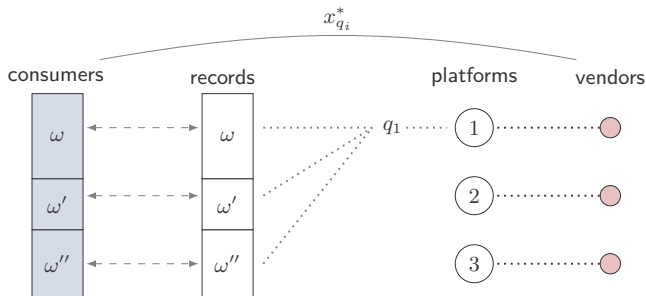
1. Each platform acquires q_i in a **competitive market** for data records

Timing:

1. Each platform acquires q_i in a **competitive market** for data records

2. Given database q_i, platform i chooses $x_{q_i}^*$
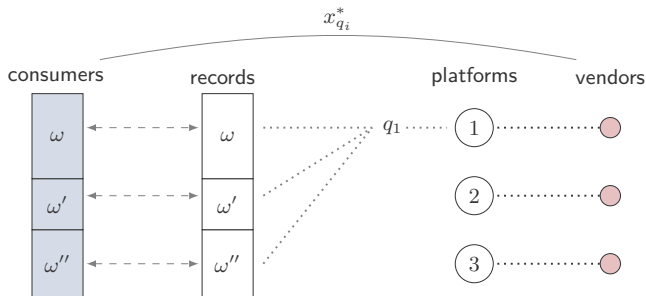
Timing:

1. Each platform acquires q_i in a **competitive market** for data records

2. Given database q_i, platform i chooses $x_{q_i}^*$

3. Payoffs realize

economy $\mathcal{E}_1$

a "status-quo" economy

In this economy:

- Consumers "expropriated" of their records: no control, imperfect compns

- Platforms trade records among each other, taking prices as given

In this economy:

- ▶ Consumers "expropriated" of their records: no control, imperfect compns

- ▶ Platforms trade records among each other, taking prices as given

Competitive Equilibrium of Economy $\mathcal{E}_1$:

Prices $p^* \in \mathbb{R}_+^\Omega$ and a data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ are an equilibrium of $\mathcal{E}_1$ if:

In this economy:

- ▶ Consumers "expropriated" of their records: no control, imperfect compns

- ▶ Platforms trade records among each other, taking prices as given

Competitive Equilibrium of Economy $\mathcal{E}_1$:

Prices $p^* \in \mathbb{R}_+^{\Omega}$ and a data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ are an equilibrium of $\mathcal{E}_1$ if:

1. Data allocation is feasible $\qquad$ for all ω, $\sum_{i \in I} q_i^*(\omega) \leq \bar{q}(\omega)$

In this economy:

► Consumers "expropriated" of their records: no control, imperfect compns

► Platforms trade records among each other, taking prices as given

---

Competitive Equilibrium of Economy $\mathcal{E}_1$:

Prices $p^* \in \mathbb{R}_+^{\Omega}$ and a data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ are an equilibrium of $\mathcal{E}_1$ if:

1. Data allocation is feasible $\qquad\qquad$ for all ω, $\displaystyle\sum_{i \in I} q_i^*(\omega) \leq \bar{q}(\omega)$

2. All markets clear $\qquad$ for all ω, $\displaystyle p^*(\omega)\Big(\bar{q}(\omega) - \sum_i q_i^*(\omega)\Big) = 0$

In this economy:

▶ Consumers "expropriated" of their records: no control, imperfect compns

▶ Platforms trade records among each other, taking prices as given

Competitive Equilibrium of Economy $\mathcal{E}_1$:

Prices $p^* \in \mathbb{R}_+^\Omega$ and a data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ are an equilibrium of $\mathcal{E}_1$ if:

1. Data allocation is feasible $\qquad\qquad$ for all ω, $\displaystyle\sum_{i \in I} q_i^*(\omega) \leq \bar{q}(\omega)$

2. All markets clear $\qquad$ for all ω, $p^*(\omega)\Big(\bar{q}(\omega) - \displaystyle\sum_i q_i^*(\omega)\Big) = 0$

3. Platforms maximize given prices $\quad q_i^* \in \arg\max_{q_i} U_i(q_i) - \displaystyle\sum_\omega p^*(\omega)q_i(\omega)$

Observation. Platform's payoff $U_i(q_i)$ depends only on acquired database, not on other platforms' databases

Observation. Platform's payoff $U_i(q_i)$ depends only on acquired database, not on other platforms' databases

Some Properties of $\mathcal{E}_1$:

- Equilibria maximize sum of platforms' payoffs

- Any platforms-optimal data allocation can be supported in eqm

Observation. Platform's payoff $U_i(q_i)$ depends only on acquired database, not on other platforms' databases

Some Properties of $\mathcal{E}_1$:

- Equilibria maximize sum of platforms' payoffs

- Any platforms-optimal data allocation can be supported in eqm

Problems:

Consumers are excluded and platforms do not internalize how their decisions (i.e., q_i^* and $x_{q_i}^*$) affect consumers

Welfare of consumers and platforms is "third best"

economy $\mathcal{E}_2$

give consumers control?

Consumer initially <u>owns</u> her data record.

▶ She decides whether to sell it, and to which platform

▶ If record is sold, consumer is paid market price $p^*(\omega)$

▶ If not, consumer enjoys some reservation utility ($\approx$ value of privacy)

Consumer initially <u>owns</u> her data record.

▶ She decides whether to sell it, and to which platform

▶ If record is sold, consumer is paid market price $p^*(\omega)$

▶ If not, consumer enjoys some reservation utility ($\approx$ value of privacy)

Competitive Equilibrium of Economy $\mathcal{E}_2$: informal

Prices $p^* \in \mathbb{R}^\Omega$ and data allocation $q^* \in \mathbb{R}_+^{\Omega \times (I+1)}$ are an equilibrium if:

Consumer initially <u>owns</u> her data record.

- ▶ She decides whether to sell it, and to which platform

- ▶ If record is sold, consumer is paid market price $p^*(\omega)$

- ▶ If not, consumer enjoys some reservation utility   ($\approx$ value of privacy)

Competitive Equilibrium of Economy $\mathcal{E}_2$: informal

Prices $p^* \in \mathbb{R}^\Omega$ and data allocation $q^* \in \mathbb{R}_+^{\Omega \times (I+1)}$ are an equilibrium if:

1. Data allocation is feasible

2. All markets clear

3. All platforms maximize given p^*

Consumer initially <u>owns</u> her data record.

- ▶ She decides whether to sell it, and to which platform

- ▶ If record is sold, consumer is paid market price $p^*(\omega)$

- ▶ If not, consumer enjoys some reservation utility ($\approx$ value of privacy)

Competitive Equilibrium of Economy $\mathcal{E}_2$: informal

Prices $p^* \in \mathbb{R}^\Omega$ and data allocation $q^* \in \mathbb{R}_+^{\Omega \times (I+1)}$ are an equilibrium if:

1. Data allocation is feasible

2. All markets clear

3. All platforms maximize given p^*

4. Consumers maximize given p^* and behavior of other consumers

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x_{q_i}^*$ depends on entire database q_i: ω' consumers affects ωs payoff

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

— In general, equilibria are inefficient

Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

— In general, equilibria are inefficient

Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

— In particular, equilibria can be especially inefficient:

Welfare in $\mathcal{E}_2$ can be even lower than in $\mathcal{E}_1$!

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

− In general, equilibria are inefficient

  Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

− In particular, equilibria can be especially inefficient:

  Welfare in $\mathcal{E}_2$ can be even lower than in $\mathcal{E}_1$!

− Even in best-case scenario of $\mathcal{E}_2$, equilibria are still inefficient

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

— In general, equilibria are inefficient

  Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

— In particular, equilibria can be especially inefficient:

  Welfare in $\mathcal{E}_2$ can be even lower than in $\mathcal{E}_1$!

— Even in best-case scenario of $\mathcal{E}_2$, equilibria are still inefficient   second-best

Observation. Why does a consumer's payoff depend on what other consumers do?

▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

— In general, equilibria are inefficient   below second-best

  Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

— In particular, equilibria can be especially inefficient:

  Welfare in $\mathcal{E}_2$ can be even lower than in $\mathcal{E}_1$!

— Even in best-case scenario of $\mathcal{E}_2$, equilibria are still inefficient   second-best

Observation. Why does a consumer's payoff depend on what other consumers do?

- ▶ $x^*_{q_i}$ depends on entire database q_i: ω' consumers affects ωs payoff

Properties of $\mathcal{E}_2$:

— In general, equilibria are inefficient   below second-best

  Consumer is compensated for value her record generates for platform, not for value it generates for other consumers

— In particular, equilibria can be especially inefficient: below third-best

  Welfare in $\mathcal{E}_2$ can be even lower than in $\mathcal{E}_1$!

— Even in best-case scenario of $\mathcal{E}_2$, equilibria are still inefficient   second-best

economy $\mathcal{E}_3$

trading how data is used

Why is $\mathcal{E}_2$ at most second-best?

Once i acquires q_i, it can use it as it wants. How data is <u>used</u> is not contractible in $\mathcal{E}_2$ (akin to moral hazard)

Why is $\mathcal{E}_2$ at most second-best?

  Once i acquires q_i, it can use it as it wants. How data is <u>used</u> is not
  contractible in $\mathcal{E}_2$ (akin to moral hazard)

We further enrich the economy:

▶ Consumer can sell her record for a <u>specific use</u> $x_i(a, \omega) \geq 0$

Why is $\mathcal{E}_2$ at most second-best?

> Once i acquires q_i, it can use it as it wants. How data is <u>used</u> is not
> contractible in $\mathcal{E}_2$ (akin to moral hazard)

We further enrich the economy:

▶ Consumer can sell her record for a <u>specific use</u> $x_i(a, \omega) \geq 0$

▶ A richer price system is needed:

> $p_i(a, \omega)$ – platform & use specific prices for each record

Why is $\mathcal{E}_2$ at most second-best?

> Once i acquires q_i, it can use it as it wants. How data is <u>used</u> is not
> contractible in $\mathcal{E}_2$ (akin to moral hazard)

We further enrich the economy:

▶ Consumer can sell her record for a <u>specific use</u> $x_i(a, \omega) \geq 0$

▶ A richer price system is needed:

> $p_i(a, \omega)$ – platform & use specific prices for each record

Competitive Equilibrium of Economy $\mathcal{E}_3$:

See paper   Follows Arrow (1969), Laffont (1976)

Observation: Since consumer trades directly how her record is used:

– She does not care how platform uses <u>other</u> records

– "Moral hazard" is no longer an issue

Observation: Since consumer trades directly how her record is used:

- She does not care how platform uses <u>other</u> records

- "Moral hazard" is no longer an issue

Properties of $\mathcal{E}_3$

- Equilibria are first-best efficient

Observation: Since consumer trades directly how her record is used:

   – She does not care how platform uses <u>other</u> records

   – "Moral hazard" is no longer an issue

Properties of $\mathcal{E}_3$

   – Equilibria are first-best efficient

   – Every first-best allocation can be supported as an equilibrium

Observation: Since consumer trades directly how her record is used:

– She does not care how platform uses <u>other</u> records

– "Moral hazard" is no longer an issue

Properties of $\mathcal{E}_3$

– Equilibria are first-best efficient

– Every first-best allocation can be supported as an equilibrium

Problems with $\mathcal{E}_3$?

Its realism

Observation: Since consumer trades directly how her record is used:

- She does not care how platform uses <u>other</u> records

- "Moral hazard" is no longer an issue

Properties of $\mathcal{E}_3$

- Equilibria are first-best efficient

- Every first-best allocation can be supported as an equilibrium

Problems with $\mathcal{E}_3$?

Its realism $\rightsquigarrow$ **Open Questions:** how to (partially) decentralize in practice?

conclusions

Summary

A framework to study competitive markets for personal data:

$\mathcal{E}_1$. In status quo, multiple sources of inefficiencies jointly lead to mkt failure

$\mathcal{E}_2$. Giving people control over their data does not correct all inefficiencies and, in fact, can backfire

$\mathcal{E}_3$. Efficiency achieved by opening many more markets $\rightsquigarrow$ how practical?

thank you