# Data and Privacy: Tools from Information Design

Jacopo Perego
Columbia University

January 2025

# Economics of Data



Personal data has become key input of the modern economy

**Economics of Data**:

- Implications of collecting, trading, using personal data on economics outcomes, e.g., consumer welfare, market power, etc.
- Institutions, e.g., data markets, privacy laws, data unions, etc.

# Economics of Data



Personal data has become key input of the modern economy

**Economics of Data**:

- Implications of collecting, trading, using personal data on economics outcomes, e.g., consumer welfare, market power, etc.
- Institutions, e.g., data markets, privacy laws, data unions, etc.

Spanning fields: micro and macro, theoretical and empirical, field and lab

A space where micro theory can aim to be sophisticated and practically relevant

# Economics of Data



Personal data has become key input of the modern economy

**Economics of Data**:

- Implications of collecting, trading, using personal data on economics outcomes, e.g., consumer welfare, market power, etc.
- Institutions, e.g., data markets, privacy laws, data unions, etc.

Spanning fields: micro and macro, theoretical and empirical, field and lab

A space where micro theory can aim to be sophisticated and practically relevant

**Today ⇒ A Methodological Point**

How/why tools from **information design** constitute a building block to study questions concerning the **economics data** (plus some future directions)

# model

# Model

The typical **Information Design** setting:

- A finite set $\Omega$ of payoff states (e.g., demand condition)
- A common prior belief of about $\omega$, denoted $q \in \Delta(\Omega)$
- One "designer," $N$ agents (e.g., e-com platform and merchants)
- Finite action sets $A_0 \times A_1 \times \ldots \times A_N =: A$ (e.g., price, feature, quality)
- Payoffs: $v : A \times \Omega \to \mathbb{R}$ for the designer; $\pi_i : A \times \Omega \to \mathbb{R}$ for each agent

# Model

The typical **Information Design** setting:

- A finite set $\Omega$ of payoff states (e.g., demand condition)
- A common prior belief of about $\omega$, denoted $q \in \Delta(\Omega)$
- One "designer," $N$ agents (e.g., e-com platform and merchants)
- Finite action sets $A_0 \times A_1 \times \ldots \times A_N =: A$ (e.g., price, feature, quality)
- Payoffs: $v : A \times \Omega \to \mathbb{R}$ for the designer; $\pi_i : A \times \Omega \to \mathbb{R}$ for each agent
- Base Game: $G = (\Omega, q, N, (A_i, u_i)_{i \in N})$

# Model

The typical **Information Design** setting:

- A finite set $\Omega$ of payoff states (e.g., demand condition)
- A common prior belief of about $\omega$, denoted $q \in \Delta(\Omega)$
- One "designer," $N$ agents (e.g., e-com platform and merchants)
- Finite action sets $A_0 \times A_1 \times \ldots \times A_N =: A$ (e.g., price, feature, quality)
- Payoffs: $v : A \times \Omega \to \mathbb{R}$ for the designer; $\pi_i : A \times \Omega \to \mathbb{R}$ for each agent
- Base Game: $G = (\Omega, q, N, (A_i, u_i)_{i \in N})$
- An information structure is a pair $(S, \pi)$ s.t. $S = S_1 \times \ldots \times S_N$ (finite) and $\pi : \Omega \to \Delta(S)$

# Model

- Base game + information structure: $(G, (S, \pi)) \Rightarrow$ a **Bayesian game**
- Solution concept: BNE

# Model

- Base game + information structure: $(G, (S, \pi)) \Rightarrow$ a **Bayesian game**
- Solution concept: BNE

The **information-design** problem:

$$V(q) = \max_{(S,\pi)} \max_{\sigma \in \mathsf{BNE}(G,(S,\pi))} \sum_{\omega,s,a} v(a,\omega)\sigma(a|s)\pi(s|\omega)q(\omega)$$

# Applications

- Online marketplaces:

  Platform runs marketplace where buyers and sellers trade goods. It designs algorithm providing info to parties about value of trade

# Applications

- Online marketplaces:

  Platform runs marketplace where buyers and sellers trade goods. It designs algorithm providing info to parties about value of trade

- Ad auctions

  Platform runs (fixed-format) auctions to sell impressions. It designs algorithm that allows bidders to condition bid on users' characteristics

# Applications

- Online marketplaces:

  Platform runs marketplace where buyers and sellers trade goods. It designs algorithm providing info to parties about value of trade

- Ad auctions

  Platform runs (fixed-format) auctions to sell impressions. It designs algorithm that allows bidders to condition bid on users' characteristics

Limiting assumptions of the standard model:

Commitment power, equilibrium selection, inherently static

# Obedient Recommendation Mechanisms

The ID problem can be equivalently formulated as choosing an obedient direct **recommendation mechanism** $x : \Omega \to \Delta(A)$: Bergemann and Morris '16, *TE*

$$V(q) = \max_{x:\Omega\to\Delta(A)} \sum_{\omega,a} v(a,\omega)x(a|\omega)q(\omega)$$

such that, for all $i \in N, a_i, a_i' \in A_i$

$$\sum_{\omega,a_{-i}} \Big(\pi_i(a_i, a_{-i}, \omega) - \pi_i(a_i', a_{-i}, \omega)\Big)x(a_i, a_{-i}|\omega)q(\omega) \geq 0$$

# Obedient Recommendation Mechanisms

The ID problem can be equivalently formulated as choosing an obedient direct **recommendation mechanism** $x : \Omega \to \Delta(A)$: Bergemann and Morris '16, *TE*

$$V(q) = \max_{x:\Omega\to\Delta(A)} \sum_{\omega,a} v(a,\omega)x(a|\omega)q(\omega)$$

such that, for all $i \in N, a_i, a_i' \in A_i$

$$\sum_{\omega,a_{-i}} \Big(\pi_i(a_i, a_{-i}, \omega) - \pi_i(a_i', a_{-i}, \omega)\Big)x(a_i, a_{-i}|\omega)q(\omega) \geq 0$$

A finite-dimensional linear program $\Rightarrow$ Quite tractable

The typical object of interest in the ID literature: Characterize the optimal $x$

# a different perspective

# Using a Database

Information Design as the problem of how to "optimally use a **database**"

# Using a Database

Information Design as the problem of how to "optimally use a **database**"

Consider a population of consumers, each with unobserved type $\theta \in \Theta$

For each consumer, there is a **data record** of her personal characteristics, which is informative about her $\theta$

# Using a Database

Information Design as the problem of how to "optimally use a **database**"

Consider a population of consumers, each with unobserved type $\theta \in \Theta$

For each consumer, there is a **data record** of her personal characteristics, which is informative about her $\theta$

- It's the realization of an exogenous signal, denoted by $\omega$, which induces a posterior belief about $\theta$
- $v(a, \omega)$ and $\pi_i(a, \omega)$ denote the *expected payoffs* conditional on $\omega$

# Using a Database

Information Design as the problem of how to "optimally use a **database**"

Consider a population of consumers, each with unobserved type $\theta \in \Theta$

For each consumer, there is a **data record** of her personal characteristics, which is informative about her $\theta$

- It's the realization of an exogenous signal, denoted by $\omega$, which induces a posterior belief about $\theta$
- $v(a, \omega)$ and $\pi_i(a, \omega)$ denote the *expected payoffs* conditional on $\omega$

A **database** of data records: $q = \big(q(\omega)\big)_{\omega \in \Omega} \in \mathbb{R}_+^{\Omega}$

# Same Problem

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega,a} v(a,\omega) x(a|\omega) q(\omega)$$

such that, for all $i \in N, a_i, a_i' \in A_i$

$$\sum_{\omega, a_{-i}} \Big( \pi_i(a_i, a_{-i}, \omega) - \pi_i(a_i', a_{-i}, \omega) \Big) x(a_i, a_{-i}|\omega) q(\omega) \geq 0$$

# Same Problem

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega,a} v(a,\omega) x(a|\omega) q(\omega)$$

such that, for all $i \in N, a_i, a_i' \in A_i$

$$\sum_{\omega, a_{-i}} \Big( \pi_i(a_i, a_{-i}, \omega) - \pi_i(a_i', a_{-i}, \omega) \Big) x(a_i, a_{-i}|\omega) q(\omega) \geq 0$$

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_+^{\Omega} \to \mathbb{R}$

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_+^{\Omega} \to \mathbb{R}$

   **Why is this interesting?**

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_+^{\Omega} \to \mathbb{R}$

   **Why is this interesting?**

   - $V(q)$ microfounds the intermediary's preference over databases, which is a key building block of any model of a data economy

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_{+}^{\Omega} \to \mathbb{R}$

   **Why is this interesting?**

   - $V(q)$ microfounds the intermediary's preference over databases, which is a key building block of any model of a data economy
   - Importance of the interpretation

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_{+}^{\Omega} \to \mathbb{R}$

   **Why is this interesting?**

   - $V(q)$ microfounds the intermediary's preference over databases, which is a key building block of any model of a data economy
   - Importance of the interpretation

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_{+}^{\Omega} \to \mathbb{R}$

**Examples of Questions** from Galperti, Levkun, Perego (2023, *Restud*)

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_+^{\Omega} \to \mathbb{R}$

   **Examples of Questions** from Galperti, Levkun, Perego (2023, *Restud*)

   – What is the value for the intermediary of each single data record in $q$?

   A benchmark for compensating individuals for their data

# New Questions

From this interpretation, several questions arise:

1. Characterize properties of the value function $V : \mathbb{R}_+^{\Omega} \to \mathbb{R}$

   **Examples of Questions** from Galperti, Levkun, Perego (2023, *Restud*)

   – What is the value for the intermediary of each single data record in $q$?

   A benchmark for compensating individuals for their data

   – What are the properties of the "demand function" for data

   Law of demand — as $q(\omega) \nearrow$, value of $\omega$-records $\searrow$

   Indifference curves — When are records complement/substitute?

   Merging databases: When is $V(q + q') > V(q) + V(q')$?

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Why is this interesting?**

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Why is this interesting?**

   – Who supplies the data records in practice? Often, the intermediary needs to obtain the data records from the consumers themselves, by promising a service and/or a financial compensation

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Why is this interesting?**

   - Who supplies the data records in practice? Often, the intermediary needs to obtain the data records from the consumers themselves, by promising a service and/or a financial compensation
   - This leads us to models of **"markets for data"**

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Why is this interesting?**

   - Who supplies the data records in practice? Often, the intermediary needs to obtain the data records from the consumers themselves, by promising a service and/or a financial compensation
   - This leads us to models of **"markets for data"**

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Examples of Markets**

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Monopsony: The intermediary has all the bargaining power

Paradigm: Information design with elicitation (obedience + truthtelling constraints)

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Monopsony: The intermediary has all the bargaining power

Paradigm: Information design with elicitation (obedience + truthtelling constraints)

Variations: Cheap talk or verifiable disclosure? Transfers or not?

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Monopsony: The intermediary has all the bargaining power

Paradigm: Information design with elicitation (obedience + truthtelling constraints)

Variations: Cheap talk or verifiable disclosure? Transfers or not?

Galperti and Perego, 2023 (*AEA P&P*) discussed a few examples

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Monopsony: The intermediary has all the bargaining power

Paradigm: Information design with elicitation (obedience + truthtelling constraints)

Variations: Cheap talk or verifiable disclosure? Transfers or not?

Galperti and Perego, 2023 (*AEA P&P*) discussed a few examples

IMHO: Still underexplored framework

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Monopsony: The intermediary has all the bargaining power

Paradigm: Information design with elicitation (obedience + truthtelling constraints)

Variations: Cheap talk or verifiable disclosure? Transfers or not?

Galperti and Perego, 2023 (*AEA P&P*) discussed a few examples

IMHO: Still underexplored framework

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

   **Examples of Markets**

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Competitive Markets: Competing intermediaries with no bargaining power

Intermediaries pay consumers for their data records

Competitive price for each type of data record clears the market: supply of data records = demand

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Competitive Markets: Competing intermediaries with no bargaining power

Intermediaries pay consumers for their data records

Competitive price for each type of data record clears the market: supply of data records = demand

$\rightarrow$ Failure of first welfare theorem: Consumers don't internalize how selling their data affects the intermediary's optimal recommendation mechanism

# New Questions

From this interpretation, several questions arise:

2. We can make $q$ endogenous

**Examples of Markets**

Competitive Markets: Competing intermediaries with no bargaining power

Intermediaries pay consumers for their data records

Competitive price for each type of data record clears the market: supply of data records = demand

$\rightarrow$ Failure of first welfare theorem: Consumers don't internalize how selling their data affects the intermediary's optimal recommendation mechanism

Galperti, Liu, Perego '24 (For more details on this, come to my talk!)

Sunday @8am, Session title: *Competitive Implications of Data Sharing*

# conclusions

# Conclusions

- A large literature in **information design** provides natural framework and powerful tools to study questions concerning the **economics of data**
- Information design as a production problem:

  *Input:* Personal Data $\rightarrow$ *Output:* Optimal Information

- A new perspective: How does changing inputs affects economic outcomes (through changes in optimal information)?
- This is key for studying: The value of data; The effects of privacy protection policies; The role of data unions; etc.
- In my view, this is a natural direction for this literature: lots of open questions and high demand for better theory

# an example

# Example: Properties of $V$

# Example: Properties of $V$



A **merchant** sells its product through an e-commerce **platform**

The platform is used by group of **consumers**, each with independent valuation for the product

# Example: Properties of $V$



A **merchant** sells its product through an e-commerce **platform**

The platform is used by group of **consumers**, each with independent valuation for the product

For each consumer, platform exogenously owns a database $q$ of data records

Each data record is informative about the corresponding consumer's valuation for the merchant product

# Example: Properties of $V$



A **merchant** sells its product through an e-commerce **platform**

The platform is used by group of **consumers**, each with independent valuation for the product

For each consumer, platform exogenously owns a database $q$ of data records

Each data record is informative about the corresponding consumer's valuation for the merchant product

Only two **types** of records:

- $\omega_L$ reveals consumer has valuation $1$
- $\omega_H$ reveals consumer has valuation $2$

Platform's **database** contains:

3 million such records

6 million such records

# Example: Properties of $V$



A **merchant** sells its product through an e-commerce **platform**

The platform is used by group of **consumers**, each with independent valuation for the product

For each consumer, platform exogenously owns a database $q$ of data records

Each data record is informative about the corresponding consumer's valuation for the merchant product

Only two **types** of records:

- $\omega_L$ reveals consumer has valuation $1$
- $\omega_H$ reveals consumer has valuation $2$

Platform's **database** contains:

3 million such records

6 million such records

# Example: Properties of $V$



A **merchant** sells its product through an e-commerce **platform**

The platform is used by group of **consumers**, each with independent valuation for the product

For each consumer, platform exogenously owns a database $q$ of data records

Each data record is informative about the corresponding consumer's valuation for the merchant product

Only two **types** of records:

- $\omega_L$ reveals consumer has valuation $1$
- $\omega_H$ reveals consumer has valuation $2$

Platform's **database** contains:

3 million such records

6 million such records

# Example



Platform is an **intermediary** that provides the merchant with **information** about each consumer, and thus can influence the price it charges to them

Merchant chooses price $a \in A$ given information received. He maximizes profits:

$$\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$$

Suppose platform choose information to maximizes consumer's surplus

$$v(a, \omega) = \max\{\omega - a, 0\}$$

# Example



Platform is an **intermediary** that provides the merchant with **information** about each consumer, and thus can influence the price it charges to them

Merchant chooses price $a \in A$ given information received. He maximizes profits:

$$\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$$

Suppose platform choose information to maximizes consumer's surplus

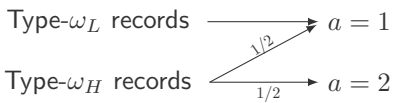
$$v(a, \omega) = \max\{\omega - a, 0\}$$

**Question:** How much **value** does platform derive from each record in $q$?

# Example



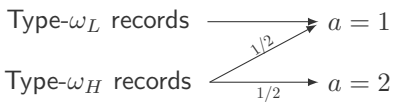
The (unique) optimal recommendation mechanism $(x_q : \Omega \to \Delta(A))$ is:



# Example



The (unique) optimal recommendation mechanism $(x_q : \Omega \to \Delta(A))$ is:

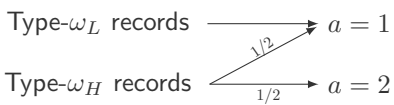



Thus, platform's expected **payoff** from each record is $u_q^*(\omega) = \begin{cases} 0 & \text{if } \omega_L \\ \frac{1}{2} & \text{if } \omega_H \end{cases}$

# Example



The (unique) optimal recommendation mechanism $(x_q : \Omega \to \Delta(A))$ is:



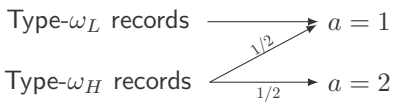

Thus, platform's expected **payoff** from each record is $u_q^*(\omega) = \begin{cases} 0 & \text{if } \omega_L \\ \frac{1}{2} & \text{if } \omega_H \end{cases}$

Are $\omega_L$-records worthless?

# Example



The (unique) optimal recommendation mechanism $(x_q : \Omega \to \Delta(A))$ is:




Thus, platform's expected **payoff** from each record is $u_q^*(\omega) = \begin{cases} 0 & \text{if } \omega_L \\ \frac{1}{2} & \text{if } \omega_H \end{cases}$

Are $\omega_L$-records worthless? No, **value of data records** is $\psi_q^*(\omega) = \begin{cases} 1 & \text{if } \omega_L \\ 0 & \text{if } \omega_H \end{cases}$

# Example



Some observations:

# Example



Some observations:

- Most valuable records are those yielding lowest payoff. Why?
- $\omega_L$ generates no payoff for platform but "helps" lowering merchant price, thus allowing $\omega_H$ to earn positive surplus

# Example



Some observations:

- Most valuable records are those yielding lowest payoff. Why?
- $\omega_L$ generates no payoff for platform but "helps" lowering merchant price, thus allowing $\omega_H$ to earn positive surplus
- Payoff $u^*$ gives *biased* account of the value created by a record **only if** platform withholds information from merchant

$$\psi_q^*(\omega) = u_q^*(\omega) + t^*(\omega)$$

# Example



Some observations:

- Most valuable records are those yielding lowest payoff. Why?
- $\omega_L$ generates no payoff for platform but "helps" lowering merchant price, thus allowing $\omega_H$ to earn positive surplus
- Payoff $u^*$ gives *biased* account of the value created by a record **only if** platform withholds information from merchant

$$\psi_q^*(\omega) = u_q^*(\omega) + t^*(\omega)$$

- **Approach.** Platform uses inputs (data records) to produce outputs (recommendations). GLP '23 use LP duality to **characterize** the **values** of these inputs, namely $\psi_q^*(\omega)$ – platform's willingnes to pay for an additional $\omega$-record