Privacy, Social Data, and Competition

Alessandro Bonatti

MIT Sloan

NBER Privacy Tutorial

October 2022
Economic theory of privacy is decades old.

Comprehensive treatment in Acquisti et al. (2016).

Since: unprecedented collection and diffusion of individual-level data.

Large digital platforms (Amazon, Facebook, Google, Alibaba, JD, Tencent): information gatekeepers and competition managers.
The Main Tradeoff

- Data from past and concurrent transactions (both on and off the platform): 
  ⇒ **surplus creation** through better matching of consumers and sellers.

- Potential for exploitation of individual data:
  - **surplus extraction** from greater market power;
  - surveillance, leakages, fraud, misinformation...

**Platform aspect** of privacy requires new theory: models, questions, results.
Today

- How do different consumers’ privacy choices interact?

- Is there a trade-off between privacy and competition?

- How can privacy regulation help, and how can it backfire?
Data must be sourced from multiple users.
Data can be monetized through multiple producers (firms, merchants).
Consumers and producers may be able to meet off-platform.
Consumer and Producer

Representative consumer interacts with a single producer (or “firm”).

Consumer preference type $\theta \in \Theta \subset \mathbb{R}$.

Consumer $\theta$’s utility function when firm takes action $a$:

$$u(\theta, a).$$

Firm chooses $a$ (ad, video, message, product, price) to match consumer type:

$$a^* = \mathbb{E}[\theta].$$


Market Segmentations

Commonly known prior distribution $F_0(\theta)$.

Firm receives an informative signal $s \in S$.

Signals induce a segmentation (Yang, 2022; Bonatti and Villas-Boas, 2022)

$$S = \{ (\pi_s, F_s) \} \subset S,$$

i.e., a mixture with weights $\pi_s$ over distributions $F_s$ that satisfy

$$\int_S F_s(\theta) \pi_s dS = F_0(\theta), \quad \forall \theta \in \Theta.$$

Interpretation: $\pi_s$ probability of signal $s$; and $F_s$ posterior beliefs under $s$.

Equivalently, market segmentation with sizes $\pi_s$ and compositions $F_s(\theta)$. 
Value of Privacy

Expected surplus of consumers in segment $s$ (i.e., conditional on signal $s$):

$$V(F_s) = \int \theta u(\theta, a^*(F_s)) dF_s(\theta).$$

Expected consumer surplus under segmentation $S$ (i.e., ex ante):

$$U(S) \triangleq \mathbb{E}_s[V(F_s)] = \int_s V(F_s)\pi_s ds.$$  

Expected consumer surplus under prior information (i.e., full privacy):

$$U(\emptyset) \triangleq V(F_0).$$

Any segmentation $S$ is a mean-preserving spread of $F_0$.

Proposition (Value of Privacy)

If $V(\cdot)$ is strictly concave (convex), consumers like (hate) privacy.
Data Acquisition (Easy)

If consumer participates → data revealed → transferred ⋆ to producer.

Equivalent to buying data S from consumer and reselling to producer.

Consumer surplus is $U(S)$ if participating, and zero otherwise. Why zero? Platform is necessary—low search costs, better service quality. . .

If $U(S) > U(∅) > 0$, all good. Data intermediation ⇒ Pareto improvement.

If $U(∅) > U(S) > 0$, consumer loses. Privacy loss = unobserved price.
If consumer participates → data $S$ revealed → transferred* to producer.

Equivalent to buying data $S$ from consumer and reselling to producer.

Consumer surplus is $U(S)$ if participating, and zero otherwise.

Why zero? Platform is necessary—low search costs, better service quality...

If $U(S) > U(∅) > 0$, all good. Data intermediation ⇒ Pareto improvement.

If $U(∅) > U(S) > 0$, consumer loses. *Privacy loss = unobserved price.*
Data Acquisition (Hard)

Consumer chooses whether to reveal information or remain anonymous.

Platform can compensate consumers for information (e.g., better quality).

Platform can extract the producer’s entire value of information.

Property rights over data $\Rightarrow$ Efficient trade? (“Coase Theorem meets privacy”)

At least two problems: **moral hazard** and **data externalities**.
Data Externalities

consumer

consumer

digital platform

producer

consumer

[Diagram of data externalities involving consumers, digital platform, and producer.]
Central feature of individual data is its social dimension.

Data about an individual user is informative about similar users.

Social nature of data generates a data externality not signed a priori.

Data externality can reduce cost of acquiring data from consumers.

Choi et al. (2019), Acemoglu et al. (2022), Ichihashi (2021b), Bergemann et al. (2022).
Consumer Surplus (\# of signals)

\[ U \]

- \( N \):  5  10  15  20
- \( U \): 1.82 1.83 1.84 1.85 1.86 1.87
Intermediation of Social Data

Suppose platform offers $t_i$ for data $S_i$. Consumer $i$ participates iff

$$t_i + U_i(S) \geq U_i(S_{-i}).$$

**Definition (Data Externality)**

Data externality imposed by consumers $-i$ on consumer $i$,

$$DE_i(S) \triangleq U_i(S_{-i}) - U_i(\emptyset).$$

**Proposition (Profitability of Intermediation)**

*Intermediation of data $S$ is profitable iff, for all $i$,*

$$\Delta W_i(S) - DE_i(S) \geq 0$$

Two sources of platform profits: total surplus creation ($\Delta W > 0$, efficient), and negative data externalities ($DE < 0$, inefficient).
Optimal Data Intermediation

Platform-optimal data sharing ≠ complete data sharing:
- uniform price rather than personalized prices;
- personalized rather than uniform product recommendations.

Still, far from socially efficient allocation of data:
- consumers compensated for individual harm, but not for social harm;
- socially efficient anonymization, not intermediation decisions;
- cost of acquiring information vanishes, gains persist as market grows.
Social Data–Discussion

*Digital Privacy Paradox:* consumers require negligible compensation.

- Randomized experiments (Athey et al., 2017).
- Evidence on effects of GDPR (Aridor et al., 2020).

> Privacy is not private, because the effectiveness of these and other private or public surveillance and control systems depends upon the pieces of ourselves that we give up [...] (Zuboff, 2019)

Individual-level regulation is unlikely to restore efficiency.

What about competing platforms?

- Ichihashi (2021a): not a straightforward question.
- “Privacy fixing” as a new anticompetitive concern.
Data Monetization

- Consumer
- Digital Platform
- Producer

Diagram shows relationships between consumer, digital platform, and producer(s).
Monetizing Data—Direct Sale

What if a platform sold consumer data directly?

- Negative externalities downstream—exclusive sales are more profitable (Admati & Pfleiderer, 1986).

- One informed firm vs. uninformed competitors—value of information is an equilibrium object (Bonatti et al., 2022).

- “Selling wine without bottles”—zero marginal cost of reproduction, profitable resale market for data (Barlow, 1994; Shapiro and Varian, 1999).

- Data depreciates but not instantaneously obsolete—can only charge for the additional information (Bergemann et al., 2018).

Monetizing Data—Indirect Sale

What about targeted advertising? Does it count as selling data?

- Consider Google / Amazon search ads.
- Advertisers buy a slot on a keyword search results page.
- Advertiser tailors message to consumer’s search (＝“type”).
- Search engine could sell data about individual searches directly, but leverages the data to sell access qualified eyeballs instead.

Far (better idea and) larger market than direct sales...
(Admati and Pfleiderer, 1990; Bergemann and Bonatti, 2019.)
Selling Access to Consumers Solves:

- the data exclusivity problem by offering a scarce number of slots;
- the problem of competition under asymmetric information structures;
- the resale and rental problems by never giving out the data;
- the quality measurement problem through conversion metrics.

Where your ads may appear
Reach a larger or new audience with Google Display Network targeting.

The easy-to-use platform for hard-to-believe results.
Create campaigns in minutes and enjoy full control over setup, management, and measurement with our easy-to-use self-service platform.

Monetize your mobile game.
You build great apps. We help build your business.
Selling Access: Privacy vs. Competition

- de Cornière and de Nijs (2016); Bergemann and Bonatti (2022); Bergemann, Bonatti, and Wu (2023).
$J$ sellers and a unit mass of consumers.

Consumer $\theta = (\theta_1, \ldots, \theta_j, \ldots, \theta_J)$ has value $\theta_j$ for the product of firm $j$.

Sellers offer horizontally differentiated products (no cost).

$\lambda \in [0, 1]$ use a platform that runs ads in order to find a seller.

$1 - \lambda$ consumers buy directly from sellers, face search costs $\sigma$ à la Diamond (positive, arbitrarily small, first search is free).

Platform observes all types $\theta$; consumers have arbitrarily precise beliefs $m$. 
On Platform: Managed Campaigns

Platform offers a **single** advertising slot per consumer.

Consumer type $\theta \sim \textit{targeting category}$: ads condition on her type.

Formally, the platform:

1. Charges a fixed fee $t$ to participating sellers (e.g., campaign budget).
2. Specifies which $j$ gets to advertise to which consumers $\theta$.
3. Reveals to the consumer her $\theta_j$ for the advertised product $j$.
4. Allows seller $j$ to advertise a personalized price $p_j(\theta)$.

**Proposition (Optimal Mechanism)**

*The platform shows the efficient seller $j^* = \arg \max_j \theta_j$ among all those participating in the mechanism.*
Managed Campaigns

About automated bidding

Automated bidding takes the heavy lifting and guesswork out of setting bids to meet your performance goals. Unlike Manual CPC bidding, there's no need to manually update bids for specific ad groups or keywords. Google Ads automatically sets bids for your ads based on that ad's likelihood to result in a click or conversion that helps you achieve a specific goal for your business.

Different types of automated bidding strategies can help you increase clicks, visibility and conversions. Automated bid strategies learn as they go, using information about a bid's performance to inform future bids. Learn how to determine a bid strategy based on your goals.

This article describes different business goals and the automated bid strategy that best achieves each goal.

Note: If you'd like to automate your bidding specifically for a Shopping campaign, read About automated bidding for Shopping campaigns.

About automated bidding for Shopping campaigns

Automated bid strategies for Shopping help you optimize your advertising spend. Using advanced machine learning, they monitor your campaign's performance and set a bid in every auction to help you achieve your goals.

This article explains the different automated bid strategies that are available for Shopping campaigns and how to choose the right one for you.

Benefits

- You can focus on high level goals and allow Smart Bidding to set the right bid for you. Whether you're trying to drive more visitors to your site or more revenue to your business, automated bidding allows you to start concentrating on overall performance of the campaign and less on how to set the perfect bid for each product group.
- Your campaign's historical performance and future goals are always taken into account. Just enter the performance goals you want to achieve and Smart Bidding does the rest.
- Use dynamic bids to get the most out of your budget. Dynamic bids automatically adjust bids based on factors like product performance and your business goals.
### Managed Campaigns

**Example - Search Term Report**

Google's search terms move will make millions in ad spend invisible to advertisers

The change removes visibility into more than 20% of search terms, one agency finds.

*Slava Moroz* on September 3, 2020 at 3:58 pm

This morning, I negated a word that cost a campaign more than $3 for the one click it received in a brand campaign last week. I didn't add the whole query, just one irrelevant word that triggered a brand keyword. Going forward, I might not even see that type word or know if it showed up across multiple low-volume queries.

As we reported yesterday, Google has notified advertisers the search terms report will “only include terms that were searched by a significant number of users.” It has given no details about what “significant” means. The company told us the reason for the change is "to maintain our standards of privacy and strengthen our protections around user data."

Unsurprisingly, the move has angered advertisers.

---

**Example (1/5) - Drivers of optimizations are more automated**

<table>
<thead>
<tr>
<th>Before</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Structure granular</td>
<td>● Structure aggregated</td>
</tr>
<tr>
<td>● Targeting Selection</td>
<td>● Broad targeting Selection</td>
</tr>
<tr>
<td>● Manual Bid</td>
<td>● Automated Bid</td>
</tr>
<tr>
<td>● Query mining</td>
<td>● Query mining limited visibility</td>
</tr>
<tr>
<td>● Ad Testing</td>
<td>● Ad Testing on autopilot</td>
</tr>
</tbody>
</table>
→ Showrooming as in Wang and Wright (2020) and Teh and Wright (2022).
Symmetric Equilibrium

Off platform, the Diamond (1971) paradox:
- \(1 - \lambda\) off-platform consumers with beliefs \(m\) face search costs \(\sigma > 0\);
- they expect symmetric menus and visit \(\hat{j} = \arg \max_j m_j\) only.

If the platform has any informational advantage:
- \(\lambda\) on-platform consumers infer that \(\theta_j^* = \max_j \theta_j\);
- they expect symmetric menus off-platform, both on and off path.

Proposition (Consideration Sets)

Every online consumer \(\theta\) only compares the displayed seller \(j^*\)'s personalized (on-platform) and posted (off-platform) prices, \(p_{j^*}(\theta)\) and \(p_{j^*}^0\).
...or in fewer words...
Search Patterns: Example

Consumer: beliefs $m$

- $m_1 = \max_j m_j$
- $\theta_2 = \max_j \theta_j$

Firm 1's price $p_1^0$
Firm 2's price $p_2^0$
Firm 2's offer $p_2(\theta)$
Interpretations

With a better-informed platform, equivalent interpretation:

- each brand has \((1 - \lambda)/J\) loyal (imperfectly informed) customers already shopping off-platform;
- the remaining \(\lambda\) consumers are not currently shoppers—they do not recognize any brands without the platform’s data;
- these consumers can be turned into shoppers by informative advertising.

This result requires an (arbitrarily small) informational advantage:

- Without advantage vs. sellers: platform cannot make money.
- Without advantage vs. buyers, platform does not control outside options—consumers’ beliefs determine where they search off platform.
Results

1. Platform sells prominence, enables trade under symmetric information, induces higher total surplus and higher prices.

2. Off-platform sales channel provides outside options to consumers.

3. Platform’s informational advantage narrows consumers’ search options.

4. The growth of a platform’s database (through more consumers or better data) reduces outside options and leads to higher prices.

⇒ Data and participation externalities interact!
Privacy and Competition

Auto-bidding is privacy preserving: advertisers only learn ROI. They don’t even know how much they bid for each category.

Only the platform holds the data. Reduced risk of leakages and spillovers. (Fainmesser et al., 2022; Jullien et al., 2020; Tucker, 2018).

A single firm uses the information at a time.

Managed advertising campaigns restrict competition by design.

Privacy sounds anti-competitive. Not so clear with data-driven mergers.

See Marthews and Tucker (2019) for more…
Data-Driven Mergers

If they stayed separate, more competition, and arguably more privacy too, because they wouldn’t merge the datasets. See also Chen et al. (2022).
A Lot of Work Left!

1. Competing data platforms
   (Ichihashi, 2021a, De Corniere and Taylor, 2020).

2. Data combination and federated learning
   (huge stats+CS+metrics lit; Bergemann et al., 2023.)

3. Evaluation of regulatory interventions

Equally (if not more) important dimensions:

- Political economy, e.g., Beraja et al. (2022) on industrial competitiveness vs.
  government surveillance.

- Fairness of algorithms, differential privacy.

- Special status of health data.
Linear PD Example

Parametrized example with utility $u(\theta, a) = (\theta + \lambda a)^2$ and $\lambda \in [-1, 1]$.

Special case $\lambda = -1/2$ is outcome-equivalent to linear price discrimination:

$$u(\theta, p) = \max_q \{\theta q - pq - q^2/2\} = (\theta - p)^2/2$$

$$p^* = \arg \max_p \{p(\theta - p)\} = \mathbb{E}[\theta]/2.$$
Proof of Concept

Surplus of segment $s$

$$V(F_s) = \int_\theta (\theta - \mathbb{E}_{F_s}[\theta])^2 \, dF_s(\theta)$$

Write as

$$V(F_s) = \mathbb{E}_{F_s}[\theta^2] - \frac{3}{4} (\mathbb{E}_{F_s}[\theta])^2$$

First term is linear in probabilities; second term is convex.

$V(\cdot)$ is a concave function.

More generally, if $\lambda < (>)0$, any MPS hurts (helps) consumers.

More on mkt segmentation: Bergemann et al. (2015) and Elliott et al. (2020).
References I


References II


References V


References VI


