Designing First-Party Data Marketplaces

Klajdi Hoxha* Bing Liu†

*Stanford GSB, †Stanford Econ

Canadian Economic Theory Conference

March 12, 2025

Motivation

- Privacy regulations like CCPA, ATT framework from Apple and the GDPR enforced by the EU are *limiting third-party cookie advertising*
    - resulted in smaller audiences for advertisers to engage with
- eCommerce businesses have shifted emphasis to *1st-party data* collection as a sustainable long-term strategy
    - 1st-party data: information collected directly from customers through store's own channels and interactions like e-mail, purchase history, account information, etc.
- However, 1st-party data are more difficult to generate targeted consumers for businesses and thus collaborations are mutually beneficial

$\Rightarrow$ Opportunity for eCommerce platforms to design new marketplaces and *trade* these data (in)efficiently at a profit

---

CCPA: California Consumer Privacy Act; **ATT**: App Tracking Transparency; **GDPR**: General Data Protection Regulation.

Running Example: Shopify Audiences

SHOPIFY AUDIENCES

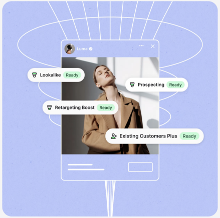
Drive up to 2x more retargeting conversions

Find more customers and increase your orders. Audiences uses commerce insights only available on Shopify to help you maximize your ad performance on top platforms.*

Install Audiences

Not on Shopify Plus? Upgrade now

*Ad campaigns using the latest version of Shopify Audiences Retargeting Boost lists have demonstrated as much as 2 times more orders for every marketing dollar spent on retargeting compared to the next-best available tactic. Results may vary.



Meta Google Pinterest Snap Inc. TikTok CRITEO

Data and privacy

To use Shopify Audiences, you must opt in to share your data to the Audiences data co-op. The co-op is a pool of data from all merchants participating in Shopify Audiences.

The Shopify Audiences algorithms analyze this co-op data to identify millions of customer behavior patterns. These signals help match a customer's intent with the unique attributes of your store. They determine how likely a customer is to purchase from you.

Likely customer prospects are added to custom audiences that you can use to improve your ad targeting. The audiences are hashed to keep the data safe and secure. The audiences are exported directly to the ad platforms. You can't view or access the customer data in the audiences.

Shopify Audiences is built to allow you to comply with applicable privacy laws in the United States and Canada. Ensure you are compliant with applicable privacy laws when sharing data used by Shopify Audiences. Learn more about privacy law compliance.

"*Shopify Audiences … is **a pool of data** from all merchants participating…*"

"*To use Shopify Audiences, you **must opt in** to share your data.*"

"*Shopify Audiences algorithms … help **match** a **customer**'s intent with the unique attributes of your **store**.*"

"*Shopify Audiences algorithms … determine **how likely** a customer is to purchase from you.*"

"*Shopify Audiences is built to allow you to **comply** with applicable privacy laws…*"

Our Question

How should a **profit maximizing** eCommerce platform (e.g. Shopify) design such (billion-dollar) marketplace **?**

Shopify Audiences: Frequently asked questions

Can I exclude competitors from using my data for their audiences?

For privacy reasons, Shopify doesn't disclose which merchants use Shopify product offerings, including Shopify Audiences. If you're using Shopify Audiences, then the data that you contribute will be used to benefit other participating merchants. You can't access information about another merchant or their customers, such as their names.

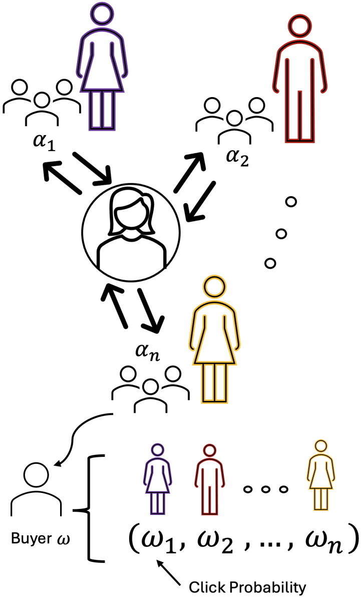
MODEL

Model

- A mediator and sellers $i \in N = \{1, 2, \ldots, n\}$
- Heterogeneous buyers: representative buyer ω has attribute (click probabilities) $\omega = (\omega_i)_{i \in N} \in \Omega$
- Unit mass $\alpha \in \Delta(\Omega)$ of buyers and initial database of seller i is $\alpha_i := (\alpha_i^\omega)_{\omega \in \Omega}$ where

$$\sum_{i \in N} \alpha_i^\omega = \alpha(\omega), \quad \forall \omega \in \Omega.$$

- Seller i has *private* profit margin θ_i of placing an ad to a buyer ω yielding expected profit $\theta_i \omega_i$



Model Assumptions ◂ back

- **Primitive assumptions**:
  - Each buyer ω has unit-demand
  - Sellers sell one unit of an indivisible commodity to each customer
  - Click probabilities $\omega_i \in \Omega_i = \{l_i, h_i\}$ where $0 \leq l_i < h_i \leq 1$
- **Distributional assumptions**:
  - type profile $\theta = (\theta_i)_{i \in N} \in \Theta$ where θ_i is drawn iid. from some $F[0,1]$ that admits a continuous and strictly positive density f with non-decreasing virtual functions

$$\phi_i^B(\theta_i) = \theta_i - (1 - F(\theta_i))/f(\theta_i) \quad \text{and} \quad \phi_i^S(\theta_i) = \theta_i + F(\theta_i)/f(\theta_i),$$

    i.e. virtual value and virtual cost, respectively.
- **Informational assumptions**:
  - The model primitives Θ and F and $(\alpha_i)_i$'s are common knowledge

Mediator's Problem

- **Commits** to a *direct mechanism* (q, x, t):
  - procurement rule $q = (q_i^\omega)_{i,\omega}$ s.t. $\theta \mapsto q_i^\omega(\theta)$
  - re-allocation rule $x = (x_i^\omega)_{i,\omega}$ s.t. $\theta \mapsto x_i^\omega(\theta)$
  - payment rule $t = (t_i)_i$ s.t. $\theta \mapsto t(\theta)$

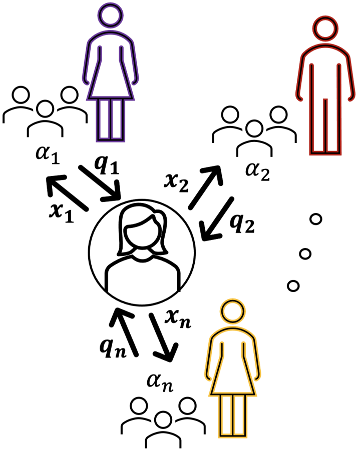
  subject to:

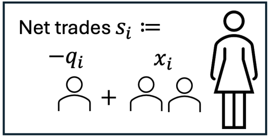
$$q_i^\omega(\theta) \le \alpha_i^\omega \text{ and } \sum_{i\in N} x_i^\omega(\theta) \le \sum_{i\in N} q_i^\omega(\theta), \quad \forall \theta, \omega \quad \text{(fsb.)}$$

- Type θ_i seller gets a *net* payoff $u_i(q, x, t)$ given by

$$u_i(q_i, x_i, t_i, \theta_i) = \theta_i \sum_{\omega\in\Omega} \omega_i s_i^\omega(\theta) - t_i(\theta)$$

- **Mediator's Objective**: Maximize expected profits $\sum_{i\in N} \mathbb{E}[t_i(\theta)]$ subject to interim IC, IR and (fsb.)

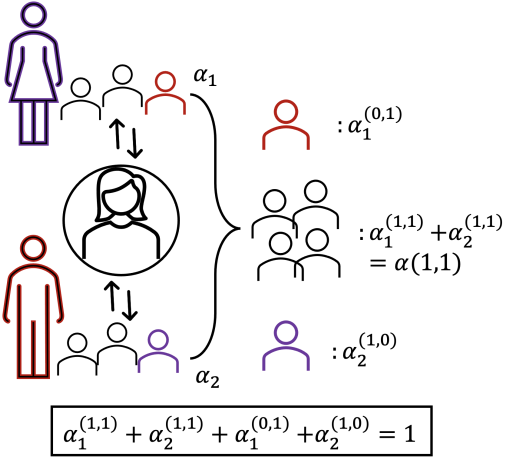




Stylized Example

Stylized Example: Exclusive and Common Data Marketplace

- Two sellers: $i = 1, 2$
- Symmetric click probabilities: $\omega_i \in \{0, 1\}$
  - initial mass $\alpha \in \Delta\big((0,0),(0,1),(1,0),(1,1)\big)$
- Types $(1,0)$ and $(0,1)$ are **exclusive** buyers and $(1,1)$ types are **common** buyers
- WLOG $\alpha(0,0) = 0$, $\alpha_1^{(1,0)} = 0$ and $\alpha_2^{(0,1)} = 0$
  - i.e. $\alpha(1,0) = \alpha_2^{(1,0)}$ and $\alpha(0,1) = \alpha_1^{(1,0)}$



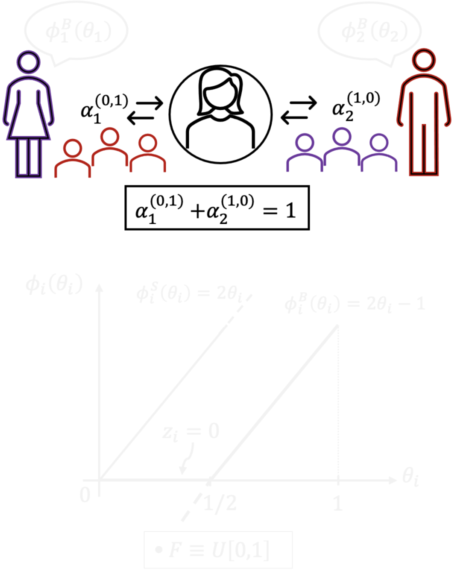
Stylized Example: Benchmark Cases

a) Each seller owns only exclusive buyers of the competitor.

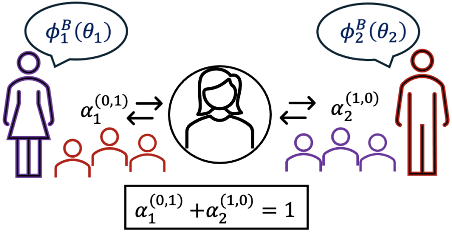
b) Only one seller owns data which are only common buyers.

c) Both sellers own a share of only common buyers data.

Benchmark Case a): $\alpha(1,1) = 0$



Benchmark Case a): $\alpha(1,1) = 0$

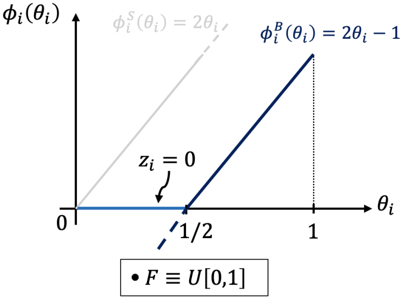


- Analogous to **monopoly pricing** setting!

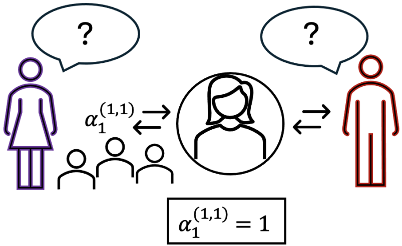
Optimal Mechanism ($F \equiv U[0,1]$)

Procure all data (at no cost) and sell each separately at the posted price $p_i^* = 1/2$.

- Excludes from buying all types with virtual value $\phi_i^B(\theta_i) < 0\,(= z)$ since their true valuation $\theta_i < 1/2$.

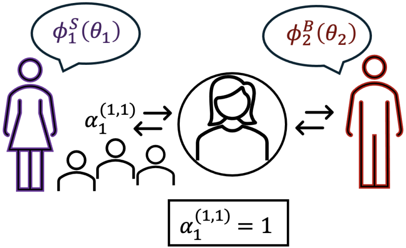


Benchmark Case b): $\alpha_1^{(1,1)} = 1$



- $F \equiv U[0,1]$

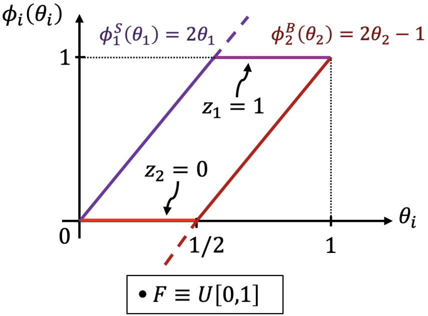
Benchmark Case b): $\alpha_1^{(1,1)} = 1$



- Analogous to **bilateral trading** model!

Optimal Mechanism

Trade if and only if $\phi^B(\theta_2) \geq \phi^S(\theta_1)$.

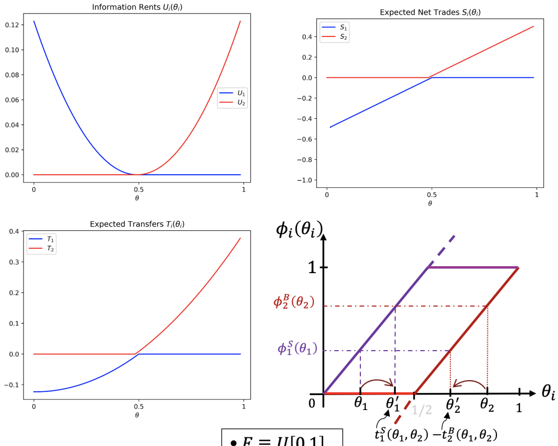


- For uniform distribution $F \equiv U([0,1])$, trade if and only if $\theta_2 - \theta_1 \geq 1/2$.

Benchmark Case b): $\alpha_1^{(1,1)} = 1$

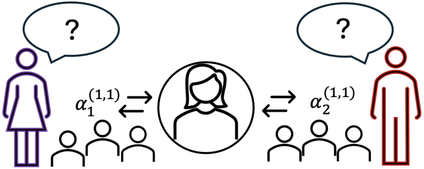
- Information rents highest for low cost seller 1 and high value buyer 2.
- One way to implement is by the following ex-post transfers:

$$t_1^S(\theta_1,\theta_2) = \sup\{\theta_1' : \phi^B(\theta_2) \geq \phi^S(\theta_1')\}$$
$$t_2^B(\theta_1,\theta_2) = \inf\{\theta_2' : \phi^B(\theta_2') \geq \phi^S(\theta_1)\}.$$



• $F \equiv U[0,1]$

Benchmark Case c): $\alpha_1^{(0,1)} = \alpha_2^{(1,0)} = 0$

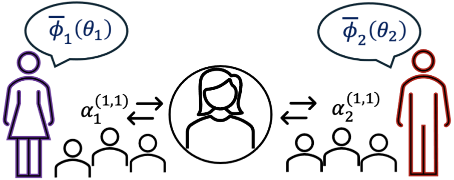


$$\boxed{\alpha_1^{(1,1)} = \beta \,,\ \alpha_2^{(1,1)} = 1 - \beta}$$

- $F \equiv U[0,1], \beta = 1/2$

Benchmark Case c): $\alpha_1^{(0,1)} = \alpha_2^{(1,0)} = 0$

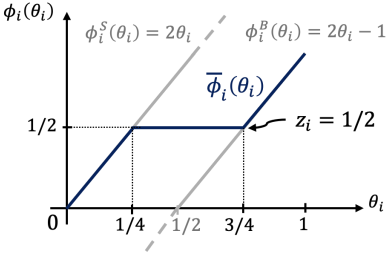
- Analogous to **partnership dissolution** model!
- **Countervailing incentives**: a high (low) type is more likely to buy (sell) and thus has an incentive to under(over)-report
  - types in the "middle" are least sure if they will sell or buy $\Rightarrow$ weakest incentives to misreport



$\alpha_1^{(1,1)} = \beta\ ,\ \alpha_2^{(1,1)} = 1 - \beta$

Optimal Mechanism ($F \equiv U[0,1], \beta = 1/2$)

Allocate all data to seller with strictly higher $\overline{\phi}_i(\theta_i)$ and no trade otherwise.

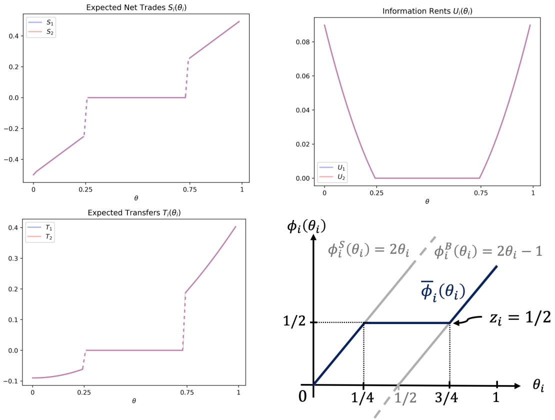


• $F \equiv U[0,1], \beta = 1/2$

- **Alternatively**, procure all data and allocate to seller with highest $\overline{\phi}_i(\theta_i)$ and break ties uniformly.

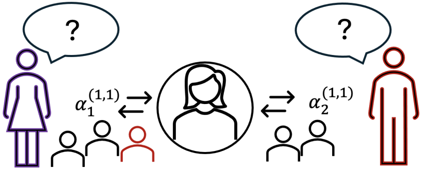
Benchmark Case c): $\alpha_1^{(0,1)} = \alpha_2^{(1,0)} = 0$

- Information rents highest for low cost "sellers" or high value "buyers".
- No trade if $\theta_1, \theta_2 \in [1/4, 3/4]$ or $\overline{\phi}_i(\theta_i) = z_i = 1/2$
  - Still, the volume of trades exceed those in b) since $\theta_1, \theta_2 \in [1/4, 3/4]$ imply $\theta_2 - \theta_1 \leq 1/2$.
- No simple ex-post implementation



• $F \equiv U[0,1], \beta = 1/2$

Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$



$$\boxed{\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \frac{\gamma}{2} \text{ and } \alpha_1^{(0,1)} = 1-\gamma}$$

- $F \equiv U[0,1], \gamma = 0.8$

Stylized Example: New Case

- **Next:** What if at least one seller owns a composition of common and exclusive buyers?

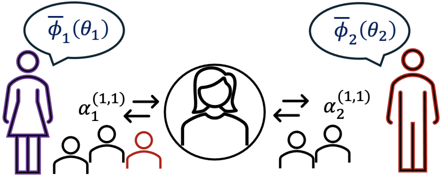
Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$

- Let $\underline{\theta}(\gamma) = \frac{1}{4} - \frac{1-\gamma}{2\gamma}$ and $\overline{\theta}(\gamma) = \frac{3}{4} - \frac{1-\gamma}{2\gamma}$ for $\gamma > 2/3$, and otherwise 0 and $1/2$, respectively.

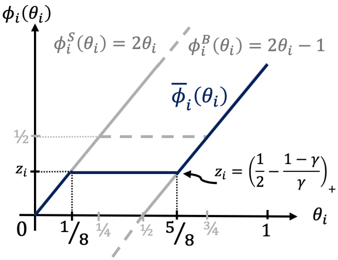
Optimal Mechanism $(F \equiv U[0,1])$

(i) WLOG procure everything first.
(ii) Ironing range $z_i(\gamma) = (1/2 - (1-\gamma)/\gamma)_+$ for $i = 1, 2$.
(iii) Allocate all common data to seller with strictly higher $\overline{\phi}_i(\theta_i)$, if any.
(iv) Allocate all exclusive lists to seller 2 iff $\overline{\phi}_2(\theta_2) > 0$.
(v) If ties, i.e. θ_i's in $[\underline{\theta}(\gamma), \overline{\theta}(\gamma)]$, break in favor of seller 1 with probability $\min\{1/2 + (1-\gamma)/\gamma, 1\}$.

- **Remark.** Sellers are treated *symmetrically* with regards to virtual types $\overline{\phi}_i$!



$\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \frac{\gamma}{2}$ and $\alpha_1^{(0,1)} = 1-\gamma$



• $F \equiv U[0,1], \gamma = 0.8$

Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$

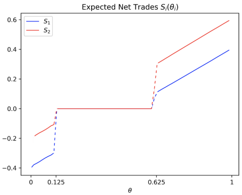
- Decompose $U_1(\theta_1) = U_1^C(\theta_1)$ and $U_2(\theta_2) = U_2^C(\theta_2) + U_2^E(\theta_2)$

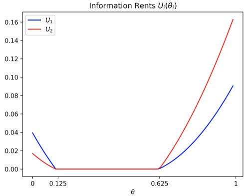
- $\theta_i \geq \overline{\theta}(\gamma)$: ("Buyer") rents from net **positive** trades of the common data (U^C) are:

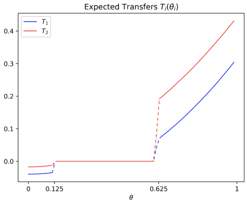
$$U_i^C(\theta_i) = \frac{\gamma}{2} \times \int_{\overline{\theta}(\gamma)}^{\theta_i} (2\hat{\theta}_i - 1)\, d\hat{\theta}_i \quad (\textit{incr.} \uparrow \theta_i)$$

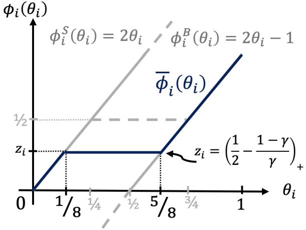
whereas seller 2's rents from **always buying** exclusive data (U^E) are:

$$U_2^E(\theta_2) = (1-\gamma) \times \int_{\overline{\theta}(\gamma)}^{\theta_2} 1\, d\hat{\theta}_2$$



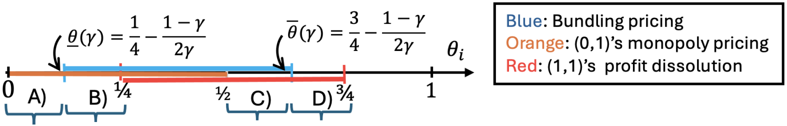






• $F \equiv U[0,1], \gamma = 0.8$

Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$



- $\theta_i \geq \overline{\theta}(\gamma)$: Use decomposition of $U_i(\theta_i)$'s to write expected transfers $T_i(\theta_i) \geq 0$ as:

$$T_1(\theta_1) = \underbrace{\frac{\gamma}{2}\Big(\theta_1^2 - \big(1-\overline{\theta}(\gamma)\big)\overline{\theta}(\gamma)\Big)}_{T_1^C(\theta_1)\geq 0} \quad \& \quad T_2(\theta_1) = \underbrace{\frac{\gamma}{2}\Big(\theta_2^2 - \big(1-\overline{\theta}(\gamma)\big)\overline{\theta}(\gamma)\Big)}_{T_2^C(\theta_2)\geq 0} + \underbrace{(1-\gamma)\overline{\theta}(\gamma)}_{T_2^E(\theta_i)\geq 0}$$

- Compare to **benchmark models** if $(1,1)$'s and $(0,1)$'s were sold separately:

$$\hat{T}_1(\theta_1) = \frac{\gamma}{2}\left(\theta_1^2 - \left(1-\frac{3}{4}\right)\frac{3}{4}\right) \quad \& \quad \hat{T}_2(\theta_1) = \frac{\gamma}{2}\left(\theta_2^2 - \left(1-\frac{3}{4}\right)\frac{3}{4}\right) + (1-\gamma)\frac{1}{2}$$

- Less profits in region C) from net **zero** expected trades.
- More profits in region D) from net **positive** common data trades and, for $\theta_i \geq 3/4$, both per-unit *bundling prices* (T_i^C/S_i^C and T_i^E/S_i^E) are **higher**.

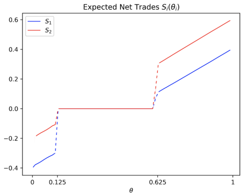
Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$

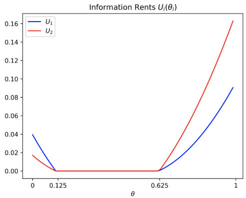
- $\theta_i \leq \underline{\theta}(\gamma)$: ("Seller") rents from net **negative** trades of the common data (U^C):

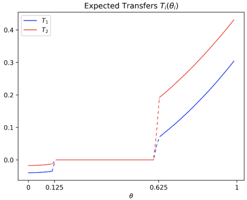
$$U_i^C(\theta_i) = \frac{\gamma}{2} \times \int_{\theta_i}^{\underline{\theta}(\gamma)} (1-2\hat{\theta}_i)\, d\hat{\theta}_i \quad (\textit{incr.} \downarrow \theta_i)$$

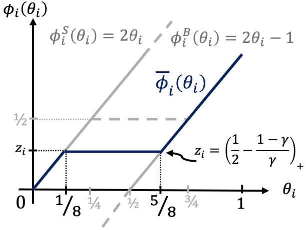
whereas seller 2's rents from **always buying** exclusive data (U^E) are:

$$U_2^E(\theta_2) = -(1-\gamma) \times \int_{\theta_2}^{\underline{\theta}(\gamma)} 1\, d\hat{\theta}_2 \quad (<0)$$



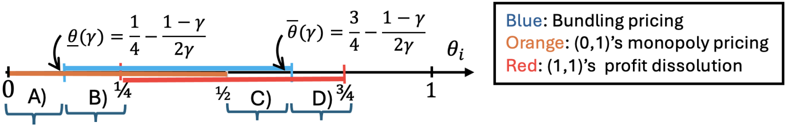






• $F \equiv U[0,1], \gamma = 0.8$

Stylized Example: $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = 1-\gamma$



- $\theta_i \leq \underline{\theta}(\gamma)$: Use decomposition of $U_i(\theta_i)$'s to write expected transfers $T_i(\theta_i) \leq 0$ as:

$$T_1(\theta_1) = \underbrace{(\gamma/2)\left(\theta_1^2 - \left(1-\underline{\theta}(\gamma)\right)\underline{\theta}(\gamma)\right)}_{T_1^C(\theta_1)\leq 0} \quad \& \quad T_2(\theta_1) = \underbrace{(\gamma/2)\left(\theta_2^2 - \left(1-\underline{\theta}(\gamma)\right)\underline{\theta}(\gamma)\right)}_{T_2^C(\theta_2)\leq 0} + \underbrace{(1-\gamma)\underline{\theta}(\gamma)}_{\boldsymbol{T_2^E(\theta_i)\geq 0}}$$

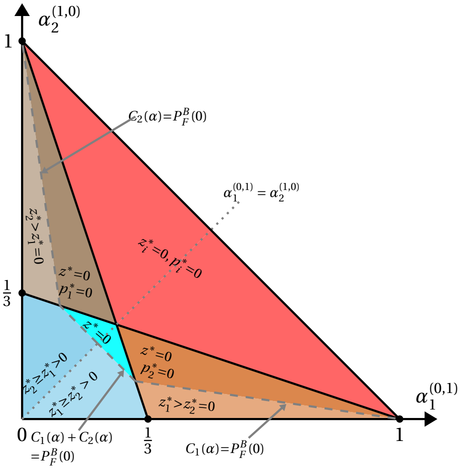
- Compare to **benchmark models** if $(1,1)$'s and $(0,1)$'s were sold separately:

$$\hat{T}_1(\theta_1) = (\gamma/2)\left(\theta_1^2 - (1-1/4)\,(1/4)\right) \quad \& \quad \hat{T}_2(\theta_1) = (\gamma/2)\left(\theta_2^2 - (1-1/4)\,(1/4)\right) + 0$$

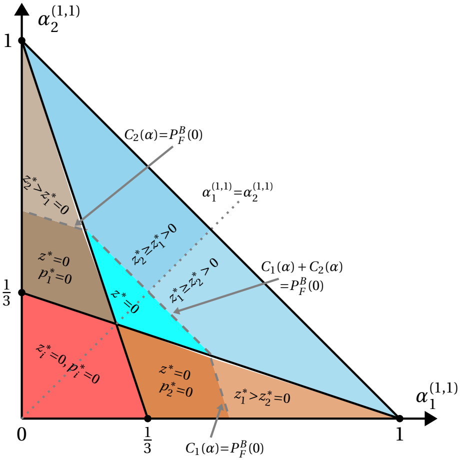
- Net zero trades in region B) and thus zero compensated transfers; compare to $\hat{T}_i(\theta_i) < 0$.
- More profits in region A) since the *per-unit compensated transfers* are **lower**.
- More profits in region A) from exclusive data trades, **overpriced** at per-unit $\underline{\theta}(\gamma) \geq \theta_2$!

Stylized Example: Optimal Mechanism (Graphically)

details



(a) Symmetric Common Data: $\alpha_1^{(1,1)}=\alpha_2^{(1,1)}$



(b) Symmetric Exclusive Data: $\alpha_1^{(0,1)}=\alpha_2^{(1,0)}$

Withholding (Ex-Ante) Data

Stylized Example: Withholding (Ex-Ante) Data

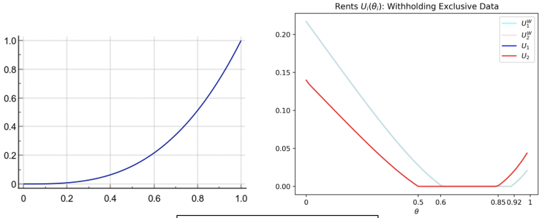
- **Example.** Let $\alpha_1^{(1,1)} = \alpha_2^{(1,1)} = \gamma/2$ and $\alpha_1^{(0,1)} = \alpha_2^{(1,0)} = (1-\gamma)/2$, where $\gamma = 0.8$

> **Observation.** Withholding the exclusive data may benefit or hurt sellers.

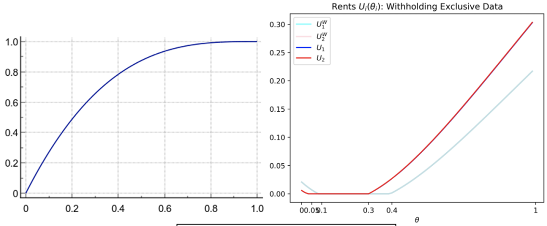
- In i) withholding data **increases** ex-ante expected rents from $U_i = \mathbb{E}[U_i(\theta_i)] \approx 3.5$ to $U_i^W = \mathbb{E}[U_i(\theta_i)] \approx 6$.

- In ii) withholding data **decreases** ex-ante expected rents from $U_i = \mathbb{E}[U_i(\theta_i)] \approx 8.5$ to $U_i^W = \mathbb{E}[U_i(\theta_i)] \approx 6$.

$\Rightarrow$ An issue if platform lacks commitment power!



i) $F \equiv Beta(3,1), \gamma = 0.8$



ii) $F \equiv Beta(1,3), \gamma = 0.8$

Shopify Audiences: Withholding (Ex-Ante) Data

Shopify Audiences: Frequently asked questions

Can I exclude specific customers from my audiences?

For privacy reasons, you can't find information about the identity or contact information of customers included in an audience, and you can't add or exclude customers to a specific audience.

Though you can't include or exclude customers for specific audiences, you can exclude them from all audiences. To do so, use the **Customer opt-out** settings in the Shopify Audiences app.

- *"...you can't add or exclude customers to a specific audience."*
- *"Though... you can exclude them from all audiences."*

General Result

Optimal Mechanism

Theorem 1 (informal)

There exists a set of ironing parameters z and tie-breaking rule p which leaves zero expected net trades ($S_i^{z,p}(\theta_i) = 0$) for $\theta_i \in [\underline{\theta}(z), \overline{\theta}(z)]$. (WLOG) Procure all data and allocate all of ω buyers to the agent with highest **weighted** virtual type $\omega_i \overline{\phi}_i(\theta_i, z)$, where ties are broken according to p.

Other stylized examples

- $\Omega_i = \{0,1\}$ but $N \uparrow$
- Heterogeneous products: $N = 2$ but $\Omega_1 = \{0,1\} \neq \Omega_2 = \{p,1\}$
  - think of ω_i's as quality differentiation

$\Rightarrow$ full characterization for each case is work in progress!

Remaining...

- **Related Literature**
- **Conclusion and Future Research**

Related Literature

- **Bilateral Trading:** Myerson and Satterthwaite (1983)
- **Partnership Dissolution:** Cramton et al. (1987), Loertscher and Wasser (2019)
- **Bundling:** Yang (2023)
  - our model: one-dimensional heterogeneity too but additive preferences.
- **Mechanism Design with Limited Commitment:** Bester and Strausz (2001)

Conclusion and Future Research

Conclusion:

- The optimal mechanism sells data in *bundles* to extract higher rents.
- By treating sellers symmetrically with regards to their (virtual) valuations, irrespective of the differences in initial endowments, makes the bundling strategy more profitable.
- (Ex-ante) sellers may be better-off withholding some of the data.

Future research:

- Extensions to the current model—many directions! revisit
- If designer lacks commitment power, look at robust (wrt. initial endowments) profit maximizing mechanisms?
- What can be implemented in dominant strategies?
- Which mechanisms maximize expected total gains from trade?

→ Reach out: *hklajdi@stanford.edu* or *bingliu@stanford.edu*

THANK YOU

Appendix

Stylized Example: Optimal Mechanism back

Proposition 1 (Optimal Mechanism)

There exists an (essentially) unique optimal mechanism characterized by (in)efficient trading regions z_1^*, z_2^* and random priority rule p^* which satisfy the followings:

1. If $\alpha \in \mathscr{R}_+$ then there exists a unique $z_1^* \geq z_2^* \in [0,1]$ such that
   1. if $C_1(\alpha) + C_2(\alpha) > P_F^B(0)$ then $z_2^* > 0$. Moreover, for $z > 0$ uniquely defined by $P_F^B(z) + P_F^S(z) = C_1(\alpha) + C_2(\alpha)$,
      1. if $P_F^B(z) - P_F^S(z) \geq C_1(\alpha) - C_2(\alpha)$ then $z_1^* = z_2^* = z$ and the unique tie-breaking rule p^* solves $p_1^* = \Big(1 + \big(C_1(\alpha) - C_2(\alpha)\big)\big(P_F^B(z) - P_F^S(z)\big)\Big)/2$.
      2. if $P_F^B(z) - P_F^S(z) < C_1(\alpha) - C_2(\alpha)$ then $z_1^* > z_2^*$ such that $z \in [z_1^*, z_2^*]$ and uniquely solve $P_F^B(z_1^*) = C_1(\alpha)$ and $P_F^S(z_2^*) = C_2(\alpha)$.
   2. if $C_1(\alpha) + C_2(\alpha) \leq P_F^B(0)$ then $z_1^* = z_2^* = 0$ and the unique random priority rule p^* satisfies $p_1^* = C_1(\alpha)/P_F^B(0)$ and $p_2^* = C_2(\alpha)/P_F^B(0)$.
2. If $\alpha \in \mathscr{R}_1$ then $z_2^* = 0$ and $z_1^* = (P_F^B)^{-1}(C_1(\alpha)) > 0$ if $C_1(\alpha) > P_F^B(0)$, and otherwise $z_1^* = 0$ and $p_1^* = C_1(\alpha)/P_B^F(0)$.
3. If $\alpha \in \mathscr{R}_-$ then $z_1^* = z_2^* = 0$ and $p^* \equiv 0$.

In particular, the expected optimal re-allocation of exclusive lists and common lists are uniquely given by $X_1^{(1,0)}(\hat{\theta}_1) = \min\{\alpha_2^{(1,0)}, \alpha_1^{(1,1)}\}$ and $X_2^{(0,1)}(\hat{\theta}_2) = \min\{\alpha_1^{(0,1)}, \alpha_2^{(1,1)}\}$.