Balancing User Privacy and Personalization

Malika Korganbekova

The University of Chicago Booth School of Business

(joint with Cole Zuber, Wayfair LLC)

October 14, 2024

Online Platforms personalize content

- Personalize advertising (Wernerfelt et al., 2023)
- Personalize prices (Dube and Misra, 2022)´
- Personalize **content/product recommendations** **(this paper)**

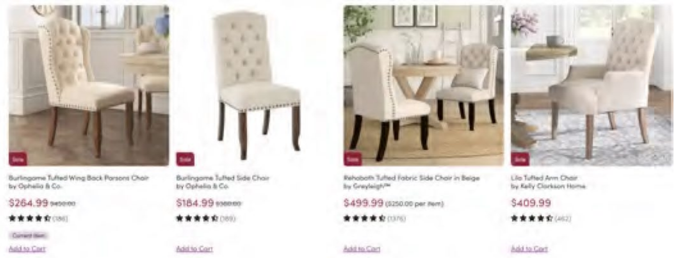
Retailers' recommendations: Furniture

Browsing History ⇒

Retailers' recommendations: Furniture

Browsing History ⇒

Recommendations ⇒



WHAT MAKES PERSONALIZATION POSSIBLE?

Tracking ⇒ Individual-level data

REGULATORS ARE CONCERNED ABOUT USER PRIVACY

- Consumers do not know what or how their data is used

- Competition & Markets Authority (UK) and FTC (US)

“Personalisation can be harmful ... These harms often occur through **the manipulation of consumer choices, without the awareness of the consumer.**”

Regulators are concerned about user privacy

- Consumers do not know what or how their data is used
- Competition & Markets Authority (UK) and FTC (US)

“Personalisation can be harmful ... These harms often occur through **the manipulation of consumer choices, without the awareness of the consumer.**”

Browsers within and across websites cookie-based

⇓

tracking limit online consumer tracking

RESEARCH QUESTIONS

1 Does **personalization hurt** consumers and sellers on the platform?

2 How do **privacy restrictions** that limit consumer tracking impact: >

different types of consumers, sellers, and the platform

3 How can platforms **mitigate the losses** from privacy restrictions?

METHODOLOGY

1 DoesStep 1: Field experiment**personalization hurt** consumers and sellers on the platform?

2 How domarketplace?Step 2: Re-train Personalization Algorithm +

Structural model of search**privacy restrictions** that limit consumer

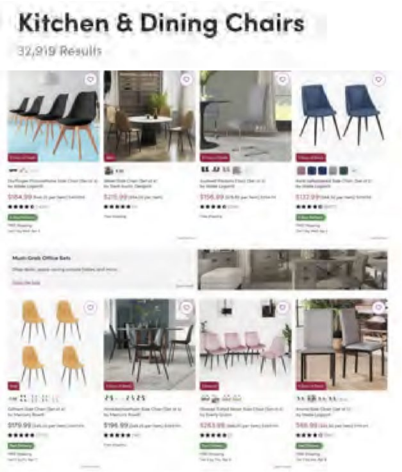
tracking impact the 3 Step 3: Propose Probabilistic Algorithm +

Structural model of searchHow can platforms **mitigate the losses** from

privacy restrictions?

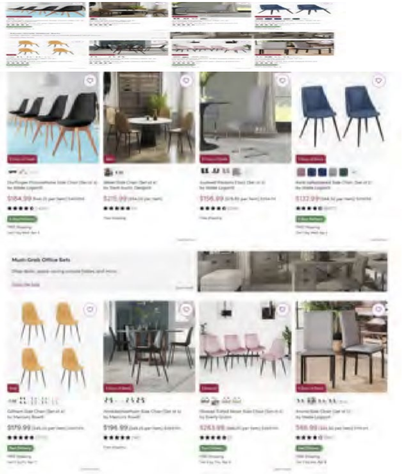
WAYFAIR PERSONALIZES PRODUCT RANKINGS

LARGE ONLINE MARKETPLACE OF FURNITURE

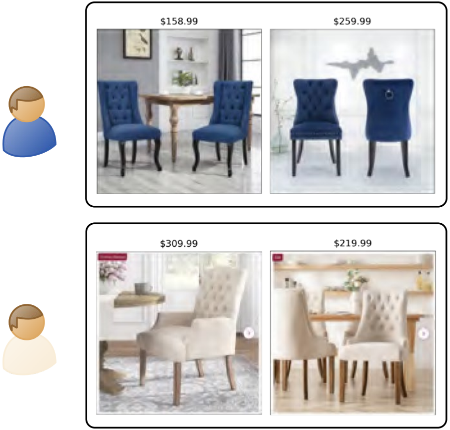

Kitchen & Dining Chairs
32,919 Results

WAYFAIR PERSONALIZES PRODUCT RANKINGS

LARGE ONLINE MARKETPLACE OF FURNITURE



PersonalizedRankings



PERSONALIZATION ALGORITHM

Input

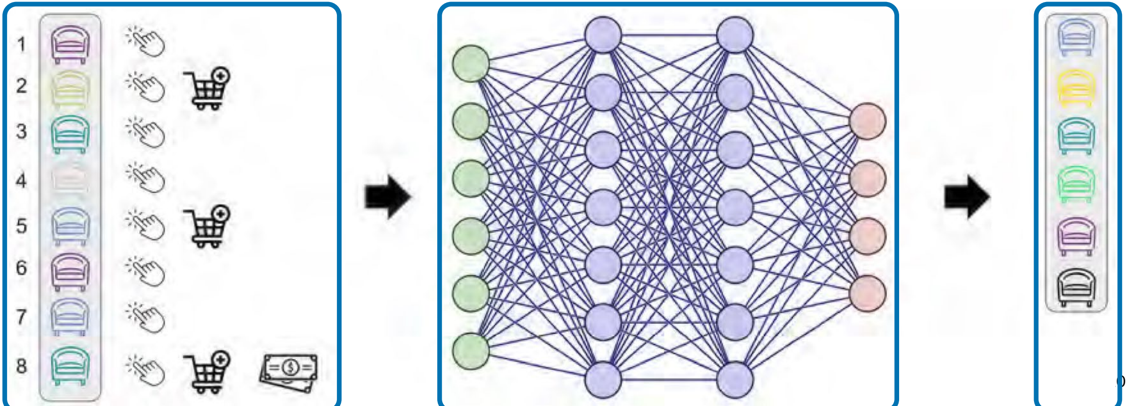
Browsinghistory

Algorithm

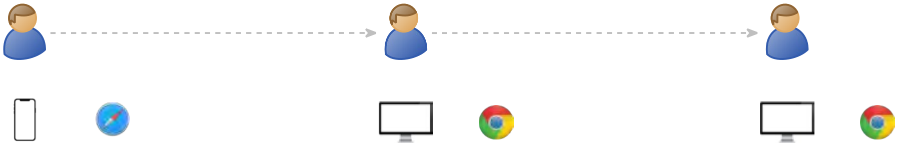
DeepLearning

Output

Rankings



PLATFORM TRACKS CONSUMERS ACROSS SESSIONS



1 Consumer Login + Deterministic Matching (40%)

2 Online tracking technologies **(1st-party and 3rd-party cookies)** (60%)

Data

- All consumers during 2018-2022
- **Device, browser, source of traffic** upon arrival
- **Product rankings** at each consumer-pageload

- **Pixel-level clickstream:** entire history of consumer actions (clicks, scrolling)
- **Transactions:** final purchase decisions and product returns
- **Retail price and seller-set wholesale prices** at a daily level

RESEARCH QUESTIONS

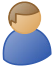
[1]Step 1: Field experimentDoes **personalization hurt** consumers and sellers on the platform?

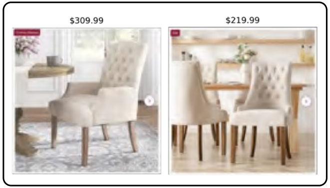
LARGE-SCALE FIELD EXPERIMENT

- >9 million consumers between January 2020 and December 2021

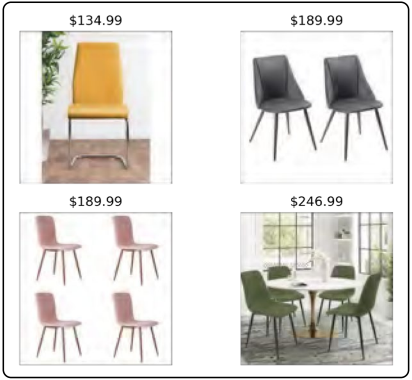
Treatment	Control

PersonalizedRankings





Non-



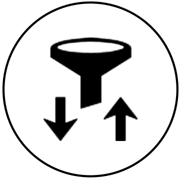
Personalized (Bestseller) Rankings

PERSONALIZATION LEADS TO LOWER SEARCH COSTS

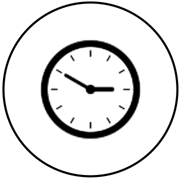
EXPERIMENT RESULTS

Compared to non-personalized group, personalized group consumers ...

scroll less



filter less



purchase faster

purchase

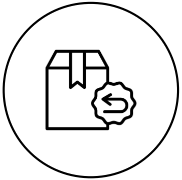
Baseline Additional Results

PERSONALIZATION BENEFITS CONSUMERS

EXPERIMENT RESULTS

↓ -83 products*** ↓ -1.9%*** ↓ -2 days*** ↑ +1.4%***

Compared to non-personalized group, in personalized group ...



purchase price

↑ +0.5%***

product return

↓ -10%***

repeat purchase

↑ +2.3%***

Table Ratings Review Fragmentation

BENEFITS: PLATFORM VS SELLERS

EXPERIMENT RESULTS

Platform

- Revenue ↑ +2%
- Profit ↑ +1.3%

Sellers

> ↑ repeat purchases

~~Less experi~~ **Experience**enced sellers benefit

- ↑ +3.2-4.1% revenue

  Tabl

- ↑ +19% visibility

Niche Mid-niche sellers ↑ +2.2%

revenue 2D embeddings Nicheness Illustration

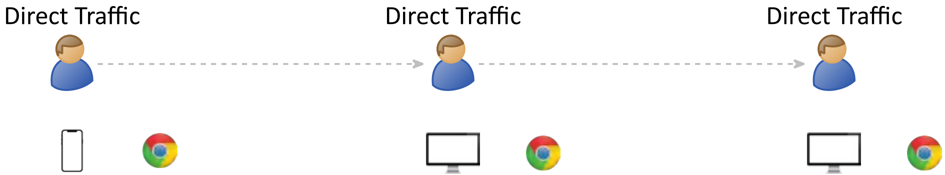
Examples

Research Questions

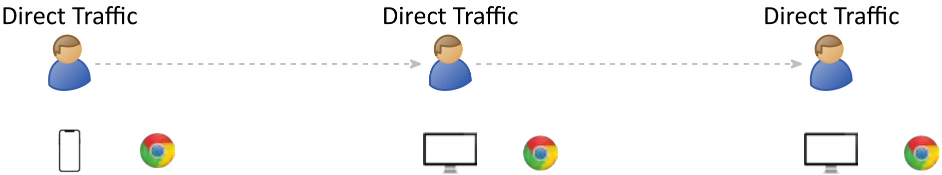
1 Does **personalization hurt** consumers and sellers on the platform? ✓

2 How domarketplace?Step 2: Re-train Personalization Algorithm + Structural model of search**privacy restrictions** that limit consumer tracking impact the

PLATFORM TRACKS CONSUMERS ACROSS SESSIONS



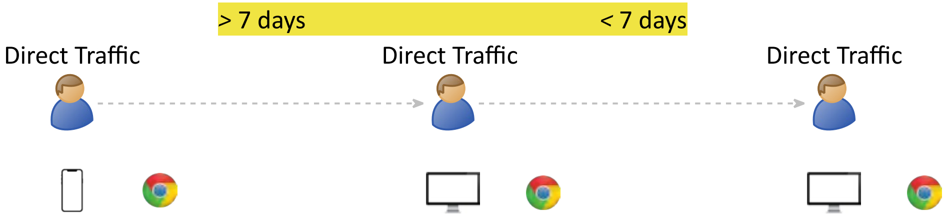
PLATFORM TRACKS CONSUMERS ACROSS SESSIONS



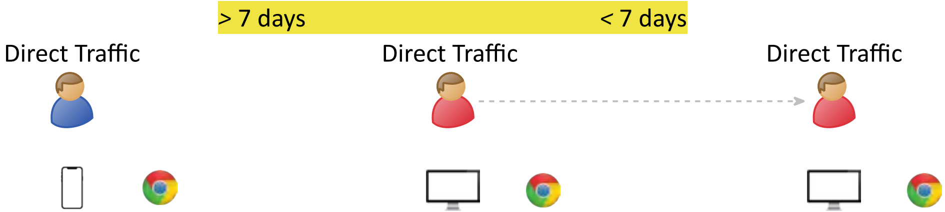
Connecting browsing sessions

- 40% of consumers login
- 60% rely on online tracking technologies (1st-party and 3rd-party cookies)

SAFARI DATA RESTRICTIONS: 1ST PARTY COOKIE EXPIRATION



SAFARI DATA RESTRICTIONS: 1ST PARTY COOKIE EXPIRATION

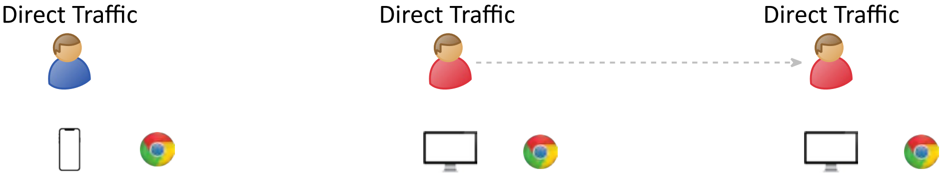


SAFARI DATA RESTRICTIONS: 1ST PARTY COOKIE EXPIRATION

> 7 days	< 7 days

17 / 26

SAFARI DATA RESTRICTIONS: 1ST PARTY COOKIE EXPIRATION



- ~30% of consumers arrive from Safari browser

Safari data restrictions: 1st party cookie expiration

Personalization Algorithm

Input

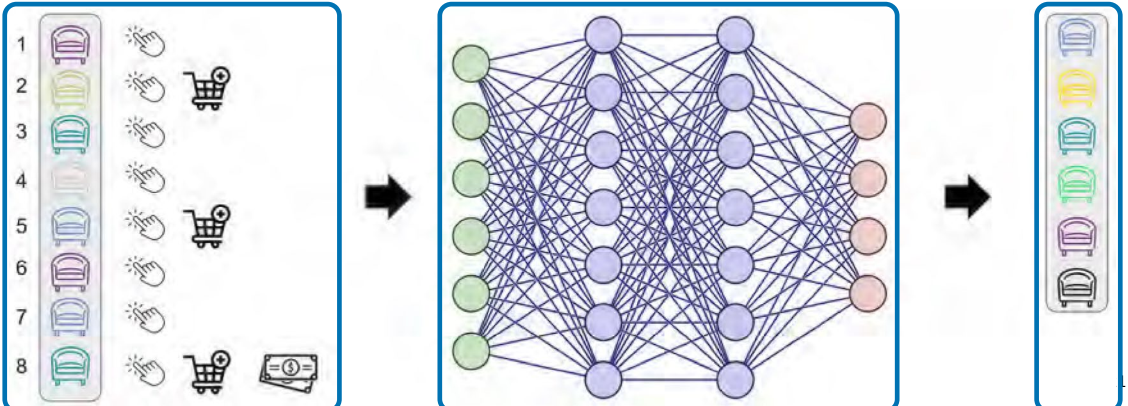
Browsinghistory

Algorithm

DeepLearning

Output

Rankings



Overview of the Counterfactual Analysis

Re-training the Algorithm

Input

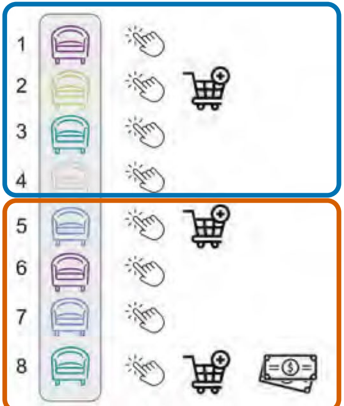
Browsinghistory

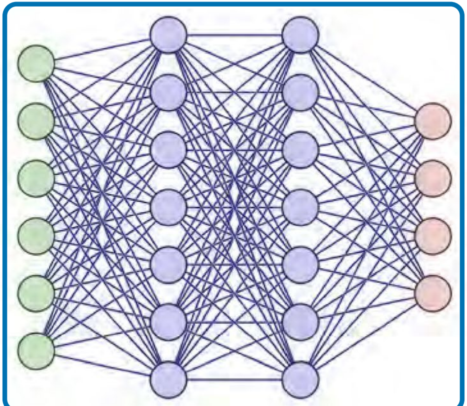
Algorithm

DeepLearning

Output

Rankings





Overview of the Counterfactual Analysis

Input

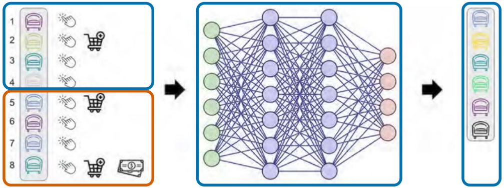
Browsinghistory

Algorithm

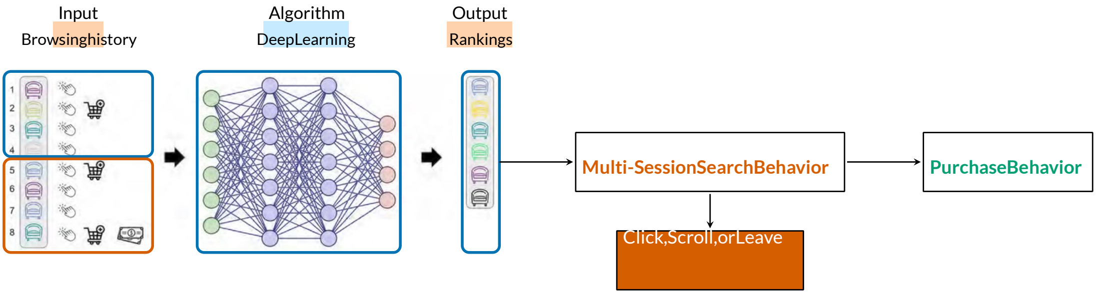
DeepLearning

Output

Rankings

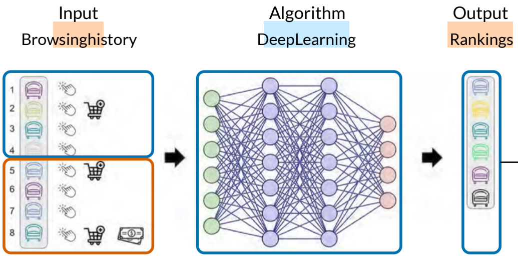


Overview of the Counterfactual Analysis



Overview of the Counterfactual Analysis

OVERVIEW OF THE MODEL STRUCTURE



Multi-SessionSearchBehavior → PurchaseBehavior

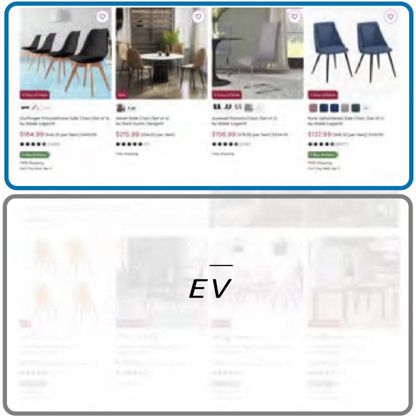
Click,Scroll,orLeave

Outcomesofinterest:changesin

- consumer choices
- seller revenue
- platform revenueandprofit

Overview of the Counterfactual Analysis

v_1 v_2 v_3 v_4



click $\longrightarrow$

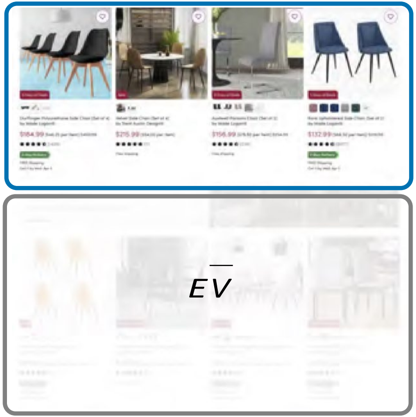
Awareness Set

Clicking cost

$$\max E[v_j | j \in A_t] - c_{0\, j \in A_t}$$

Overview of the Model Structure

V1 **V2** **V3** **V4**



click

Awareness Set

Clicking cost

$$\max E[v_j \mid j \in A_t] - c_{0\, j \in A_t}$$

scroll

Outside Awareness Set

Scrolling cost

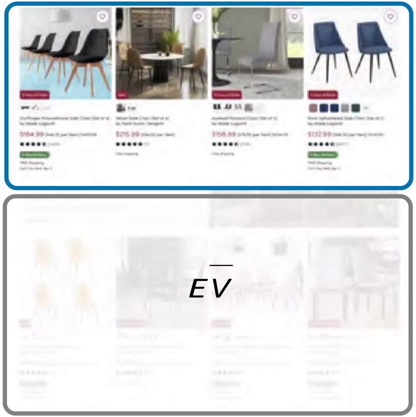
$$\max E[v_j \mid j \notin A_t] - c_{s\, j \notin A_t}$$

leave

purchase

Overview of the Model Structure

V1 **V2** **V3** **V4**



click →

scroll →

leave →

purchase

Awareness Set ← → Clicking cost

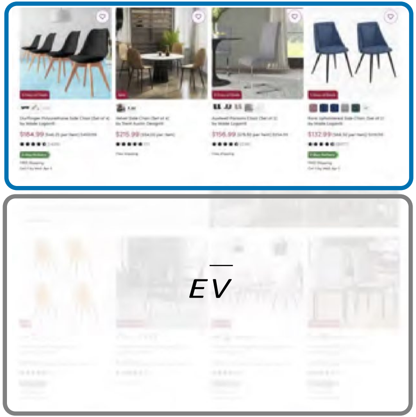
$$\max E[v_j | j \in A_t] - c_{0\, j \in A_t}$$

Outside Awareness Set ← → Scrolling cost

$$\max E[v_j | j \notin A_t] - c_{s\, j \notin A_t}$$

OVERVIEW OF THE MODEL STRUCTURE

V1 **V2** **V3** **V4**



click

scroll

leave

purchase

$v_0 = 0 \longrightarrow$ Outside option

Awareness Set

Clicking cost

$$\max E[v_j \,|\, j \in A_t] - c_0 \; {}_{j \in A_t}$$

Outside Awareness Set

Scrolling cost

21 / 26

Overview of the Model Structure

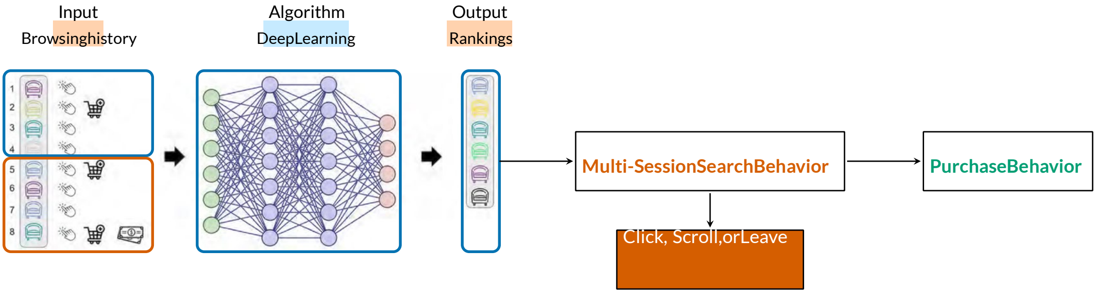
V1 *V2* *V3* *V4*

$$\max E[v_j \mid j \notin A_t] - c_s \quad j \notin A_t$$

$v_0 = 0 \longrightarrow$ Outside option

$v^* \longrightarrow$ Best value so far

Overview of the Counterfactual Analysis



Input
Browsinghistory
Algorithm
DeepLearning
Output
Rankings
1
2
3
4
5
6
7
8
Multi-SessionSearchBehavior
PurchaseBehavior
Click, Scroll,orLeave

Impact of Safari data restrictions

Counterfactual model estimates

IMPACT OF RESTRICTIONS

COUNTERFACTUAL

SAFARI DATA

MODEL ESTIMATES

Research Questions

1 Does **personalization hurt** consumers and sellers on the platform? ✓

2 How do privacy restrictions that limit consumer tracking impact the marketplace? ✓ 3 How can platforms **mitigate the losses** from privacy restrictions?

Step 3: Propose Probabilistic Algorithm + Structural model of search

Probabilistic Identity Recognition

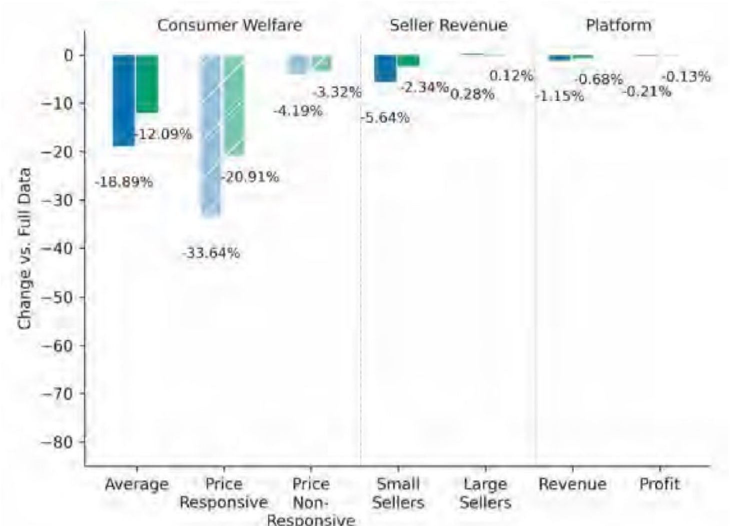
Algorithm developed in Korganbekova, Zuber (2023b)

- Use device-level IP address and behavioral (clickstream) data on Wayfair to **predict the user labels**
- XGBoost algorithm

Details

PROBABILISTIC RECOGNITION CAN IMPROVE OUTCOMES

COUNTERFACTUAL MODEL ESTIMATES


Consumer Welfare
Seller Revenue
Platform
Change vs. Full Data
0
−10
−20
−30
−40
−50
−60
−70
−80
-18.89%
-12.09%
-33.64%
-20.91%
-4.19%
-3.32%
-5.64%
-2.34%
0.28%
0.12%
-1.15%
-0.68%
-0.21%
-0.13%
Average
Price Responsive
Price Non-Responsive
Small Sellers
Large Sellers
Revenue
Profit

Key Takeaways

- Personalization benefits consumers, sellers, and the platform
- Privacy restrictions primarily hurt
  - smaller sellers
  - high search cost and price responsive consumers

- Platforms can partially mitigate these losses using probabilistic identityrecognition

Thank You!

Feedback welcome: malika.korganbekova@chicagobooth.edu

Appendix

CONSUMER OUTCOMES

	Logistic				*OLS*		
	Clicks (1)	Add-to-cart (2)	Basket page (3)	Converted (4)	Log(Revenue) (5)	Purchases (6)	Log(Profit) (7)
Personalized	0.002 (0.012)	0.011** (0.005)	0.014*** (0.005)	0.014** (0.005)	0.021*** (0.008)	0.024*** (0.008)	0.015** (0.006)

Intercept 2.988

~~Observations635,267~~(0.008)∗∗∗0.246~~635,267~~(0.004)∗∗∗0.148~~635,267~~(0.004)∗∗∗-0.870~~635,267~~(0.004)∗∗∗
1.947~~635,267~~(0.005)∗∗∗1.095~~635,267~~(0.006)∗∗∗~~635,267~~(0.004)–

Notes.
the logistic specification and Columns (5)-(7) report the OLS specification results. Robust standard errors in parentheses. The intercept in profit Column (7) is hidden for data sensitivity reasons. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
This table reports the output from the estimation of equation 1. Data is at the consumer-level. Columns (1)-(4) report

Back

CONSUMER OUTCOMES

$$y_i = \alpha + \beta treatment_i + \varepsilon_i \tag{1}$$

	Clicks (1)	Add-to-cart (2) *Logistic*	Basket page (3)	Converted (4)	Log(Revenue) (5)

Purchases*OLS*(6)Log(Profit)(7)

** *** ** *** *** **

Personalized (0.012)0.0020.011(0.005) 0.014(0.005) 0.014(0.005) 0.021(0.008) 0.024(0.008)
0.015(0.006)

Intercept 2.988

~~Observations635,267~~(0.008)***0.246~~635,267~~(0.004)***0.148~~635,267~~(0.004)***-0.870~~635,267~~(0.004)***

1.947~~635,267~~(0.005)***1.095~~635,267~~(0.006)***~~635,267~~(0.004)¯

Notes.
the logistic specification and Columns (5)-(7) report the OLS specification results. Robust standard errors in parentheses. The intercept in profit Column (7) is hidden for data sensitivity reasons. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

This table reports the output from the estimation of equation 1. Data is at the consumer-level. Columns (1)-(4) report

Back

$$y_{ij} = \alpha + \beta treatment_i + \varepsilon_{ij}$$

	Purchase Outcomes			
	(1)	(2)	(3)	(4)
	log(price)	log(quantity)	log(revenue)	log(profit)
		0.009	0.015	0.004
	(0.001)***	(0.001)***	(0.003)***	
Personalized group				0.005 (0.003)

PLATFORM OUTCOMES

Intercept	6.022***	2.075***	7.343***	–

(2)

Observations	~~2,022,708~~(0.001)	~~2,022,708~~(0.001)	~~2,022,708~~(0.002)

Notes.
consumer-purchased product level. The intercept in Column (4) is hidden for data sensitivity reasons. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
~~2,022,708~~(0.002) This table reports the output from the estimation of

equation 2. Data is at the

Back

Probabilistic Identity Recognition

Algorithm developed in Korganbekova, Zuber (2023b)

	(y)	**Device A and Device B**
~~device~~	~~1~~	~~IP address ... Screen Size ...~~
~~2~~device	0	~~Clickstream~~IP address ... Screen Size ...

~~device~~ ~~1~~device 3device 1	4device 3	1	ClickstreamIP address ... Screen Size ... Clickstream
Device A	Device B	Same consumer	Features of

Back

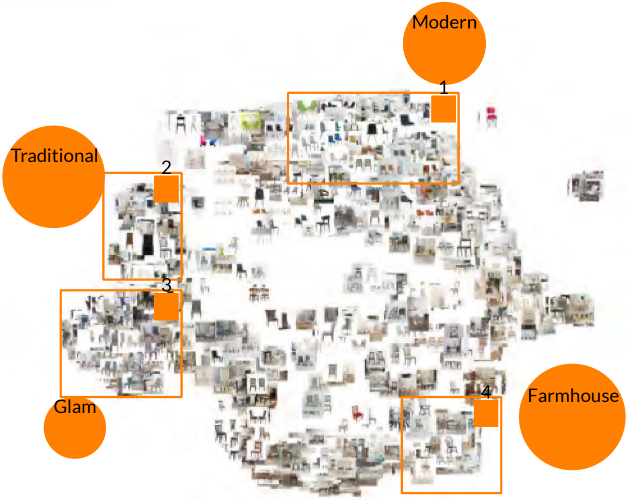


IMAGE EMBEDDIN

IDENTIFICATION

- Prior parameters from first searches: $\mu(X) = \alpha + X\beta_i$ and

!

$\kappa(X_j, X_k) =$

$$exp\left(- \sum_a \frac{(X_{ja} - X_{ka})^2}{\rho_a} \right.$$

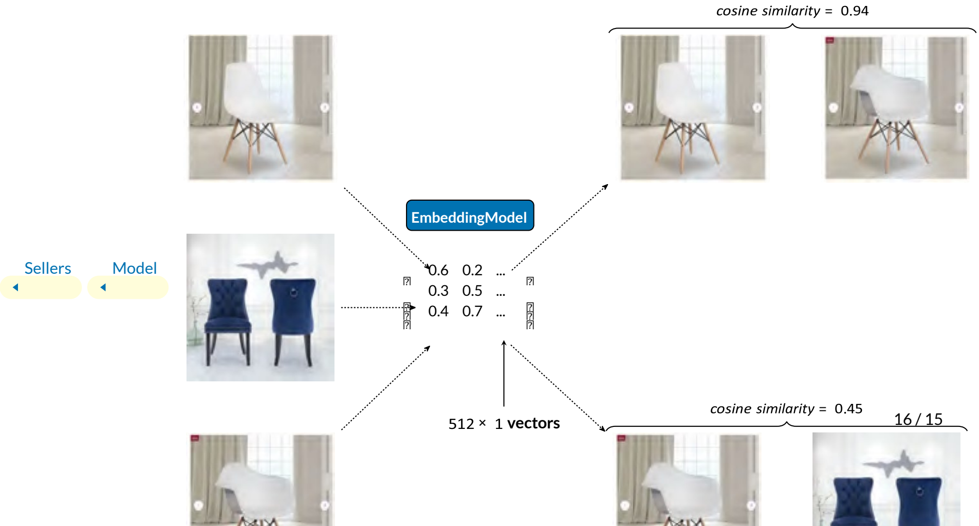
- higher variation in attributes in searched products across individuals versuswithin

$\Rightarrow \beta_i$

- Price parameters: Price experiments -12% to $+12\%$ random variation
- Scrolling cost: # of scrolls
- Clicking cost: experimental variation
- Product fixed effects ξ_j: from purchase—click probability
- Learning: consumer jumps in attribute space

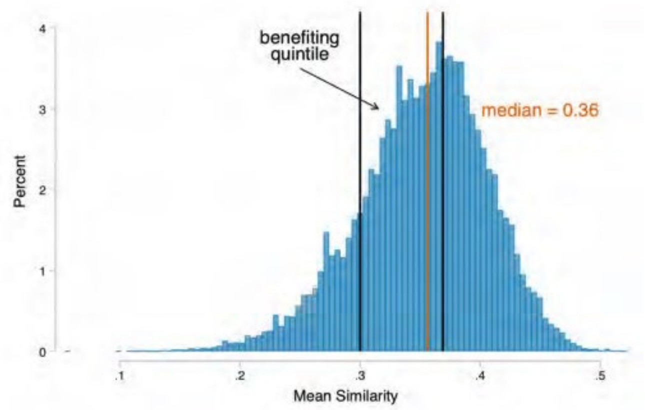
Back

Image Embeddings Idea


cosine similarity = 0.94
Sellers
Model
EmbeddingModel
0.6 0.2 ...
0.3 0.5 ...
0.4 0.7 ...
512 × 1 vectors
cosine similarity = 0.45

SELLER TYPES BY NICHENESS

Niche: low weighted cosine similarity to other products


benefiting
quintile
median = 0.36
Percent
Mean Similarity
0
1
2
3
4
.1
.2
.3
.4
.5

Sellers

RATING ANALYSIS

		Sentiment		
	rating	neutral	positive	negative
	(1)	(2)	(3)	(4)
1.treatment	0.004**	-0.013***	0.015***	-0.001
	(0.001)	(0.005)	(0.005)	(0.001)
Intercept	4.536(0.001)***	5.865(0.003)***	3.884(0.003)***	0.251(0.001)***

~~O~~ Class FE ~~bservations 1,707,183~~ ✓ ~~1,153,709~~ ✓ ~~1,153,709~~ ✓ ~~1,153,709~~ ✓

Notes.
is at the consumer-purchased product level. The intercept in Column (4) is hidden for data sensitivity reasons. Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

This table reports the output from the estimation of equation 2. Data

Back

RANDOMIZATION CHECKS

Back

~~Non-personalized~~(1)~~Personalized~~(2)~~Difference~~(3)~~p-value~~(4)

Historical Purchases ($)	886.98	937.77	-50.79 (-1.26)	0.21
Historical Quantity Bought	6.19	6.51	(-1.22) -0.32	0.22
Estimated networth	348,657.63	346,654.85	2,002.77(1.51)	0.13
Estimated income	70,847.86	70,779.64	68.22 (0.48)	0.63
Age	24.90	24.89	(0.52) 0.01	0.60
Home Value	175,067.06	175,177.83	-110.77(-0.19)	0.85
Prices searched	87.24	87.29		0.84
Gender dummy (0,1)	0.85	0.85		0.82

(-0.20) -0.05

(0.22) 0.00

Observations	319,783	315,484

“ When I saw these chairs online I thought they were probably too good to be true. The price is ridiculously low so I worried that the chairs would be cheaply made. Well they are VERY CHEAPLY MADE. I’m giving them 2 stars instead of one because they are cute and they are so inexpensive. I ordered six chairs total (2 sets from another online retailer and one set from Wayfair) and 4 of them were damaged in one way or another right out of the box. Some of it was minor damage like chipped paint and scratches and in once case the damage was severe where the composite wood seat was completely chipped up and pieces of wood were sheared off. I still liked the chairs though and forged ahead. I exchanged the severely damaged set and while I waited for those to be shipped I put the other chairs together and really found out how poorly made these are. When tightening the screws the wood actually started to split with hairline-like fractures along the sides of the seat. These fractures are visible. From far away the chairs look fine, but every single one of the six I ended up with is damaged in some minor and visible way (chipped paint, split wood, seats that don’t quite fit on the base properly). **Anyway,**

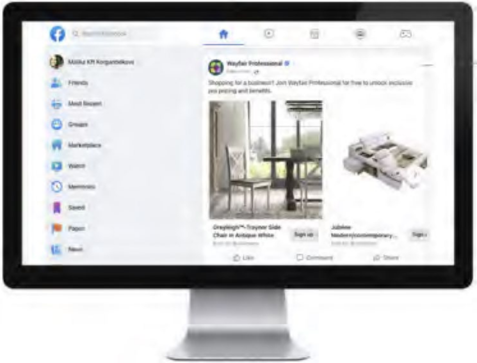
~~hoping that something so inexpensive would be acceptable in~~ quality.I would have gladly paid

more for chairs that were better made. I guess it’s my fault though for” Back

Within-website (first-party) tracking

Cross-website (third-party) tracking


wayfair
FRESH DEALS
Our Best Kitchen Brands
MOST WANTED
Outdoor Lighting Favorites for Less
Top-Rated Box Springs for Less
Hello, Rewards
The Best Outdoor Seating for Every Budget
The Best Wine Refrigerators for Every Budget


Wayfair Professional
Friends
Groups
Marketplace
Watch
Saved

- **Safari resets first-party cookies days of inactivity**
- **Chrome will block 3rd party after 7 cookies**

Back

IDENTITY FRAGMENTATION

Lin and Misra (2022) point out three types of possible biases

1. Outcome fragmentation ↑ observations but outcome variation is constant ✓ → direct attenuation

2. Exposure fragmentation ✗ omitted variables (data from other devices) → positive or negative bias

3 Spurious covariance $\text{cov}(\text{personalization exposure, device usage}) \sim \rightarrow$

positive or negative bias Results are attenuated

- checked on consumers who are 100% identified - estimates are larger

Back

Additional experimental results

Consumers in personalized group

- spend less overall time searching (-1.1%)
- give more positive reviews (+1.5%)

- no significant differences in time spent on product page vs ranking pagewithin a session

- within product pages they search more intensively
  - images (+0.7%)reviews (+0.4%)

Back