informs
https://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–22
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Welfare Effects of Personalized Rankings

Robert Donnelly,[a] Ayush Kanodia,[b] Ilya Morozov[c,*]

[a] Arena AI, New York, New York 10010; [b] Stanford University, Stanford, California 94305; [c] Northwestern University, Evanston, Illinois 60208
*Corresponding author
Contact: rndonnelly@gmail.com (RD); kanodiaayush@gmail.com (AK); imorozov@stanford.edu, https://orcid.org/0000-0003-1185-4215 (IM)

Received: March 25, 2021
Revised: January 31, 2022; October 17, 2022
Accepted: February 26, 2023
Published Online in Articles in Advance: May 15, 2023

https://doi.org/10.1287/mksc.2023.1441

Abstract. Many online retailers offer personalized recommendations to help consumers make their choices. Although standard recommendation algorithms are designed to guide consumers to the most relevant items, retailers can instead choose to steer consumers toward profitable options. We ask whether such strategic behavior arises in practice and to what extent it reduces consumers' benefits from personalized recommendations. Using data from a large-scale randomized experiment in which a large online retailer introduced personalized rankings, we show that personalization makes consumers search more and generates more purchases relative to uniform bestseller-based rankings. We then estimate a model of search and rankings and use it to reverse-engineer the retailer's objectives and to assess the effect of personalized rankings on consumer welfare. Our results reveal that although the current algorithm does put positive weight on profitability, personalized rankings still substantially increase consumer surplus. This case study suggests that online retailers may have incentives to adopt consumer-centric personalization algorithms as a way to retain consumers and maximize long-term growth.

History: Tat Chan served as the senior editor.
Supplemental Material: The data files and online appendix are available at https://doi.org/10.1287/mksc.2023.1441.

Keywords: recommender systems • product rankings • consumer search • welfare effects

1. Introduction

Many online companies use personalized recommendations to guide their users to the most relevant alternatives. Recommender systems have become ubiquitous and are now routinely used for recommending products, news articles, movies, and music. A highly cited McKinsey report estimates that recommendations drive at least 35% of consumer purchases on Amazon and 75% of viewing choices on Netflix.[1] Similarly, YouTube officials report that the site's recommendations drive over 70% of total watch time.[2] Deploying recommender systems has also become simple, and numerous services, such as Amazon Personalize, promise to help any website owner implement recommender systems with "no ML expertise required."[3] This decrease in the adoption cost, as well as the general effectiveness of recommender systems, may explain why almost 1.8 million U.S. companies were offering personalized recommendations on their websites by the end of 2021.[4]

On the surface, this shift sounds like great news for consumers. If personalized recommendations help consumers discover better-matching options, they should increase consumer surplus and lead to more transactions, thus improving the overall market efficiency (Koren et al. 2009, Bobadilla et al. 2013, Gopalan et al. 2013). However, there is a gap between the existing research on recommender systems and their use in real markets. Because recommendations strongly affect consumers' choices, companies may use them as an efficient marketing tool, much in the same way they would use personalized pricing or advertising. Anecdotal evidence supports this idea, indicating that Amazon, Taobao, Netflix, and Pandora have used personalized recommendations to reduce per-unit costs and promote profitable options.[5] In line with these anecdotes, the theoretical work on this topic suggests that companies have strong incentives to deviate from standard recommendation algorithms in order to optimize profit-related metrics (Hosanagar et al. 2008, Bourreau and Gaudin 2022). To the best of our knowledge, there is little empirical work on measuring such profit-driven distortions in real-world markets. In this paper, we measure such distortions in online retail and explore their effects on consumer welfare.

We analyze a large-scale experiment, in which a major U.S. online retailer personalized item rankings for some of its users. Personalized rankings are a common implementation of recommender systems because the order in which items are shown strongly affects users' choices (Ferreira et al. 2016, Ursu 2018). In the experiment, randomly selected users saw personalized item rankings tailored to their browsing histories, whereas others saw non-personalized rankings that ordered items by historical popularity. Using detailed click-stream data from this experiment, we show that personalized rankings induced more active consumer search and substantially increased the purchase probability. We also show that personalized rankings increase the aggregate diversity of purchased items, shifting demand toward items that were previously unpopular and generating a long tail effect in consumption (Anderson 2006).[6]

Although personalized rankings induce users to buy a wider range of items, it does not necessarily follow that they also increase consumer surplus. To measure the effect of personalization on consumer welfare, we need to estimate users' tastes and analyze the extent to which rankings direct each person toward the highest-utility items. To estimate tastes, however, we need to address three main challenges. First, we need to distinguish between items that are searched because users like them and those that are searched simply because they appear at the top of the page. This differentiation would help separate the true tastes from position effects. Second, we cannot estimate taste heterogeneity from panel data on repeat purchases in our application because repeat purchases of furniture are rare. Although we could use search data to estimate taste heterogeneity, as in Kim et al. (2010), this strategy would only partly address the issue because most users search very few items, resulting in data that are typically sparse. Third, it is difficult to capture users' tastes with only a handful of observed item attributes, especially in our setting, in which assortments are large and many choice-relevant attributes are high-dimensional (e.g., photos) or difficult to measure (e.g., design originality).

We develop an empirical framework that addresses all three challenges. Our framework is based on a simultaneous search model similar to those in De Los Santos et al. (2012) and Honka (2014). The user first decides which items to search based on observed rankings and prior knowledge of item attributes. Next, the user chooses which item to buy given the information gathered during search. We estimate this model using both personalized and non-personalized data as follows. We first estimate position effects by exploiting the fact that the non-personalized algorithm adds random noise to rankings before showing them to users. This exogenous variation, which is orthogonal to users' tastes, allows us to estimate the impact of moving an item to a higher position on the item's search and purchase rates. We extrapolate these estimates to the personalized sample, assuming that position effects are "stable" primitives that do not change under alternative ranking algorithms. After fixing position effects, we use data from the personalized sample to estimate taste heterogeneity, as well as the unknown parameters of the personalization algorithm.

Our empirical framework models how the retailer personalizes rankings for each user. We assume that when choosing rankings, the retailer balances two objectives: (a) maximizing the user utility and (b) maximizing short-term profitability. The personalization algorithm ranks items based on a weighted linear combination of these two objectives. As shown in Agarwal et al. (2012), one can view this linear specification as approximating a wider class of more complex multiobjective personalization algorithms. To proxy the short-term profitability, we use data on item markups obtained directly from the retailer. We then use data on rankings, searches, and choices from the personalized sample to estimate the weights of the two objectives. Recovering these weights is helpful for two reasons. First, the estimated weights reveal the extent to which personalized rankings are aligned with users' tastes, which directly affects the welfare estimates. Second, these weights also indicate the degree to which observed rankings are informative about individual tastes. Because the retailer has already inferred users' tastes from browsing histories using a dataset that is richer than ours, personalized rankings contain valuable information about individual tastes. We extract information about individual tastes directly from these personalized rankings by combining three different (noisy) signals: rankings, searches, and purchases. In this sense, our approach generalizes the empirical strategy of estimating individual heterogeneity from search and purchase data (Kim et al. 2010).

Another novel feature of our analysis is our model of consumer utility. Using the idea of latent factorization from computer science (Koren et al. 2009), we augment indirect utility with *latent attributes* that are observed by users, but not the researcher, and we allow users to have heterogeneous tastes for these attributes. Such a utility structure enables us to flexibly recover taste heterogeneity, even with limited data on item attributes. Several other authors have used this latent factorization approach to predict purchases (Jacobs et al. 2016, Wan et al. 2017); detect substitutes and complements among grocery items (Ruiz et al. 2020); and estimate flexible demand models with multiple unobserved product attributes (Elrod and Keane 1995, Goettler and Shacha 2001, Keane 2004, Athey and Imbens 2007, Athey et al. 2018, Donnelly et al. 2019). We contribute by applying latent factorization to a structural search model and showing how latent attribute structures can be identified from observed searches and rankings. By adopting this approach, we also contribute to the recent effort to use machine learning tools for the estimation of economic models with flexible consumer heterogeneity (Farrell et al. 2020).

Having estimated the model, we study the effect of personalized rankings on consumer surplus. To this end, we simulate consumer behavior under alternative ranking algorithms, varying the extent to which rankings are aligned with individual tastes. We first compare consumer surplus with personalization to that without personalization by simulating user behavior under the two ranking algorithms used in the experiment. Our results reveal that personalized rankings benefit both the retailer and the users. Whereas personalized rankings increase the average consumer surplus by $4 (2.5% of the average item price) compared with non-personalized rankings, they increase the retailer's expected revenues by 5.8%. We therefore do not find any evidence that the existing personalization algorithm is designed to increase short-term revenues at the expense of consumer welfare. To understand the drivers of this increase in consumer surplus, we further decompose the welfare estimates. We find that, on one hand, users incur higher search costs under personalization because they search more items. On the other hand, however, they now discover and purchase items that match their tastes much better, which more than offsets the utility wasted on search costs.

After estimating the weights that the retailer puts on utility and profitability, we explore hypothetical scenarios in which these weights are different. The estimated weights imply that the retailer puts nonzero weight on profitability, thus occasionally showing profitable, but not necessarily high-utility, items in the top positions. To put this result into perspective, we show that the retailer could have put a much higher weight on profitability, which would substantially reduce consumer surplus. Such a change could reduce users' benefits from personalization by a factor of two. At the same time, the retailer could have put zero weight on profitability, which would further increase consumer surplus by about 25%, or $1 per user. These results suggest that when resolving the trade-off between the maximization of short-term utility and short-term profits, the retailer has settled for a personalization algorithm that balances these two objectives and benefits both the retailer and the users. In other words, despite the minor profit-driven distortions documented here, users still extract substantial surplus from having access to personalized rankings.

Our work is related to three main strands of literature. The first strand is the large and active literature on recommender systems (Koren et al. 2009, Bobadilla et al. 2013, Gopalan et al. 2013). Much of this literature has focused on designing and testing recommender systems that help consumers discover relevant items, services, or information. Ever since the famous million-dollar Netflix challenge in 2009, the task of designing recommender systems has been formulated as a matrix completion problem, in which missing elements of the user-item taste matrix have to be predicted from limited historical data (Jannach et al. 2016). Driven by this formulation, researchers focused primarily on designing and studying recommender systems that maximize consumer-centric objectives (e.g., ratings or clicks). In this paper, we emphasize that companies may have strong incentives to modify these standard algorithms in practice in order to maximize profitability. Consistent with this argument, there has been increasing interest in recent years in developing profit-aware recommender systems that balance user-centric and profit-centric objectives (Panniello et al. 2016, Jannach and Adomavicius 2017, L'Ecuyer et al. 2017, Abdollahpouri et al. 2020). The recent literature suggests that such profit-aware recommender systems are already deployed in practice (Wang et al. 2022a, b). Practitioners call for additional research that would make next-generation recommendation systems account for product profitability (Underwood 2019), which raises further concerns about the impact on consumer welfare. In our case study, we find evidence of profit-driven distortions, but their magnitude is too small to reduce consumer welfare. This finding informs the ongoing discussion about whether regulators should control the range of personalization algorithms that online retailers can use (Koene et al. 2019).

Our paper also contributes to the recent literature on estimating the welfare effects of online personalization. The impact of personalization on consumers has been the subject of ongoing data privacy debate. Dubé and Misra (2023) present an empirical case study of personalized pricing at a large digital firm. Although they find that third-degree price discrimination can theoretically reduce consumer surplus, they also show that most consumers benefited from personalization. Goli et al. (2021) argue that firms can instead adopt "personalized versioning" strategies that personalize service quality, rather than price. Using results from a field experiment in which Pandora varied advertising intensity across users, they find that personalization slightly reduced the average consumer welfare. The personalized rankings we study can be viewed as an alternative approach for retailers to engage in "implicit" price discrimination. The retailer may, for example, identify price-insensitive users and show them relatively expensive items, leading them to pay more and preventing them from finding their optimal match. The possibility of such manipulations might justify increasing monitoring from regulators and consumer-protection groups, especially given that such practices are harder to detect than personalized prices.[7] We contribute to this discussion by considering a case study that allows us to explore whether personalized rankings in online retail are aligned with consumers' preferences.

Finally, our work is related to the recent marketing literature on optimal rankings. Several authors emphasized that companies face a trade-off between increasing consumer utility and maximizing profits when optimizing rankings (Ursu 2018, Choi and Mela 2019, Compiani et al. 2021, Zhou and Zou 2021). For example, Choi and Mela (2019) empirically show that ranking items by transaction revenue in an online marketplace tends to reduce consumer utility, whereas ranking items by utility may reduce profits. Compiani et al. (2021) document a similar tension between maximizing utility and maximizing profitability. They also develop a near-optimal ranking algorithm capable of improving both profits and consumer surplus. As a whole, this literature studies hypothetical rankings that retailers could have potentially adopted. By contrast, we analyze the actual implementation of a personalized ranking algorithm, which allows us to bridge the gap between the theory of optimal ranking algorithms and the implementation of these algorithms in practice.

2. Data and Institutional Details

Studying the effects of personalized rankings is generally challenging, as doing so requires a dataset that meets several criteria. First, it must include the exact rankings shown to each user. Without such data, it would be difficult to estimate the effect of rankings on choices, describe the current personalization algorithm, or study whether rankings steer users toward profitable options. Second, the degree of personalization must vary across users for reasons unrelated to their preferences. Such variation would help us address an endogeneity problem that can occur because the retailer, in general, will be more likely to personalize rankings for active users with long browsing histories, and we could mistakenly conclude that personalization makes users more active during their search. We now describe a dataset obtained from a large online retailer that meets these criteria.

2.1. Field Experiment

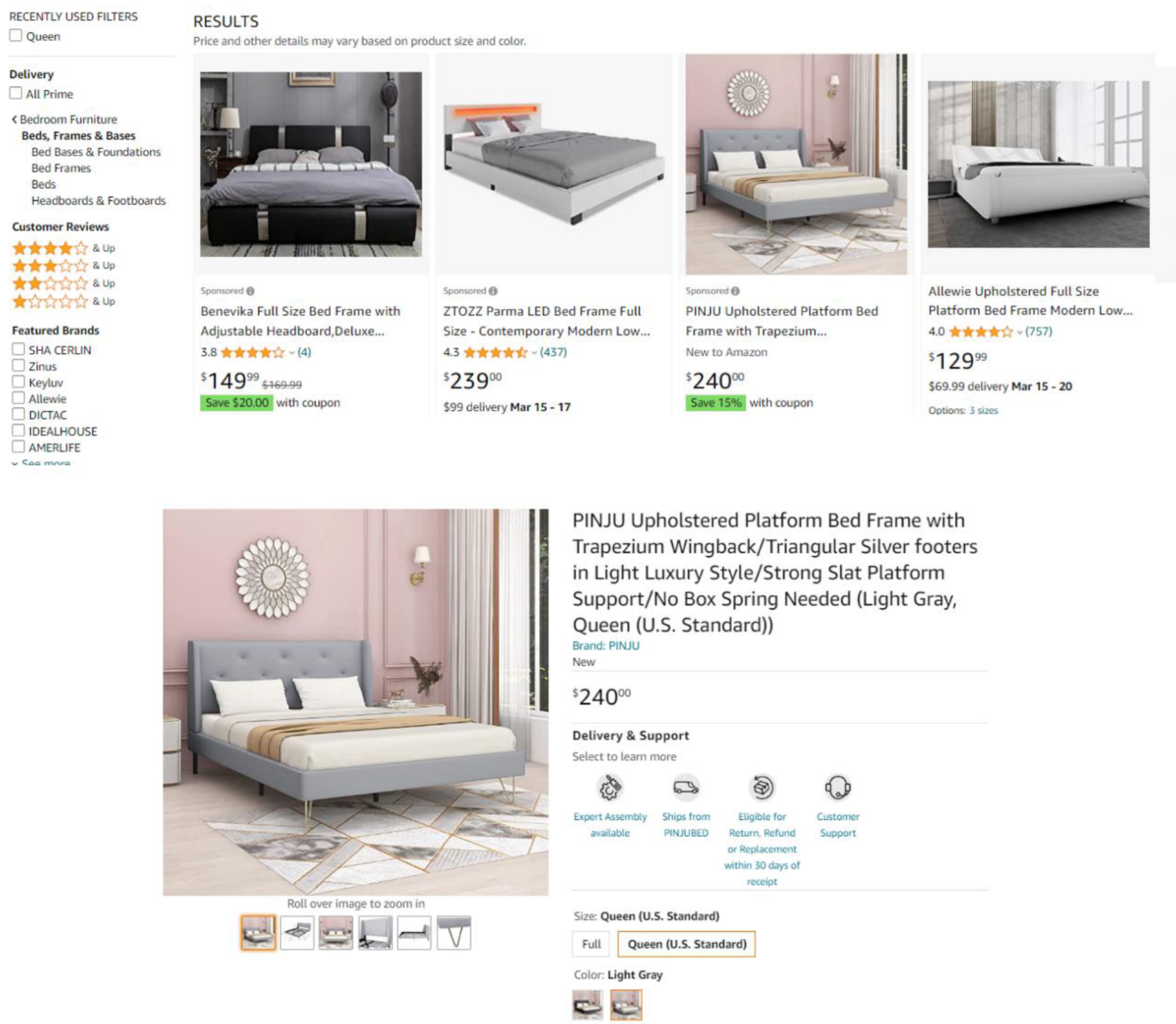
We have obtained a dataset from a large U.S. online retailer. The dataset describes the following randomized experiment. In 2019, the retailer randomized all new website users into two groups: *personalized* and *non-personalized*. With a probability of 95%, new website users were assigned to a personalized group in which the company used a proprietary algorithm to personalize item rankings in each category. The remaining 5% of new users were assigned to a non-personalized group, for whom items were sorted by historical popularity. The random assignment was implemented at the user level, so a given user would observe either always-personalized or always-non-personalized rankings in all product categories. We describe both ranking algorithms in more detail below.

Our dataset documents user behavior during the two-month period between September 9, 2019, and November 13, 2019, and it contains 1,930,992 users who visited the category of beds at least once during this period. All users in this sample made their first visit to the website at some point in 2019. We focus on the category of beds because it is a category that features a large assortment of differentiated items, presenting users with a challenging search problem. For each user in our data, we observe their experimental assignment (personalized or non-personalized), the number of times they visited the category within the two-month period, the item rankings they encountered upon each visit, the items they searched and purchased, and their socio-demographic data (e.g., age, gender, income, and living area). One limitation of our data is that we do not observe users' browsing histories before their first visit to this retailer, so we cannot directly verify that users in the two experimental groups have similar histories. Nevertheless, we were able to verify that users in these two groups are statistically similar in terms of demographic variables and general shopping habits (see Online Appendix A.1).

Users typically start their visits by opening the relevant category page (see the top panel of Figure 1). On this page, users observes item prices, pictures, and some basic attributes of items such as size, material, and style. Importantly, the order in which items are displayed in this list varies across the two experimental groups. Although we do not know the exact algorithm the retailer used for each group, we do observe the rankings each user encountered during the visit. After seeing the initial item list, users can then *search* specific items by exploring their product pages. These product pages reveal additional information about items, including textual descriptions, additional photos, and customer reviews (see the bottom panel of Figure 1). The visit ends when users either buy an item or leave the category without buying.

In the personalized group, the retailer uses a proprietary algorithm, which can be viewed as a modified collaborative filtering algorithm (Koren et al. 2009). The algorithm infers users' tastes from two kinds of data: browsing histories from the same product category of beds; and browsing histories from other related categories, such as night tables, armchairs, and mattresses. For example, if a user previously searched beds with a modern style, the algorithm will use these data to infer style preferences and show this user more beds that have a modern style in subsequent visits. It also makes cross-category inferences: If the user has recently searched queen-size mattresses, the algorithm will put more weight on displaying queen-size beds. Additionally, people who use search queries (e.g., "wooden beds") encounter more personalization, as their rankings are further personalized based on the search query itself. Lastly, users may encounter different amounts of personalization depending on the amount of information the retailer has about them. A completely new user who does not have an associated browsing history would see popularity-based rankings identical to those in the non-personalized group. In this sense, one can view our estimate below as an intention-to-treat estimate. Although the retailer attempts to personalize rankings for everyone, it can only

Figure 1. (Color online) Examples of a Category Page (Top Panel) and a Product Page (Bottom Panel)



Note. To preserve the retailer's confidentiality, we show examples from another online store that has a similar layout.

do so effectively for users with sufficiently informative browsing histories.

In the non-personalized group, users see uniform rankings that sort items by historical popularity. To construct these rankings, the algorithm computes the historical popularity based on the number of clicks and purchases each item attracted in the past few months. The algorithm also adjusts these raw popularity indices in order to promote sponsored or new items. We observe all such promotion instances in the data. Importantly, before showing items to users, the algorithm infuses rankings with exogenous variation by adding random noise to the computed popularity indices. In our empirical strategy, we exploit this exogenous variation to estimate the causal effect of rankings on users' choices. Finally, having made all these adjustments, the non-personalized algorithm displays items in order of decreasing popularity indices.

Our sample includes 1,930,992 users who visited the category of beds at least once during the observation period. Because many users visit the category of beds more than once, we select only the earliest category visit for each user. Choosing one visit per user allows us to adopt a clean definition of the item rankings that were displayed to each user. However, to compute the outcome variables, such as searches and purchases, we use data from the first visit and all subsequent visits. That is, we measure the effect of rankings that a user saw in the first visit on all subsequent searches and purchases. A limitation of these data is that we do not observe any long-term outcomes, such as which users came back to the website later in order to visit other product categories. We cannot, therefore, assess the extent to which personalized rankings can help the retailer to increase customer retention and maximize long-term profits.

2.2. The Effects of Personalized Rankings

We first compare user behavior in the personalized and non-personalized groups. Because users are randomly assigned to these two groups, this comparison yields the causal estimates of personalization effects, which serve as a starting point for our analysis. Later, in Sections 3 and 4, we will estimate a search model and study how the observed personalization effects translate into consumer welfare changes. Column (2) in Table 1 describes user behavior in the non-personalized group. Out of all category visitors, 31.7% of users search at least one item during the visit. Conditional on searching at least one item, an average user searches only 2.4 items, indicating limited consumer search. Users search more actively during visits that end in purchases: They search, on average, 2.2 items in the visits with purchases, but only 0.7 items before leaving the category without a purchase. Our model in Section 3 will rationalize this difference by assuming that some users have a stronger preference for buying a bed on this retailer's website (or a worse outside option), which induces them to search more and purchase with a higher probability. Lastly, only 2.3% of users make a purchase. Although it is common to see conversion rates of less than 5% in online retail, this scarcity of purchase data does suggest that we need to rely on search and ranking data for the estimation of users' preferences.

We compare this behavior to that in the personalized rankings group (see column (2) in Table 1). With personalized rankings, users are twice as likely to search at least one item (64.4% versus 31.7%, $p < 0.0001$). In addition, they search more than twice as many items, on average (1.92 versus 0.77, $p < 0.0001$), and they are 9% more likely to make a purchase (with purchase rates of 2.54% versus 2.33%, $p < 0.0001$). These results suggest that personalized rankings successfully display highly relevant items, which encourages users to search more and makes them more likely to find a good match. We also observe that personalization makes users more likely to search items located in higher positions ($p = 0.011$), suggesting that these positions display more appealing items relative to non-personalized rankings.[8] Table 1 also shows that personalization increases the expected revenue per user by 4.5%, although the difference is only marginally significant ($p = 0.01$). At the same time, we do not find any strong evidence that personalized rankings nudge users toward expensive items. If anything, users in the personalized group purchase slightly cheaper items, although the observed decrease of 4% is economically small. The change in revenues is therefore mostly driven by higher transaction volume, rather than higher prices paid by the users.

As expected, personalized rankings redistribute demand from bestsellers to less popular items, as illustrated in Figure 2. This shift likely occurs because many unpopular items, which were previously overlooked, are now frequently recommended to users who might find them to be a good match. As personalization boosts the demand for these unpopular options, it increases the variety of purchased items, thus generating a long tail effect in consumption (Anderson 2006).

Altogether, these results support the idea that personalization allows users to discover appealing items that would otherwise be difficult to find. Anecdotally, we find that personalized rankings increase the sales of beds with unusual designs and uncommon combinations of attributes. Several bed styles become substantially more popular under personalization, including *Slick and Chic Modern*, *Glam*, *Modern Contemporary*, and *Bold and Eclectic Modern*. Most beds of these styles have unusual shapes, provocative colors, and highly original designs. It is possible that these items strongly appeal to

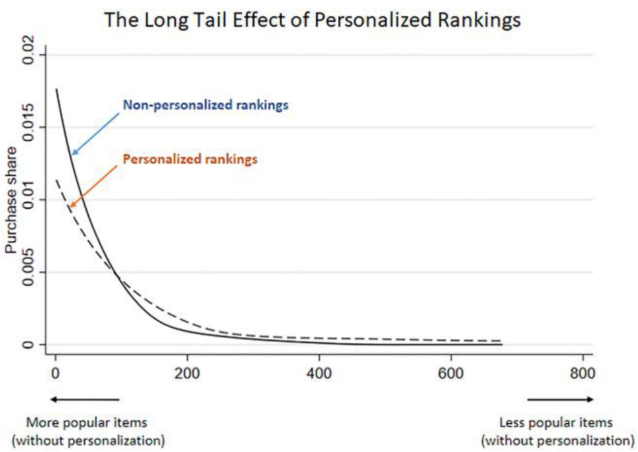
Table 1. The Effect of Personalized Rankings on Users' Search and Purchase Behavior and on the Retailer's Revenues

			Difference (person.-nonperson.)		
Variable	Nonperson. rankings	Personalized rankings	Change %	*t*-statistic	*p*-value
Searched any item	0.317	0.644	103.2	217.2	<0.0001
Searches (total)	0.77	1.92	149.4	143.3	<0.0001
Searches (if positive)	2.44	2.98	22.1	36.8	<0.0001
Searches (if purchase)	2.21	3.52	59.3	20.9	<0.0001
Searches (if no purchase)	0.74	1.88	154.1	142.3	<0.0001
Opened second page	0.141	0.182	29.1	39.2	<0.0001
Purchase rate	0.0233	0.0254	9.0	4.2	<0.0001
Revenue per user	100%[a]	104.5%	4.5	1.7	0.0997
Price searched items	100%[a]	116.0%	16.0	16.0	<0.0001
Price purchased item	100%[a]	96.0%	−4.0	−1.7	0.0878
Position searched items	21.7	20.4	−6.1	−2.5	0.0114
Position purchased item	14.9	16.9	12.9	0.7	0.4884

Notes. N = 1,930,992 users (N^p = 1,824,654 personalized (person.), N^h = 106,338 non-personalized (nonperson.)). Because users are randomly assigned to these two groups, the difference between the values in columns (1) and (2) estimates the causal effect of personalized rankings. The last two columns report t-statistics and p-values from the corresponding two-tailed tests of mean differences.

[a]To preserve data confidentiality, these three rows report revenues and prices only relative to their values in the non-personalized sample.

Figure 2. (Color online) Personalized Rankings Generate the Long Tail Effect in Purchases



Notes. The graph shows the distribution of purchase shares across 700 most popular items in our data for two groups of users, personalized (dashed line) and non-personalized (solid line). On the horizontal axis, we order items by their popularity in the non-personalized group, with the most popular items at the left of the graph. We measure popularity by the number of times each item was purchased in the non-personalized sample. The graph shows that personalized rankings redistribute demand from popular bestsellers toward relatively unpopular items.

some consumers, while leaving others indifferent. At the same time, personalization reduces demand for more "mainstream" furniture styles, including *Beachy*, *Modern Rustic*, *Traditional*, *American Traditional*, and *Ornate Traditional*. Most beds of these styles are standard beds in neutral colors that are made of wood and use common upholstery materials. Overall, these contrasting examples suggest that the personalized ranking algorithm successfully identifies users with relatively unusual tastes and guides them toward appealing items. In the remainder of the paper, we quantify the effect of personalization on consumer welfare.

3. Estimating Welfare Effects of Personalized Rankings

3.1. Empirical Search Model

3.1.1. Indirect Utility. Consider a product category where N users (indexed $i = 1, \dots, N$) choose among J different items (indexed $j = 1, \dots, J$). Each user i only observes a subset of items $J_i \subseteq J$ determined by the ranking algorithm. We assume that the indirect utility user i derives from item j is given by:

$$u_{ij} = \underbrace{-\alpha_i p_{j,t(i)} + x_j'\beta_i + \xi_j'\theta_i}_{\delta_{ij}\text{ pre-search utility}} + \varepsilon_{ij}, \tag{1}$$

where $p_{j,t(i)}$ is the price of item j on the day $t(i)$ of user i's visit; x_j is a vector of K_O observed item attributes; ξ_j is a vector of K_L latent attributes; and α_i, β_i, and θ_i are user i's price sensitivity and attribute preferences. This model mimics the actual information environment in which users make choices in our application. We assume that for items in the set J_i, the user observes prices $p_{j,t(i)}$ and item attributes x_j and ξ_j without searching, which is the case in the category of beds where users observe prices and basic attributes of beds on the category page. The observed attributes x_j include the bed's style and material. We include material in x_j because it is often explicitly mentioned in the item's name (e.g., "Bandini Wood Bed" or "Chuckanut Metal Bed") or is easy to guess from the photos. Additionally, we include the bed's style in x_j, assuming that from the item's photos, users can gauge reasonably well which beds belong to similar styles.[9] It is plausible that users can infer other bed attributes from the photos (e.g., design originality and perceived comfort). We capture this possibility using latent attributes ξ_j. Because attributes ξ_j are observed to the users, but not the researcher, we will need to estimate them together with other parameters.

The term ε_{ij} is the match value of item j for user i, which captures additional information that the user acquires upon closer inspection of the product page (e.g., information about the ease of assembly and stain resistance). The user can also choose the outside option for which the presearch utility is normalized to zero, so $\delta_{i0} = 0$ and $u_{i0} = \varepsilon_{i0}$. Choosing the outside option can be interpreted as continuing search in another store or deciding not to buy a bed at all. We assume that ε_{ij} and ε_{i0} are independent and identically distributed (i.i.d.) type I extreme value, and we normalize their variance to one to fix the utility scale, so that $\sigma_\varepsilon^2 = 1$.

3.1.2. Consumer Search. We assume that users search according to the simultaneous search model (Chade and Smith 2006). We choose this model for its computational convenience. As we detail below, this model allows us to derive analytical solutions for the probability that the user chooses to search a specific set of items. These analytical solutions simplify likelihood estimation and make it possible for us to estimate the model with latent factors ξ_j and flexible consumer heterogeneity in α_i, β_i, and θ_i. This model is different from the sequential search model of Weitzman (1979), in which we would need to numerically integrate the likelihood with respect to different feasible search paths. The simultaneous search model has been extensively used in the empirical literature for estimating demand under information frictions (De Los Santos et al. 2012, Honka 2014, Moraga-González et al. 2015, Honka et al. 2017, Armona et al. 2021). Additionally, in Online Appendix F, we provide simulation results suggesting that switching to a sequential search model would likely have only marginal effects on our key welfare estimates.

Users search as follows. They observe the presearch utilities of items $\delta_{ij} = -\alpha_i p_{j,t(i)} + x_j'\beta_i + \xi_j'\theta_i$, but they have to search in order to learn the match values ε_{ij}. We assume that users must visit the product page of item j in order to learn the match value ε_{ij} and that this value is completely revealed to them upon opening that page. The user chooses how many and which items to search, thus committing to a specific search set $S \subseteq J_i$. For each item included in the set S, the user has to pay a search cost c_{ij}, which we parameterize below. Once the search set S is selected, the user examines all items in this set, learns their exact utilities u_{ij}, and chooses the highest-utility item.

Given the assumed distribution of match values ε_{ij}, the expected net benefit to user i of searching all items in the set S, denoted by m_{iS}, is the difference between the expected maximum utility of items in this set and the total cost of searching these products:

$$m_{iS} = E\left[\max_{j\in S}\{u_{ij}\}\right] - \sum_{j\in S} c_{ij} = \log\left(\sum_{j\in S}\exp(\delta_{ij})\right) - \sum_{j\in S} c_{ij}, \tag{2}$$

where the outside option is always included in S and has zero presearch utility and zero search cost ($\delta_{i0} = 0$, $c_{i0} = 0$). It is optimal for user i to choose the search set S that maximizes the expected net benefit m_{iS}. Following De Los Santos et al. (2012), we smooth the choice set probabilities by adding a mean-zero stochastic noise term η_{iS} to the expected benefit m_{iS} of each potential search set. The stochastic term η_{iS} might be interpreted as reflecting errors in an individual's assessment of the net expected gain of searching the set S. As we will demonstrate below, smoothing choice probabilities significantly simplifies estimation.[10] In practice, we do not lose much from adding an idiosyncratic error term; at the same time, we gain substantial robustness to measurement error in observed searches.

Assuming η_{iS} is i.i.d. type I extreme value distributed with scale parameter σ_η, the probability that the user i finds it optimal to search the set S conditional on observing the ranking R is then:

$$P_{iS\,|\,R} = \frac{\exp(m_{iS}/\sigma_\eta)}{\sum_{S'\subseteq J_i}\exp(m_{iS'}/\sigma_\eta)}, \tag{3}$$

where J_i is the set of items the user observes under the ranking R. Note that the summation in the denominator runs across all possible sets S' that the user could have formed from the set of displayed items J_i. Because the number of possible search sets is very large, in the actual estimation, we approximate this sum using simulation (see Section 3.3 for details).

Having resolved the uncertainty about match values ε_{ij}, the user chooses the highest-utility option within the selected set S. Given that match values are distributed i.i.d. type I extreme value, the probability of user i choosing item j is given by the standard logit formula:

$$P_{ij\,|\,S} = \frac{\exp(\delta_{ij})}{\sum_{k\in S}\exp(\delta_{ik})} \quad \text{for} \quad j \in S. \tag{4}$$

The purchase probability is zero for items outside the set S, because we assume the user has to search an item before buying it.

The outside option plays an important role, as it enables us to model users who either search but do not buy anything, or decide not to search at all after seeing item rankings. We could, in principle, remove data on users without purchases and write a simpler model in which users must always make a purchase. Because less than 10% of users make purchases, however, this simplification would mean discarding over 90% of the data on searches and rankings that are highly informative about taste heterogeneity. An advantage of our model, therefore, is that it enables us to use more of the available data in estimation. It also allows us to understand whether personalized rankings can persuade some users to make a purchase by showing them highly relevant items.

Following Ursu (2018), we model position effects by assuming that the search cost c_{ij} depends on the position in which item j is displayed to user i. We assume that the item in the first position has a search cost c, which we interpret as a *baseline search cost*. The items in the subsequent positions have search costs $c + p_2, c + p_3, \ldots, c + p_{48}$, where we interpret p_r as the *position effect* for position r (for $r = 2, \ldots, 48$). In theory, we could estimate all position effects $p_2, \ldots, p_{48}$ nonparametrically, thus recovering a flexible position effect curve. In practice, we simplify the problem by estimating only position effects $p_{2,\ldots}p_{10}$ for the first 10 positions, and we separately estimate p_{11} which captures the average position effect for all remaining positions. Estimating more flexible search costs functions makes little difference for our qualitative results, mostly because the vast majority of searches and purchases land on the first ten positions. Because items closer to the bottom of the page are more difficult to locate and search, we expect them to have higher estimated position effects p_r (and therefore higher search costs c_{ij}). Lastly, we assume that rankings affect choices only indirectly via search costs, but not directly by influencing utilities u_{ij}. This assumption follows the results in Ursu (2018), who shows that, in a randomized experiment, rankings shift search probabilities, but do not directly affect purchase probabilities conditional on search.

3.1.3. Latent Factor Approach. Our empirical framework combines a simultaneous search model from economics and marketing with the latent factorization idea from computer science (Koren et al. 2009). In this model, users have heterogeneous preferences for both observed item attributes x_j and multiple latent attributes ξ_j. It helps to think about this combination in the context of the discrete choice modeling literature. The idea of specifying a choice model with this level of flexibility can be traced back at least to McFadden (1981), but few authors have followed this approach. Some notable exceptions are papers that used the latent factor approach to predict purchases (Jacobs et al. 2016, Wan et al. 2017), detect substitutes and complements among grocery items (Ruiz et al. 2020), and estimate flexible demand models in which tastes depend on multiple unobserved attributes (Elrod and Keane 1995, Goettler and Shachar 2001, Keane 2004, Athey and Imbens 2007, Athey et al. 2018, Donnelly et al. 2019). Our innovation is to use the latent factorization approach within a structural search model.[11]

Combining a search model with a flexible utility structure allows us to get the best of two worlds. On the one hand, the added flexibility helps us capture rich substitution patterns. By modeling heterogeneity as a function of multiple observed and latent attributes, we effectively impose a convenient low-dimensional structure on the covariance matrix of utilities u_{ij}. This feature makes our approach more flexible than that of Berry et al. (1995), which mostly relies on observed attributes x_j and includes only one latent attribute ξ_j entering utility functions of all users with the same coefficient. As discussed in Section 3.4, this added flexibility proves critical for recovering taste heterogeneity from data on rankings, searches, and purchases. On the other hand, the structure of the search model enables us to construct a series of counterfactuals exploring how users would have changed their behavior under alternative ranking algorithms. It also helps us interpret estimated tastes and search costs as structural parameters, thus giving us a meaningful and interpretable measure of consumer surplus.

3.2. Model of Rankings

3.2.1. Non-personalized Rankings. We also model how the retailer selects rankings for each user, both in the personalized and non-personalized groups. Our main goal here is to capture the primary aspects of both ranking algorithms. Although we do not know the exact algorithms the company uses, we know from our private conversations with the retailer which variables serve as key inputs to each of them. Our general strategy is to specify a family of algorithms with that structure and estimate the unknown parameters directly from the observed rankings.

We model non-personalized rankings as follows. To construct rankings for user i, the retailer sorts items in the order of decreasing indices v_{ij}, showing only $\overline{R}$ items with the highest index values. This process generates the effective choice set J_i of user i. Because more than 90% of users only look at the first page in our application, we simplify the analysis by assuming that the user does not see other items outside of the set J_i. We then set $\overline{R}$ to the actual number of items the retailer shows on the first page of the category list. The index v_{ij} is given by:

$$v_{ij} = \tilde{w} \cdot \underbrace{(-\overline{\alpha} p_{j,t(i)} + x_j'\overline{\beta} + \xi_j'\overline{\theta})}_{\overline{u}_{ij}\text{ mean utility}} + \tilde{\gamma}' z_{j,t(i)} + \mu_{ij}^h \quad \textit{if } i \in \text{Non-personalized}, \tag{5}$$

where $\overline{\alpha}, \overline{\beta}, \overline{\theta}$ are users' mean tastes, $\tilde{w}$ is the weight that the retailer puts on the mean utility, $t(i)$ is the day of user i's category visit, $z_{j,t(i)}$ is the $M \times 1$ vector of time-varying item-specific characteristics, $\tilde{\gamma}$ is the $M \times 1$ vector of regression coefficients, and μ_{ij}^h is a stochastic term distributed i.i.d. type I extreme value with scale parameter σ_μ^h. The vector $z_{j,t(i)}$ in our application includes indicators for new and sponsored items, which accounts for periods when the retailer promotes such items at the top of the page. The stochastic term μ_{ij}^h captures the exogenous variation that the algorithm is adding to the non-personalized rankings.

To estimate position effects, we use exogenous variation in rankings generated by the non-personalized algorithm. This algorithm computes item popularity indices that capture how frequently users searched and purchased a given item in the past few months. Assuming that user tastes remain stable over time, we can capture the historical popularity of items in Model (5) using the average tastes $\overline{\lambda}$.[12] The algorithm then adjusts popularity indices v_{ij} for new and sponsored items (captured by the $\gamma' z_{j,t(i)}$ term), and it adds some random noise before showing rankings to users. Therefore, the algorithm infuses non-personalized rankings with exogenous variation. Our general strategy is to isolate such variation using the stochastic term μ_{ij}^h, which allows us to observe how the same item moves across positions for reasons unrelated to unobserved demand shocks. This variation enables us to estimate position effects by measuring to what extent shifting an item to a higher position increases its search and purchase probabilities. The main identifying assumption we rely on is that, conditional on the average utilities $\overline{u}_{ij}$ and covariates $z_{j,t(i)}$, the idiosyncratic changes in rankings μ_{ij}^h are independent from the tastes of individual users u_{ij}.

Given the extreme value distribution of stochastic terms μ_{ij}^h, the probability that the retailer chooses the ranking R has a closed-form solution given by the so-called "exploded logit" formula (Punj and Staelin 1978). Because the scale parameter σ_μ^h is not separately identified from coefficients $\tilde{w}$ and $\tilde{\gamma}$, we divide all terms in Equation (5) by σ_μ^h, so that the stochastic term μ_{ij}^h has a variance of one and $w = (\tilde{w}/\sigma_\mu^h)$ and $\gamma = (\tilde{\gamma}/\sigma_\mu^h)$ are the normalized coefficients that we will recover during estimation. Without loss of generality, assume that the retailer shows item $j = 1$ in the first position, item $j = 2$ in the second position, and so on until position $\overline{R}$ is filled with item $j = \overline{R}$. Then, the probability that the retailer finds ranking R is optimal is given by:

$$P_{iR} = \prod_{r=1}^{\overline{R}} \frac{\exp(w\overline{u}_{ir} + \gamma' z_{r,t(i)})}{\sum_{J \geq k \geq r} \exp(w\overline{u}_{ik} + \gamma' z_{k,t(i)})}, \tag{6}$$

where $\overline{u}_j$ is the mean utility of item j from Equation (5). The product in (6) is taken across all $\overline{R}$ positions displayed to the user. The first term of this product is the probability that item $j = 1$ has the highest index v_{ij} out of all J items, the second term is the probability that item $j = 2$ has the highest index v_{ij} among the remaining $J - 1$ items, and so on. We note that taking a log of the likelihood in (6) yields a sum of R terms, each of which is a standard logit probability.

3.2.2. Personalized Rankings. We also model how the retailer selects personalized rankings. Here, we pursue two objectives. First, to reverse engineer the existing personalization algorithm, we need to specify a class of possible algorithms and estimate the unknown parameters from the data. Second, we use this model to extract information about heterogeneous tastes from the personalized rankings shown to different users. Our model recognizes that rankings partly reveal information that the retailer has acquired about each user's tastes from their browsing histories. Therefore, ranking data potentially contain rich information about taste heterogeneity. Our approach contributes to the prior literature by incorporating these data into estimation, while at the same time recognizing that the retailer may modify personalized rankings based on criteria other than utility maximization (e.g., profitability). This aspect of the model distinguishes our approach from the prior work, which primarily recovered taste heterogeneity using panel data or repeat purchases (Rossi et al. 1996, 2012) or individual search data (Kim et al. 2010).

To model personalized rankings, we assume, as before, that the retailer sorts items in the order of decreasing indices v_{ij} and shows only $\overline{R} = 48$ highest-index items to user i. The indices v_{ij} are defined as:

$$v_{ij} = \tilde{w}_u \delta_{ij} + \tilde{w}_\pi \pi_{ij} + \mu^p_{ij} \quad \textit{if } i \in \text{Personalized}, \qquad (7)$$

where $\delta_{ij} = -\alpha_i p_{j,t(i)} + x'_j \beta_i + \xi'_j \theta_i$ is the presearch utility defined in (1); π_{ij} is the variable measuring the profitability of selling the item j to user i, which we explain below; $\tilde{w}_u$ and $\tilde{w}_\pi$ are weights the retailer puts on the two objectives (maximizing utility and profitability); and μ^p_{ij} is a stochastic term distributed i.i.d. type I extreme value with the scale parameter σ^p_μ. We do not include the match value ε_{ij} in the ranking index v_{ij}, thus implicitly assuming that the retailer knows as much about this match value as the user. Although in reality, the retailer might partly infer match values from browsing histories, including ε_{ij} in Equation (7) would make estimation computationally burdensome as it would prevent us from using a convenient closed-form likelihood function in (8).

The term μ^p_{ij} reflects that the retailer does not perfectly observe the tastes δ_{ij} and personalizes rankings based on a noisy estimate. Suppose the retailer does not observe δ_{ij} and only receives a noisy signal centered on δ_{ij} with some variance σ^2_s. This signal might come from a prediction algorithm that utilizes historical data to predict the tastes of individual users for a given item. Having received such a noisy signal, the retailer forms a posterior belief about δ_{ij} and ranks items accordingly. Under the assumption that the signal variance σ^2_s is the same for all items and users, we can show that: (a) The variance of μ^p_{ij} in the ranking Equation (7) is equal to the signal variance σ^2_s, and (b) the utility weight $\tilde{w}_u$ in (7) is the "true" weight the retailer puts on maximizing utility, adjusted by the signal variance σ^2_s (see Online Appendix C for details).

This microfoundation clarifies several points about our modeling approach. First, apart from reflecting the retailer's incentives, the weight $\tilde{w}_u$ also reflects the retailer's ability to infer individual tastes. A low estimated utility weight $\tilde{w}_u$ implies that the retailer either places low weight on maximizing utility or places high weight on utility, but is unable to personalize rankings due to noisy inference. For example, if the signal variance σ^2_s is large, then personalized rankings will be misaligned with preferences, and we will estimate a high variance σ^p_μ and a low utility weight $\tilde{w}_u$ attenuated toward zero. Second, we assume the uniform signal variance σ^2_s across users. In reality, however, the amount of personalization each user gets depends on how much the retailer knows about them. It would be better to allow the signal variance σ^2_s to be lower for users with longer browsing histories, but doing so is not feasible in our application given that we do not observe browsing histories. In this sense, our model captures the average amount of uncertainty the retailer faces when learning individual tastes.[13] An important direction for future research would be to model how the retailer personalizes rankings, while knowing a lot about some users and little about others.

One practical question is how to compute the profitability π_{ij} in the ranking Equation (7). In the main specification, we proxy profitability with item-specific markups, so that $\pi_{ij} = \text{markup}_j$. To construct this profitability metric, we use data on average item markups obtained directly from the retailer, which leads to an intuitive ranking model. Depending on weights $\tilde{w}_u$ and $\tilde{w}_\pi$, the retailer may highlight high-utility items (items with high δ_{ij}), highlight profitable items (items with high π_{ij}), or balance between the two objectives. In Online Appendix G.2, we show that our welfare results are robust to using alternative profitability metrics π_{ij}.

Given the distribution of stochastic terms μ^p_{ij} in Equation (7), we can once again use the exploded logit formula to compute the likelihood of observing specific rankings R. As before, the scale parameter σ^p_μ is not separately identified from weights $\tilde{w}_u$ and $\tilde{w}_\pi$, so we fix the variance of the stochastic term to one and estimate the normalized weights $w_u = \tilde{w}_u / \sigma^p_\mu$ and $w_\pi = \tilde{w}_\pi / \sigma^p_\mu$. The probability that the retailer finds it optimal to choose the ranking R is then equal to:

$$P_{iR} = \prod_{r=1}^{\overline{R}} \frac{\exp(w_u \delta_{ir} + w_\pi \pi_{ir})}{\sum_{J \geq k \geq r} \exp(w_u \delta_{ik} + w_\pi \pi_{ik})}, \qquad (8)$$

where, as with non-personalized rankings, the product runs across all $\overline{R}$ positions displayed to the user.

The ranking model in (7) approximates the retailer's actual personalized ranking algorithm. It is relatively inconsequential to assume a linear ranking index v_{ij}, given that one can often express complex multiobjective recommendation systems as a simpler algorithm that sorts items based on a weighted linear combination of objectives.[14] At the same time, we simplify the actual algorithm in several ways. First, due to data limitations, we consider only two potential objectives of the retailer: maximizing short-term utility δ_{ij} and maximizing short-term profitability π_{ij}. Although such a model does not capture the retailer's other incentives (e.g., maximizing long-term profits, maximizing advertising revenues), it does allow us to study the trade-off between utility-centric and profit-centric rankings. Second, we do not consider the problem of choosing "truly optimal" rankings in which the retailer chooses one ranking out of all possible item orderings. Such a model would be impossible to solve because the number of possible rankings is astronomically large. To circumvent this issue, we instead consider a simplified model in which the retailer sorts items by simple indices v_{ij}, which we parameterize and recover from the data. This model can be viewed as an approximation of the full optimization problem. By considering this simplified ranking rule, we obtain a more practical model that is straightforward enough to estimate and yet sufficiently flexible to help us study the retailer's incentives.

Despite its apparent simplicity, the ranking model in (7) is capable of capturing a wide range of personalization strategies. First, note that by changing the weights, the retailer can either show utility-based rankings (high w_u), show profitability-based rankings (high w_π), or adopt an interior solution that puts nonzero weights on both objectives. Choosing a nonzero weight w_π is also not equivalent to moving high-markup items to the top of the page for all users. When $w_\pi > 0$, high-markup items will only be shown to users who are sufficiently likely to purchase them—that is, users who derive high utility δ_{ij} and, therefore, have a high purchase rate s_{ij}. For example, the retailer will only show high-markup items to users who value other attributes x_j that this item offers.

3.3. Maximum Likelihood Estimation

Suppose we have data on N users and observe each user's ranking R_i, selected search set S_i, and purchase decision $y_i \in S_i$. Our goal is to recover the heterogeneous tastes $\lambda_i = (\alpha_i, \beta_i, \gamma_i)$, search cost c, position effects $(p_2, \ldots, p_{11})$, scale parameter σ_η, and ranking parameters (γ, w, w_u, w_π). We start by computing the log-likelihood of observing data for a specific user i with the taste profile λ_i:

$$\log L_i(\lambda_i;\omega) = \log P_i(R_i) + \log P_i(S_i \,|\, R_i) + \log P_i(y_i \,|\, S_i), \tag{9}$$

where ω is a vector of all unknown parameters, $P_i(R_i)$ is the likelihood of observing the rankings R_i, $P_i(S_i|R_i)$ is the likelihood of choosing a search set S_i given the rankings shown to user i, and $P_i(y_i|S_i)$ is the likelihood of a purchase $y_i \in S_i$ given the search set S_i. We compute the exact values of log-likelihoods for purchases and rankings, $\log P_i(y_i|S_i)$ and $\log P_i(R_i)$, using the derived closed-form solutions in (4), (6), and (8). In addition, we approximate the search likelihood, $\log P_i(S_i|R_i)$, as described below. We then estimate parameters ω and types λ_i by maximizing the total log-likelihood across N users in the dataset, computing it as $Log\,L(y,S,R) = (1/N)\sum_{i=1}^{N} \log L_i(\lambda_i;\omega)$.

The main challenge comes from computing the term $\log P_i(S_i|R_i)$, which contains a summation over a very large number of possible search sets S':

$$\log P_i(S_i \,|\, R_i) = (m_{iS}/\sigma_\eta) - \log \sum_{S' \subseteq J_i} \exp(m_{iS'}/\sigma_\eta). \tag{10}$$

This summation is computationally burdensome, so we need to approximate it. However, we cannot easily form an unbiased estimator of this log-likelihood because of the nonlinearity introduced by the logarithm. To solve this issue, we follow Ruiz et al. (2020) and apply the one-versus-each bound from Titsias (2016) to write:

$$\log P_i(S_i \,|\, R_i) \geq \sum_{S' \subseteq J_i} \log \sigma\left(\frac{m_{iS} - m_{iS'}}{\sigma_\eta}\right), \tag{11}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function. Therefore, instead of maximizing the log-likelihood function in (10), we can maximize a lower bound of this function. Although replacing the objective function with its lower bound may introduce bias, in practice, our simulations show that the bound is sufficiently tight and that we are able to recover the true values of parameters when using it in estimation.

Because applying this bound moves the logarithm function under the summation, we can form an unbiased estimator of the summation via subsampling. Specifically, we sample potential search sets S' from the universe of all possible search sets on J_i, and we approximate the right-hand-side expression in (11) using the following unbiased estimator:

$$\frac{|\mathcal{S}_i|}{|\mathcal{B}_i|} \sum_{S' \in \mathcal{B}} \log \sigma((m_{iS} - m_{iS'})/\sigma_\eta), \tag{12}$$

where $\mathcal{S}_i$ is the set of all possible search sets the user i could form given rankings R_i, and $\mathcal{B}_i \subseteq \mathcal{S}_i$ are (feasible) alternative sets we sample that are different from the set S_i actually selected by user i. In Online Appendix D, we explain how we select the alternative search sets $\mathcal{B}_i$ in actual estimation.

3.4. Identification

3.4.1. Position Effects $p_2, \ldots, p_{11}$. We identify position effects using the exogenous variation in rankings displayed in the non-personalized group. Our model of non-personalized rankings in (5) uses mean utilities $\overline{u}_{ij}$ to capture the average item popularity, and it uses time-variant item characteristics z_{jt} to capture promotions of new and sponsored items. This specification allows us to isolate the residual within-item variation in rankings by capturing it with random shocks μ^h_{ij}. By design, in the retailer's non-personalized algorithm these shocks are conditionally independent from user-specific tastes u_{ij}; therefore, the isolated variation induces exogenous shifts in item rankings r_{ij}, which we use to identify position effects $p_2, \ldots, p_{11}$.

In our application, we recover position effects from non-personalized data, and we then extrapolate these estimates to the personalized sample. For this extrapolation to work, we assume that the position effects $p_2, \ldots, p_{11}$ parameters are "stable" primitives that do not change under alternative ranking algorithms.[15] To extrapolate the estimates, we divide the estimated position effects by the average price coefficient $\overline{\alpha}$ to convert them into dollar terms. We then fix the parameters $p_2, \ldots, p_{11}$ throughout our estimation in the personalized sample so that the dollarized position effects match the pre-estimated values.

This approach to estimating position effects differs from the prior literature, in which researchers used a regression-discontinuity approach (Narayanan and Kalyanam 2015), a propensity-score estimation (Schnabel et al. 2016), or experiments in which rankings are fully randomized (Ferreira et al. 2016, Ursu 2018). Whereas these researchers aimed to maximize the exogenous variation in rankings, we want our analysis to recognize that rankings are endogenous in order to infer tastes from personalized rankings. We therefore use a hybrid approach that exploits the exogenous variation in non-personalized rankings, while also recognizing that personalized rankings are endogenous with respect to users' tastes and can therefore be used as a source of preference information. A potential advantage of our approach over fully randomized experiments is that random rankings might show users irrelevant and unappealing products, thus making them think the ranking algorithm is unreliable and affecting their beliefs—and, therefore, choices—in unpredictable ways.

3.4.2. Search Cost c and Scale Parameter σ_η. The baseline search cost c is identified from the average number of items searched by users. Conditional on the selected search set S_i, the purchase decision depends only on tastes, but not on search costs. Therefore, we can separate search costs from taste parameters by using conditional purchase probabilities. Additionally, we identify the scale parameter σ_η from the extent to which the model can explain the observed searches using item prices p_{jt} and attributes x_j and ξ_j. The estimated value of σ_η, then, indicates whether the observed search sets provide useful information for estimating taste heterogeneity.

3.4.3. Tastes α_i and β_i. We now explain how we identify mean tastes, as well as taste heterogeneity from rankings, searches, and purchases in the personalized sample. We identify the mean tastes $\overline{\beta}$ from two key sets of moments: (a) users' propensity to search and purchase items with certain values of attributes x_j, and (b) the retailer's propensity to display items with certain values of attributes x_j in the top positions of personalized rankings. Consider an example in which most users prefer wooden beds over beds made of other materials. The retailer will then frequently fill the top positions on the page with wooden beds, and most users will often search wooden beds, even when these beds are not displayed at the top. Therefore, conditional on position effects pre-estimated in Section 3.4.1, we identify the mean tastes $\overline{\beta}$ from the average patterns in search and ranking data.

The identification of the mean price coefficient $\overline{\alpha}$ is somewhat different because the price of a given item may vary over time. We therefore identify the average price sensitivity $\overline{\alpha}$ from the extent to which temporary price discounts increase search and purchase probabilities of the discounted items and shift these items to higher positions in rankings. Recovering $\overline{\alpha}$ enables us to translate personalization benefits into dollar terms, thus making our estimates of consumer surplus interpretable. One may worry that within-item price variation correlates with unobserved temporal demand shocks—for example, seasonal demand fluctuations or temporary advertising campaigns. However, our conversations with the retailer revealed that such endogeneity is not a significant issue in the product category we analyze. As it turns out, much of the temporal price variation we observe comes from a sequence of price experiments, which affected almost half of all the items in our dataset.[16] We have also verified that the temporary price discounts we observe in the data rarely coincide with obvious candidates for demand shifters, such as weekends and holidays. Based on these observations, we assume throughout our analysis that prices p_{jt} are exogenous with respect to unobserved taste shocks.

Next, we identify the heterogeneity in tastes λ_i from individual-level data on searches and rankings. To see how searches inform taste heterogeneity parameters, suppose there is substantial heterogeneity in price sensitivities α_i and preferences for specific materials $\beta_i^{Material}$ (e.g., wood, metal). Because users make search decisions based on the observed prices p_{jt} and attributes x_j of items, this preference heterogeneity will generally lead to highly polarized search decisions. For example, pricesensitive users may decide to almost exclusively search cheap items and ignore other alternatives, whereas users who value high-quality materials may focus on searching expensive wooden beds, while showing little interest in beds made of cheaper materials. As a result, a typical consumer will choose a search set S_i that contains relatively homogeneous items (i.e., with similar prices p_{jt} and attributes x_j), even though these search sets will look substantially different across consumers. As we show in Online Appendix B, such concentration of search sets is observed for most of the attributes x_j in our data. We can therefore identify the heterogeneity in tastes λ_i from the extent to which items searched by the same user are more similar to each other in terms of prices p_{jt} and attributes x_j than items searched by different users.

Another set of moments we use to identify heterogeneity is based on the observed personalized rankings. According to the ranking model in (7), the retailer infers each user's tastes δ_{ij} from historical data and personalizes rankings accordingly. Therefore, the fact that the retailer shows different rankings to different users is itself informative about taste heterogeneity. For instance, consider a pair of beds j and k that have the same style (e.g., modern). For users who previously revealed a strong preference for modern furniture, the retailer may choose rankings that place both beds j and k at the top. By contrast, the retailer may remove both of these beds from the rankings of users who have revealed themselves to have either neutral or negative tastes for modern furniture. We can therefore identify heterogeneity in users' tastes for style β_i^{Style} from the extent to which personalized rankings consistently display beds of the same style next to each other. Following the same logic for different item attributes, we can identify taste heterogeneity from co-ranking patterns in the data.[17] In Online Appendix B, we document such co-ranking patterns for most of the item attributes in our data.

This identification argument distinguishes our approach from prior work, which primarily recovered taste heterogeneity using panel data or repeat purchases (Rossi et al. 1996, 2012) or search data (Kim et al. 2010). Panel data on purchases are unavailable in our application, as is the case for most durable goods. Additionally, a typical user in our data only searches a few items, which limits how much we can learn about these users' tastes from search data.[18] By contrast, data on personalized rankings contain information about all items shown to a user, rather than only items the user searched. Because displayed rankings partly reveal information the retailer has learned about users' tastes from their browsing histories, ranking data potentially contain rich information about taste heterogeneity. Our approach incorporates these data into estimation, while at the same time recognizing that rankings might be imperfectly correlated with user utilities due to profit-driven distortions.

3.4.4. Latent Attributes ξ_j and Heterogeneity of Tastes θ_i. We now discuss identification of latent attributes and associated taste heterogeneity. Consider a pair of beds j and k that are frequently searched together and often displayed together in the top positions of personalized rankings. These beds must appeal to users with similar tastes. Suppose, however, that these co-search and co-ranking patterns do not correlate with any of the item attributes x_j that we observe in the data. Given the structure of the utility model, we then infer that beds j and k are located close to each other in the space of latent attributes ξ_j. We can then rationalize the data patterns above by assuming that the two beds have similar values of a certain latent attribute and that a significant share of users has strong tastes for this attribute. Therefore, we identify latent attributes ξ_j from co-search and coranking patterns that cannot be explained by the similarity of observed attributes x_j.

We emphasize that it is not possible to identify specific parameters of the latent factorization $\xi_j'\theta_i$. To see this, note that if a latent attribute ξ_j^m is perfectly collinear with some observed attribute x_j^k, it is then impossible to separate the corresponding taste parameters θ_i^m and β_i^k from each other. The latent attributes ξ_j are also interchangeable and therefore suffer from the label-switching problem. But although we cannot identify separate parameters in this factorization, we can identify (a) the sum $\xi_j'\overline{\theta}$, which essentially plays a role of an item-specific residual similar to Berry et al. (1995); and (b) the pairs of items for which the utilities are correlated, which helps identify the correlation structure of utilities u_{ij}.

Fortunately, we only need to identify the averages $\xi_j'\overline{\theta}$ and pairwise correlations of item utilities. The main goal of our counterfactual analysis is to examine whether users benefit from personalized rankings. To answer this question, we need to predict the items that users would have searched and purchased had they been assigned to a non-personalized group and encountered different choice sets J_i and rankings r_{ij}. We must, therefore, infer the items that users would have substituted to, and our ability to make this inference relies directly on our knowledge of the utility averages and correlations. Therefore, we can identify the key counterfactual quantities, even though we cannot identify specific components of the latent factorization.

3.4.5. Ranking Algorithm Weights w_u and w_π. To identify the weights w_u and w_π in the Ranking Algorithm (7), we contrast the actual tastes of users with the personalized rankings they are shown. It helps to think about this estimation problem as recovering two objects: the utility weight w_u and the ratio of weights w_π/w_u. We identify the utility weight w_u from the extent to which users see the highest-utility items at the top positions. This weight will generally be high if the retailer knows users' tastes well and shows rankings that are highly aligned with these tastes. We note that the estimated value of the weight w_u will also show to what extent personalized rankings contain relevant information for recovering users' tastes. Next, to identify the ratio w_π/w_u, we examine whether highly profitable items (i.e., high π_{ij}) are more frequently shown in the top positions than would be optimal under utility-based rankings with $w_\pi = 0$. For example, if the retailer displays expensive, high-margin beds more often than warranted by the mean price sensitivity $\overline{\alpha}$, the size of the discrepancy indicates how much weight the retailer puts on profitability relative to utility.

4. Estimation Results

We estimate our model using the maximum likelihood approach described in Section 3.3. The main computational challenge is dealing with a large number of users N and items J. To ease the computational burden, we estimate a non-personalized model using a random subsample of $N^{hold} = 50,000$ users in the non-personalized group. We then construct an equal-sized personalized subsample of $N^{pers} = 50,000$ users by randomly drawing them from the personalized group. We use the remaining data from both groups, which are not included in the estimation samples, for out-of-sample model validation (see Online Appendix G for details).

The selection of items for estimation presents a practical challenge. Because the raw data include tens of thousands of items, it would be too ambitious to attempt the estimation of latent factors ξ_j for all these items, especially given that most items are almost never searched or purchased in our sample. On one hand, we need to restrict the sample to a more manageable set of items. On the other hand, we do not want to drop too many items because many items have nonzero demand in the personalized sample (see Figure 2). Dropping them, therefore, would eliminate a disproportionate number of searches and purchases from the personalized sample. To strike this balance, we keep only $J = 243$ most popular items, where we define popularity as the market shares of items across the two user groups, personalized and non-personalized. This assortment corresponds to a set of all items that have market shares above 0.1% in our data. With this selection criteria, we retain 70.1% of searches and 88.1% of purchases from the original dataset (see Table 11 in Online Appendix G for details). Because the estimation becomes too burdensome with more than $J = 243$ items, this need to select relatively few items is a limitation in our empirical approach.

4.1. Position Effects

We start by estimating position effects from non-personalized data. To this end, we use the Log-Likelihood Function (9), while assuming that rankings are determined as in Equation (5). Table 2 presents the key estimates from this estimation, while Figure 3 (left panel) interprets the estimated search costs c and position effects $p_2, \ldots, p_{11}$ by converting them into the implied dollar values of search costs by position. We make several observations based on these estimates. First, we obtain relatively large estimates of search costs that vary between \$13.50 and \$31.40, depending on the position. Given the predicted search probabilities for different positions, users pay a search cost of about \$26 among items they actually search. This high estimated search cost is natural, given that about 70% of users in the non-personalized sample do not search at all, and those who do search rarely click on more than two or three items (see Table 1). The high estimated search cost is also in line with the previous literature, in which researchers often obtain high estimates of search costs in datasets with limited search (Chen and Yao 2016, De los Santos and Koulayev 2017, Ursu 2018).

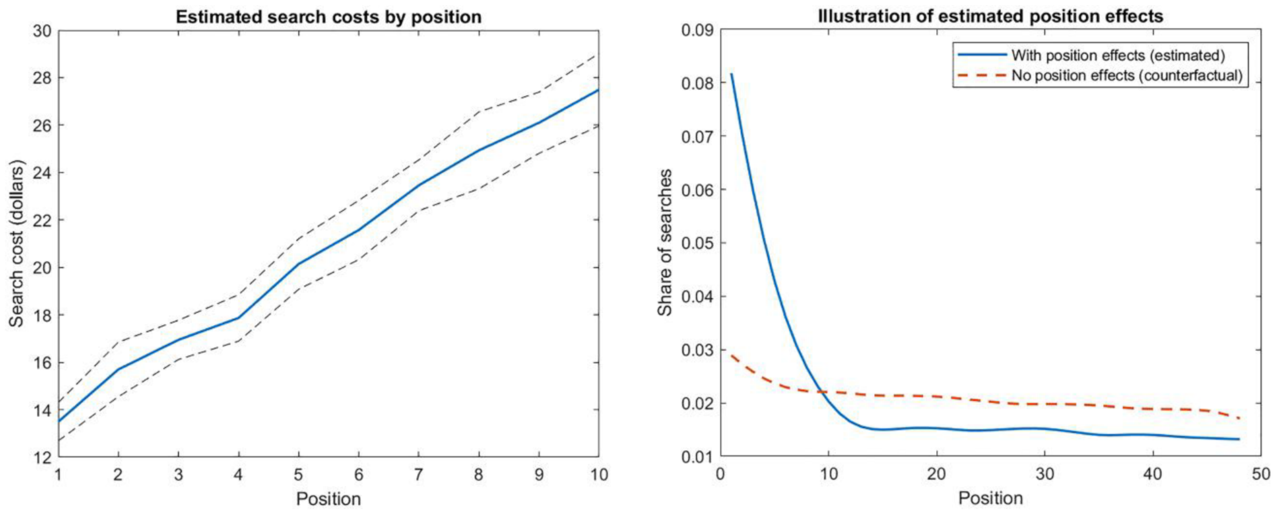
Second, we find evidence of strong position effects. Whereas this search cost is \$13.50 for the first position, it increases by about \$1.80 per position and doubles by position 10, reaching \$27.5. Figure 3 (right panel, solid line) illustrates these estimated position effects by plotting the predicted distribution of searches across

Table 2. Selected Parameter Estimates from the Model of Non-personalized Rankings

Parameter	Coef.	Est.	S.E.
Search costs			
Position 1 (baseline)	c	0.043	(0.001)
Position 2	$c+p_2$	0.050	(0.002)
Position 3	$c+p_3$	0.050	(0.001)
Position 4	$c+p_4$	0.053	(0.002)
Position 5	$c+p_5$	0.061	(0.002)
Position 6	$c+p_6$	0.066	(0.002)
Position 7	$c+p_7$	0.073	(0.002)
Position 8	$c+p_8$	0.078	(0.003)
Position 9	$c+p_9$	0.083	(0.002)
Position 10	$c+p_{10}$	0.088	(0.001)
Positions 11-48	$c+p_{11}$	0.098	(0.000)
Taste parameters			
Price coefficient	$\overline{\alpha}$	−0.317	(0.054)
Scale parameter	σ_η	0.036	(0.001)
Algorithm parameters			
Weight on avg util.	w	0.452	(0.071)
New item boost	γ_1	0.059	(0.003)
Spons. item boost	γ_2	1.193	(0.013)
N users		50,000	
J items		243	
LogL		−210.1	

Notes. We obtain these estimates using a maximum likelihood estimation procedure developed in Section 3.3. The sample used in this estimation includes N = 50,000 users randomly selected from the non-personalized group. We observe these users choosing among J = 243 unique items. The reported *LogL* value corresponds to the average log-likelihood value at the optimum, where the average is taken across N users in the sample. Coef., coefficient; est., estimate; S.E., standard error; spons., sponsored; util., utility.

Figure 3. (Color online) Estimated Search Costs and Implied Position Effects



Notes. The graph on the left shows the estimates of search costs for the first 10 positions of the page, which we obtained from the non-personalized model (see Table 2). The solid line in the graph shows point estimates, while the dashed lines indicate the 95% confidence intervals. The graph on the right shows the distribution of searches across positions for the estimated model (the solid line) and compares it to the distribution of searches in a counterfactual model from which we have removed position effects (the dashed line).

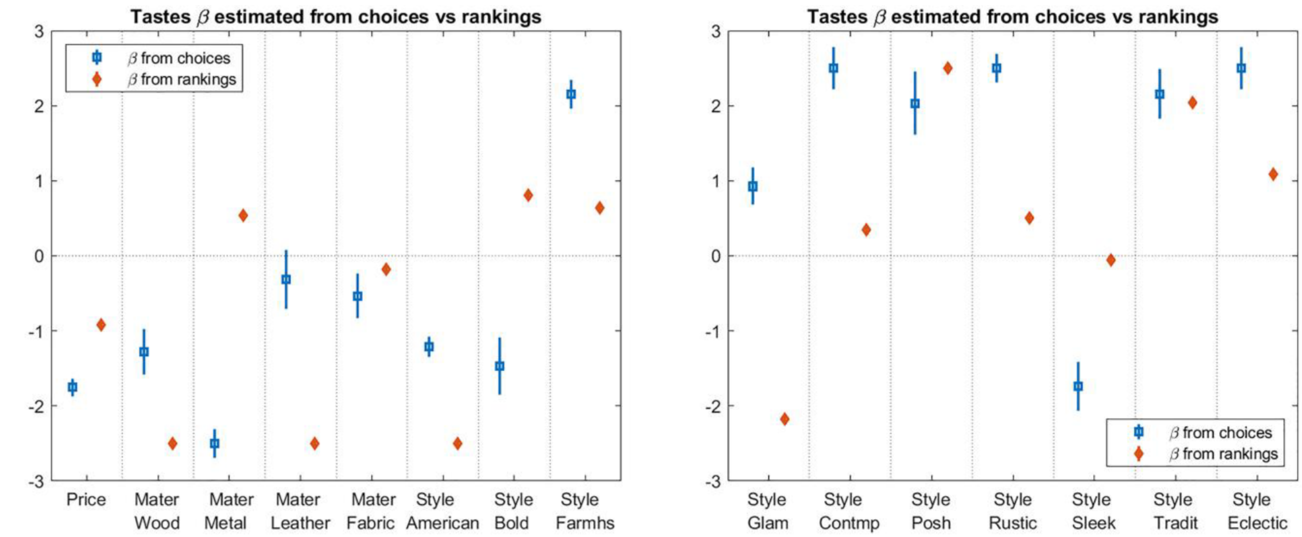
positions. An item shown in the first position is about twice as likely to be searched than an item in position 5 and three times more likely to be searched than items in positions 10–15. Moving an item from position 1 to position 48 at the bottom of the page reduces its search probability more than 10 times. The same figure (right panel, dashed line) also shows the predicted search probabilities when we fix search costs for all positions at the average level of $26. The comparison reveals that shutting down position effects generates a more even distribution of searches across positions and reduces the search rates of items at the top five positions by a factor of two to three. Note, however, that higher positions still attract more searches, reflecting the fact that these positions are generally filled with more appealing items. The strong estimated position effects are consistent with the results of our randomized experiment, which reveals that personalized rankings strongly affect users' search and purchase behavior (see Section 2.2).

4.2. Retailer's Incentives

Armed with the estimates of position effects, we now present preliminary evidence that personalized rankings display items in an order that's aligned with the users' actual tastes. We also show that this alignment is imperfect, possibly because the algorithm puts some weight on showing profitable items. To this end, before estimating the full model of search and rankings from Section 3.3, we estimate the average user tastes $\overline{\lambda}$ in two different ways. First, we estimate tastes $\overline{\lambda}$ using only data on displayed personalized rankings, ignoring the actual choices. For this estimation, we only use the ranking likelihood $\log P_i(R_i)$ based on the personalized ranking Equation (7), and, for the moment, we set the profitability weight w_π to zero; we relax this assumption below. The resulting estimates, denoted as $\overline{\lambda}_R$, indicate which item attributes are favored by the current personalization algorithm. Second, we estimate tastes $\overline{\lambda}$ from the observed searches and purchases by using only two terms of the likelihood function in (9): the purchase likelihood $\log P_i(y_i | S_i)$ and the search likelihood $\log P_i(S_i | R_i)$. Importantly, we fix the search costs c and position effects $p_2, \ldots, p_{11}$ using the estimates we obtain from non-personalized data in Section 4.1. Without fixing these parameters at their pre-estimated levels, we would be unable to separate tastes from position effects because personalized rankings do not contain relevant exogenous variation (see Section 3.4). This second estimation procedure helps us infer the "actual" user tastes, denoted as $\overline{\lambda}_U$. By comparing the two estimates $\overline{\lambda}_U$ and $\overline{\lambda}_R$, we can then contrast the attributes users are seeking with the attributes that are promoted by the personalized rankings. If the retailer uses personalized rankings to steer users away from high-utility items and toward profitable options, the implied tastes $\overline{\lambda}_R$ should then deviate from the actual user tastes $\overline{\lambda}_U$.

Figure 4 shows the estimates $\overline{\lambda}_U$ and $\overline{\lambda}_R$ for 16 observed attributes that enter the utility function (1).[19] These attributes include price as well as multiple indicator variables capturing the bed's material and style. Even though the two sets of estimates were obtained from two completely different types of data, the estimated tastes $\overline{\lambda}_U$ and $\overline{\lambda}_R$ are remarkably aligned. The correlation between the two sets of point estimates is 0.51. Beds with certain attributes—for example, "Style Farmhouse" or "Style Eclectic"—are both appealing to users

Figure 4. (Color online) Comparison of Estimated Mean User Tastes Obtained Only from Choices vs. from Rankings



Notes. Both graphs show the estimates of mean tastes $\overline{\lambda}$ obtained from two estimators: (1) using only data on searches and purchases (the first estimate on the left for each attribute); and (2) using only data on personalized rankings without accounting for search and choice (the second estimate on the right for each attribute). The bars reflect 95% confidence intervals.

(positive $\overline{\lambda}_U$) and often shown in personalized rankings (positive $\overline{\lambda}_R$). In turn, beds with attributes, such as "Material Wood" and "Style American," are relatively undesirable to users (negative $\overline{\lambda}_U$), and the current algorithm rarely shows them at the top (negative $\overline{\lambda}_R$).

Despite this general alignment, we observe that for several attributes, personalized rankings deviate from the actual tastes of users. Notably, we estimate somewhat lower price coefficient from rankings than choices, implying that the retailer is more likely to show expensive items than would be implied by the actual price sensitivity of users. We also obtain estimates of different signs for several attributes, including "Material Metal" and "Style Bold." Although both attributes are undesirable, according to the estimated tastes, the current algorithm often shows items with these attributes at the top. Overall, we find that the attributes of items displayed by the personalized rankings are correlated (but are not perfectly aligned) with the actual tastes of users. It is then natural to ask whether and to what extent these deviations from the actual tastes affect consumer welfare.

4.3. Welfare Effects of Personalized Rankings

We estimate the full model with search and rankings by using the likelihood function in (9) and the sample of $N^{pers} = 50,000$ users with personalized rankings. For this estimation, we use data on searches, purchases, and the personalized rankings that were displayed to each user. We fix the position effects $p_2, \ldots, p_{11}$ at the level at which we pre-estimated them from non-personalized data in Section 4.1. Table 3 shows the point estimates and bootstrap standard errors for the key parameters in the full model. We estimate the search cost parameter to be $\hat{c} = 0.031\,(0.006)$. Because this estimate is about 30% lower than $\hat{c} = 0.043\,(0.001)$ in the non-personalized sample (see Table 2), our estimates suggest that personalized rankings reduce search costs—for example, by making it more enjoyable to search.

Turning to the estimated ranking algorithm, we estimate the weights in Equation (7) to be $\hat{w}_u = 1.286\,(0.482)$ for utility and $\hat{w}_\pi = 0.620\,(0.062)$ for profitability. So, whereas personalized rankings are mostly aligned with user's utilities δ_{ij}, the retailer also puts considerable weight on profitability, thus occasionally showing items that are expensive, but not necessarily the most

Table 3. Selected Parameter Estimates from the Model of Personalized Rankings

Parameter	Coef.	Est.	S.E.
Taste parameters and search costs			
Baseline search cost	c	0.031	(0.006)
Price coefficient	$\overline{\alpha}$	−0.253	(0.070)
Scale parameter	σ_η	0.028	(0.017)
Personalization algorithm parameters			
Utility weight	w_u	1.286	(0.482)
Profitability weight	w_π	0.620	(0.062)
N users	50,000		
J items	243		
LogL	−196.9		

Notes. We obtain these estimates using a maximum likelihood estimation procedure developed in Section 3.3. The sample used in this estimation includes N = 50,000 users randomly selected from the personalized group. We observe these users choosing among J = 243 unique items. The reported *LogL* value corresponds to the average log-likelihood value at the optimum, where the average is taken across N users in the sample. Coef., coefficient; est., estimate; S.E., standard error.

appealing to users. We now ask how this profit-driven distortion affects consumer welfare. To this end, we generate a sample of users from the estimated distribution of tastes λ_i, and we simulate their behavior under two ranking algorithms: personalized and non-personalized. For the personalized scenario, we generate rankings using the estimated model in (7), whereas in the non-personalized scenario, we generate rankings from the model in (5) that we estimated from the non-personalized sample. In both scenarios, we define the ex ante consumer surplus of each user λ_i as follows:

$$CS_i(\lambda_i) = \frac{1}{\alpha_i}\sum_{R_i \in \mathscr{R}} P_{R_i} \sum_{S_i \in \mathscr{S}} P_{iS_i|R_i}\left(\log\left(\sum_{j \in S_i}\exp(\delta_{ij})\right) - \sum_{j \in S_i} c_{ij}\right), \tag{13}$$

where the first summation is over all possible rankings $\mathscr{R}$, the second summation is over all possible search sets $\mathscr{S}$, and the expression in the brackets is the expected net utility from searching all items in the set S_i. We approximate these two summations using simulation—that is, we draw rankings R_i and search sets S_i for each user i and average across the resulting net utility values. Having computed the consumer surplus $CS_i(\lambda_i)$ for each user, we then average across users to compute the expected consumer surplus CS. Below, we also decompose CS into the effect of total incurred search costs, the purchase probability, and the utility obtained conditional on a purchase (see Online Appendix D for the formal derivations).

Table 4 shows the welfare results. We find that personalized rankings increase the average consumer surplus by \$4.02, an increase of approximately 18% compared with non-personalized rankings. This surplus increase constitutes around 2.5% of the average item price in this category. Although this effect may seem small, one should keep in mind that the probability of purchase in the estimated model is less than 5%. Therefore, the surplus increase conditional on purchasing is almost 50% of the average item's price. Extrapolating these estimates to all users in the personalized sample of $n = 1{,}930{,}992$, we find that personalized rankings generate a substantial total surplus of \$7.77 million.

Table 4. The Effect of Personalized Rankings on Consumer Surplus

	ΔCS per user		% of total	ΔCS
Surplus measure	Estimate (\$)	S.E.	Change	Aggregate (\$)
Total surplus change	4.02	(1.50)	100.0	7.77M
(a) Search costs	−5.44	(1.22)	36.5	−10.50M
(b) Purch. prob.	1.32	(0.61)	8.8	2.54M
(c) Match value	8.15	(2.78)	54.7	15.73M

Notes. We compute the change in the expected ex ante surplus in the first row using Equation (13). In the remaining rows, we decompose the total surplus change into the effects of changing (a) the total search costs paid by the user, (b) the purchase probability (purch. prob.), and (c) the expected utility conditional on marking a purchase (see Equation (14) in Online Appendix D for details). The aggregate estimates of the surplus change are computed by multiplying the average ΔCS estimate by the total number of users in the personalized sample ($N = 1{,}930{,}992$). M, million; S.E., standard error.

The remaining rows in Table 4 show the decomposition of this surplus change. About 36.5% of the change is driven by search costs: Even though personalization somewhat reduces search costs, users search more items and, therefore, incur higher total search costs. This surplus loss is more than offset by positive effects generated by a higher purchase probability (8.8% of the change) and the fact that, conditional on a purchase, users discover and buy better-matching items (54.7% of the change).

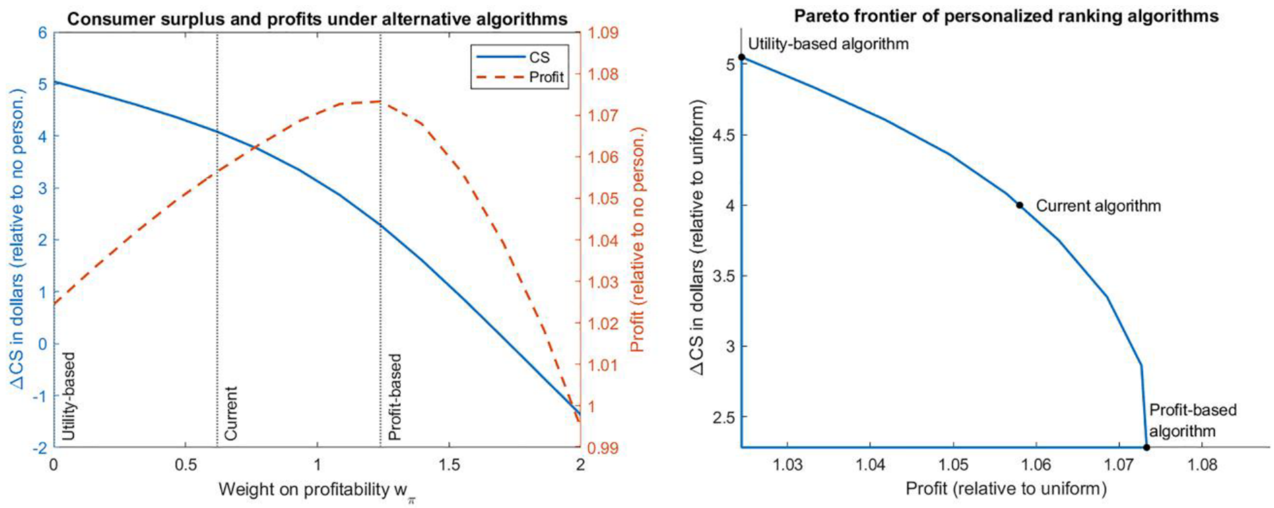
We emphasize that the observed welfare increase is not automatically implied by our experimental results. In Table 1, we show that under personalization, users search more items and purchase with a higher probability. Although one might be tempted to conclude that consumer surplus must increase, this change need not happen if the retailer optimizes rankings for both utility and profitability. In fact, in Online Appendix E, we construct an example in which the retailer distorts personalized rankings to the point at which consumer surplus decreases, even though users still search and purchase more than without personalization.

To understand why CS increases in our application, consider Figure 9 in Online Appendix G. The figure shows that personalization uniformly increases the click rates of all top positions, suggesting that consumers now see more appealing items at the top. Given this figure, the ex ante consumer surplus should increase. Even if consumers selected the same search sets S_i as they would without personalization, their expected utility would increase because the same search sets S_i would now generate a higher expected benefit m_{iS} in Equation (2), thus increasing the expected surplus. In reality, however, consumers adjust their behavior by searching more items (see Table 1). By the revealed preference argument, a consumer only switches to another search set when the net benefit of switching exceeds the search cost. Therefore, although consumers pay a higher total search cost under personalization (see row 2 in Table 4), the fact that they find it beneficial to search more suggests that their consumer surplus is also higher.

4.4. Welfare under Alternative Personalization Algorithms

Our results above show that despite profit-related distortions in the algorithm, personalized rankings are highly beneficial to users. We further illustrate this point in Figure 5, in which we explore other ranking algorithms that the retailer could have chosen. To construct these alternative algorithms, we fix the sum of the two weights at the estimated level, so that $w_u + w_\pi = 1.286 +$

Figure 5. (Color online) Consumer Surplus and Retailer Profits under Alternative Ranking Algorithms



Notes. The left figure shows ex ante consumer surplus and the retailer's expected profits under alternative personalized ranking algorithms with different profitability weights w_π. The solid line shows the expected ex-ante consumer surplus from Equation (13), the dashed line shows the expected revenue, and the vertical dashed line labeled "current" depicts the current personalization algorithm that corresponds to the estimated weights $\hat{w}_u$ and $\hat{w}_\pi$. The right figure presents the Pareto frontier of consumer surplus and profits, showing the values of the two objectives that can be achieved within the analyzed family of personalization algorithms. To preserve data confidentiality, the figures do not report any actual profit values; instead, they only report the relative values of expected profits predicted by the estimated model.

$0.620 = 1.906$, and we then change the profitability weight w_π to values above and below its estimated level ($\hat{w}_\pi = 0.620$). Because the sum of two weights is fixed, increasing the profitability weight w_π also decreases the utility weight w_u.[20] Changing the weights in this way allows us to explore the consumer surplus and profits that would have been created under alternative ranking algorithms available to the retailer. As in Section 4.3, we assume that the position effects $p_2, \ldots, p_{11}$ do not change as we vary the ranking algorithm.

In Figure 5 (left panel), we show the predicted profit and consumer surplus under different weights w_π, reporting both outcome variables relative to their values under non-personalized rankings. The blue solid line shows the expected ex ante consumer surplus from Equation (13), the red dashed line shows the expected profit, and the vertical dashed line labeled "current" depicts the current personalization algorithm that corresponds to the estimated weights $\hat{w}_u$ and $\hat{w}_\pi$. Additionally, the right panel of Figure 5 visualizes the Pareto frontier of consumer surplus and profits, showing the values of the two objectives that can be achieved within the specified family of personalization algorithms. Based on these results, we make two important observations about the current algorithm. First, the retailer could have further increased consumer surplus by putting zero weight on profitability ($w_\pi = 0$). Such a change would increase consumer surplus by about $1 per user, making it $5 more than the consumer surplus under non-personalized rankings. It would also, however, dramatically reduce the expected profits. Second, the retailer could have maximized its profits by putting a higher weight on profitability (e.g., $w_\pi \approx 1.25$), obtaining a profit about 7% higher than that under non-personalized rankings (as opposed to 5.8% in the current algorithm). Such a change would make users substantially worse off: The consumer surplus would drop by almost $2 compared with the current algorithm, thus removing half of all gains the users derive from personalization. These results suggest that when faced with a trade-off between maximizing short-term utility and short-term profits, the retailer has settled for a personalization algorithm that balances these two objectives and benefits both the retailer and the users.

4.5. Discussion of Results

Our primary motivation was to explore whether online retailers have incentives to nudge consumers toward profitable, but low-utility, items through personalized recommendations. Although such profit incentives might be strong in theory, our results suggest that they may not be particularly strong in practice. Specifically, the case study we present here shows that personalized rankings, in the way that the retailer has implemented them, have generated a substantial surplus that was to a great extent internalized by the retailer's consumers. Even though we find some evidence of profit-driven distortions, such distortions appear too weak to negatively affect consumer surplus. As a whole, our results show that in this context, personalized rankings increased overall efficiency by matching consumers to better-suited items, which, in turn, increased the volume of transactions and made both the retailer and consumers better off.

It is natural to ask whether we expect these results to generalize. For example, one may wonder to what extent other large online retailers are facing similar incentives that induce them to engage in "benevolent" personalization. The only way to answer this question conclusively would be to repeat our analysis across different retailers and product categories. Nevertheless, we conjecture that online retailers may generally have incentives to offer personalized recommendations that are beneficial for consumers. Although retailers may indeed want to optimize short-term profits by recommending profitable, high-markup items, providing helpful recommendations may be a better long-term strategy, as it may lead to more customer loyalty, better user retention, and stable long-term growth. This strategy may also help retailers avoid losing customers to a competing retailer that offers more helpful personalized recommendations. Our view is that, for these reasons, any online retail company interested in building a reputation and maximizing long-term growth will generally put substantial weight on maximizing consumer surplus. As the company matures and loses its incentives to maximize growth, it may shift its focus on monetizing the existing personalization algorithm by putting more weight on short-term profits. One testable hypothesis is that older, more established companies might implement less user-centric and more profit-centric personalized recommendations. We view testing this hypothesis as a promising area for future empirical work.

Of course, to make generalizable conclusions, one would need to study the effects of personalization more broadly, across different retailers and markets. Drawing these conclusions would be valuable both for academic researchers and for regulators who seek to increase algorithmic transparency. It is interesting to discuss how researchers could approach such a broad analysis. Our case study relies on confidential data obtained directly from the retailer, which gives us a unique opportunity to scrutinize one specific personalization algorithm. Given that large online companies, such as Pandora and Ziprecruiter, are increasingly allowing researchers to use their data and analyze their personalization strategies, we are hopeful that our analysis can be replicated and extended in collaboration with other companies as well.[21] Additionally, our view is that personalization algorithms can (and should) also be studied in contexts in which it is not feasible to access proprietary data. State-of-the-art scraping tools that imitate website visits of users with different profiles (e.g., from different locations, with different browsing histories) could allow for this analysis, or researchers could crowdsource data directly from website users. In this vein, a recent report from the European Parliament has called for additional research on "reverse engineering the black-box algorithms."[22] And several groups of researchers have attempted to monitor YouTube's recommendation algorithm by repeatedly scraping video recommendations (Faddoul et al. 2020, Roth et al. 2020). One could apply similar tools to study personalized recommendations in online retail. We hope that these initial efforts to monitor the existing recommendation algorithms, as well as the case study we present here, will pave the road for future studies of the effects of personalization on the online experience of consumers.

5. Conclusion

In this paper, we explored whether personalized rankings in online retail benefit consumers. We began by analyzing the results of a randomized experiment, which revealed that personalized rankings make consumers search more, increase the purchase probability, and redistribute demand toward relatively unpopular items. We then developed and estimated a choice model, which enabled us to recover heterogeneous users' tastes from data on searches, purchases, and observed rankings, and which helped us to empirically separate tastes from position effects. Having estimated this model, we showed that personalized rankings substantially increase both consumer surplus and the retailer's revenues. We therefore did not find any strong evidence that the retailer in our case study used personalized rankings to increase profitability at the expense of reducing consumer welfare.

Of course, our analysis is narrow, in the sense that we focus on a specific retailer and product category. Although this limited sample definition helps us to develop a realistic choice model and formulate an accurate model of rankings, we cannot immediately extrapolate our results to other product categories or other retailers. Yet, our analysis is sufficiently general to help us draw several broader lessons. By being systematic about discussing the main empirical challenges, we have provided a conceptual framework that may help researchers estimate demand in contexts with many products, few observed attributes, and access to search and ranking data. Going forward, researchers may apply this framework to datasets from different retailers and categories to establish whether the personalization effects we document here generalize to other contexts.

One intriguing question is why personalized recommendations make such a large difference, even though the retailer is already offering its users multiple search tools (e.g., search queries, sorts, filters). We believe that this benefit occurs, in part, because the existing search tools and personalized recommendations cater to different types of consumers. Search tools may be especially helpful to consumers who broadly know what they are looking for and simply need to hone in on a set of products with specific attributes (e.g., blue sectional sofas). By contrast, personalized recommendations are most helpful to consumers who do not know how to narrow the scope of their search and are simply looking to explore available options. To further understand how benefits from personalization are distributed across consumers, future research may need to document the prevalence of these consumer types and quantify their benefits from personalization.

Acknowledgments

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report. The retailer has the right for a limited review of the manuscript prior to its circulation, during which it can request removing its intellectual property or trade secrets from the paper. This paper was previously circulated with the title "The Long Tail Effect of Personalized Rankings." The authors thank a group of researchers whose insightful comments and suggestions have greatly improved the paper: Eric Anderson, Bart Bronnenberg, Hana Choi, Jean-Pierre Dubé, Brett Gordon, Rafael Greminger, George Gui, Wes Hartmann, Ella Honka, Yufeng Huang, Malika Korganbekova, Sanjog Misra, Harikesh Nair, Sridhar Narayanan, Navdeep Sahni, Stephan Seiler, Anna Tuchman, and Caio Waisman, as well as seminar participants at Stanford University, Kellogg School of Management, Chicago Booth, Virtual Quant Marketing Seminar, European Quant Marketing Seminar, and Consumer Search Digital Seminar. The authors also thank James Ryan for his excellent research assistance. All errors are our own.

Endnotes

[1] See the report "How Retailers Can Keep Up with Consumers" (MacKenzie et al. 2013).

[2] The 70% statistic was quoted by Chief Product Officer Neal Mohan at a CES panel discussion in 2018. See the article "YouTube's AI Is the Puppet Master over Most of What You Watch" (Solsman 2018).

[3] Source: Amazon Personalize (https://aws.amazon.com/personalize/).

[4] We obtained this statistic from builtwith.com, a platform that tracks internet technologies used by over 673 million websites.

[5] Pandora famously conducted "steering experiments," in which recommendations were tilted toward the music tracks owned by the company (McBride 2014). Similarly, Netflix's recommendations tend to favor in-house productions, for which the streaming platform does not have to pay additional fees (Carr 2013). Amazon researchers have revealed that their email recommendation algorithm is more likely to favor more profitable recommendations (Mangalindan 2012). Finally, Taobao prioritizes recommending products with moderate prices to balance conversion rates and commissions (Zhou and Zou 2021).

[6] Holtz et al. (2020) also find that personalized podcast recommendations on Spotify increase the diversity of selected podcasts. By contrast, Lee and Hosanagar (2019) find that personalized recommendations in online retail reduce the aggregate diversity of consumption and shrink the total market share of relatively unpopular items.

[7] In this vein, the recent report of the European Parliament has called for additional research on "reverse engineering the black-box algorithms" (Koene et al. 2019). The European Commission has also commissioned a pilot study that monitors personalization algorithms via web scraping (Faddoul et al. 2020).

[8] Theoretically, personalization might either increase or decrease the average position of searched items. Consumers might become more likely to search items ranked higher in the list because they no longer need to scroll down as much to discover appealing items. On the other hand, they may become more likely to search items ranked lower in the list because scrolling down reveals, on average, more appealing options. See Online Appendix A.2 for a theoretical example of this ambiguity.

[9] This assumption does not literally mean that all users can perfectly infer from photos which beds are" Scandinavian style" and which are" Modern Style." Instead, what is meant by this assumption is that, when consumers see two Scandinavian-style beds, they perceive them as visually similar. This perception makes consumers likely to compare the two beds with each other, making these two items close substitutes.

[10] Alternatively, we could add a stochastic term directly to the presearch utility δ_{ij} (Honka and Chintagunta 2015, Ursu 2018). However, the estimation procedure would then require computing search likelihoods by simulating the search behavior of artificial users, which is computationally expensive. We would also need to artificially smooth the resulting frequency estimators, which might generate bias of an unknown sign and magnitude (Chung et al. 2023).

[11] In a parallel effort, Armona et al. (2021) also estimate a simultaneous search model with latent attributes. They do not incorporate ranking effects into their estimation and rely on a different inequality-based estimation technique. They also use the estimated model to predict competition and analyze mergers, thus focusing on a completely different research question.

[12] Ideally, we would have data on the exact inputs that went into the non-personalized ranking algorithm on each day (e.g., the exact historical popularity indices of items used by the algorithm). We were, unfortunately, unable to obtain such data. As a robustness check, we experimented with estimating our model on October–November data, while approximating historical popularity using purchases made by users in September, and we arrived at similar qualitative results.

[13] If the true signal variance σ_s^2 differs across consumers, our estimates might be biased, but the sign of the bias is ambiguous. On the one hand, our model would underestimate personalization gains for low-variance users, whose rankings are almost perfectly aligned with their individual tastes. On the other hand, we would overestimate personalization gains for high-variance users, not realizing that their rankings contain almost no personalization. The net bias will then be case-specific and will depend on the distribution of uncertainty across users.

[14] Agarwal et al. (2012) show that, if the primal objective function is convex (e.g., due to preference for fairness), the Lagrangian duality formulation returns a ranking function that is a weighted linear combination of the multiple objectives. Wang et al. (2022a) use this idea to optimize a multiobjective recommendation system for a major food-delivery marketplace.

[15] This stability condition is implicitly assumed by most of the existing work on estimating demand models with rankings (Ursu 2018, Compiani et al. 2021, Greminger 2022).

[16] At the time, the retailer was conducting a series of price experiments. At each point in time, the company would randomize the prices of items in a selected group of products, while also setting optimal prices of all other items based on a proprietary algorithm.

[17] Consistent with this identification argument, our estimated model predicts the highest correlation of utilities for those pairs of items that are frequently searched together and are displayed together to the same users. For example, we predict the correlation of utilities 0.4–0.5 for pairs of items that are searched together by at least 100 users in the data, whereas we estimate the correlation below 0.05 for pairs of items that are never searched together.

[18] This is a salient feature of most consumer search datasets. For example, that the average user searches 2.3 options when booking hotel rooms (Chen and Yao 2016), clicks on 1.8–2.4 items when shopping for cosmetics products (Morozov et al. 2021), visits 1.1–2.0 auto dealerships (Yavorsky and Honka 2020), requests 2–3 quotes for auto insurance (Honka 2014), and examines 2.3 vehicles when comparing used cars (Gardete and Antill 2019).

[19] Because $\overline{\lambda}_U$ is only identified with respect to the normalized scale σ_ε, and $\overline{\lambda}_R$ is identified with respect to another scale σ_μ^p, we cannot directly compare tastes $\overline{\lambda}_R$ and $\overline{\lambda}_U$ with each other. We therefore normalize each taste vector, dividing it by the estimated taste parameter $\overline{\lambda}_k$ for one of the attributes that is not shown in the figure.

[20] To gain formal intuition behind this counterfactual, consider the initial formulation of the ranking model in Equation (7). Suppose the raw weights $\tilde{w}_u$ and $\tilde{w}_\pi$ are related to the estimates, such that $\tilde{w}_u = \hat{w}_u/(\hat{w}_u + \hat{w}_\pi)$ and $\tilde{w}_\pi = \hat{w}_\pi/(\hat{w}_u + \hat{w}_\pi)$, so the weights add up to one. By normalizing the variance to $\sigma_\mu^2 = 1/(\hat{w}_u + \hat{w}_\pi)^2$, one obtains a ranking equation that is equivalent to the estimated model. We can then think about the main counterfactual as changing the weight $\tilde{w}_\pi$ while maintaining that the weights $\tilde{w}_u$ and $\tilde{w}_\pi$ sum up to one.

[21] Dubé and Misra (2023) study personalized pricing using a proprietary dataset from Ziprecruiter, whereas Goli et al. (2021) use Pandora's data to study a related problem of personalized product versioning.

[22] See "A Governance Framework for Algorithmic Accountability and Transparency" (Koene et al. 2019).

References

Abdollahpouri H, Adomavicius G, Burke R, Guy I, Jannach D, Kamishima T, Krasnodebski J, Pizzato L (2020) Multistakeholder recommendation: Survey and research directions. *User Model. User-Adapted Interaction* 30:127–158.

Agarwal D, Chen B-C, Elango P, Wang X (2012) Personalized click shaping through Lagrangian duality for online recommendation. *SIGIR'12 Proc. 35th Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery, New York), 485–494.

Anderson C (2006) *The Long Tail: Why the Future of Business Is Selling Less of More* (Hachette Books, New York).

Armona L, Lewis G, Zervas G (2021) Learning product characteristics and consumer preferences from search data. Preprint, submitted June 3, https://dx.doi.org/10.2139/ssrn.3858377.

Athey S, Imbens GW (2007) Discrete choice models with multiple unobserved choice characteristics. *Internat. Econom. Rev.* 48:1159–1192.

Athey S, Blei D, Donnelly R, Ruiz F, Schmidt T (2018) Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers Proc.* 108:64–67.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowledge-Based Systems* 46:109–132.

Bourreau M, Gaudin G (2022) Streaming platform and strategic recommendation bias. *J. Econom. Management Strategy* 31(1):25–47.

Carr D (2013) Giving viewers what they want. *New York Times* (February 24), https://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html.

Chade H, Smith L (2006) Simultaneous search. *Econometrica* 74:1293–1307.

Chen Y, Yao S (2017) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 63(12):4345–4365.

Choi H, Mela CF (2019) Monetizing online marketplaces. *Marketing Sci.* 38(6):948–972.

Chung JH, Chintagunta PK, Misra S (2023) Estimation of sequential search models. Working paper, The Chinese University of Hong Kong, Hong Kong.

Compiani G, Lewis G, Peng S, Wang W (2021) Online search and product rankings: A double index approach. Preprint, submitted August 4, https://dx.doi.org/10.2139/ssrn.3898134.

De los Santos B (2018) Consumer search on the internet. *Internat. J. Indust. Organ.* 58:66–105.

De los Santos B, Koulayev S (2017) Optimizing click-through in online rankings with endogenous search refinement. *Marketing Sci.* 36(4):542–564.

De Los Santos BI, Hortacsu A, Wildenbeest M (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *Amer. Econom. Rev.* 102:2955–2980.

Donnelly R, Ruiz FR, Blei D, Athey S (2019) Counterfactual inference for consumer choice across many product categories. Preprint, submitted June 6, https://arxiv.org/abs/1906.02635.

Dubé J-P, Misra S (2023) Personalized pricing and consumer welfare. *J. Polit. Econom.* 131(1):131–189.

Elrod T, Keane MP (1995) A factor-analytic probit model for representing the market structure in panel data. *J. Marketing Res.* 32(1):1–16.

Faddoul M, Chaslot G, Farid H (2020) A longitudinal analysis of YouTube's promotion of conspiracy videos. Preprint, submitted March 6, https://arxiv.org/abs/2003.03318.

Farrell MH, Liang T, Misra S (2020) Deep learning for individual heterogeneity. Preprint, submitted October 28, https://arxiv.org/abs/2010.14694.

Ferreira P, Zhang X, Belo R, Godinho de Matos M (2016) Welfare properties of recommender systems: Theory and results from a randomized experiment. Preprint, submitted October 21, https://dx.doi.org/10.2139/ssrn.2856794.

Gardete PM, Antill MH (2019) Guiding consumers through lemons and peaches: A dynamic model of search over multiple characteristics. Working paper, Nova School of Business and Economics, Carcavelos, Portugal.

Goettler RL, Shachar R (2001) Spatial competition in the network television industry. *RAND J. Econom.* 624–656.

Goli A, Reiley D, Hongkai Z (2021) Personalized versioning: Product strategies constructed from experiments on Pandora. PhD thesis, University of Chicago, Booth School of Business, Chicago.

Gopalan P, Hofman JM, Blei DM (2013) Scalable recommendation with Poisson factorization. Preprint, submitted November 7, https://arxiv.org/abs/1311.1704.

Greminger RP (2022) Heterogeneous position effects and the power of rankings. Working paper, UCL School of Management, London.

Holtz D, Carterette B, Chandar P, Nazari Z, Cramer H, Aral S (2020) The engagement-diversity connection: Evidence from a field experiment on Spotify. Preprint, submitted April 13, https://dx.doi.org/10.2139/ssrn.3555927.

Honka E (2014) Quantifying search and switching costs in the US auto insurance industry. *RAND J. Econom.* 45:847–884.

Honka E, Chintagunta PK (2015) Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Sci.* 36(1):21–42.

Honka E, Hortaçsu A, Vitorino MA (2017) Advertising, consumer awareness, and choice: Evidence from the US banking industry. *RAND J. Econom.* 48:611–646.

Hosanagar K, Krishnan R, Ma L (2008) Recommended for you: The impact of profit incentives on the relevance of online recommendations. *ICIS 2008 Proc.*, 31.

Jacobs BJ, Donkers B, Fok D (2016) Model-based purchase predictions for large assortments. *Marketing Sci.* 35(3):389–404.

Jannach D, Adomavicius G (2017) Price and profit awareness in recommender systems. Preprint, submitted July 25, https://arxiv.org/abs/1707.08029.

Jannach D, Resnick P, Tuzhilin A, Zanker M (2016) Recommender systems: Beyond matrix completion. *Commun. ACM* 59:94–102.

Keane M (2004) Modeling health insurance choice using the heterogeneous logit model. MPRA Paper. Working paper, Carey Business School, Johns Hopkins, Baltimore.

Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Kim JB, Albuquerque P, Bronnenberg BJ (2016) The probit choice model under sequential search with an application to online retailing. *Management Sci.* 63(11):3911–3929.

Koene A, Clifton C, Hatada Y, Webb H, Richardson R (2019) A governance framework for algorithmic accountability and transparency. Working paper, University of Nottingham, Nottingham, UK.

Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42:30–37.

L'Ecuyer P, Maillé P, Stier-Moses NE, Tuffin B (2017) Revenue-maximizing rankings for online platforms with quality-sensitive consumers. *Oper. Res.* 65(2):408–423.

Lee D, Hosanagar K (2019) How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Inform. Systems Res.* 30:239–259.

MacKenzie I, Meyer C, Noble S (2013) How retailers can keep up with consumers. Accessed April 21, 2023, https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers.

Mangalindan J (2012) Amazon's recommendation secret. https://fortune.com/2012/07/30/amazons-recommendation-secret/.

McBride S (2014) Written direct testimony of Stephan McBride (on behalf of Pandora Media, Inc.). Accessed April 21, 2023, https://www.crb.gov/rate/16-CRB-0003-PR/statements/pandora/statement.pdf.

McFadden D (1981) Econometric models of probabilistic choice. Manski CF, McFadden D, eds. *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MD), 198–272.

Moraga-González JL, Sándor Z, Wildenbeest MR (2015) Consumer search and prices in the automobile market. IESE Research Papers D/1123, IESE Business School. Preprint, submitted May 15, https://dx.doi.org/10.2139/ssrn.2609338.

Morozov I, Seiler S, Dong X, Hou L (2021) Estimation of preference heterogeneity in markets with costly search. *Marketing Sci.* 40(5): 871–899.

Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Sci.* 34(3):388–407.

Panniello U, Hill S, Gorgoglione M (2016) The impact of profit incentives on the relevance of online recommendations. *Electronic Commerce Res. Appl.* 20:87–104.

Punj GN, Staelin R (1978) The choice process for graduate business schools. *J. Marketing Res.* 15:588–598.

Rossi PE, Allenby GM, McCulloch R (2012) *Bayesian Statistics and Marketing* (John Wiley & Sons, Chichester, UK).

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.

Roth C, Mazières A, Menezes T (2020) Tubes and bubbles topological confinement of YouTube recommendations. *PLoS One* 15:e0231703.

Ruiz FJ, Athey S, Blei DM (2020) Shopper: A probabilistic model of consumer choice with substitutes and complements. *Ann. Appl. Statist.* 14:1–27.

Schnabel T, Swaminathan A, Singh A, Chandak N, Joachims T (2016) Recommendations as treatments: Debiasing learning and evaluation. Preprint, submitted February 17, https://arxiv.org/abs/1602.05352.

Solsman JE (2018) YouTube's AI is the puppet master over most of what you watch. Accessed April 21, 2023, https://www.cnet.com/news/youtube-ces-2018-neal-mohan.

Titsias MK (2016) One-vs-each approximation to softmax for scalable estimation of probabilities. Preprint, submitted September 23, https://arxiv.org/abs/1609.07410.

Underwood C (2019) Use cases of recommendation systems in business—Current applications and methods. Accessed April 21, 2023, https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/.

Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Sci.* 37(4):530–552.

Wan M, Wang D, Goldman M, Taddy M, Rao J, Liu J, Lymberopoulos D, McAuley J (2017) Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs. *WWW'17 Proc. 26th Internat. Conf. World Wide Web* (International World Wide Web Conferences Steering Committee, Geneva), 1103–1112.

Wang Y, Tao L, Zhang X (2022a) Recommending for a three-sided food delivery marketplace: A multi-objective hierarchical approach. Working paper, Google Brain, Mountain View, CA.

Wang Y, Tao L, Zhang X (2022b) Recommending for a multi-sided marketplace with heterogeneous contents. *Proc. 16th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 456–459.

Weitzman ML (1979) Optimal search for the best alternative. *Econometrica* 47:641–654.

Yavorsky D, Honka E (2020) Consumer search in the US auto industry: The value of dealership visits. Preprint, submitted January 1, https://dx.doi.org/10.2139/ssrn.3499305.

Zhou B, Zou T (2021) Competing for recommendations: The strategic impact of personalized product recommendations in online marketplaces. Preprint, submitted January 21, https://dx.doi.org/10.2139/ssrn.3761350.