Search Algorithm, Repetitive Information, and Sales on Online Platforms.*

Yangguang Huang[†] Yu Xie[‡]

May 2022

Abstract

A prominent feature of online sales is that buyers rely on the search tools offered by platforms to process information when searching for products. We develop a model that captures how the search algorithm affects buyers' search processes, which further influences market equilibrium and welfare. If a platform adopts a highly unequal search algorithm, buyers are likely to obtain repetitive information about a small group of sellers, which causes buyers to consider fewer options and suppresses competition. By using data from food delivery platforms, we provide empirical evidence that markets with less equal distributions of store rankings in search results have higher average prices and more concentrated sales. We suggest that regulators should restrict search algorithms from showing repetitive information.

Keywords: online platform, search algorithm, repetitive information, consideration set, food delivery platform

JEL codes: D83, L11, L13, L42

*We appreciate the valuable comments and suggestions provided by Mark Armstrong, John Asker, Chiara Farronato, Ginger Jin, Julian Wright, and Daniel Xu. We also thank the seminar participants at HIT-SZ, HKUST IO Workshop, SUFE IO conference, University of Tokyo, Hitostubashi University, BU Platform Symposium, EARIE Conference, Asia-Pacific IO conference, NBER Summer Institute, and Econometric Society World Congress.

[†]Hong Kong University of Science and Technology, Hong Kong. Email: `huangyg@ust.hk`.

[‡]University of Hong Kong, Hong Kong. Email: `emmax@connect.hku.hk`.

1 Introduction

Online e-commerce platforms often host a large number of sellers that offer heterogeneous products with a tremendous amount of information. It is impossible for consumers to thoroughly search and study all available products due to information overload (Anderson and De Palma, 2009). One of the most important features of online sales is that consumers rely on the search tools provided by the platforms to search for and learn about products. Search tools such as search engines, recommender systems, and price-comparison shopbots use information technology to assist buyers in searching for products and learning about product characteristics. These search tools have strong impacts on consumer searches (Teh and Wright, 2020; Chen and Tsai, 2021; Teh, 2022). For example, Backus et al. (2014) shows that the eBay search algorithm can cause identical items to have vastly different visibility, leading to dispersion in prices and the number of bidders. Dinerstein et al. (2018) find that transaction prices on eBay fell significantly after the platform redesigned the search process to promote price competition.

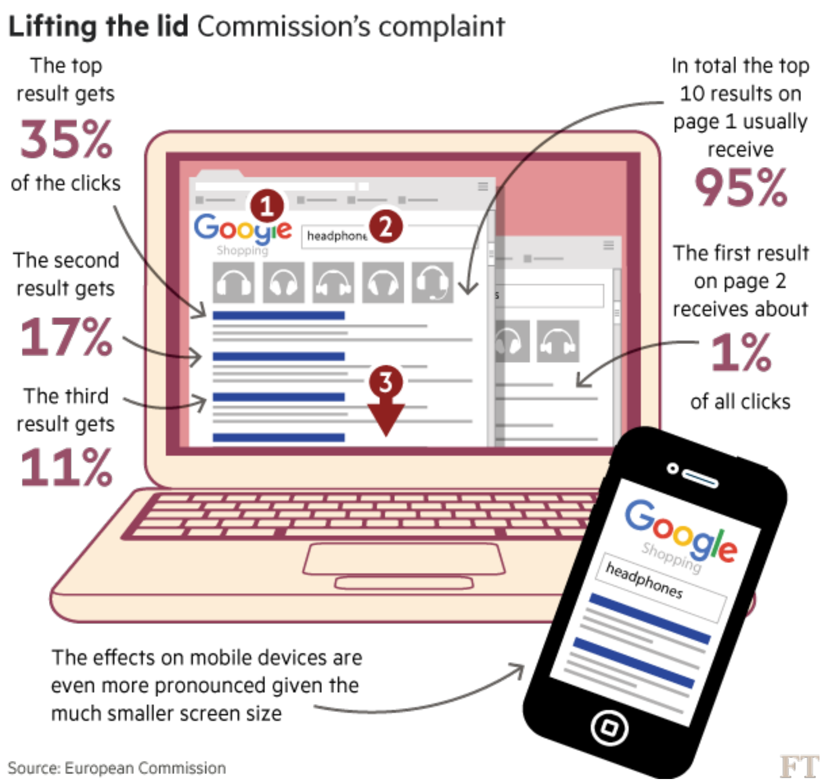


Figure 1: How Search Algorithms Affect Clicks

Because search tools are designed and operated by for-profit online platforms, they might not be designed to boost competition. Regulators and antitrust authorities have started to regulate the search tools controlled by online platforms. For example, in June 2017, the European Commission (2017) fined Google €2.42 billion for "abusing dominance as search engine by giving illegal advantage to own comparison shopping service." Figure 1 illustrates Google's ability to control users' attention (clicks). The Federal Trade Commission (2013) requires search engines to "distinguish between advertisements and search results," because sponsored searches can profoundly influence buyers' decisions and cause potential efficiency losses (Ghose and Yang, 2009). China passed a law in 2021 restricting platforms from exercising algorithmic discrimination against frequent customers.[1]

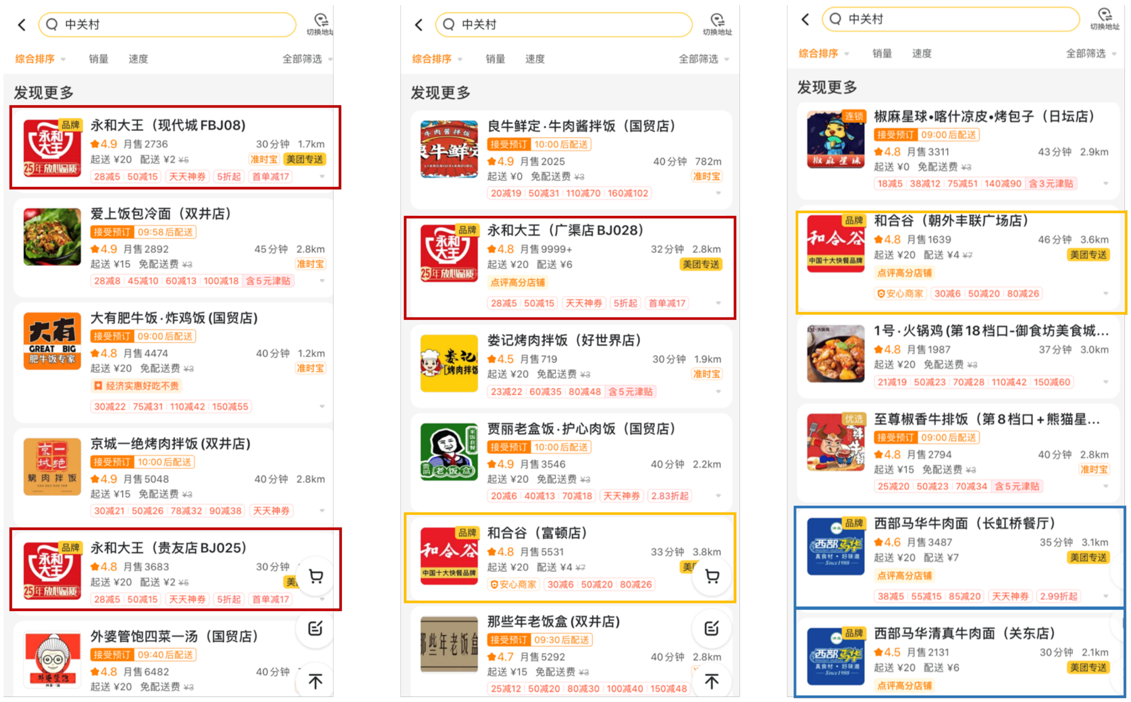
Many researchers have studied the private incentives of online platforms in designing search algorithms and the related consequences. Hagiu and Jullien (2011) note that platforms have two main motives in diverting searches: obtaining higher revenues from participating users and affecting sellers' choices in pricing or other strategic variables. A platform has a clear incentive to favor sellers who are vertically integrated with it (De Corniere and Taylor, 2019). In designing the search algorithm, platforms might bias the search results toward their own content or that of sponsored sellers (De Corniere and Taylor, 2014, 2019). Chen and Tsai (2021) find empirical evidence suggesting that Amazon's products are recommended in frequently-bought-together lists much more often than are the same products carried by third-party sellers. Search advertising also deeply affects consumers' search behaviors and social welfare (Athey and Ellison, 2011; Chen and He, 2011; Eliaz and Spiegler, 2011b; Blake et al., 2015).[2] De Corniere (2016) shows that even with neck-and-neck competition among search engines, sub-optimal sponsored links persist, and welfare can be worsened.

In this paper, we construct a model that demonstrates how the search algorithm influences the consumer search process on the platform, which further affect the sales distribution among sellers and social welfare. Buyers search for products with the search tool offered by the platform. After the search process, each buyer spends a limited amount of effort in searching and obtains a consideration set (Goeree, 2008; Honka et al., 2017) that contains several options. Then, the buyer chooses his or her favorite option from within the consideration set (Eliaz and Spiegler, 2011a).

The search algorithm determines the ranking and composition of the product information in the search results. Most search results are lists of products or sellers with rankings that determine the probability of each product/seller being considered by the buyers. The algorithm usually does not treat sellers equally. Some products may receive preferential treatment and appear frequently at the top of search results. Moreover, the same product or seller may appear multiple times in the search results. In practice, search algorithms commonly yield search results with repetitive information. For example, Figure 2 shows the results from a trial search on Meituan, the largest food delivery platform in China. The search results display the information of some chain restaurants multiple times. Because all of these chain restaurants sell identical items, the repetitive listings do not provide any additional information and may reduce the number of options considered by buyers. As the number of repetitive listings increase, a buyer obtains fewer effective options after spending a limited amount of effort in searching.

[1]See `www.xinhuanet.com/english/2021-08/17/c_1310132178.htm` and `www.scmp.com/news/china/politics/article/3145390/china-set-pass-new-law-protect-legitimate-rights-personal-data`.

[2]We do not explicitly distinguish between organic and sponsored search results in this paper. Buyers naturally discount sponsored links and generally find sponsored content to be less relevant (Jansen and Resnick, 2006).



Page 1 Page 2 Page 3

Note: This figure shows the first 18 items in a trial search using the Meituan food delivery mobile app and an anonymous account in Beijing in June 2021. The query is “Zhongguancun”, a location in Beijing. The colored frames highlight the multiple appearances of chain restaurants.

Figure 2: Repetitive Listings in Search Results

Our model shows that a highly unequal algorithm results in consumers' being more likely to obtain repetitive information about the same set of sellers. This phenomenon leads to smaller consideration sets, softens price competition, and reduces consumer surplus. The expected size of the consideration set increases as the algorithm becomes more equal in the sense of Lorenz ordering. The buyer-side surplus and total welfare improve if the platform adopts a more equal search algorithm. However, both the platform and sellers have private incentives to make the search algorithm “unequal”, which causes repetitive listings to appear more frequently in the search results.

We use the data from food delivery platforms to explore how the search algorithm affects sales in practice. Based on trial search results, we construct two variables to measure how equal and how repetitive are the search algorithms. For each store, we record the default position of the store and how many times stores with the same brand appear in the search results. We show that restaurant revenues are critically determined by the ranking in search results. Repetitive listings of the stores with the same brand help each store, on average, earn more revenue. Based on market-level regressions, we find that markets with less equal search results have higher average prices and more skewed revenue distributions.

This paper mainly contributes to the growing literature on the search design of online platforms. The model regarding how the search algorithm affects the composition and size of the consideration set is closely related to two papers. Dukes and Liu (2015) consider a model in which consumers must decide their search breadth and depth. They find that platforms have the incentive to strategically increase search costs to discourage buyers from evaluating too many sellers. Dinerstein et al. (2018) note the trade-off between navigating consumers to more desired products and promoting price competition. If the search algorithm lists homogeneous products by price, sellers face fierce price competition, but consumers are more likely to be matched to undesirable products. In contrast, if the search algorithm lists products that are more heterogeneous, price competition it softens, but consumers can more easily find desirable products. Compared to the previous studies, our model emphasize the role of repetitive listings and demonstrate how the features of search algorithm affect welfare measures.

2 Model

Consider a market of I buyers (consumers, him) and J heterogeneous sellers (firms, her) interacting on a platform with buyer-side search frictions and seller-side market power. Sellers are indexed by $j = 1, 2, ..., J$. A generic seller draws an efficiency parameter θ as her private type independently from a distribution F with a compact support $[\underline{\theta}, \overline{\theta}]$ and continuous density f. Each seller offers a product as a price-quality combination $(p, q) \in \mathbb{R}^2_+$. For a seller with efficiency parameter θ, the cost of producing y units of products at quality q is $y \cdot c(q, \theta)$, where $c(\cdot, \cdot)$ is continuous and twice differentiable with $c_q > 0$, $c_{qq} > 0$, $c_\theta < 0$, and $c_{q\theta} < 0$.

For I buyers in the market, each buyer purchases one unit of a product. Buyers have quasilinear preferences $u(p, q) = v(q) - p$, where $v(\cdot)$ is continuous and twice differentiable with $v_q > 0$ and $v_{qq} \leq 0$. Buyers face search frictions, and the platform can influence the search process through the search algorithm. After the search process, each buyer obtains a consideration set $\mathcal{C}$ that contains several products, and the buyer chooses his favorite product within the consideration set. Therefore, the quantity of demand for seller j's product is $y_j = I \cdot \Pr(j \in \mathcal{C}) \cdot \Pr(j \text{ is chosen} | j \in \mathcal{C})$. The probability of j being considered, $\Pr(j \in \mathcal{C})$, and the probability of j being the favorite option, $\Pr(j \text{ is chosen} | j \in \mathcal{C})$, are determined by the search process under the influence of the search algorithm. The seller chooses the product (p, q) by maximizing her expected profit

$$(1) \qquad \max_{p,q} [p - c(q, \theta_j)] \, y_j = [p - c(q, \theta_j)] \times I \times \Pr(j \in \mathcal{C}) \times \Pr(j \text{ is chosen} | j \in \mathcal{C}).$$

The timing of the game is as follows. The platform commits to a search algorithm at the beginning. After observing the algorithm, the sellers simultaneously draw their efficiency parameters as private information. Then, all J sellers choose p and q simultaneously. Lastly, the buyers search for a product by using the search algorithm and make a purchasing decision.

2.1 Search Algorithm and Search Process

Essentially, a search algorithm uses data as inputs and yields search results as the output. The data may include seller characteristics, past sales histories, consumer reviews, and information obtained from data brokers. How the search algorithm maps data inputs to search results can be extremely complicated and usually cannot be described intuitively. In this paper, we focus on how the output of a search algorithm affects the search process of buyers. The search results are usually displayed as an ordered list of options. The ranking of the options determines the probability of each option being sampled by buyers. The click-through rate of a seller substantially decreases based on its position in the search results (De los Santos and Koulayev, 2017; Ursu, 2018). This motivates us to model the search process as follows.

Buyers conduct fixed-sample-size searches as in Burdett and Judd (1983).[3]. In each search, the search algorithm is modeled as a vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_J)$, where σ_j denotes the probability of seller j being sampled and $\sum_{j=1}^{J} \sigma_j = 1$. Each buyer draws K samples, where K is a positive integer that measures the search intensity. By varying search intensity K, the model nests two extreme cases of uninformed buyers ($K = 1$) and fully informed buyers ($K \to \infty$) in Varian (1980).

After taking K samples, a buyer obtains search result $\boldsymbol{x} = (x_1, x_2, ..., x_J)$, where x_j is the number of samples taken from seller j. $\boldsymbol{x}$ follows a multinomial distribution with a probability mass function

$$h(\boldsymbol{x}) = \frac{K!}{x_1! x_2! \cdots x_J!} \sigma_1^{x_1} \sigma_2^{x_2} \cdots \sigma_J^{x_J}, \quad \text{for } \sum_{j=1}^{J} x_j = K.$$

Seller j is included in the consideration set if at least one sample is taken from her, such that $\mathcal{C} \equiv \{j : x_j > 0, \, j = 1, 2, ..., J\}$.

One intuitive way to understand this search process is by the Galton board illustrated in Figure 3.[4] The search process is similar to dropping K balls on a Galton board with J distinguishable bins that each represents a product. Dropping a ball represents the buyer "spending an eyeball" to view information regarding a product, and σ_j denotes the probability that a ball falls into bin j. Some bins are more likely to be reached by other bins. After dropping K balls, if at least one ball falls in bin j, product j is included in the consideration set. Multiple balls may fall into the same bin, which represents repetitive information. This presents the case that a buyer sample the same sellers multiple times, as illustrated in Figure 2. Obtaining repetitive information does not expand the consideration set.

Given this search process, the probability of having product j in the consideration set is

$$\Pr(j \in \mathcal{C}) = 1 - \Pr(j \notin \mathcal{C}) = 1 - (1 - \sigma_j)^K,$$

[3]The fixed-sample-size search is optimal when searching involves a fixed cost such that the average search cost decreases along with the number of searches (Hong and Shum, 2006; Morgan and Manning, 1985). De los Santos et al. (2012) provide empirical evidence of consumers adopting the fixed-sample-size search. In studying search algorithm design, Dinerstein et al. (2018) also assume that consumers spend a fixed amount of effort searching for options.

[4]See `mathworld.wolfram.com/GaltonBoard.html`.

which increases in K and in σ_j. Let $N \equiv |\mathcal{C}|$ denote the cardinality of the consideration set. N is a random variable that can take the value of $n = 1, 2, ..., K$. The probability that the consideration set has exactly n elements is

$$P_n \equiv \Pr(N = n) = \sum_{\boldsymbol{x} \in \mathcal{X}(n)} h(\boldsymbol{x}),$$

where $\mathcal{X}(n) = \left\{\boldsymbol{x} : \sum_{j=1}^{J} x_j = K \text{ and } \sum_{j=1}^{J} \mathbf{1}(x_j > 0) = n\right\}$. In general, $n < K$ because of the repetitive samples obtained in the search process.

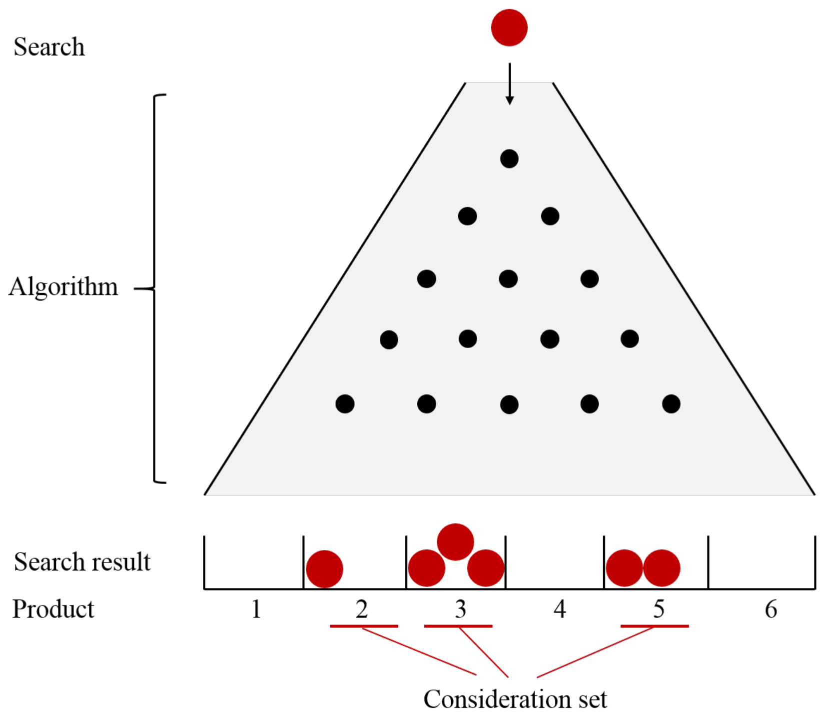


Figure 3: Illustration of the Search Process

Given a consideration set, a buyer chooses the product that yields the highest utility, $u_j = v(q_j) - p_j$. Provided that product j is in the consideration set, the probability that a buyer chooses j depends on the total number of options, that is,

$$\Pr\left(j \text{ is chosen} | j \in \mathcal{C}\right) = \sum_{n=1}^{J} \left\{P_n \times \Pr\left(j = \arg\max_{j' \in \mathcal{C}}\{u_{j'}\}\right)\right\}, \tag{2}$$

which increases in u_j.[5] The stochastic order of N determines competition intensity. The search intensity K and search algorithm $\boldsymbol{\sigma}$ affect N in the following way:

Lemma 1. *Let N_1 and N_2 denote the consideration set size under search intensity K_1 and K_2, respectively. If $K_1 > K_2$, then N_1 first-order stochastically dominates (FSD) N_2, which is denoted*

[5]The last equality is based on the assumption that the search process is not affected by product utility. This assumption can be relaxed as long as $\Pr\left(j \text{ is chosen} | j \in \mathcal{C}(K)\right)$ is increasing in u_j.

as $N_1 \succcurlyeq_1 N_2$. *It follows that* $E[N_1] > E[N_2]$.

Therefore, as the search intensity increases, consideration set size N increases in the sense of FSD; thus, the market becomes more competitive.[6]

Next, we use the results from the theory of majorization (Marshall et al., 2011) to study the impact of the search algorithm $\boldsymbol{\sigma}$. Without a loss of generality, let $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_J)$ be indexed such that $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_J$. The concept of majorization is developed to formally define a vector $\boldsymbol{\sigma}^1 = (\sigma_1^1, \sigma_2^1, ..., \sigma_J^1)$ that is "less spread out" or "more equal" than $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, ..., \sigma_J^2)$.

Definition. For $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2 \in \Delta^{J-1} \subset \mathbb{R}^J$, if $\sum_{j=1}^k \sigma_j^1 \geq \sum_{j=1}^k \sigma_j^2$ for $k = 1, 2, ..., J-1$, we say that $\boldsymbol{\sigma}^1$ *is majorized by* $\boldsymbol{\sigma}^2$, which is denoted as $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$.[7]

Intuitively, having $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$ is equivalent to the Lorenz curve of $\boldsymbol{\sigma}^1$ being above the Lorenz curve of $\boldsymbol{\sigma}^2$. Therefore, majorization is also called *Lorenz ordering*. The theory of majorization helps us to establish the relationship between the search algorithm and the size of the consideration set.

Lemma 2. *Let* N_1 *and* N_2 *denote the consideration set sizes under search algorithm* $\boldsymbol{\sigma}^1$ *and* $\boldsymbol{\sigma}^2$*, respectively. If* $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$*, then* $N_1 \succcurlyeq_1 N_2$.

Thus, when the search algorithm $\boldsymbol{\sigma}^1$ is majorized by $\boldsymbol{\sigma}^2$, the size of the consideration set decreases in the sense of FSD. Notably, the *equal-treatment algorithm*,[8] $\boldsymbol{\sigma} = (\frac{1}{J}, \frac{1}{J}, ..., \frac{1}{J})$, is majorized by all algorithms, i.e., $(\frac{1}{J}, ..., \frac{1}{J}) \prec (\sigma_1, ..., \sigma_J) \prec (0, ..., 0, 1)$. Therefore, we obtain following result.

Lemma 3. *The equal-treatment algorithm maximizes* N *in the sense of FSD.*

Lemmas 2 and 3 imply that if the search algorithm emphasizes some option(s), the buyers will obtain a smaller consideration set. We consider two examples. Let the *prominent-seller algorithm* be $\boldsymbol{\sigma}^P(a) = \left(\frac{1}{a-1+J}, \frac{1}{a-1+J}, ..., \frac{a}{a-1+J}\right)$ with $a > 1$.[9] Seller J is the prominent seller that has a higher probability being sampled by buyers than the other sellers. Let the *exponential algorithm* be $\boldsymbol{\sigma}^E(b) = \left(\frac{1^b}{B}, \frac{2^b}{B}, ..., \frac{J^b}{B},\right)$, where $B = \sum_{j=1}^J j^b$. The probability of being sampled by the buyers increases exponentially from seller 1 to seller J. As parameter a (b) increases, the prominent-seller (exponential) algorithm becomes more unequal in the sense of Lorenz ordering. That is, for $a_1 < a_2$, $\boldsymbol{\sigma}^P(a_1) \prec \boldsymbol{\sigma}^P(a_2)$; for $b_1 < b_2$, $\boldsymbol{\sigma}^E(b_1) \prec \boldsymbol{\sigma}^E(b_2)$.

Figure 4-(A) illustrates the equal-treatment, the prominent-seller, and the exponential algorithms. Figure 4-(B) shows the probability mass function of the random consideration set size

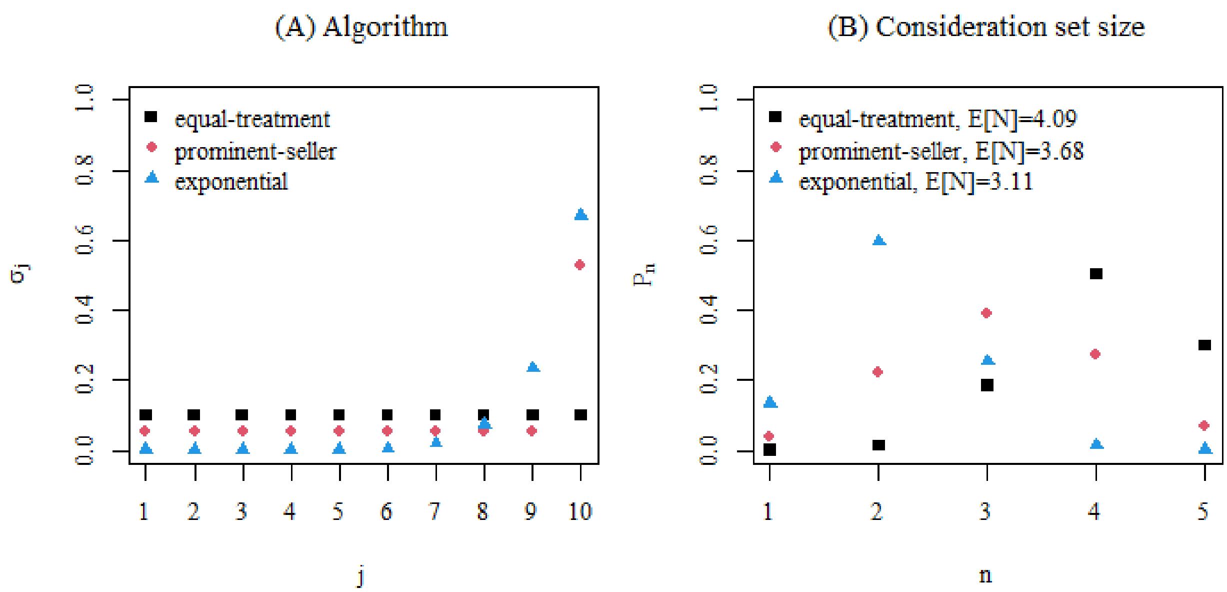
[6] In general, online platforms lower search costs and increase search intensity compared to offline sales channels. As shown in Goldmanis et al. (2010), the adoption of e-commerce enables buyers to better learn about products and to compare more options. Brynjolfsson et al. (2011) find that the internet channel exhibits a significantly less concentrated sales distribution than the traditional channel when the two channels have exactly the same catalogs and prices.

[7] The general definition of majorization does not require the vectors to belong to the $J-1$ simplex.

[8] The name "equal-treatment" is obtained from European Commission (2017) as it charges Google to "comply with the simple principle of giving equal treatment."

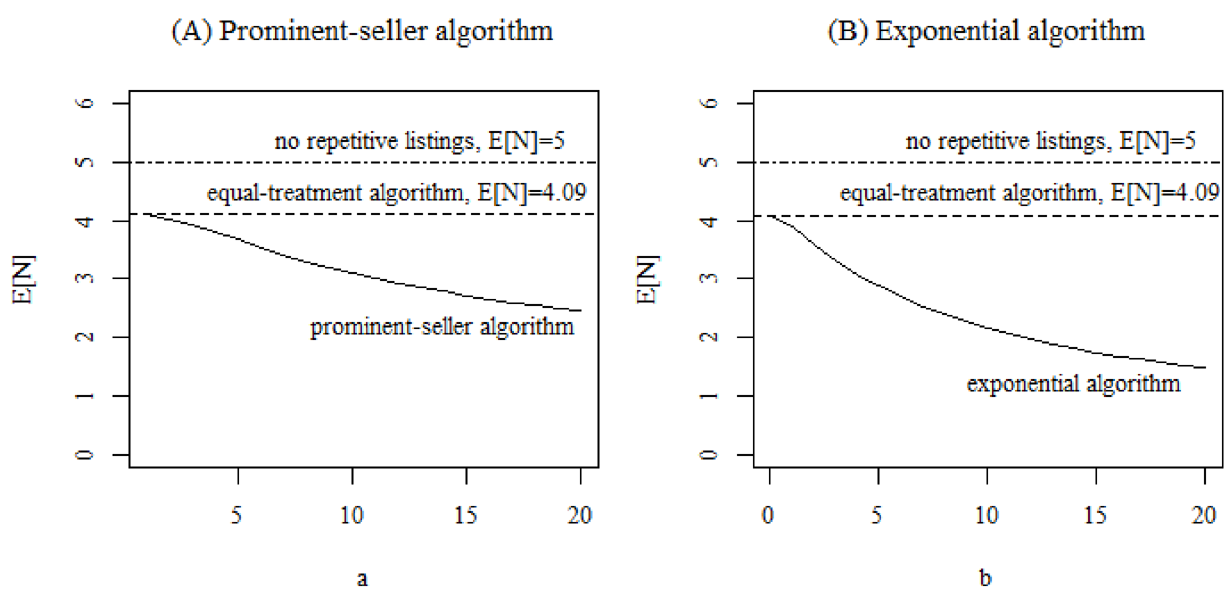
[9] The name "prominent-seller" is taken from Armstrong et al. (2009).

(P_n). Figure 5 illustrates Lemmas 2. As the prominent-seller (exponential) algorithm becomes more unequal, the expected consideration set size ($E[N]$) decreases.



Note: we set $J = 10$, $K = 5$, $\theta \sim U[0, 1]$, $c(q, \theta) = q^2/(2\theta)$, $v(q) = q$, $a = 10$, and $b = 10$.

Figure 4: Illustration of the Random Consideration Set Size



Note: we set $J = 10$, $K = 5$, $\theta \sim U[0, 1]$, $c(q, \theta) = q^2/(2\theta)$, and $v(q) = q$.

Figure 5: Search Algorithm and the Expected Consideration Set Size

2.2 Market Equilibrium and Welfare

The search intensity (K), number of products (J), and search algorithm ($\boldsymbol{\sigma}$) jointly determine the probability that product j is in the consideration set ($\Pr(j \in \mathcal{C})$) and the distribution of consideration set size (P_n). Given the above search process, a generic seller j of type θ chooses the price and quality through the maximization problem (1). Because I and $\Pr(j \in \mathcal{C})$ are not affected by the seller's choice of price and quality, the maximization problem above is equivalent to

$$\max_{p,q} \, [p - c(q,\theta)] \Pr\left(j = \arg\max_{j' \in \mathcal{C}}\{u_{j'}\}\right). \tag{3}$$

The surplus of a product with quality q is the difference between its value to the buyer and the cost of the product, i.e., $v(q) - c(q,\theta)$. Given that v is concave and c is strictly convex in q, there is a unique solution of $\max_q \{v(q) - c(q,\theta)\}$ characterized by the first-order condition $v_q(q^*) - c_q(q^*,\theta) = 0$. Define the solution as

$$q^*(\theta) = \arg\max_{q} \{v(q) - c(q,\theta)\}. \tag{4}$$

For a seller with type θ, by producing at quality $q^*(\theta)$, the product generates the largest social surplus, $w^*(\theta) = v(q^*(\theta)) - c(q^*(\theta),\theta)$. Both $q^*(\theta)$ and $w^*(\theta)$ are increasing in the efficiency parameter θ.[10] $w^*(\cdot)$ is a one-to-one mapping from type θ to surplus measure w. Given distribution function F, there is a unique distribution function G for w. G has a compact support $[\underline{w}, \overline{w}]$, where $\underline{w} = w^*(\underline{\theta})$ and $\overline{w} = w^*(\overline{\theta})$.

Then, we can rewrite the maximization problem (3) such that the seller first chooses a utility level u for her product based on the social surplus w and seeks a price-quality combination to fulfill this utility level.

$$\begin{aligned}
(3) \Leftrightarrow \max_u & \left\{ \max_{(p,q) \text{ s.t. } v(q)-p=u} [p - c(q,\theta)] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)\right\} \\
\Leftrightarrow \max_u & \left\{ \max_q \, [v(q) - c(q,\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)\right\} \\
\Leftrightarrow \max_u & \left\{ [v(q^*(\theta)) - c(q^*(\theta),\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)\right\} \\
\Leftrightarrow \max_u & \left\{ [w^*(\theta) - u] \Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)\right\}. \qquad (5)
\end{aligned}$$

Ultimately, product utility can determine whether j is chosen among the available options. Therefore, the seller will choose the efficient quality $q^*(\theta)$ in equilibrium because it maximizes the first term, $[v(q) - c(q,\theta) - u]$, without affecting the second term, $\Pr\left(u \geq \max_{j' \in \mathcal{C}}\{u_{j'}\}\right)$.

Given social surplus w, the firm will choose the utility level of the product that solves (5). We focus on the symmetric and monotone Bayesian Nash equilibrium (BNE), $u = \mu(w)$. Let $W_{(1:N)}$

[10] Because $c_{q\theta} < 0$, according to the implicit function theorem, $\frac{dq^*}{d\theta} = -\frac{-c_{q\theta}}{v_{qq} - C_{qq}} = \frac{c_{q\theta}}{v_{qq} - c_{qq}} > 0$.

denote the greatest order statistic among N independent draws from G. $G_{(1:N)}(w) = \Pr(W_{(1:N)} \leq w)$ is the distribution function of $W_{(1:N)}$, where $G_{(1:N)}(w) = [G(w)]^N$. Given that $\mu(\cdot)$ increases in w, the *choice probability* of a seller with social surplus w is

(6) $$\mathcal{P}(w) = \sum_{n=1}^{J} \left\{ P_n \times G_{(1:n-1)}(w) \right\} = E_N \left[G_{(1:N-1)}(w) \right],$$

indicating the probability that a product has the highest social surplus among N options in the consideration set. The solution to (3) is characterized in the following proposition.

Proposition 1. *A seller with efficiency parameter θ chooses the product with quality $q^*(\theta)$ given in (4) and $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$, where the product utility is*

$$\mu(w^*(\theta)) = w^*(\theta) - \frac{\int_{\underline{w}}^{w^*(\theta)} \mathcal{P}(\omega)\, d\omega}{\mathcal{P}(w^*(\theta))}.$$

Note that $\mu(w^*(\theta)) = v(q^*(\theta)) - p^*(\theta)$ is the product utility or buyer surplus offered by the seller with type θ in equilibrium. The equilibrium price divides the social surplus of a transaction into the seller's and the buyer's surpluses:

$$\underbrace{w^*(\theta)}_{\text{social surplus}} = \underbrace{[p^*(\theta) - c(q^*(\theta), \theta)]}_{\text{seller's profit}} + \underbrace{[v(q^*(\theta)) - p^*(\theta)]}_{\text{buyer's surplus}}.$$

Let $\mu(w; \boldsymbol{\sigma})$ and $p^*(\theta; \boldsymbol{\sigma})$ denote the equilibrium price and product utility given search algorithm $\boldsymbol{\sigma}$, respectively. We have the following result:

Proposition 2. *Given two search algorithms $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$ with $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$, we have $p^*(\theta; \boldsymbol{\sigma}^1) \leq p^*(\theta; \boldsymbol{\sigma}^2)$ and $\mu(w^*(\theta); \boldsymbol{\sigma}^1) \geq \mu(w^*(\theta); \boldsymbol{\sigma}^2)$.*

Proposition 2 indicates that if the search algorithm becomes more equal, the average price decreases and the average product utility increases in equilibrium. In Section 4, we find empirical evidence that markets with less equal search results have higher average prices. Therefore, the search algorithm critically determines how the seller and the buyer divide the surplus.

In equilibrium, the demand of a seller with type θ is

$$y^*(\theta) = I \times \Pr(j \in \mathcal{C}) \times \mathcal{P}(w^*(\theta)).$$

The first probability, $\Pr(j \in \mathcal{C})$, is determined by search algorithm ($\boldsymbol{\sigma}$) and search intensity (K). Seller j cannot affect this probability through her choice of price and quality.[11] Obviously, if

[11]If there are no search frictions, the consideration set includes products from all sellers. Without the random utility term the in discrete choice models, the most efficient seller will offer the best product (in terms of utility) and capture the entire market. The source of market power is rooted the in buyers' inability to search all products. Therefore, the most efficient seller cannot sell to all buyers due to search frictions.

$\Pr(j \in \mathcal{C})$ is large, seller j's sales and revenue will increase. The second probability, $\mathcal{P}(w^*(\theta))$, is the choice probability determined by efficiency parameter of the seller (θ) and consideration set size (N).

Given n options in the consideration set, a buyer purchases the product with the highest utility, which is offered by the seller with the highest social surplus among n sellers. Considering the randomness of N, we can construct *ex ante* welfare measures. The total social welfare is

$$SW = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} w dG_{(1:N)}(w) \right] = I \cdot E_N \left[E\left[W_{(1:N)} \right] \right]. \tag{7}$$

Social welfare can be decomposed into the buyer-side surplus,[12]

$$U = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} \mu(w) dG_{(1:N)}(w) \right] = I \cdot E_N \left[\mu(W_{(1:N)}) \right] \tag{8}$$

and the seller-side profit,

$$\mathit{\Pi} = I \cdot \sum_{n=1}^{J} P_n \left[\int_{\underline{w}}^{\overline{w}} \left[w - \mu(w) \right] dG_{(1:n)}(w) \right] = I \cdot E_N \left[E\left[W_{(1:N)} - \mu(W_{(1:N)}) \right] \right]. \tag{9}$$

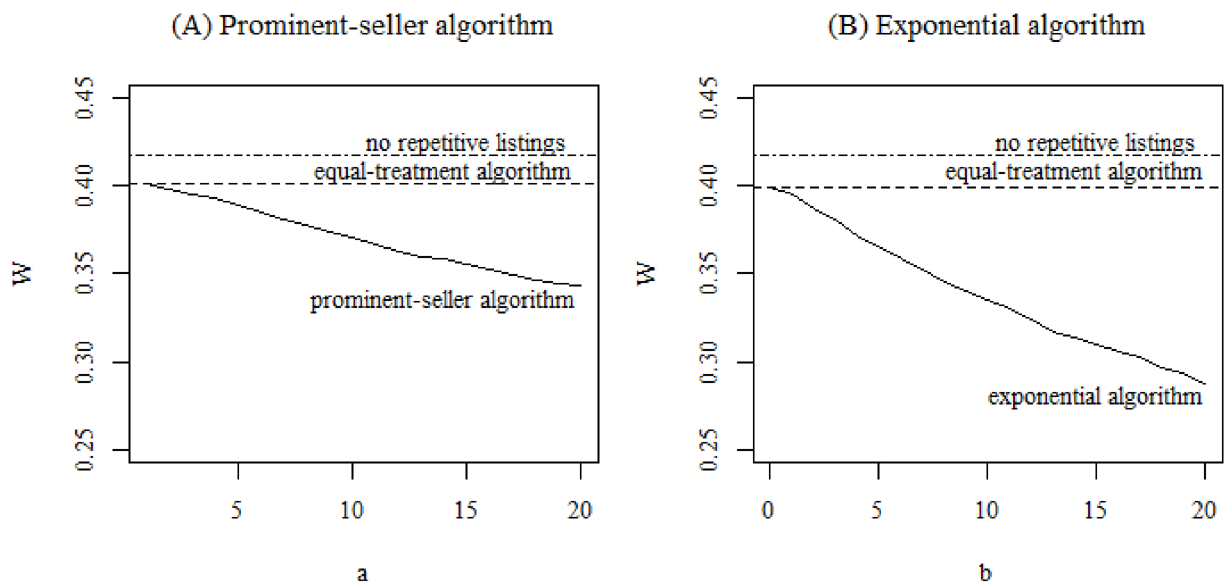
These three welfare measures depend on the search algorithm and the consideration set size:

Proposition 3. *The welfare measures have the following properties:*
(i) *If N increases in the sense of FSD, SW increases, U increases, and $\mathit{\Pi}$ decreases.*
(ii) *Given two search algorithms $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$ with $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$.*
(iii) *The equal-treatment algorithm maximizes SW and U.*

Overall market efficiency and buyer-side surplus both improve when the market becomes more competitive. Figure 6 illustrates Proposition 3. Social welfare decreases as the algorithm becomes increasingly unequal. Because the equal-treatment algorithm maximizes the expected size of consideration set (Lemma 3), it also maximizes social welfare.

However, the platform may adopt a highly unequal search algorithm. As the algorithm becomes "less equal," the consideration set size shrinks. It softens the competition among sellers (Lemma 2) and raises prices (Proposition 2). As a result, by adopting an unequal algorithm, the platform can harm social welfare and buyer-side surplus (Proposition 3). De Corniere and Taylor (2019) reaches a similar result that a biased intermediary harms consumers because the favored firm offers a product with lower utility.

[12] By assuming a fixed-sample-size search, we do not consider the saving of search costs from using the search tools in the welfare computation. With buyers searching sequentially and strategically, search design affects search costs and buyer welfare in a complicated and ambiguous way. See Chen and Zhang (2017, 2018).



Note: we set $J = 10$, $I = 1$, $K = 5$, $\theta \sim U[0,1]$, $c(q,\theta) = q^2/(2\theta)$, and $v(q) = q$.

Figure 6: Welfare Implications of the Search Algorithm

Because the equal-treatment algorithm maximizes the consideration set, it maximizes social welfare at all search intensity levels (Proposition 3). As the search algorithm favors some option(s), buyers are more likely to obtain repetitive information, and thereby consider less options and softens the competition among sellers. These results support the ruling by the European Commission (2017). Moreover, the requirement by the Federal Trade Commission (2013) for distinguishing sponsored and organic search results can help buyers search more effectively. As the search intensity increases, social welfare and buyer-side surplus improve.

2.3 No Repetitive Listing

We consider an alternative scenario that the search algorithm is not allowed to present repetitive information. When there is no repetitive listing, the sampling among J sellers becomes a sampling process without replacement. Under the fixed-sample-size search, each buyer considers exactly K options. As a result, each seller expects that she will compete with $K-1$ other sellers. The choice probability of a seller in (6) becomes $\mathcal{P}(w) = G_{(1:K-1)}(w)$.

In Figures 5 and 6, we depict the case of no repetitive listings. Compared with the baseline case, the price level is lower, the buyer can purchase a product with better quality, and social welfare becomes higher than that it is under the equal-treatment algorithm.

3 Discussion

3.1 Informative Search Algorithm

The results above are obtained under a setting in which the platform commits to a search algorithm at the beginning of the game. Because all sellers are *ex ante* symmetric, the search algorithm is uninformative in the sense that it does not reflect seller types or other characteristics (such as price and quality). The uninformative setting establishes an important benchmark for how search algorithms lead to repetitive information and affect market equilibrium and welfare.

In reality, platforms usually possess data regarding sellers and use the data as inputs in their search algorithms. Platforms can strategically design the search algorithm based on seller attributes. As a result, the search algorithm becomes informative as it promotes sellers with certain characteristics. The promotion can be personalized for different buyers based on buyer-side data, such as queries and purchase histories.

If the platform has information about product quality, it can set the search algorithm to rank an efficient seller up front.[13] However, a self-interested platform might not have the incentive to do so. Low-quality sellers may be favored by the algorithm because they pay for search advertisements.[14] In this case, platforms are likely to cause welfare loss because high-quality sellers appear less frequently in consideration sets.

Consider a simple scenario in which a search algorithm depends on realized efficiency parameters. Because sellers still solve the profit maximization problem (3), the equilibrium in Proposition 1 still holds.[15] Let the most efficient seller be promoted by the prominent-seller algorithm with parameter a. Figure 7-(A) demonstrates the welfare implications. As shown by the red dashed curve, social welfare increases in a. However, if the least efficient seller is promoted, social welfare decreases in a, as illustrated by the blue dashed curve. If the algorithm is uninformative, social welfare still decreases in a (Proposition 3). Figure 7-(B) shows a similar result for the exponential algorithm.

In practice, the platform has a natural incentive to favor sellers who pay higher commissions or are vertically integrated with it (Inderst and Ottaviani, 2012; Teh and Wright, 2020). For example, the European Commission (2017) find that "Google systematically gave prominent placement to its own in-house service and demoted rival comparison shopping services in search results, so even the most highly ranked rival service appears on average only on **page four** of Google's search results." Moreover, because buyers are less likely to purchase niche products, the platform might be inclined to promote already popular mass-market products (Bar-Isaac et al., 2012).

[13] If products are horizontally differentiated, then the search algorithm can promote products that fit the tastes of different buyers. Queries and buyer-side data can be used to infer buyer preferences.

[14] For example, in 2016, a Chinese college student died after receiving an experimental treatment for synovial sarcoma promoted by Baidu. Because of this case, the Cyberspace Administration of China imposed new restrictions on search advertisements. See `en.wikipedia.org/wiki/Death_of_Wei_Zexi`.

[15] Note that, if the search algorithm is not based on θ but on q or p, the profit maximization problem (1) is no longer equivalent to (3). The equilibrium choice of p and q in Proposition 1 will not hold because sellers will strategically choose prices and quantity levels in response to the design of the search algorithm.

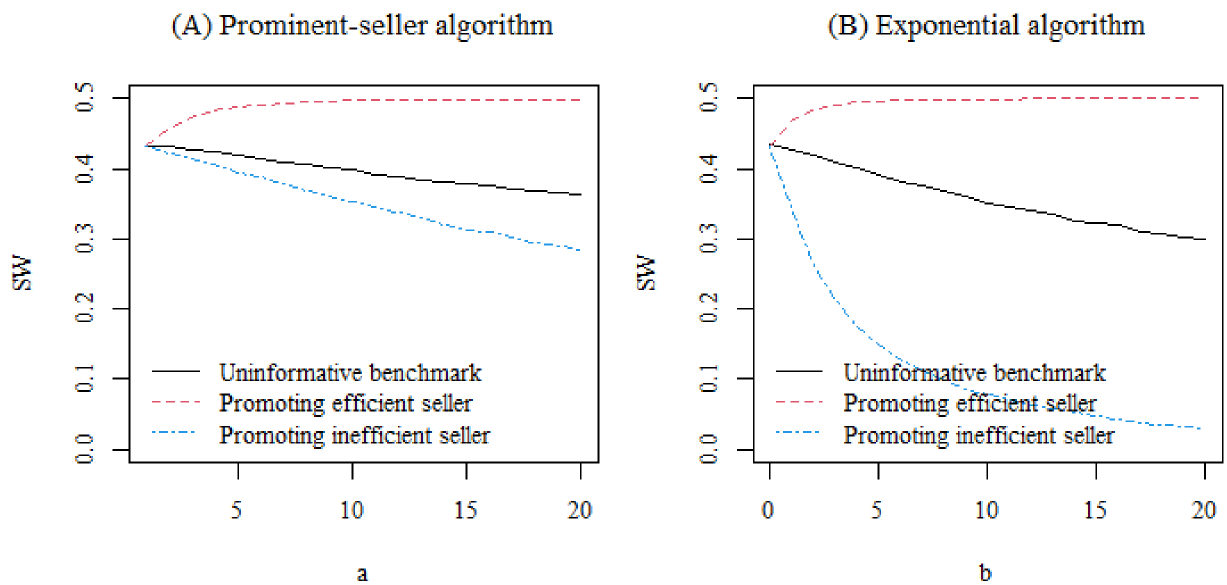


Figure 7: Informative Algorithm and Welfare

Sellers can also influence the search algorithm through search advertisements or other approaches. Search advertisements in the search results can substantially affect the search processes of buyers (Ricci et al., 2011). By paying for sponsored searches, advertising sellers not only increase their chances of being considered but also indirectly suppress competition since other products are considered less often. The platform also has a strong incentive to promote sponsored results and render repetitive search results. By filling buyer consideration sets with a few sponsored options, platforms obtain not only more advertisement revenue but also higher commissions from higher prices under the popular proportional fee scheme.

3.2 Endogenous Participation and Platform Choice

A platform can manipulate the consideration set by selecting which items to show to each buyer. In the extreme, it can only display one option and make other options never appear. However, to some extent, platforms will design algorithms to be in line with social welfare because they must attract users to participate. Endogenous participation restricts the extent to which the platform uses the search algorithm for its private interest. We can extend the model to incorporate endogenous participation based on the following timeline:

(i) The platform chooses the fee and establishes the search algorithm.

(ii) Buyers and sellers decide whether to participate in the platform.

(iii) Sellers draw their efficiency parameters and choose products as price-quality combinations.

(iv) Buyers search via the search algorithm, form consideration sets, and buy products.

The platform chooses the search algorithm by committing to a way of assigning the probability vector $\boldsymbol{\sigma}$ given J participating sellers. Under this setting, the baseline model in Section 2 is a subgame at stage (iii) with I buyers and J sellers that participate in the market. At stage (ii), all sellers are symmetric, and the expected profit of a seller is $\pi(I,J,\boldsymbol{\sigma}) = \frac{1}{J}\Pi = \frac{1}{J}\cdot I \cdot E_N\left[E\left[W_{(1:N)} - \mu(W_{(1:N)})\right]\right]$, which increases in I and decreases in J. Suppose that the platform charges a uniform membership fee $\tau \geq 0$ for each seller (Armstrong, 2006).[16] Seller participation is determined by the zero profit condition

$$\pi(I,J,\boldsymbol{\sigma}) = \tau \Leftrightarrow J = \mathcal{J}(I,\tau,\boldsymbol{\sigma}), \tag{10}$$

where $\mathcal{J}$ is an increasing function of I and a decreasing function of τ.

Suppose that buyers have heterogeneous reservation utility r, which represents the payoff of not participating in the platform or purchasing from another competing platform. At the beginning of the participation stage, each buyer draws an r from a commonly known distribution $\Gamma(\cdot)$. Γ has continuous support and is strictly increasing. From (8), the expected payoff of a generic buyer i is $u(J,\tau,\boldsymbol{\sigma}) = E_N\left[\mu(W_{(1:N)})\right]$,which increases in J. A buyer participates if $u(J,\tau,\boldsymbol{\sigma}) \geq r$. Given $\mathbb{I}$ potential buyers, the number of buyers that will participate in the platform is

$$I = \mathbb{I} \times \Pr\left(u(J,\tau,\boldsymbol{\sigma}) \geq r\right) = \mathbb{I} \times \Gamma\left(u(J,\boldsymbol{\sigma},\tau)\right) = \mathcal{I}(J,\boldsymbol{\sigma},\tau), \tag{11}$$

where $\mathcal{I}$ is an increasing function of J and a decreasing function of τ.

Because seller-side profit and buyer-side surplus are both increasing functions of the other side's participation, indirect network externalities (Rochet and Tirole, 2006) arise naturally from the model. Equations (10) and (11) constitute a typical two-sided market. Given a membership fee τ and a search algorithm $\boldsymbol{\sigma}$, the equilibrium numbers of sellers and buyers are determined by

$$\begin{cases} I = \mathcal{I}(J,\tau,\boldsymbol{\sigma}) \\ J = \mathcal{J}(I,\tau,\boldsymbol{\sigma}) \end{cases} \Rightarrow \begin{cases} J^*(\tau,\boldsymbol{\sigma}) \\ I^*(\tau,\boldsymbol{\sigma}). \end{cases}$$

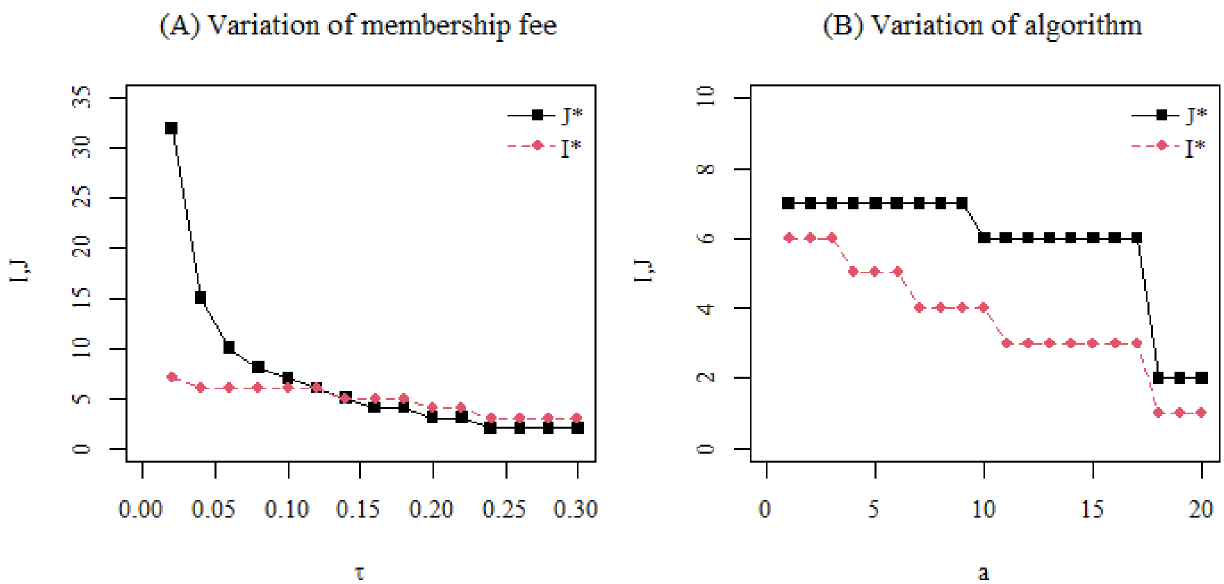
The platform's profit is $T(\boldsymbol{\sigma},\tau) = \tau \times J^*(\boldsymbol{\sigma},\tau)$.

Figure 8-(A) uses a numerical example to illustrate how equilibrium numbers of sellers and buyers decrease with the membership fee. Because the membership fee is a transfer from sellers to the platform, social welfare can still be expressed as (7). Obviously, social welfare is maximized at $\tau = 0$, but the platform has the incentive to charge a positive fee ($t > 0$). Hence, the membership fee based on the platform's private benefit is too high from a social welfare perspective.

Figure 8-(B) shows equilibrium numbers of sellers and buyers under the prominent-seller algorithm with different levels of a. With a less equal algorithm, the platform becomes less attractive to both buyers and sellers. Social welfare is maximized under the equal-treatment algorithm with $a = 1$. However, the platform can obtain advertisement income by promoting a prominent seller

[16] If the platform charges a proportional or usage fee (Rochet and Tirole, 2003), the fee will distort the quality provision, but the result will be qualitatively similar.

($a > 1$). Hence, the discrepancy between the platform's private interest and the public interest also lies in the design of the search algorithm.



Note: we set $K = 5$, $\theta \sim U[0, 1]$, $c(q, \theta) = q^2/(2\theta)$, and $v(q) = q$.

Figure 8: Sellers and Buyers Participation in Equilibrium

3.3 Policy Implications

Many previous studies about regulating online platforms have focused on fees and contract forms. For example, Chang et al. (2005) examine the case in which Australia regulators cap credit card fees; Wang and Wright (2020) investigate the implications of banning price-parity clauses; Etro (2021) investigates how seller entry affects commissions set by competing platforms; and Hagiu et al. (forthcoming) discuss whether platforms should be allowed to sell on their own marketplaces.

Recently, the anti-competitive role of search algorithms starts to receive increasing attention. Teh and Wright (2020) find that steering can cause prices to increase. Chen and Tsai (2021) note the antitrust concerns of Amazon's using the recommendation algorithm to steer buyers toward its own products. Teh (2022) demonstrates how platforms can use the recommendation algorithm as a tool for information design to extract consumer surplus. Casner (2020) finds that platforms have the incentive to combine seller recommendation and search obfuscation, which softens competition and increases prices. In this paper, we emphasize that, even if the algorithm recommends high-quality sellers, it may still result in anti-competitive effect due to repetitive listings.

In light of the model, the general goal of regulating search algorithms is to induce them to promote a large number and variety of alternative options (Proposition 3). Displaying a sufficient number of available options is an important competitive force that protects buyers' interests. Our policy recommendation is that regulators can consider restricting or prohibiting search algorithms from presenting the same sellers multiple times in the (first few pages of) search results. This regulatory goal does not conflict with platforms' presenting high-quality sellers or the most relevant options on top. If platforms can be restricted from presenting repetitive information, sellers ranked at the top of search results would face stronger competitive pressure and lower prices (Figures 5 and 6).

Note that implementing regulations on search algorithms involve an important challenge: search algorithms are not directly observable. Regulations on fee levels and contract forms are straightforward because they are observable. For example, Australia imposed a credit card interchange fee cap after the regulatory decisions (Chang et al., 2005); Booking.com and Expedia removed some price-parity clauses after being investigated by antitrust authorities (Wang and Wright, 2020). In contrast, search algorithms are high-dimensional, complicated, and hidden; thus, it is difficult (if not impossible) for regulators to monitor algorithms and enforce regulations. Even if regulators require platform operators to report their algorithm and provide historical data, such requests are subject to manipulation and concealment.

Nevertheless, some key features of a search algorithm based on systematic trial searches. In fact, the investigation of Google by the European Commission (2017) was based on trial search results. One way to implement regulations on search algorithms is to conduct trial searches regularly and announce such practice to the public.[17] Based on trial search results and data regarding buyer behavior (e.g., click-through rates), the modeling framework developed in this paper can be used to estimate the probability of each option being considered and the average size of consideration sets. Regulators can then approximately predict the effect of regulatory proposals and evaluate their welfare implications.

4 An Empirical Study of Food Delivery Platforms

We conduct a short empirical exercise to demonstrate how search algorithms affect online sales using data from Chinese food delivery platforms. We do not have administrative data from the platforms. Instead, we construct the data by scraping and trial searches. In particular, we keep track of repetitive listings and store rankings in the trial search results. These variables measure the amount of repetitive information and how equal the search algorithm is. We find that markets with less equal search results tend to have higher prices, which supports Proposition 2.

4.1 Industry Background

Online food delivery platforms allow buyers to search for food from nearby restaurants and place orders through mobile applications or websites. Much like other e-commerce platforms, as buyers

[17] Although search results are often personalized based on buyers characteristics, regulators can use multiple fictitious accounts to conduct trial searches. Regulators can also focus on monitoring search results by using an anonymous or guest account. Then, buyers who are concerned about their data being used against their interests can use search tools with an anonymous account.

open the app, a list of restaurants appears. Buyers can scroll down for more options and refine the search results by queries or filters based on the food category, delivery time, style, and price. By clicking a restaurant's picture on the list, the buyer goes to the store's page. The page contains a food menu with images, prices, estimated delivery times, ratings, reviews, and other textual descriptions. Figure 9 shows an example of a store page.

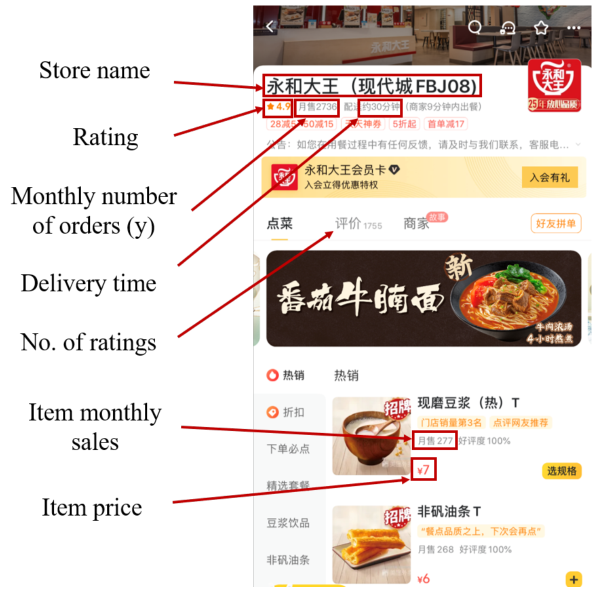


Figure 9: A Store Page from Meituan App

The catering industry in China experienced drastic changes since food delivery platforms became popular in 2013. The market size of takeout food in China grew from less than US$10 billion in 2013 to more than US$37 billion in 2017. In 2018, there were more than 256 million active users of online food delivery services in China, which covers more than 1,300 cities. There were three major food delivery platforms, namely, Meituan, Ele.me, and Baidu.[18] Meituan and Ele.me together accounted for more than 80% of all takeout food transactions.[19] Many restaurants operated on multiple platforms, while most buyers used only one platform.[20]

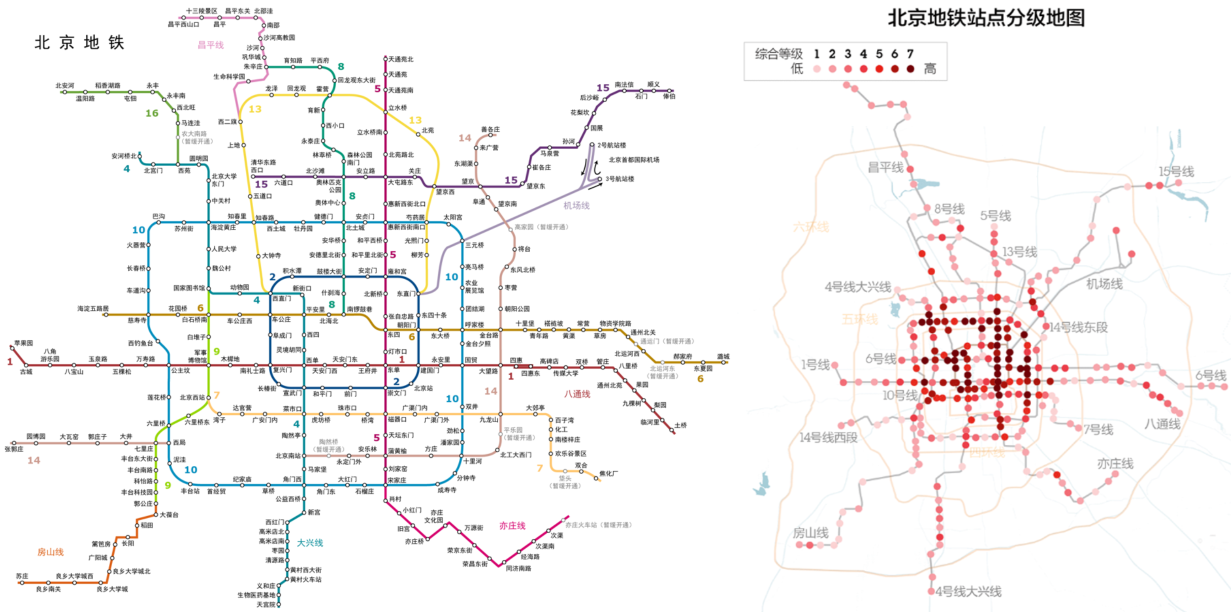
With food delivery platforms, buyers can access more information about takeout food, but the search process is now guided by the search algorithms provided by the platforms. On the one hand, small restaurants have the opportunity to develop their reputations and attract buyers who search carefully. Given sufficient taste heterogeneity, reducing the search costs empowers small and

[18] Meituan (`waimai.meituan.com`) was backed by Tencent. Alibaba was the major investor in Ele.me (`www.ele.me`). The food delivery service offered by Baidu (`waimai.baidu.com`) was facilitated by the major internet search engine in China. See `www.scmp.com/business/companies/article/2111163/dinner-your-door-inside-chinas-us37-billion-online-food-delivery`. In August 2017, Ele.me acquired the food delivery business of Baidu. In October 2018, the Baidu platform was rebranded as `star.ele.me`.

[19] Sources: `www.itrustdata.cn`.

[20] Almost 60% of the restaurants on Ele.me also operated on Meituan, but only 7.6% of buyers actively used both Ele.me and Meituan. This two-sided market structure fits the case of the competitive bottleneck (Armstrong, 2006) in which platforms compete for a larger installed base of single-homing buyers, and grant them market power in addressing multi-homing sellers.

high-quality sellers, which leads to less concentrated sales. On the other hand, big sellers with well-known brands and chain stores could be further strengthened by platforms.[21] The overall effect of food delivery platforms on sales concentration is ambiguous.



Note: The right panel shows the economic activeness of subway stations. Sources: DTCJ (`www.dtcj.com`).

Figure 10: Beijing Subway Stations and Their Activeness Levels

Food delivery platforms serve as a good data source for studying online platforms' influence on sales distribution because restaurants are segregated by geographic locations.[22] Therefore, there are many market-level observations with variations in their seller composition. Moreover, buyers' search processes are heavily influenced by the search algorithms. Once buyers turn on the mobile app, the algorithm will recommend a list of nearby restaurants without taking any queries. This restaurant list is mainly based on buyer locations.[23]

4.2 Data

Our data cover restaurants that operate on three food delivery platforms in Beijing in August 2018. Beijing has a large and highly active takeout food market with more than 100 thousand active restaurants and more than 10 million active users on these platforms. The data consist of all restaurants that provide services in the vicinity of all 296 subway stations in Beijing. Figure 10 shows a map of Beijing's subway stations.

[21] In 2018, the top five sellers on Ele.me were Kentucky Fried Chicken, McDonald's, Pizza Hut, Burger King, and Yoshinoya (Sources: `www.cbndata.com`).

[22] Food delivery services are restricted by the physical locations of restaurants because the food must be in good condition after delivery. The retention of heat critically depends on the delivery distance.

[23] It can also depend on other data such as purchase histories and personal characteristics. However, for takeout food, location is the most important factor.

Table 1 summarizes the store-level data. A store is defined as a restaurant page on a food delivery platform. Figure 9 shows an example of a restaurant page on the Meituan Mobile App. A physical restaurant might operate on multiple platforms, and each of them is treated as a different store. For each store, we observe its monthly number of orders, which is treated as the quantity y. Moreover, for each item on the menu, we observe its price and monthly sales, from which we compute the weighted average price p and *revenue*.[24] On average, a store earns a monthly revenue of CNY 10,289 by operating on one food delivery platform. We observe the overall *rating* of the store on a scale from 1 to 5 and the total number of ratings (*N.rating*). For most stores, we also observe the minimum price for delivery (*deliv.min.p*), delivery fee (*deliv.fee*), and the average time of delivery (*deliv.time*) in minutes.

Table 1: Summary Statistics of Store-level Data

Variable	Obs	Overall Mean	St. Dev.	Min	Max
Ele.me	102,037	0.360	0.480	0	1
Meituan	102,037	0.392	0.488	0	1
y	102,037	436.635	916.256	1	16,712
p	102,037	32.049	97.133	0.010	9,999
revenue	102,037	10,288.680	37,780.600	0.010	2,225,805
rating	91,340	4.587	0.309	1.800	5
N.rating	98,297	338.883	1,407.747	0	61,644
deliv.min.p	92,129	24.362	22.567	0	2,000
deliv.fee	98,016	6.690	6.184	0	205
deliv.time	96,188	38.956	12.851	0	565
weight.position	102,037	199.814	168.093	1	707.083
repeat.listing	102,037	1.502	1.626	1	31
J	102,037	154.821	160.569	1	761
activeness	102,037	3.581	2.055	1	7
N.brand.stores	102,037	45.651	104.396	1	640
N.brand.stores.plat	102,037	19.617	44.296	1	328

Note: *Ele.me* and *Meituan* are indicators for stores on Ele.me and Meituan platforms, respectively. y is the monthly number of orders. p is the weighted average price of orders. *revenue* is the monthly revenue of the store. *rating* is the store rating. *N.rating* is the number of ratings and reviews. *deliv.min.p*, *deliv.fee*, and *deliv.time* are the minimum price for delivery, delivery fee, and average delivery time, respectively. *weight.position* is the weighted average position of the store in search results. *repeat.listing* is the average number of stores with the same brand in the search results. J is the number of stores in the market. *activeness* measures the consumer size of the market. *N.brand.stores* is the number of stores with the same brand across all three platforms. *N.brand.stores.plat* is the number of stores with the same brand on the store's platform. All monetary data are in CNY.

By textual analysis, we find that the stores belong to 34,823 brands (restaurant chains that share the same brand). Based on brand identities, *N.brand.stores* counts the number of stores of each restaurant chain on all three platforms. For each store, *N.brand.stores.plat* counts the number of stores with the same brand on its platform. In the data, the largest restaurant chain is Shaxian

[24] For example, there are three items on the menu of a store with prices 2, 3, and 4. The monthly average sales of the three items are 15, 12, and 5, respectively. The revenue is $2 * 15 + 3 * 12 + 4 * 5 = 86$. The weighted average price is $p = 86/(15 + 12 + 5) = 2.6875$.

Xiaochi with 640 stores in total on the three platforms. Table 2 shows that approximately 10% of the brands have more than 5 stores but comprise 76% of the total revenue in the entire market.

Table 2: Number of Stores and Revenue Share by Brand

N.brand.stores	Obs	Percentage	Total revenue	Revenue share
≥1	34,823	100%	1,049,826,347	100%
≥2	14,235	40.88%	974,365,741	92.81%
≥3	7,739	22.22%	901,015,003	85.83%
≥5	3,660	10.51%	801,825,095	76.38%
≥10	1,211	3.48%	590,139,462	56.21%
≥50	170	0.49%	282,892,568	26.95%

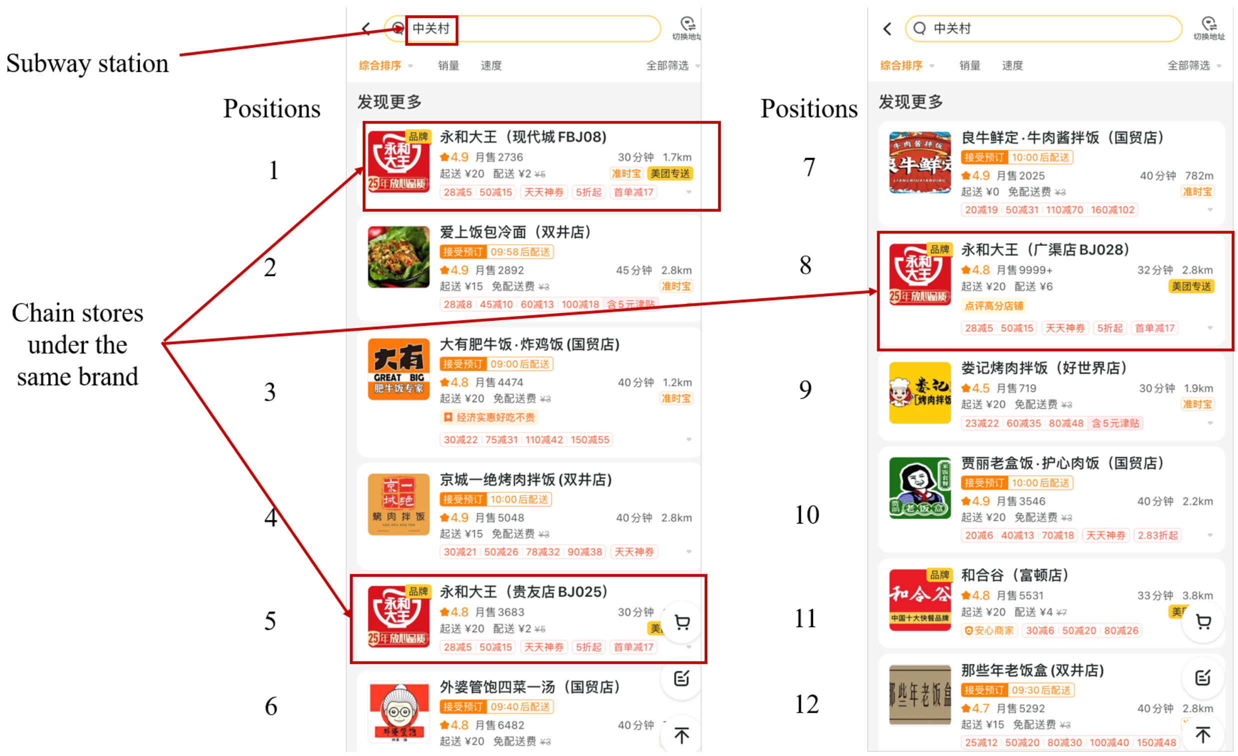


Figure 11: Illustration of a Search Result

We construct two variables, *weight.position* and *repeat.listing*, for each store based on the search results from trial searches. In August 2018, we chose 30 random times between 10:00 AM and 2:00 AM and used a computer program to record the positions (rankings) of all stores as appeared in the default search results by location (subway station), category, and platform.[25] Figure 11 illustrates the positions of stores in a search result. The store that appears at the top of the search result has *position*= 1; the store that appears in second place has *position*= 2, and so forth. Because

[25]The ranking of stores is likely to be personalized based on the transaction history and demographic information of consumers. Unfortunately, we do not have administrative data that contains click-through records and personalized search results of individual consumers as in De los Santos and Koulayev (2017); Dinerstein et al. (2018) and Ursu (2018).

many stores appear in several markets, we compute the weighted average position of each store by using the *activeness* of different markets as weights (*weight.position*). A store with a larger *weight.position* is less likely to be considered by consumers.

As illustrated in Figures 2 and 11, the search results frequently include multiple stores of the same brand. These chain stores provide the same food items with the same prices set by the central office of the brand. For each store, we count the number of stores with the same brand in the search results and construct the variable *repeat.listing* as the average across trial searches.

We define a market as a unique combination of location, category,[26] and platform. Table 3 provides the summary statistics of the market-level data. Let J denote the number of stores in each market. The variable *activeness* is a big-data based index measuring the activeness of economy in the region provided by DTCJ (`www.dtcj.com`), which is illustrated in Figure 10. We use *activeness* to control for the consumer size of different markets. For each market, we compute the aggregate number of monthly orders (*market.y*), total number of ratings and reviews (*market.N.rating*), and the average price of delivery orders (*market.ave.p*). We measure the sales concentration of a market by the Gini coefficient of store revenues (*Gini.revenue*)[27] and the market share of the top 20% of stores (*top-20 share*).[28] For each market, we count the number of brands with more than 10 stores on this platform (*N.brand.stores.plat*$\geq$ 10) and across the three platforms (*N.brand.stores*$\geq$ 10) based on brand identities. These two variables measure the presence of large chain restaurants in the market.

Table 3: Summary Statistics of Market-level Data

Variable	Obs	Mean	St. Dev.	Min	Max
J	7,442	210.1	244.3	1	761
market.y	7,442	72,770	128,242	1	1,596,397
market.N.rating	7,442	123,532	232,539	0	1,533,184
market.ave.p	7,442	61.82	65.29	1.945	1,708
Gini.revenue	7,442	0.640	0.152	0	0.937
top-20 share	7,442	0.693	0.151	0	0.998
Gini.position	7,442	0.155	0.084	0	0.578
N.brand.stores$\geq$ 10	7,442	69.24	82.943	0	584
N.brand.stores.plat$\geq$ 10	7,442	41.67	52.48	0	462
activeness	7,442	3.382	1.970	1	7

Note: J is the number of stores. *market.y* is the total number of monthly orders among all stores in the market. *market.N.rating* is the total number of ratings among all stores. *market.ave.p* is the average price of orders across all stores. *Gini.revenue* is the Gini coefficient of *weight.position* of all stores in the market. *top-20 share* is the revenue share of top 20% stores. *N.brand.stores*$\geq$ 10 is the number of brands with more than 10 stores in the market. *N.brand.stores*$\geq$ 10 is the number of brands with more than 10 stores on the platform. *activeness* measures the consumer size of the market.

[26] There are 12 main food categories, such as fast food, formal meals, and dessert and drink.

[27] The Gini coefficient is widely used as a measure of sales concentration (Fleder and Hosanagar, 2009, 2007; Brynjolfsson et al., 2011). We compute it using the method in Gaswirth (1972).

[28] We select 20% because the Pareto principle (`en.wikipedia.org/wiki/Pareto_principle`) suggests that the top 20% of sellers can capture 80% of the market (Brynjolfsson et al., 2011).

For each market, we compute the Gini coefficient of *weight.position* for all stores in the market (*Gini.position*) and use it to measure how unequal the search algorithm is. Under the equal-treatment search algorithm, the ranking of the stores is random. As a result, the *weight.position* of all stores will be close to one another, and the *Gini.position* will be close to zero. By contrast, if the search algorithm is highly unequal, some stores will always appear on top of the search results, while some other stores will always appear at the bottom. The large difference of the *weight.position* across stores result in a large *Gini.position* for the market. Figure 12 show that *Gini.revenue* and *market.ave.p* are positively correlated with *Gini.position*.

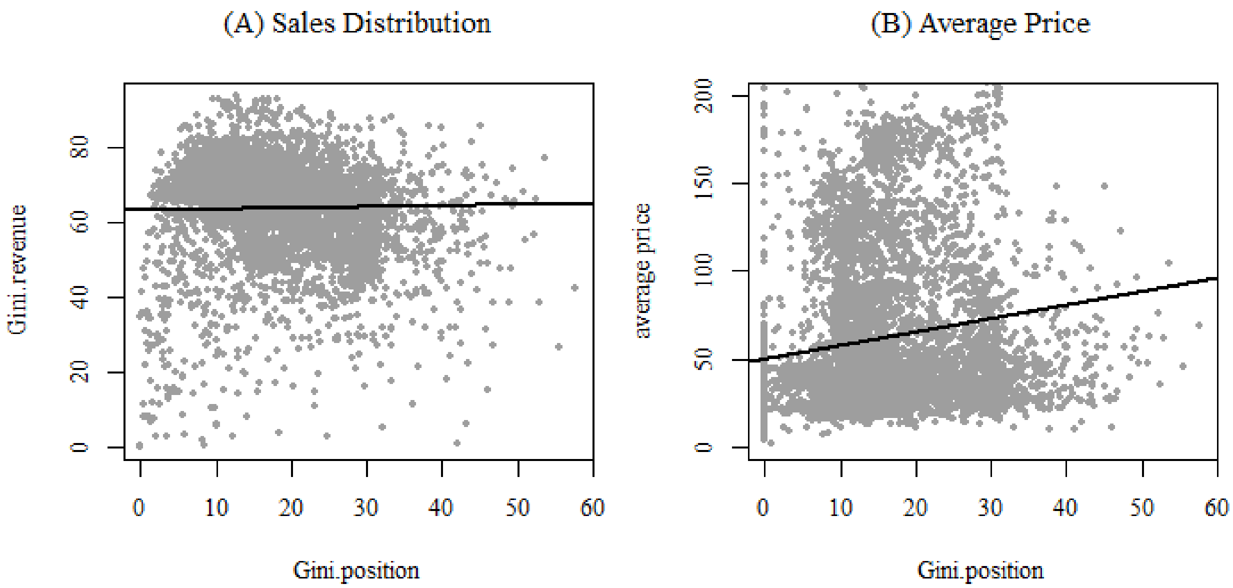


Figure 12: Search Result Distribution versus Sales Distribution and Average Price

4.3 Results

Table 4 reports the results from store-level regressions by using *revenue* as the dependent variable. The main coefficients of interest are *weight.position* and *repeat.listing*. We find that the estimated coefficients for *weight.position* in regressions (2), (4), and (5) suggest that being ranked one position lower leads to more than a CNY20 monthly revenue reduction.[29] The scale is similar to that of Ursu (2018), who finds that the average position effect is USD1.92 on Expedia. This indicates that consumers heavily rely on the default search results when making purchases on food delivery platforms. In regressions (3)-(5), the coefficient estimates for *repeat.listing* are significantly positive. This indicates that chain stores appearing multiple times in the search results tend to earn more revenue. One possible reason is that chain stores are more popular in general.

[29] We assume that the effect is linear. In reality, the difference between position 1 and 10 can be much larger than the difference between position 101 and 110.

Table 4: Store-level Regression Results

	Dependent variable: *revenue*				
	(1)	(2)	(3)	(4)	(5)
weight.position		-24.891^{***}		-24.917^{***}	-25.528^{***}
		(0.617)		(0.617)	(0.621)
repeat.listing			124.992^{**}	152.713^{***}	174.632^{***}
			(50.326)	(49.806)	(57.934)
rating	$5{,}372.117^{***}$	$4{,}940.388^{***}$	$5{,}361.862^{***}$	$4{,}927.406^{***}$	$4{,}946.423^{***}$
	(257.352)	(254.899)	(257.376)	(254.920)	(254.810)
N.rating	1.792^{***}	1.588^{***}	1.788^{***}	1.583^{***}	1.574^{***}
	(0.048)	(0.048)	(0.049)	(0.048)	(0.048)
deliv.min.p	-10.729^{**}	-9.587^{**}	-10.772^{**}	-9.638^{**}	-9.818^{**}
	(4.448)	(4.402)	(4.448)	(4.402)	(4.400)
deliv.fee	-91.671^{***}	-73.462^{***}	-93.592^{***}	-75.790^{***}	-73.679^{***}
	(17.028)	(16.857)	(17.045)	(16.873)	(16.867)
deliv.time	88.089^{***}	110.149^{***}	88.541^{***}	110.723^{***}	106.932^{***}
	(7.824)	(7.762)	(7.826)	(7.764)	(7.785)
activeness	-54.946	394.647^{***}	-51.534	399.287^{***}	426.775^{***}
	(88.484)	(88.269)	(88.491)	(88.277)	(88.294)
Ele.me	$6{,}438.702^{***}$	$7{,}592.732^{***}$	$6{,}417.866^{***}$	$7{,}568.484^{***}$	$7{,}512.145^{***}$
	(298.212)	(296.492)	(298.320)	(296.581)	(296.553)
Meituan	$15{,}974.620^{***}$	$18{,}628.780^{***}$	$15{,}995.400^{***}$	$18{,}656.960^{***}$	$18{,}933.080^{***}$
	(321.912)	(325.282)	(322.010)	(325.394)	(327.155)
N.brand.stores					-1.034
					(0.773)
J					6.284^{***}
					(0.745)
Category FE	Y	Y	Y	Y	Y
Station FE	Y	Y	Y	Y	Y
Observations	77,310	77,310	77,310	77,310	77,310
R^2	0.119	0.137	0.119	0.138	0.138

Note: For all regression results reported in this paper, * indicates significance at 10%; ** indicates significance at 5%; and *** indicates significance at 1%. The dependent variable *weight.position* is the weighted average position of the store. *N.brand.stores.plat* and *N.brand.stores* are the number of stores sharing the same brand on the platform and across three platforms, respectively.

Our main empirical results are based on market-level regressions. In Table 5, regressions (1) and (2) use the *Gini.revenue* in the market as the dependent variable. In regressions (3) and (4), the dependent variable is the market share of the top 20% stores. The coefficients of *Gini.position* are all significantly positive. This indicates that markets with less equal search results tend to have more skewed sales distributions (Figure 12-A). Table 5 also shows that as the market accumulates more sales (*market.y*) and more reviews (*market.N.rating*), the sales concentration decreases. One reason could be the information released through sales and reviews helps consumers learn the quality of small independent restaurants and rely less on brands.

Table 6 uses the market average price (*market.ave.p*) as the dependent variable. The coefficient estimates of *Gini.position* are significantly positive in all specifications. This indicates that markets with less equal search results have higher average prices (Figure 12-B). This is consistent with

Proposition 2.

The coefficients estimates of J are positive in Tables 5 and 6. Thus, having more stores in a market cannot smooth the sales distribution and cause prices to reduce. One explanation is that consumers' consideration sets do not expand when having more stores available in the market. The skewed-distributed store positions in search results reduce the competitive pressure faced by stores. Therefore, to protect consumers, regulators should not only monitor the number of competing sellers but also whether the search algorithm can promote competition by letting consumers easily reach a sufficiently large number of options.

Table 5: Market-level Regression on Inequality of Sales Distribution

Dependent variable	*Gini.revenue*×100		*top-20 share*×100	
	(1)	(2)	(3)	(4)
Gini.position×100	0.493***	0.490***	0.445***	0.442***
	(0.043)	(0.043)	(0.048)	(0.048)
J	0.016***	0.017***	0.013***	0.018***
	(0.003)	(0.002)	(0.003)	(0.002)
activeness	−10.588***	−10.137***	−6.612**	−6.366**
	(2.365)	(2.370)	(2.650)	(2.656)
N.brand.stores≥10	0.048***		0.046***	
	(0.009)		(0.009)	
N.brand.stores.plat≥10		0.080***		0.052***
		(0.010)		(0.010)
market.N.rating	−0.00001***	−0.00001***	−0.00001***	−0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
market.y	−0.00002***	−0.00002***	−0.00002***	−0.00002***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
market.ave.p	0.005	0.006	0.004	0.004
	(0.007)	(0.007)	(0.008)	(0.008)
Ele.me	−7.952***	−7.646***	−10.946***	−10.838***
	(0.463)	(0.472)	(0.499)	(0.509)
Meituan	−12.313***	−12.462***	−10.300***	−10.380***
	(0.846)	(0.847)	(0.904)	(0.906)
Category FE	Y	Y	Y	Y
Station FE	Y	Y	Y	Y
Observations	7,442	7,442	7,442	7,442
R^2	0.361	0.363	0.265	0.265

Note: The dependent variable *Gini.revenue* is the Gini coefficient of revenue of stores operating in the market; *top-20 share* is the share of total revenue of the top 20% largest stores in the market. *N.brand.stores≥10* and *N.brand.stores.plat≥10* count the number of large chain stores (with more than 10 stores overall and on this platform, respectively) operating in the market.

Table 6: Market-level Regression Results on Market Average Price

	Dependent variable: *market.ave.p*			
	(1)	(2)	(3)	(4)
Gini.position×100	0.764***	0.619***	0.705***	0.716***
	(0.107)	(0.148)	(0.159)	(0.160)
J			0.038***	0.045***
			(0.008)	(0.008)
activeness			1.941	0.996
			(7.709)	(7.710)
N.brand.stores≥10			−0.069***	
			(0.023)	
N.brand.stores.plat≥10				−0.143***
				(0.026)
Ele.me		13.955***	10.946***	10.028***
		(0.620)	(0.880)	(0.980)
Meituan		42.825***	44.231***	44.527***
		(3.380)	(3.748)	(3.709)
Category FE	N	Y	Y	Y
Station FE	N	Y	Y	Y
Observations	7,442	7,442	7,442	7,442
R^2	0.010	0.396	0.397	0.397

Note: The dependent variable *market.ave.p* is the weighted average price of all delivery orders across all stores operating in the market.

5 Conclusion

As sales shift from offline to online, platforms obtain the power to influence buyers' search behavior through the design of search algorithms. We develop a novel model that captures how a search algorithm affects buyers' search processes, which further affects market equilibrium and welfare. The model shows that adopting a highly unequal search algorithm causes buyers to obtain more repetitive information and consider fewer options. As a result, sellers can charge higher prices. The interests of buyers can be further jeopardized if the search algorithm promotes low-quality products. Based on data constructed from trial searches, we find empirical evidence that markets with unequal search results tend to have higher average prices.

Search algorithms are important tools by which online platforms exercise their market power. Regulators must exercise due diligence in monitoring search algorithms and implement policies to help buyers obtain a large variety of options. The modeling framework proposed in this paper provides a tractable way of analyzing search algorithms that can potentially be used for welfare analyses, antitrust investigations, and regulation of online platforms.

References

Anderson, S. P. and A. De Palma (2009). Information congestion. *RAND Journal of Economics 40*(4), 688–709.

Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics 37*(3), 668–691.

Armstrong, M., J. Vickers, and J. Zhou (2009). Prominence and consumer search. *RAND Journal of Economics 40*(2), 209–233.

Athey, S. and G. Ellison (2011). Position auctions with consumer search. *Quarterly Journal of Economics 126*(3), 1213–1270.

Backus, M. R., J. U. Podwol, and H. S. Schneider (2014). Search costs and equilibrium price dispersion in auction markets. *European Economic Review 71*, 173–192.

Bar-Isaac, H., G. Caruana, and V. Cuñat (2012). Search, design, and market structure. *American Economic Review 102*(2), 1140–1160.

Blake, T., C. Nosko, and S. Tadelis (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica 83*(1), 155–174.

Brynjolfsson, E., Y. Hu, and D. Simester (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science 57*(8), 1373–1386.

Burdett, K. and K. L. Judd (1983). Equilibrium price dispersion. *Econometrica*, 955–969.

Casner, B. (2020). Seller curation in platforms. *International Journal of Industrial Organization 72*, 102659.

Chang, H., D. S. Evans, S. D. D. Garcia, et al. (2005). The effect of regulatory intervention in two-sided markets: An assessment of interchange-fee capping in Australia. *Review of Network Economics 4*(4), 1–31.

Chen, N. and H.-T. Tsai (2021). Steering via algorithmic recommendations. *Working Paper*.

Chen, Y. and C. He (2011). Paid placement: Advertising and search on the internet. *Economic Journal 121*(556), F309–F328.

Chen, Y. and T. Zhang (2017). Entry and welfare in search markets. *Economic Journal 128*(608), 55–80.

Chen, Y. and T. Zhang (2018). Intermediaries and consumer search. *International Journal of Industrial Organization 57*, 255–277.

De Corniere, A. (2016). Search advertising. *American Economic Journal: Microeconomics 8*(3), 156–88.

De Corniere, A. and G. Taylor (2014). Integration and search engine bias. *RAND Journal of Economics 45*(3), 576–597.

De Corniere, A. and G. Taylor (2019). A model of biased intermediation. *RAND Journal of Economics 50*(4), 854–882.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review 102*(6), 2955–80.

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

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2018). Consumer price search and platform design in internet commerce. *American Economic Review 108*(7), 1820–59.

Dukes, A. and L. Liu (2015). Online shopping intermediaries: The strategic design of search environments. *Management Science 62*(4), 1064–1077.

Eliaz, K. and R. Spiegler (2011a). Consideration sets and competitive marketing. *Review of Economic Studies 78*(1), 235–262.

Eliaz, K. and R. Spiegler (2011b). A simple model of search engine pricing. *Economic Journal 121*(556), F329–F339.

Etro, F. (2021). Platform competition with free entry of sellers. *Working Paper*.

European Commission (2017). Antitrust: Commission fines Google EUR2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service. `ec.europa.eu/commission/presscorner/detail/en/IP_17_1784`.

Federal Trade Commission (2013). FTC consumer protection staff updates agency's guidance to search engine industry on the need to distinguish between advertisements and search results. www.ftc.gov.

Fleder, D. and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science 55*(5), 697–712.

Fleder, D. M. and K. Hosanagar (2007). Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*, pp. 192–199. ACM.

Gaswirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Review of Economics and Statistics 54*, 306–316.

Ghose, A. and S. Yang (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science 55*(10), 1605–1622.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Goldmanis, M., A. Hortaçsu, C. Syverson, and Ö. Emre (2010). E-commerce and the market structure of retail industries. *Economic Journal 120*(545), 651–682.

Hagiu, A. and B. Jullien (2011). Why do intermediaries divert search? *RAND Journal of Economics 42*(2), 337–362.

Hagiu, A., T.-H. Teh, and J. Wright (forthcoming). Should platforms be allowed to sell on their own marketplaces? *RAND Journal of Economics*.

Harstad, R. M., J. H. Kagel, and D. Levin (1990). Equilibrium bid functions for auctions with an uncertain number of bidders. *Economics Letters 33*(1), 35–40.

Hong, H. and M. Shum (2006). Using price distributions to estimate search costs. *RAND Journal of Economics 37*(2), 257–275.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *RAND Journal of Economics 48*(3), 611–646.

Inderst, R. and M. Ottaviani (2012). Competition through commissions and kickbacks. *American Economic Review 102*(2), 780–809.

Jansen, B. J. and M. Resnick (2006). An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce web searching. *Journal of the Association for Information Science and Technology 57*(14), 1949–1961.

Marshall, A. W., I. Olkin, and B. C. Arnold (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer Science & Business Media.

McAfee, R. P. and J. McMillan (1987). Auctions with a stochastic number of bidders. *Journal of Economic Theory 43*(1), 1–19.

Morgan, P. and R. Manning (1985). Optimal search. *Econometrica*, 923–944.

Ricci, F., L. Rokach, and B. Shapira (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pp. 1–35. Springer.

Rochet, J.-C. and J. Tirole (2003). Platform competition in two-sided markets. *Journal of the European Economic Association 1*(4), 990–1029.

Rochet, J.-C. and J. Tirole (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, 645–667.

Schur, I. (1923). Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft 22*(9-20).

Teh, T.-H. (2022). Platform governance. *American Economic Journal: Microeconomics*.

Teh, T.-H. and J. Wright (2020). Intermediation and steering: Competition in prices and commissions. *American Economic Journal: Microeconomics*.

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

Varian, H. (1980). A model of sales. *American Economic Review 70*(4), 651–59.

Wang, C. and J. Wright (2020). Search platforms: Showrooming and price parity clauses. *RAND Journal of Economics 51*(1), 32–58.

Wong, C.-K. and P. Yue (1973). A majorization theorem for the number of distinct outcomes in n independent trials. *Discrete Mathematics 6*(4), 391–398.

Appendix

Proof of Lemma 1. Let $\mathcal{C}(K)$ denote the consideration set with search intensity K. After taking K_1 samples, we have a probability distribution of N. Then, consider taking one more sample. In this sampling, the buyer obtains product j. If $j \in \mathcal{C}(K_1)$, $\mathcal{C}(K_1+1) = \mathcal{C}(K_1)$ and $\Pr(N \le n) = \Pr(|\mathcal{C}(K_1+1)| \le n)$. If $j \notin \mathcal{C}(K_1)$, $\mathcal{C}(K_1+1) = \mathcal{C}(K_1) \bigcup \{j\}$. In this case, $|\mathcal{C}(K_1+1)| = J+1$ and $|\mathcal{C}(K_1+1)| > |\mathcal{C}(K_1)|$.

Because $K_1 < J$, the latter event occurs with positive probability, i.e, $\Pr(j \notin \mathcal{C}(K_1)) > 0$,

$$\begin{aligned}\Pr(|\mathcal{C}(K_1+1)| \le n) &= \Pr(|\mathcal{C}(K_1)| \le n) - \Pr(|\mathcal{C}(K_1+1)| = n+1) \\ &= \Pr(|\mathcal{C}(K_1)| \le n) - \Pr(j \notin \mathcal{C}(K_1)) < \Pr(|\mathcal{C}(K_1)| \le n).\end{aligned}$$

With this inequality, we can easily observe the FSD property in the lemma. □

Proof of Lemma 2. We first introduce the concept of Schur-convex function. A real-valued function $\phi : \mathcal{A} \subset \mathbb{R}^j \to \mathbb{R}$ is considered *Schur-convex*[30] on $\mathcal{A}$ if $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$ implies $\phi(\boldsymbol{\sigma}^1) \le \phi(\boldsymbol{\sigma}^2)$.

By Proposition E.11.b. of Marshall et al. (2011),[31] suppose that an experiment with J possible outcome is repeated K times. The number N of distinct outcomes is a random variable representing the nonzero components of multinomial random vector $\boldsymbol{x}$. Then, $\psi(\boldsymbol{\sigma}) = \Pr(N \le n|\boldsymbol{\sigma})$ is a Schur-convex function of $\boldsymbol{\sigma}$ for all n. By the definition of a Schur-convex function, $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2 \Rightarrow \Pr(N \le n|\boldsymbol{\sigma}^1) \le \Pr(N \le n|\boldsymbol{\sigma}^2) \Leftrightarrow N_1 \succcurlyeq_1 N_2$. □

Proof of Proposition 1. Consider a symmetric and monotone BNE $u = \mu(w)$. Let $\mu^{-1}(\cdot)$ denote the inverse of μ. Provided that all other sellers follow the BNE, the probability of the generic seller j being chosen is

$$\begin{aligned}\Pr\left(u \ge \max_{j' \in \mathcal{C}}\{\mu(w_{j'})\}\right) &= \Pr\left(\mu^{-1}(u) \ge \max_{j' \in \mathcal{C}}\{\mu^{-1}(\mu(w_{j'}))\}\right) \\ &= \Pr\left(\mu^{-1}(u) \ge \max_{j' \in \mathcal{C}}\{w_{j'}\}\right) \\ \text{By (2)} &= \sum_{n=1}^{J}\left\{P_n \times \Pr\left(\mu^{-1}(u) \ge W_{(1:n-1)}\right)\right\} \\ &= \sum_{n=1}^{J}\left\{P_n \times G_{(1:n-1)}\left(\mu^{-1}(u)\right)\right\} \equiv \mathcal{P}\left(\mu^{-1}(u)\right).\end{aligned} \tag{12}$$

In equilibrium, $\mathcal{P}\left(\mu^{-1}(u)\right) = \mathcal{P}(w)$ represents the choice probability of a seller with $w = w^*(\theta)$.

[30] It is named after Schur (1923), who conducts the first systematic study of order-preserving functions for majorization. The term "convex" originates from the property of a convex function. Given a random variable σ that can have different realizations, $\sigma_1, \sigma_2, .., \sigma_n$, for all convex functions $g : \mathbb{R} \to \mathbb{R}$, we have $\sum_{i=1}^n g(\bar{\sigma}) \le \sum_{i=1}^n g(\sigma_i)$.

[31] This proposition is originally derived by Wong and Yue (1973).

A generic seller with w chooses u by solving

$$\max_u (w-u)\mathcal{P}\left(\mu^{-1}(u)\right) = \max_u (w-u) \sum_{n=1}^{J} \left\{ P_n \times G_{(1:n-1)}\left(\mu^{-1}(u)\right) \right\}.$$

This problem is equivalent to a first-price auction with an uncertain number of bidders. We derive its BNE following McAfee and McMillan (1987) and Harstad et al. (1990). The first-order condition yields

$$\frac{d\mathcal{P}\left(\mu^{-1}(u)\right)}{du}(w-u) - \mathcal{P}\left(\mu^{-1}(u)\right) = 0,$$

which is an ordinary differential equation. In equilibrium, $\mu^{-1}(u) = w$, $u = \mu(w)$; thus,

$$\frac{\mathcal{P}'(w)}{\mu'(w)}(w - \mu(w)) - \mathcal{P}(w) = 0.$$

which implies $\mathcal{P}(w)\mu'(w) + \mathcal{P}'(w)\mu(w) = w\mathcal{P}'(w)$, and thus,

$$\frac{d}{dw}\left[\mathcal{P}(w)\mu(w)\right] = w\mathcal{P}'(w).$$

Given the boundary condition $\mu(\underline{w}) = 0$ and integrating on both sides,

$$\mathcal{P}(w)\mu(w) = \int_{\underline{w}}^{w} \omega\mathcal{P}'(\omega)\,d\omega = w\mathcal{P}(w) - \int_{\underline{w}}^{w} \mathcal{P}(\omega)\,d\omega.$$

The solution is

$$\mu(w) = \frac{\int_{\underline{w}}^{w} \omega\mathcal{P}'(\omega)\,d\omega}{\mathcal{P}(w)} = w - \frac{\int_{\underline{w}}^{w} \mathcal{P}(\omega)\,d\omega}{\mathcal{P}(w)}, \quad w \in [\underline{w}, \overline{w}].$$

$\mu(w)$ is obviously increasing in w, so it is a proper monotone BNE.

With fixed quality at $q^*(\theta)$, to offer a product with utility $\mu(w^*(\theta))$, the seller sets prices at $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$. □

Proof of Proposition 2. Let N_1 and N_2 be the random consideration set sizes under search algorithm $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$, respectively. By Lemma 2, because $\boldsymbol{\sigma}^1 \prec \boldsymbol{\sigma}^2$, $N_1 \succcurlyeq_1 N_2$. We can express $\mu(w)$ as a conditional expectation taken upon the random consideration set size N:

$$\begin{aligned}
\mu(w) &= \frac{\int_{\underline{w}}^{w} \omega\mathcal{P}'(\omega)\,d\omega}{\mathcal{P}(w)} = \frac{\int_{\underline{w}}^{w} \omega \frac{d\left[\sum_{n=1}^{J}\left\{P_n \times G_{(1:n-1)}(w)\right\}\right]}{d\omega} d\omega}{\mathcal{P}(w)} \\
&= \frac{\int_{\underline{w}}^{w} \omega \left[\sum_{n=1}^{J}\left\{P_n \times g_{(1:n-1)}(w)\right\}\right] d\omega}{\mathcal{P}(w)} = \sum_{n=1}^{J} P_n \times \frac{\int_{\underline{w}}^{w} \omega g_{(1:n-1)}(w)\,dw}{\mathcal{P}(w)} \\
&= \sum_{n=1}^{J} P_n \times E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right] = E_N\left\{E\left[W_{(1:N-1)} | W_{(1:N-1)} < w\right]\right\}.
\end{aligned}$$

$E\left[W_{(1:N-1)}|W_{(1:N-1)} < w\right]$ is an conditional expectation of $W_{(1:N-1)}$. By the property of order statistics, the maximum among a larger sample must have a higher expectation, so $E\left[W_{(1:N-1)}|W_{(1:N-1)} < w\right]$ is an increasing function of N. By the property of stochastic order, because $N_1 \succcurlyeq_1 N_2$, we obtain

$$\mu(w^*(\theta);\boldsymbol{\sigma}^1) = E_{N_1}\left\{E\left[W_{(1:N_1-1)}|W_{(1:N-1)} < w\right]\right\} \geq E_{N_2}\left\{E\left[W_{(1:N_2-1)}|W_{(1:N_2-1)} < w\right]\right\} = \mu(w^*(\theta);\boldsymbol{\sigma}^2).$$

Intuitively, the seller "bids" more aggressively in an "auction" with more rival sellers.

It follows that $p^*(\theta;\boldsymbol{\sigma}^1) \leq p^*(\theta;\boldsymbol{\sigma}^2)$ because $p^*(\theta) = v(q^*(\theta)) - \mu(w^*(\theta))$ and $q^*(\theta)$ does not depend on N. □

Proof of Proposition 3. We first show part (i). By the property of stochastic order, if $N_2 \succcurlyeq_1 N_1$, $E\left[\varphi(N_2)\right] \geq E\left[\varphi((N_1)\right]$ for any (weakly) increasing function φ.

(1) Social welfare is, $SW_N = I \cdot E_N\left[E\left[W_{(1:N)}\right]\right]$. Define

$$\xi(N) = E\left[W_{(1:N)}\right] = \int_{\underline{w}}^{\overline{w}} w dG_{(1:N)}(w).$$

By the property of order statistics, the maximum among a larger sample must have a higher expectation, so $\xi(\cdot)$ is increasing in N. Because $N_1 \succcurlyeq_1 N_2$, we obtain $E_{N_1}\left[\xi(N_1)\right] \geq E_{N_2}\left[\xi(N_2)\right]$. Therefore,

$$SW_{N_1} = I \cdot E_{N_1}\left[\xi(N_1)\right] \geq I \cdot E_{N_2}\left[\xi(N_2)\right] = SW_{N_2}.$$

Thus, when N increases in the sense of FSD, social welfare SW increases.

(2) Buyer-side utility is $U_N = I \cdot E_N\left[E\left[\mu(W_{(1:N)})\right]\right]$. Define $\zeta(N) = E\left[\mu(W_{(1:N)})\right]$. Because both $\mu(\cdot)$ and $W_{(1:N)}$ are increasing in N, $\zeta(\cdot)$ is increasing in N. Hence, $N_1 \succcurlyeq_1 N_2$ implies $E_{N_1}\left[\zeta(N_1)\right] \geq E_{N_2}\left[\zeta(N_2)\right]$, which further implies

$$U_{N_1} = I \cdot E_{N_1}\left[\zeta(N_1)\right] \geq I \cdot E_{N_2}\left[\zeta(N_2)\right] = U_{N_2}.$$

(3) Seller-side profit is $\mathit{\Pi}_N = I \cdot E_N\left[E\left[W_{(1:N)} - \mu(W_{(1:N)})\right]\right]$. Define $\varsigma(N) = E\left[W_{(1:N)} - \mu(W_{(1:N)})\right]$. Recall that $\mu(w) = w - \int_{\underline{w}}^{w} \mathcal{P}(\omega)\, d\omega / \mathcal{P}(w)$, so $\varsigma(N) = E\left[\int_{\underline{w}}^{W_{(1:N)}} \mathcal{P}(\omega)\, d\omega / \mathcal{P}\left(W_{(1:N)}\right)\right]$ is the bid shading in a first-price auction with a uncertain number of bidders. By Theorems 1 and 2 from Harstad et al. (1990), $\varsigma(N)$ decreases in N.

Given that $N_1 \succcurlyeq_1 N_2$, $E_{N_1}\left[\varsigma(N_1)\right] \leq E_{N_2}\left[\varsigma(N_2)\right]$, which implies

$$\mathit{\Pi}_{N_1} = I \cdot E_{N_1}\left[\varsigma(N_1)\right] \leq I \cdot E_{N_2}\left[\varsigma(N_2)\right] = \mathit{\Pi}_{N_2}.$$

This completes the proof of part (i). Along with Lemmas 2 and 3, we can directly obtain parts (ii) and (iii). □