Data, Competition and Digital Platforms*

Dirk Bergemann† Alessandro Bonatti‡

December 15, 2021

Preliminary and Incomplete

Abstract

We propose a comprehensive model of digital commerce in which data and heterogeneity are defining features. A digital platform enables matching of consumers and advertisers online. Each consumer has heterogenous preferences across heterogeneous advertisers. The advertisers can tailor their products to the preferences of the consumer. Each consumer can access the product line of each advertiser online or offlline.

The platform can improve the quality of the match formation through its past and present data collection. The platform monetizes its services by digital advertising that are sold in (generalized) second price auctions.

We derive the equilibrium surplus sharing between consumer, advertisers and the platform in the digital economy. We evaluate how different data governance rules affect the creation and sharing of the social surplus. We contrast the unrestricted use of data with contextual and cohort restricted use of data. We show that privacy enhancing data governance rules, such as federated learning of cohorts can increase the competition among the advertisers and lead to welfare gains, and in particular to private gains to the platform and the consumers.

KEYWORDS: Data, Privacy, Data Governance, Digital Advertising, Competition, Digital Platforms, Digital Intermediaries, Personal Data, Matching, Price Discrimination

JEL CLASSIFICATION: D18, D44, D82, D83.

*We acknowledge financial support through NSF Grant SES-1948692, the Omidyar Network and the Sloan Foundation. We thank audiences at the FTC Microeconomic Conference 2021 and the Asia-Pacific Industrial Organization Conference 2021 and in particular, our discussant Vasiliki Skreta for helpful comments and suggestions.

†Department of Economics, Yale University, New Haven, CT 06511, dirk.bergemann@yale.edu

‡Sloan School of Management, MIT, Cambridge, MA 02142, bonatti@mit.edu

1 Introduction

The role of data in shaping competition in online markets is a critical issue for both economics and policy. In particular, the ways in which consumer data can be used to create and extract surplus online reflects a rapidly evolving landscape, due to several factors: (i) regulation such as the GDPR and CPRA intends to assign formal control rights over data to consumers; (ii) technological innovation allows the seamless integration of multiple data sources (apps, devices, IoT) and enables ad delivery across different media (mobile, apps, video, home); (iii) tech companies develop privacy strategies and suggest their own privacy initiatives (e.g., Google's FLoC, or Apple's no-tracking preferences in the most recent iOS).

Digital platforms such as Amazon, Google, and Facebook operate as matching engines that connect viewers and advertisers. The value and the quality of the matches they create depends to a large extent on the quantity and quality of the data available to these engines. The data is typically collected from the individual viewers and, to a lesser extent, from the advertisers. The most common format for the digital platform to monetize its service is through auctions for digital advertising. As a result, the data allows better pairing of viewers and advertisers, but also informs the choice of the products that the advertiser offers to each consumer.

In this paper, we develop a model of an intermediated online market and trace out how the presence of a digital platform acting as a data intermediary changes the distribution of the surplus among the relevant market participants. Our approach consists of providing a tractable and flexible framework to study the Internet economy where (a) platforms monetize data through ad auctions, and (b) different privacy regimes can be compared. The key tradeoff in our model is that superior information on the platform improves match quality, but also creates the potential for price discrimination through product steering. Consistent with evidence from online marketplaces, consumers who are perceived to be of high value do not receive higher prices for the same goods; instead, they receive personalized offers to buy higher-quality and higher-priced goods. In this context, the presence of an off-platform sales channel serves as a restraints for sellers to practice this form of price discrimination on the platform.

We consider an auction model for online advertising under alternative data governance regimes. Amonopolist platform auctions sponsored links to advertisers who alsof second-degree price discrimination on two distinct but related, where consume sell their products on their own websites. The winning bidder's problem is that the consumer can move from the platform to individual websites. Thus, we combine elements of nonlinear pricing and

market segmentation, where the consumer's choice of sales channel limits the scope for price discrimination through product steering.

In particular, each consumer has a vector of preferences over the characteristics of the sellers' products. The available data can then identify the most valuable match in terms of the advertiser and the specific product of the advertiser. The digital platform creates value by identifying valuable matches in terms of advertisers and in terms of products. The quality of the data therefore influences the quality of the match and the ability of the producer to receive a share of the surplus from the consumer.

Our setting therefore naturally lends itself two analyze two significant issues in the economics of digital platforms: data-governance regimes, and (self-)preferential treatment. First, we ask how data governance in the data mediated markets influences the creation and distribution of social surplus. We discuss three important data policies: contextual data, cohort-based data (e.g., as the outcome of federated learning), and separation of matching and data platform. Second (in ongoing work), we ask how preferential treatment by the platform changes the efficiency and surplus generated on the platform. For example, Google may enable to steer the viewer to its own websites rather than competing services. Similarly, Amazon may rank its own private label ahead of the independent sellers.

Related Literature In our earlier work Bergemann, Bonatti, and Gan (2022), we highlighted the relevance of this question for theory, as we showed that the simple assignment of ownership rights to consumers over their data is insufficient to bring about the efficient use of their information. The wedge between equilibrium and efficient uses of information is a data externality that consumers impose on each other. Unlike other famous examples, this does not have a clear sign.

The connection to data externalities papers (including Choi, Jeon, and Kim (2019) and Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2021)) is that the growth of a platform's database (through the participation of other consumers) influences the ability to match products to tastes, but also affects each consumer's outside option—whatever the firms offer to consumers off the platform. This theme is also present in the platform model of Kirpalani and Philippon (2021). In this work, we trace out the implications of these (data and network) externalities for product line design, under alternative privacy policies.

The mechanisms at work in our paper are also related to a growing literature on showrooming, product lines, and multiple sales channels. Prominent contributions on these topics include Wang and Wright (2020), Miklós-Thal and Shaffer (2021), Bar-Isaac and Shelegia (2020), and Idem (2020).

At a broad level, the paper relates to information structures in advertising auctions, e.g., Bergemann, Brooks, and Morris (2021), and to the literature on nonlinear pricing, market segmentation, and competition, including Bergemann, Brooks, and Morris (2015), Elliott, Galeotti, and Koh (2020), and Bonatti (2011).

Finally, even though our current analysis does not explicitly discuss self-preferencing by a monopoly platform, our framework can easily speak to this question and relate to a number of recent papers, including Hagiu, Teh, and Wright (2020), Padilla, Perkins, and Piccolo (2020), Lee (2021), Muir (2021), Lam (2021), Musolff (2021).

2 Model

We introduce the payoff of consumers and advertiser. Then we describe how data and additional information enables improved matches. Finally we describe the interaction between online and offline markets.

2.1 Payoffs

Consider J sellers of differentiated product lines. Each seller j can produce a good of quality q_j at a cost

$$c(q_j) = q_j^2/2.$$

There is a unit mass of consumers i with unit demand. The consumer's type is her willingness pay for the product of each firm j :

$$\theta_i = (\theta_{i1}, ..., \theta_{ij}, ..., \theta_{iJ}) \in \mathbb{R}_+^J.$$

In particular, consumer i has value

$$u_i(\theta_i, q_j) = \theta_{ij} q_j$$

for a product of quality q_j produced by firm j.

Neither the consumers nor the producers initially know the consumer's valuation. Consumers' valuations θ_{ij} are i.i.d. across i, j with distribution F and

$$\mathbb{E}_F[\theta_{ij}] = \mu.$$

In addition, each consumer i observes a vector-valued signal s_i about her type θ_i. We will work directly with the posterior expectation

$$m_{ij} = \mathbb{E}[\theta_{ij} \mid s_i].$$

In particular, consumers' beliefs m_{ij} are also assumed i.i.d. with distribution G. Clearly, F is a mean-preserving spread of G. We will further assume that F and G have the same support.[1]

Finally, we assume that the platform has access to extensive data that allows it to observe each consumer's $\theta_i \in \mathbb{R}^J$ perfectly. Importantly, the goods being sold are not experience or inspection goods. Thus, learning θ (by consumers or producers) requires access to the platform's data.

2.2 Matching with Product Steering

A fraction $1 - \lambda$ consumers (exogenously, for now) buy off the platform, e.g. from the merchant's own websites or physical stores. Consumers who buy off the platform face search costs (beyond the first firm) as in the Diamond model. As a result, a consumer i with beliefs m_i visits seller $j^* = \arg\max_j m_{ij}$. Each seller offline only knows the prior distribution of the consumer's type (and of its highest order statistic), and will therefore need to elicit the consumer's beliefs through a nonlinear quality-price schedule

$$p_j^0(m), q_j^0(m). \tag{1}$$

In contrast, a fraction $\lambda \in [0, 1]$ of all consumers types uses a monopolist digital platform to find a seller. The platform provides *organic* information that communicates θ_i to each individual consumer. In addition, the platform runs an auction to match consumers and sellers. More specifically, the platform sells a single advertising slot for each consumer in a second-price auction. In this auction, the consumer's type θ is a *targeting category,* i.e., the producer has access to the same information as the consumer. In particular, all sellers can condition their bids on the consumer's type $b_j(\theta)$, and the winning bidder can offer a tailored product $q_j(\theta)$ at price $p_j(\theta)$ to each online consumer. Thus, the platform creates the opportunity for price discrimination through *product steering.*

[1] We could equivalently assume that the platform owns strictly more informative, yet imperfect signals θ of the consumer's unknown underlying true type. The baseline analysis remains unchanged, as the platform owns all the information available in the market. In extensions, we could model an increase in the precision of the platform's signal through a further mean-preserving spread in the distribution F.

The on- and off-platform markets are connected. In particular, if a consumer visits the platform, she learns her entire type $\theta \in \mathbb{R}^J$, and can then easily find any producer off the platform. In particular, the consumer will be able to buy from the offline schedule (1) posted by the winning bidder j, or from any other bidder.

Thus, in an efficient equilibrium of the advertising auction, the platform provides a match-quality advantage based on superior information. This advantage is profitable for sellers, who bid for the right to make an offer to the consumer. However, this advantage the platform gives the producer relative to the off-platform matches disappears once the price is quoted. In other words, the producer's ability to product steer and price discriminate on the platform is limited by the presence of the off-platform channel.

Figure 1 describes possible search patterns for the online vs. offline consumers: consumer i who shops offline may visit firm $j = 1$ that has the highest expected value m_{i1}. If consumer i shopped online instead, she would learn that $\theta_{i2} > \theta_{i1}$ and would then either accept firm 2's offer, or shop offline from firm 2's menu.

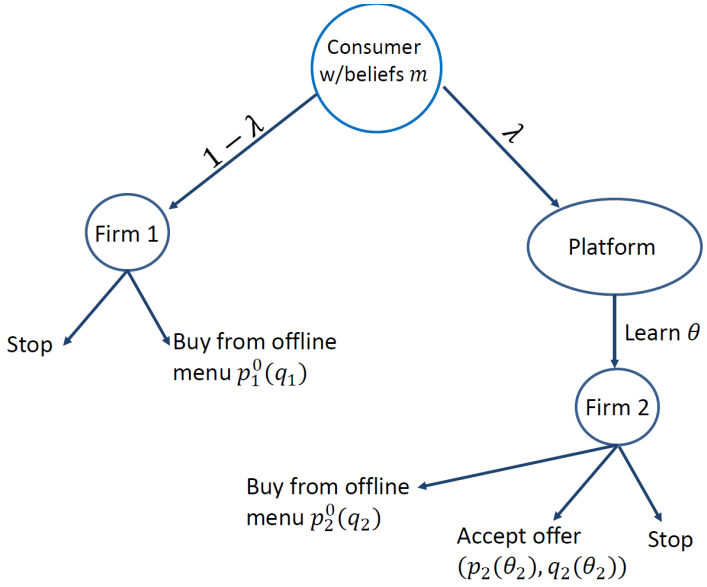


Figure 1: Sample search patterns for consumer i online and offline. Accessing the platform allows the consumer to buy from $\arg\max_j \theta_{ij}$ instead of $\arg\max_j m_{ij}$

2.3 Offline and Online Markets

We consider a sequential game that captures the greater flexibility that automated bidding and algorithmic pricing offer relative to off-platform sales. Therefore, the timing is as follows:

1. All sellers set prices and qualities for all offline consumers, $p_j^0(m)$ and $q_j^0(m)$ as in (1).

2. A consumer visits the platform, which reveals her type θ to the consumer and to all sellers. Sellers place bids $b_j(\theta)$ in a second-price auction for a single sponsored link to be shown to consumer θ.

3. Upon winning the auction, seller j offers a single product of quality $q_j(\theta)$ at price $p_j(\theta)$ to the consumer.

4. After receiving this offer, consumer θ can buy on the platform from seller j or off platform from any seller.

3 Equilibrium Product Lines

We begin to describe our results by providing an overview of the economic forces at play. On platform, each seller knows the consumer's type. While the seller must still induce the consumer not to shop elsewhere (i.e., not to practice "showrooming"), the on-platform interaction takes place under symmetric information.

Off the platform, each firm j faces the consumers that value its product the most based on their information m, but must elicit their willingness to pay. In particular, seller j could offer the optimal Mussa and Rosen (1978) tariff $(q^0(m), p^0(m))$ for the distribution $G^J(m)$ of the highest-order statistic. However, the resulting information rent $U^0(m)$ provided to the consumers has an additional shadow cost: it makes buying offline more attractive for the online consumers.

3.1 Showrooming

More formally, the ad auction creates surplus by matching each i to the best j. As we have symmetric information after the auction, the winning bidder extracts a lot of the surplus. This occurs not through personalized price discrimination, but through **product steering**. At the same time, the winning bidder faces the "showrooming constraint"

$$\theta \cdot q_j(\theta) - p_j(\theta) \geq \max_m \left[\theta \cdot q_j^0(m) - p_j^0(m)\right] \text{ for all } \theta. \tag{2}$$

This is a necessary condition for making any sales online. (The consumer could also buy offline from other sellers $-j$ but that will not happen in equilibrium.)

However, if firm j offers an incentive compatible menu off the platform, each on-platform consumer would also report their type truthfully when showrooming. Thus, consumer θ

effectively chooses between two products, and the showrooming constraint (2) reduces to

$$U_j(\theta) := \theta q_j(\theta) - p_j(\theta)) \geq \theta q_j^0(\theta) - p_j^0(\theta)) =: U_j^0(\theta). \tag{3}$$

The showrooming constraint prevents the winning bidder from extracting the entire surplus from online consumers. Because the online transaction takes place under symmetric information, it is immediate to show that the high bidder offers each online consumer θ a single product at the socially efficient quality level

$$q^*(\theta) = \theta,$$

as well as a discount that satisfies (3) with equality.

Proposition 1 (Profits in the Online Auction)
Suppose seller j offers the offline menu (p_j^0, q_j^0) with the associated rent function U_j^0. Then seller j's profit in the auction for online type θ is given by:

$$\pi(\theta, U_j^0) = \theta^2/2 - U_j^0(\theta).$$

Sellers bid for sponsored links knowing what their profit margins will be on each type. In the second-price advertising auction, seller j therefore bids

$$b_j(\theta) = \pi(\theta, U_j^0).$$

In other words, Lemma 1 shows how information rents offline determine market shares, profits, and auction revenues online.

3.2 Symmetric Equilibrium

We now characterize a symmetric equilibrium in nonlinear prices offline and trace out its implications for online quantities and prices. We adopt the more compact notation (q, U) for an incentive compatible menu (so that prices are given by $p = \theta q - U$). Now suppose every firm $k \neq j$ offers the menu $(\hat{q}, \hat{U})$, and consider firm j's best reply. If firm j offers (q_j^0, U_j^0), we can compute its market share of online consumers of each type $\theta_{ij} = \theta$. In particular, define cutoff $x(\theta)$ by

$$\theta^2/2 - U_j^0 = x^2/2 - \hat{U}(x),$$

and note that the probability firm j wins the auction for a consumer with valuation θ for its product is given by $F^{J-1}(x(\theta))$. The payment for consumer θ at the auction is then determined by the highest component θ_{ik} with $k \neq j$.

Combining the offline and online portions of its profits, seller j's best response problem to the menus offered by its competitors is given by

$$\begin{aligned} \max_{q^0, U^0} (1-\lambda) \int_0^1 \left(\theta q^0(\theta) - (q^0(\theta))^2/2 - U^0(\theta)\right) G^{J-1}(\theta) \mathrm{d}G(\theta) \\ + \lambda \int_0^1 \int_0^{x(\theta)} \left(\theta^2/2 - U^0(\theta) - \pi(s, \hat{U})\right) \mathrm{d}F^{J-1}(s) \mathrm{d}F(\theta). \end{aligned} \tag{4}$$

$$\begin{aligned} \text{s.t. (IR)} \quad & U^0(\theta) \geq 0 \\ \text{(IC)} \quad & U^{0\prime}(\theta) = q^0(\theta) \end{aligned}$$

Maximizing (4) over rent and quality functions U^0 and q^0 using standard optimal-control methods, and imposing symmetry at the optimum, we obtain the following characterization of the optimal menus.

Proposition 2 (Equilibrium Menus)
There exists a unique symmetric equilibrium. The equilibrium quality levels on and off the platform are given by

$$\begin{aligned} q^*(\theta) &= \theta, \\ q^0(\theta) &= \max\left\{0, \theta - \frac{1 - \lambda F^J(\theta) - (1-\lambda) G^J(\theta)}{(1-\lambda) J G^{J-1}(\theta) g(\theta)}\right\}. \end{aligned} \tag{5}$$

Furthermore,

$$U^*(\theta) = U^0(\theta) = \int_0^\theta q^0(m) dm.$$

The equilibrium quality provision offline has several intuitive properties. First of all, the efficient quality is sold to each consumer i on platform, on the basis of her favorite firm, *i.e.*, $\max_j\{\theta_{ij}\}$. Conversely, matching is inefficient off the platform because it is based on insufficient information, i.e., on beliefs m instead of fundamentals θ.

Moreover, the consumer's private information off the platform requires the firms to provide information rents. The rents left to each type m are, as usual, increasing in the quality provided to all lower types. If there were no online channel, the producer would set q_0 at the Mussa and Rosen (1978) level. However, the possibility of showrooming introduces an opportunity cost of off-platform sales: by leaving positive rents off the platform, each seller

must also provide rents on the platform:

$$U^*(\theta) = U^0(\theta) > 0 \text{ iff } q^0(\theta) > 0.$$

3.3 Online vs. Offline Quality

As a result of the shadow cost of showrooming, the offline quality schedule q^0 is further distorted downward, relative to both the efficient level and the monopoly schedule under second-degree price discrimination. In particular, we can rewrite (5) in Proposition 2 as

$$q^0(\theta) = \theta - \frac{1 - G^J(\theta)}{JG^{J-1}(\theta)g(\theta)} - \frac{\lambda}{1-\lambda}\frac{1 - F^J(\theta)}{JG^{J-1}(\theta)g(\theta)}, \tag{6}$$

where the first two terms identify the optimal (offline only) quality level for the type distribution G^J, which is the distribution of the highest order statistic out of J (belief) variables. The last term represents the additional discounts that offline quality provision imposes on online sales. Figure 2 illustrates the optimal quality schedule relative to the efficiency and Mussa and Rosen (1978) benchmarks.

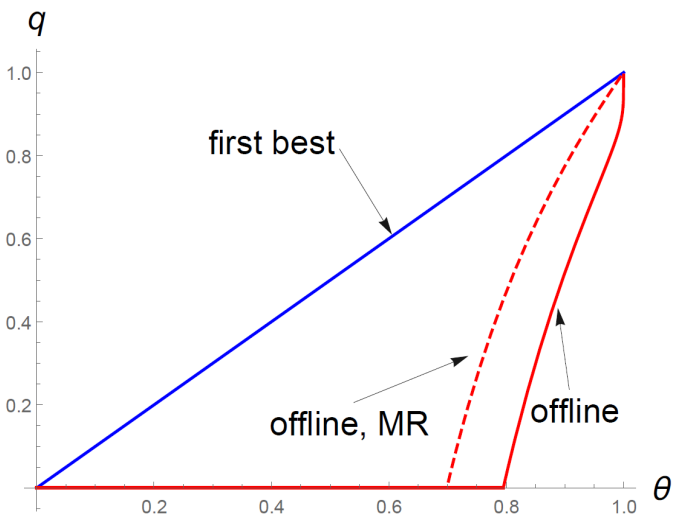


Figure 2: Equilibrium and Benchmark Quality Levels, $\lambda = 1/2, J = 5, G(m) = m, F(\theta) = Beta(\theta, 1/4, 1/4)$

The formulation of the optimal offline menu (6) allows us to establish several intuitive properties of the equilibrium. Figure 3 compares quality and price offline and online as a function of θ for the case of uniformly distributed m and Beta-distributed θ, and Corollary 1 formalizes this result.

Corollary 1 (On- vs. Off-Platform Prices)

1. *For each type θ, prices and quality levels are higher online than offline*

$$p(\theta) \geq p^0(\theta), \quad q(\theta) \geq q^0(\theta).$$

2. *The nonlinear pricing schedule online $p(q)$ lies weakly below the nonlinear schedule offline $p^0(q)$.*

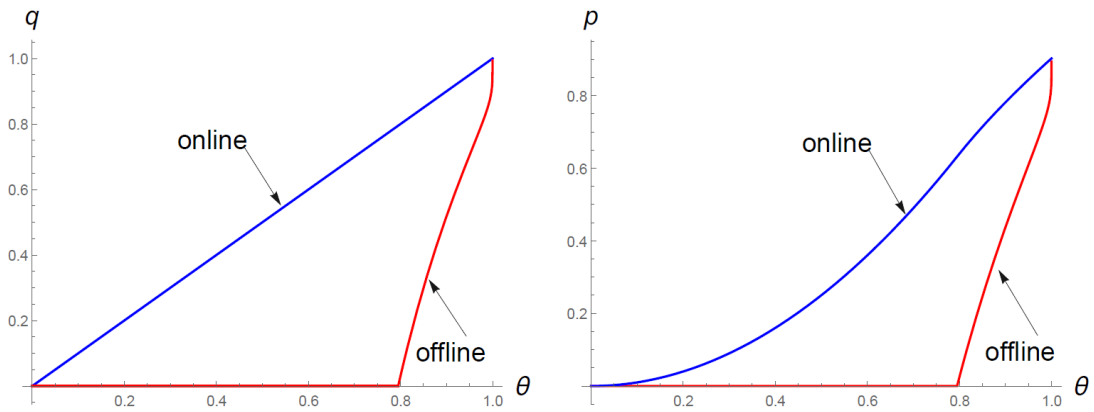


Figure 3: Equilibrium Quality and Price Levels, $\lambda = 1/2, J = 5, G(m) = m, F(\theta) = Beta(\theta, 1/4, 1/4)$

However, while each type receives a better product at a higher price on the platform relative to offline, each seller is forced to introduce "online only" discounts by the threat of showrooming. Figure 4 shows the nonlinear pricing schedules for the same parameter values.

We conclude this section by stating some intuitive comparative statics on the effects of a larger platform on online and offline quality and prices.

Proposition 3 (Platform Size)

1. *$q^0(\theta)$ is decreasing in the fraction of online consumers λ for all θ.*

2. *There exists $\bar{\lambda} < 1$ such that $q^0(\theta) \equiv 0$ for all $\lambda \geq \bar{\lambda}$.*

3. *Expected consumer surplus online and offline is decreasing in λ.*



Figure 4: Nonlinear tariffs: every offline variety q is sold at a lower price online.

4 Welfare Analysis

In order to close the model, we can derive the equilibrium total (ex ante) surplus levels. In particular, total surplus is given by

$$TS = \lambda \mathbb{E}_{F^J}[\theta^2/2] + (1-\lambda)\mathbb{E}_{G^J}[\theta q^0(\theta) - q^0(\theta)^2/2].$$

Consumer surplus is given by

$$CS = \lambda \mathbb{E}_{F^J}[U^0(\theta)] + (1-\lambda)\mathbb{E}_{G^J}[U^0(\theta)].$$

Finally, the platform's revenue is given by

$$\lambda \int_0^1 b(\theta) \mathrm{d}[JF(\theta)^{J-1}(1-F(\theta))],$$

which allows to efficiently compute expected seller profits by difference.

4.1 Consumer Surplus

The effect of platform size on consumer surplus (Proposition 3) is clear: as the platform grows large, the offline market is eventually not served at all, and consumers make zero surplus both offline and online. Conversely, it is reasonable to conjecture that consumer surplus is not monotone in the number of sellers participating in the platform. When J increases, each

consumer on the platform enjoys higher expected value from the better matches, but also a lower quality provision offline. This is because every producer distorts its offline quality q_0 schedule knowing that, conditional on a consumer visiting its store, her type is more likely to be high. Figure 5 (left) illustrates this intuition.

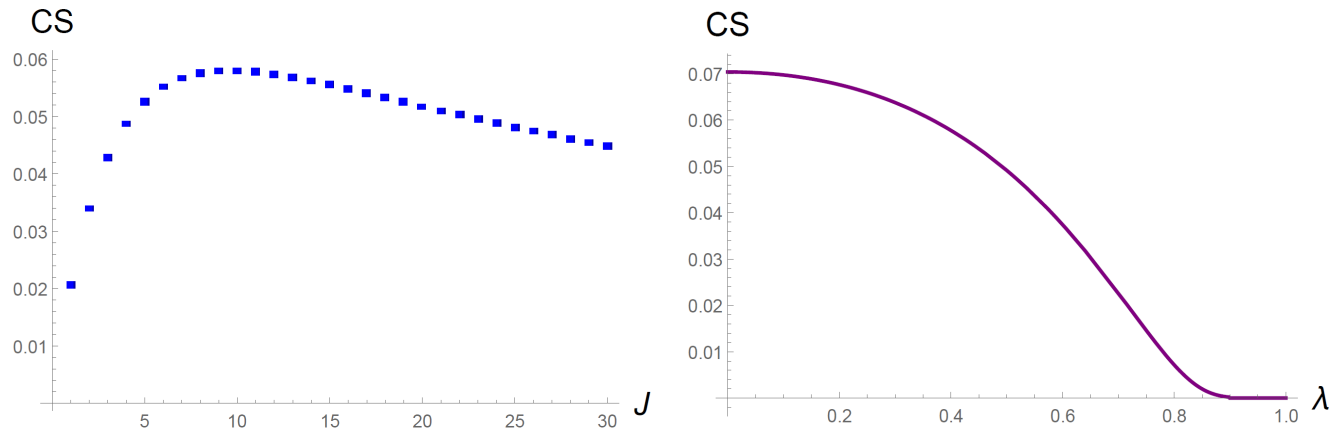


Figure 5: Consumer surplus: higher J improves match quality but reduces information rents (left); a higher λ reduces quality offline until eventual shutdown (right).

4.2 Producer Surplus

Producer surplus reflects a similar tradeoff: a higher J improves the average match quality (both online and offline), but also intensifies competition on the advertising market. Conversely, a higher λ reduces offline quality (and hence prices), in a way that more than offsets the shift of consumers toward the more lucrative online market. When λ is sufficiently high, however, the offline market is inactive ($q^0 \equiv 0$) and producer surplus increases in λ.

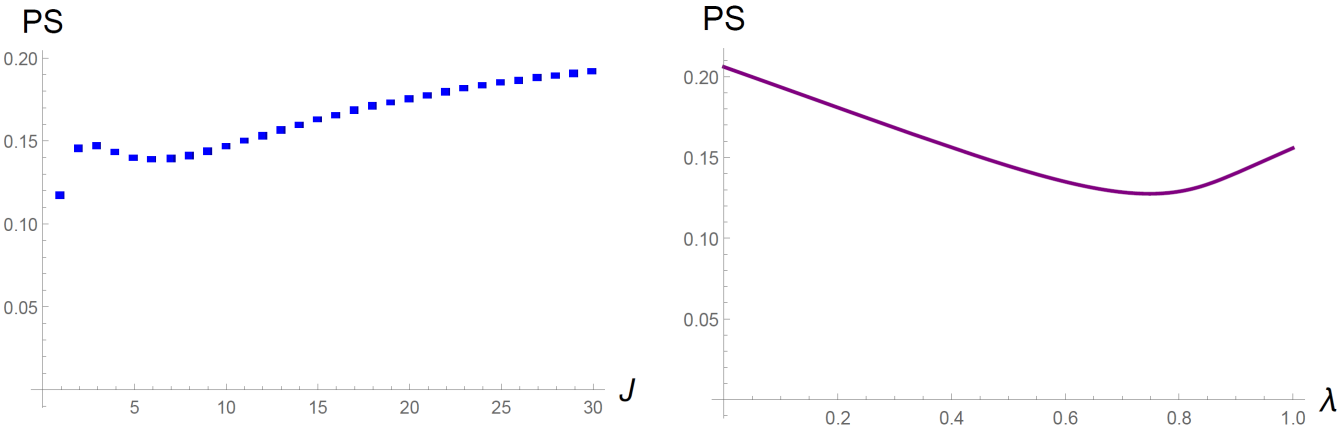


Figure 6: Producer surplus as a function of J and λ.

4.3 Platform Revenue

As we have shown, for sufficiently high λ, all firms set $q^0 = 0$ for all m. For any higher $\lambda > \bar{\lambda}$, the platform captures a constant fraction of the total surplus generated. This fraction is a function of the number of sellers J. As J grows without bound, that fraction goes to 1, i.e., the platform appropriates all of the (efficient) total surplus. Figure 7 illustrates.

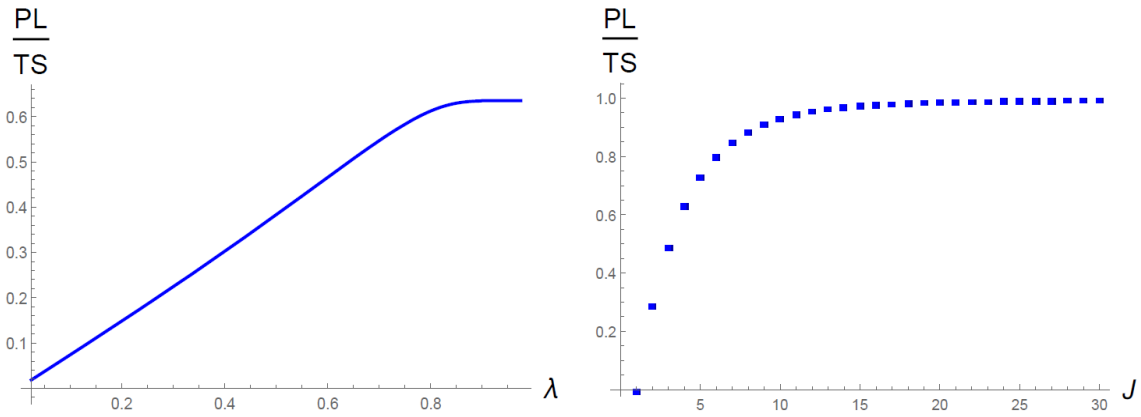


Figure 7: Expected platform revenue, for $J = 4$ (left) and $\lambda = 1$ (right).

5 Endogenous Participation

So far, consumers' search behavior (on or off the platform) was assumed exogenous. It is immediate to see, however, that consumers are weakly better off on the platform. While for each type we have

$$U^0(\theta) = U(\theta),$$

we also know that $F \succ_{mps} G$, which implies $\mathbb{E}_{F^J}\theta > \mathbb{E}_{G^J}\theta$ and hence

$$\mathbb{E}_{F^J}U > \mathbb{E}_{G^J}U,$$

because incentive compatibility requires the function U to be increasing and convex.

In order to relax this assumption, we introduce a small cost of using the platform. This additive cost ε represents for example the privacy cost of running an online search. It is not related to the purchase decision, but rather to the "footprints" that consumers leave online.

At this stage, we assume $\varepsilon \sim H(\varepsilon)$ is independent of the consumer's type and beliefs, and that it is realized before consumers learn their types. For simplicity, we further assume that consumers make their on- vs. off-platform search choice before learning their m, so that

they evaluate the surplus on the two markets from an ex ante perspective, but knowing their ε.

Figure 8 (left) shows expected consumer surplus online and offline as a function of platform size λ. Figure 8 (right) shows the gap in expected consumer utility as a function of λ, and the critical cost level

$$\varepsilon = H^{-1}(\lambda)$$

that yields every platform size.

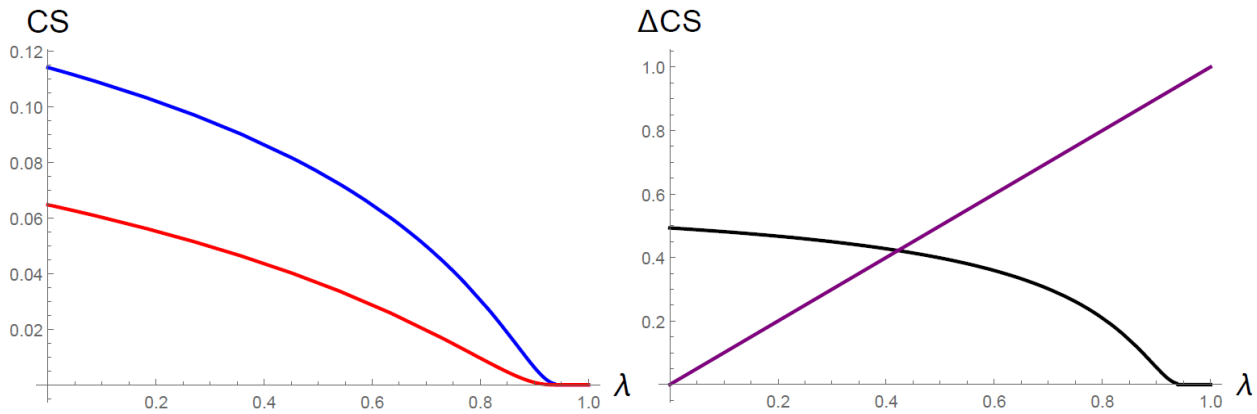


Figure 8: Left: On-off platform gap in consumer surplus decreases with λ. Right: $\varepsilon \sim [0, 1/50]$. Platform participation decisions are strategic substitutes. Unique equilibrium.

Finally, Figure 9 shows that, for small supports of ε, the equilibrium platform size λ^* is increasing in the number of sellers J.

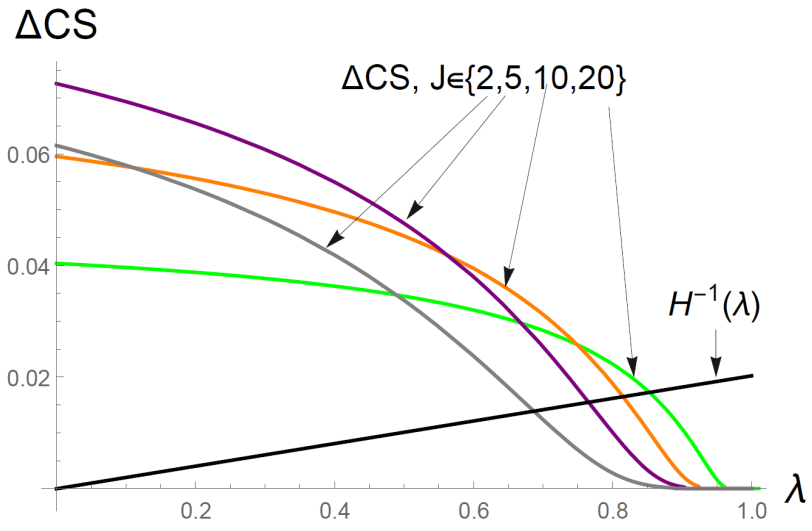


Figure 9: $\varepsilon \sim [0, 1/50]$. Equilibrium λ^* increases with J.

6 Data Governance

So far, we have imposed no restrictions on platform's ability to share data θ with bidders. In this section, we compare two alternative data-governance regimes: (a) the "contextual ads" regime, where the platform cannot gather any additional information on the consumer; and (b) the "cohort-based" ads, where the platform can share any information with the consumers but only coarse information with the bidders.

6.1 Information Design in Auction

6.1.1 Contextual Ads

If the platform has no superior information, it must target ads based on input from the consumer only. Nonetheless, the platform still holds considerable power, because it can auction off the exclusive right to sell a product to the consumer under symmetric information. As earlier, the presence of an offline channel disciplines the sellers' ability to extract the entire willingness to pay m of each consumer, and consequently limits the platform's revenues.

The comparison between the full model and the contextual model of ads then reduces to the question of whether a less informative distribution (G) is preferable by the auctioneer to the true distribution of types (F) . Under the contextual model, in fact, the sellers' bids can only be measurable with respect to m, not θ. As a consequence, allowing targeted ads under G generates worse matches and lower surplus, but reduces the perceived differentiation in the J sellers' products, which intensifies competition at the auction.

Figure 10 confirms the Levin and Milgrom (2010) "conflation" intuition, whereby using inferior information and generating a lower total surplus is revenue-superior when there are only a few bidders.

6.1.2 Cohort-Based Ads

Now suppose information about each individual θ_i is shared with the online consumer i but not with the sellers bidders. In contrast, the platform announces the ranking of consumer valuations θ_{ij} to all the bidder. Thus, the platform creates a discrete set of "auction types"– one for each *cohort* of consumers, and each consumer within a cohort ranks the J sellers in the same way. This governance regime is reminiscent of Google's FLoC proposal to replace third-party cookies.

Under this regime, the efficient matching of firms to consumers is still possible (if the auction outcome is efficient). However, online consumers still have private information after

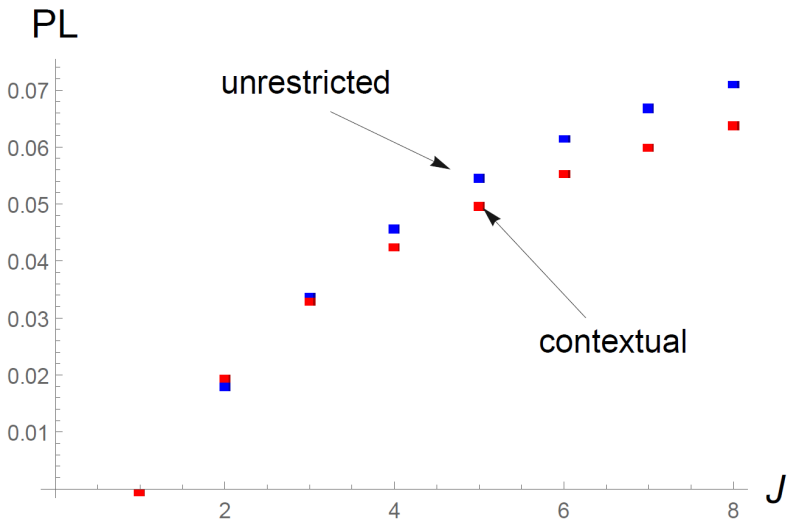


Figure 10: Contextual vs. Unrestricted Targeting

the auction. Unlike in the baseline case, with cohort-based the winning seller only knows the distribution of the consumer's type based on the order statistics implied by her cohort. As a result, the winning seller faces a screening problem with type-dependent participation constraints. In equilibrium, each seller offers a menu of products, whereby the $i \longrightarrow j$ matching is efficient, but qualities online are distorted downward. A priori, this has an ambiguous effect on surplus and platform revenue.

We first characterize the symmetric equilibrium menus under cohort-based ads, then discuss the role of market thickness. In particular, we show that if the distribution of the highest θ dominates that of the highest m in likelihood ratio, then each seller gives up on market segmentation. Indeed, each firm offers the same quality and price to each consumer type, both on and off the platform.

Proposition 4 (Cohort-Based Menus)
Assume $F^J \succ_{lr} G^J$ over all θ for which virtual values are positive under both distributions. In the unique symmetric equilibrium, each firm offers quality levels

$$q^0(\theta) = q^*(\theta) = \max\left\{0, \theta - \frac{1 - \lambda F^J(\theta) - (1-\lambda)G^J(\theta)}{\lambda J F^{J-1}(\theta) f(\theta) + (1-\lambda) J G^{J-1}(\theta) g(\theta)}\right\}.$$

In particular, q^* is the Mussa and Rosen (1978) quality level for mixture with weights $(\lambda, 1-\lambda)$ of the distributions of the highest order statistics of θ and m, respectively. Thus, cohort-based ads yield higher quality provision offline, but lower quality online, relative to the baseline model with full disclosure θ.

Corollary 2 (Quality and Consumer Surplus)

1. *Quality provision under cohort-based ads satisfies*

$$q^*(\theta) \in \left[q_F^0(\theta), \theta\right]$$

2. *Consumer surplus is higher under cohort based ads for all types:*

$$U^*(\theta) \geq U_F^0(\theta).$$

A critical implication of Proposition 4 is that all consumers are now better off as a consequence of the greater quality provision offline. Indeed, total surplus can be higher, too. Figure 11 traces out the implications of this data governance regime for platform revenues as a function of J.

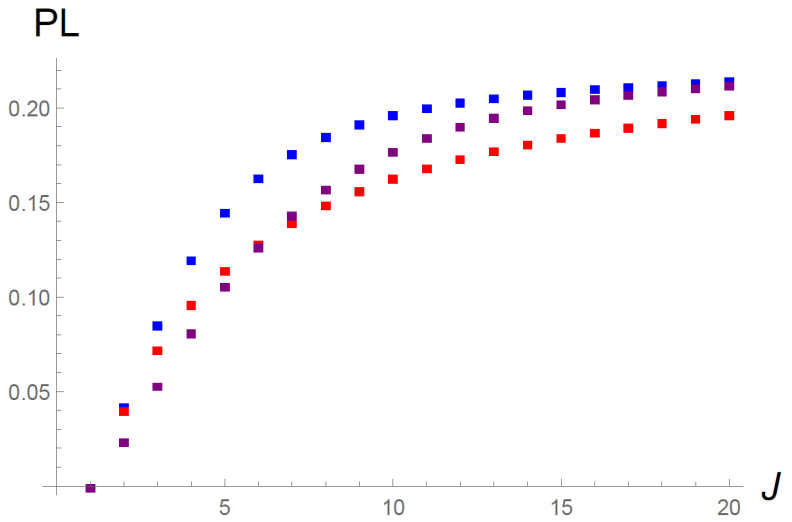


Figure 11: Platform revenue under cohort-based model (purple) dominates contextual for large J, approaches unrestricted model.

7 Conclusion

We have developed a flexible framework of product line design and pricing where each consumer can choose to buy on or off the platform. The two channels are linked by the showrooming mechanism—if the consumer sees an online ad by a producer, she can easily find that producer off the platform too. Thus, the match-quality advantage the platform gives the producer relative to the off-platform match disappears once the price is quoted. As a result, the producer's ability to price discriminate on the platform is limited by the presence of the off-platform channel.

We have also shown that the growth of a platform's database (through the participation of a larger number of consumers) influences each consumer's outside option and can reduce surplus for all online and offline consumers alike.

Our model is stylized and simplified along many dimensions, some of which are readily amenable to extensions. For example, differentiated products are also heterogeneous in their offline vs. online presence. In some cases, very little happens off platform (e.g., niche products sold only on Amazon can be modeled as $\lambda = 1$). In other cases, organic results are right below sponsored links and the on- and off-platform channels interact more heavily.

Overall, our paper emphasizes the need for a more complete picture of the role of data in managing competition in the Internet economy. First, product design and price decisions interact with data governance modes (e.g., with the rules by which a platform can share its data). Second, external data providers (as yet unmodeled) create value on and off the platform by providing better matching, more precise steering, and higher bids. All of these forces have subtle effects on consumer surplus, which we explore in ongoing work.

8 Appendix

The appendix collects the remaining proofs for the results presented in the main body of the text.

Proof of Proposition 2. Consider seller j's best response problem

$$\begin{aligned}\max_{q^0,U^0}(1-\lambda)\int_0^1\left(\theta q^0(\theta)-(q^0(\theta))^2/2-U^0(\theta)\right)G^{J-1}(\theta)\mathrm{d}G(\theta)\\+\lambda\int_0^1\int_0^{x(\theta)}\left(\theta^2/2-U^0(\theta)-\pi(s,\hat{U})\right)\mathrm{d}F^{J-1}(s)\mathrm{d}F(\theta),\end{aligned}$$

where

$$x(\theta):\theta^2/2-U^0(\theta)=x^2/2-\hat{U}(x).$$

The necessary pointwise conditions for q^0 and U^0 can be obtained from the control problem with Hamiltonian

$$H=(1-\lambda)g\left(\theta q^0-\left(q^0\right)^2/2-U^0\right)G^{J-1}+\lambda f\int_0^{x(\theta)}\left(\theta^2/2-U^0-\pi(s,\hat{U})\right)\mathrm{d}F^{J-1}(s)+\gamma q_0.$$

At a symmetric equilibrium, the necessary conditions are given by

$$\begin{aligned}(1-\lambda)G^{J-1}g\left(\theta-q^0\right)+\gamma&=0,\\(1-\lambda)G^{J-1}g+\lambda F^{J-1}f&=\dot{\gamma},\\\gamma(1)&=0.\end{aligned}$$

(Note that the integrand in the second term of the Hamiltonian is nil for $\theta=x(\theta)$ and hence the effect of U^0 on profits via market shares x is second-order.)

We then obtain

$$\begin{aligned}\gamma&=\frac{1}{J}\left((1-\lambda)G^J+\lambda F-1\right),\\q^0&=\theta-\frac{1-(1-\lambda)G^J-\lambda F}{(1-\lambda)G^{J-1}g}.\end{aligned}$$

This ends the proof. ■

Proof of Proposition 4. We construct an equilibrium where each firm j sets their offline menu to maximize profits given that it expects to win the online auctions for the types θ that rank j highest. Therefore, consider the joint optimization problem over menus (q,U)

and (q^0, U^0) when facing distributions F^J and G^J, respectively, under the showrooming constraint. Firm j solves:

$$\begin{aligned} & \max_{q,q^0,U,U^0} \left[\lambda \int \left(\theta q - c\left(q\right) - U\right) dF^J\left(\theta\right) + \left(1-\lambda\right) \int \left(\theta q^0 - c\left(q^0\right) - U^0\right) dG^J\left(\theta\right)\right] \\ \text{s.t.} \quad & \dot{U} = q, \ \dot{U}^0 = q^0 \\ & U \geq U^0 \geq 0. \end{aligned}$$

We now show that if F^J likelihood-ratio dominates G^J, then the solution to the above nonlinear pricing problem is given by

$$q\left(\theta\right) = q^0\left(\theta\right) = \theta - \frac{1 - \left(1-\lambda\right) G^J - \lambda F}{J\lambda F^{J-1} f + J\left(1-\lambda\right) G^{J-1} g}. \tag{7}$$

In this case, we verify that the solution satisfies the optimality conditions (which are sufficient because the problem is linear in q, concave U, and additively separable in these two variables). In particular, the Hamiltonian is given by

$$\begin{aligned} H &= \lambda\left(\theta q - c\left(q\right) - U\right) F^{J-1} f + \left(1-\lambda\right)\left(\theta q^0 - c\left(q^0\right) - U^0\right) G^{J-1} g \\ &\quad + \gamma q + \gamma^0 q^0 + \bar{\gamma}\left(U - U^0\right). \end{aligned}$$

The optimality conditions for this problem are

$$\begin{aligned} 0 &= \left(\theta - q\right) \lambda F^{J-1} f + \gamma \\ 0 &= \left(\theta - q^0\right)\left(1-\lambda\right) G^{J-1} g + \gamma^0 \\ \dot{\gamma} &= \lambda F^{J-1} f - \bar{\gamma} \\ \dot{\gamma}^0 &= \left(1-\lambda\right) G^{J-1} g + \bar{\gamma} \\ \bar{\gamma} &\geq 0 \\ 0 &= \bar{\gamma}\left(U - U^0\right). \end{aligned}$$

When $q = q^0$ as in (7), we obtain

$$\begin{aligned} \gamma &= -\frac{\lambda J F^{J-1} f}{\left(1-\lambda\right) J G^{J-1} g + \lambda J F^J f}\left(1 - \left(1-\lambda\right) G^J - \lambda F^J\right) \\ \gamma_2 &= -\frac{\left(1-\lambda\right) J G^{J-1} g}{\left(1-\lambda\right) J G^{J-1} g + \lambda J F^J f}\left(1 - \left(1-\lambda\right) G^J - \lambda F^J\right) \end{aligned}$$

and differentiating both with respect to θ, we obtain

$$\bar{\gamma} = \frac{\lambda(1-\lambda)\left(1-(1-\lambda)G^J-\lambda F\right)}{\left((1-\lambda)JG^{J-1}g+\lambda JF^{J-1}f\right)^2} J\left(\frac{\mathrm{d}F^{J-1}f}{\mathrm{d}\theta}G^J - \frac{\mathrm{d}G^{J-1}g}{\mathrm{d}\theta}F^J\right),$$

which is positive if and only if dF^J/dG^J is increasing in θ, which means F^J likelihood-ratio dominates G^J. Finally, note that under the likelihood ratio condition, the monopoly quality schedule for distribution F^J lies weakly below the schedule for G^J.

Finally, we sketch the argument why no firm wishes to deviate from the menu (7), and bid in such a way to win more auctions than its competitors. In order to win an auction for which θ_j is not the highest value, firm j would need to bid more than its competitors, while satisfying its *competitors'* no showrooming constraint. This is because consumer i would learn that there is another firm with $\theta_{ik} > \theta_{ij}$, and could potentially buy from k. This means firm j must outbid firm k when facing a worse distribution (i.e., not the highest order statistic). But firm k is already bidding its optimal profits under (7), which by definition are the highest achievable profits that leave rent $U = U^0$ to the consumer. ■

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2021): "Too Much Data: Prices and Inefficiencies in Data Markets," *American Economic Journal: Microeconomics*, p. forthcoming.

Bar-Isaac, H., and S. Shelegia (2020): "Search, Showrooming, and Retailer Variety," Discussion paper, CEPR Discussion Paper No. DP15448.

Bergemann, D., A. Bonatti, and T. Gan (2022): "The Economics of Social Data," *RAND Journal of Economics*, forthcoming.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

——— (2021): "Search, Information and Prices," *Journal of Political Economy*, 129, 2275–2319.

Bonatti, A. (2011): "Brand-Specific Tastes for Quality," *International Journal of Industrial Organization*, 29, 562–575.

Choi, J., D. Jeon, and B. Kim (2019): "Privacy and Personal Data Collection with Information Externalities," *Journal of Public Economics*, 173, 113–124.

Elliott, M., A. Galeotti, and A. Koh (2020): "Market Segmentation through Information," Discussion paper, Cambdrige University.

Hagiu, A., T.-H. Teh, and J. Wright (2020): "Should Amazon be allowed to sell on its own marketplace," Discussion paper, Boston Universtity.

Idem, B. (2020): "Optimal Marketplace Design," Discussion paper, Pennsylvania State University.

Kirpalani, R., and T. Philippon (2021): "Data Sharing and Market Power with Two-Sided Platforms," Discussion Paper 28023, NBER.

Lam, H. T. (2021): "Platform Search Design and Market Power," Discussion paper, Northwestern University.

Lee, C. (2021): "Optimal Recommender System Design," Discussion paper, University of Pennsylvania.

Levin, J., and P. Milgrom (2010): "Online Advertising: Heterogeneity and Conflation in Market Design," *American Economic Review, Papers and Proceedings*.

Miklós-Thal, J., and G. Shaffer (2021): "Input price discrimination by resale market," *RAND Journal of Economics*.

Muir, E. (2021): "Contracting and vertical control by a dominant platform," Discussion paper, Stanford University.

Musolff, L. (2021): "Entry into Two-Sided Markets Shaped By Platform-Guided Search," Discussion paper, Princeton University.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Padilla, J., J. Perkins, and S. Piccolo (2020): "Self-preferencing in markets with vertically-integrated gatekeeper platforms," *Available at SSRN 3701250*.

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