# Data, Competition, and Digital Platforms*

Dirk Bergemann† Alessandro Bonatti‡

March 28, 2022

Preliminary and Incomplete—Please Do Not Cite

**Abstract**

We propose a comprehensive model of digital commerce in which data and heterogeneity are defining features. A digital platform matches consumers and advertisers online. Each consumer has heterogenous preferences for each advertiser's brand, and the advertisers can tailor their product lines to the preferences of the consumer. Each consumer can access each seller's products online or offline. The digital platform can improve the quality of the match through its data collection, and monetizes its data by selling digital advertising space in (generalized) second-price auctions.

We derive the equilibrium surplus sharing between consumer, advertisers and the digital platform. We evaluate how different data-governance rules affect the creation and distribution of the surplus. We contrast the unrestricted use of data with contextual and cohort-restricted uses of data. We show that privacy-enhancing data-governance rules, such as those corresponding to federated learning, can increase the competition among the advertisers and lead to welfare for the digital platform and for 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 2021 FTC Microeconomic Conference, the 2021 Asia-Pacific Industrial Organization Conference, the Luohan Academy Webinar, and the UniBG Industrial Organization Winter Symposium. We are grateful to our discussant Vasiliki Skreta for helpful comments and suggestions, and to Andrew Koh for stellar research assistance.

†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 Privacy Sandbox, 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 the digital platform monetizes the data 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 a digital platform acting as a data intermediary changes the distribution of the surplus among the 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.

**Modeling Choices** Data and additional information enable improved matches. In particular, information is used to match brands, products, and consumers. Instead of modeling horizontal product customization, we consider quality provision in a vertical setting. In particular, each consumer has heterogeneous preferences over the characteristics of the sellers' products. The available data can then identify the most valuable consumer-advertiser match, and the most valuable product offered by that advertiser. In other words, information about the consumer's type generates value by allowing the platform to show consumers their favorite brand, and by letting the brands themselves target the consumer with an offer from their product line. In turn, consumers expect the brand with a sponsored link to generate the highest value, so they contemplate buying (online or offline) from that brand.

However, each firm also has a pool of loyal consumers, who are imperfectly informed about their type. Because of search costs, these consumers will only look at one brand's offers. The more the firm wants to trade with loyal offline consumers, the less flexibility is has to offer targeted online promotions.

We extend our auction model for online advertising to encompass alternative data governance regimes. A monopolist platform auctions off sponsored links to advertisers who practice second-degree price discrimination on two distinct but related channels. 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.

Our setting 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. As in other data-externalities papers, including Choi, Jeon, and Kim (2019) and Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2021), 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. This theme is also present in the platform model of Kirpalani and Philippon (2021). In our work, we trace out the implications of 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). In particular, Anderson and Bedre-Defolie (2021) introduce the self-preferencing problem by letting the platform choose whether to be hybrid, i.e., to sell its private label products.

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), Kang and Muir (2021), Lam (2021), Musolff (2021). Some of this work is described as "partial mechanism design" or "mechanism design with a competitive fringe," see Philippon and Skreta (2012), Tirole (2012), Calzolari and Denicolo (2015), and Fuchs and Skrzypacz (2015). In these papers, the platform is limited in its ability to monopolize the market as the firms have access to an outside option. Our setting shares some of the same features but in an oligopoly environment where firms compete for heterogeneous consumers. Furthermore, the firms choose their product menus understanding that customers may appear through two different sales channels, online vs. offline, and that they may have distinct information online vs. offline.

# 2 Model

We consider a retail platform that manages both competition and information, and $J$ sellers (brands) that offer 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 with unit demand. The consumer's type is her willingness pay for the product of each firm $j$ :

$$\theta = (\theta_1, ..., \theta_j, ..., \theta_J) \in \mathbb{R}_+^J.$$

In particular, consumer $\theta$ has value

$$u(\theta, q_j) = \theta_j q_j$$

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 $\theta_j$ are i.i.d. across consumers and firms, with marginal distribution $F(\theta_j)$. In addition, each consumer observes a vector-valued signal $s$ about her type $\theta$. We work directly with the posterior expectation

$$m_j = \mathbb{E}[\theta_j \mid s].$$

In particular, consumers' beliefs $m_j$ are also assumed i.i.d. with marginal distribution $G$. Clearly, $F$ is a mean-preserving spread of $G$. Finally, we make the further assumption that $F$ and $G$ have identical supports.[1]

Finally, we assume that the platform has access to extensive data that allows it to observe each consumer's $\theta \in \mathbb{R}^J$ perfectly. Importantly, the goods being sold are not experience or inspection goods. Thus, in order to learn $\theta$, consumers and producers must gain access to the platform's data.

## 2.1 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 $\theta_i$ 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 $\theta$ 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

[1] We could equivalently assume that the platform owns strictly more informative, yet imperfect signals $\theta$ 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 $F$.

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*.

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.

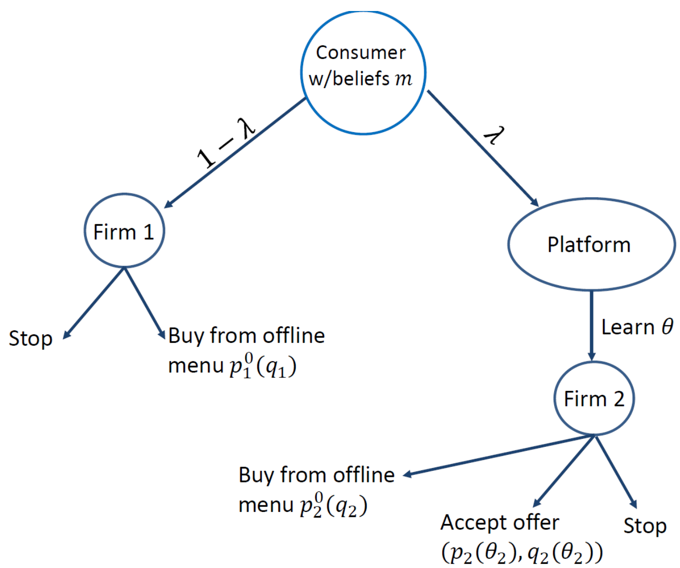


FIGURE 1: Sample search patterns for consumer $i$ online and offline. Accessing the platform allows the consumer to buy from $\arg\max_j \theta_{ij}$ instead of $\arg\max_j m_{ij}$

## 2.2 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 $\theta$ 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 $\theta$.

3. The winner $j$ offers a single product of quality $q_j(\theta)$ at price $p_j(\theta)$ to the consumer.

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

# 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 information is symmetric between the consumer and the winning bidder after the auction, the winning bidder can extract a lot of the surplus. Surplus extraction does not occur through personalized price discrimination, but through product steering. The only limit on surplus extraction by the winning bidder is given by 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 $\theta$

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 $\theta$ 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 $\theta$ is given by:*

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

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 $\theta_j = \theta$. 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 $\theta$ for its product is given by $F^{J-1}(x(\theta))$. The payment for consumer $\theta$ at the auction is then determined by the highest component $\theta_{ik}$ 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

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

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.*, $\max_j\{\theta_j\}$. Conversely, matching is inefficient off the platform because it is based on insufficient information, i.e., on beliefs $m$ instead of fundamentals $\theta$.

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 and 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.

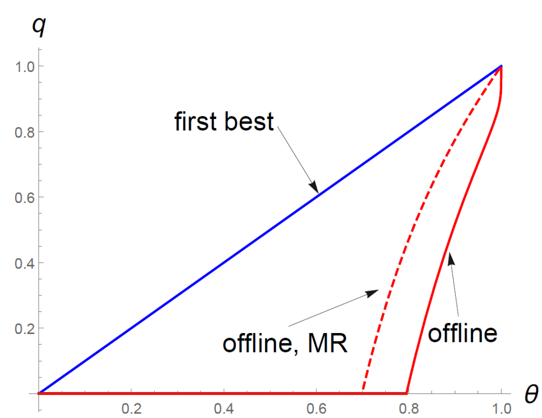


FIGURE 2: Quality Levels, $\lambda = 1/2, J = 5, G(m) = m, F(\theta) = Beta(\theta, 1/4, 1/4)$

The formulation of the optimal offline menu (6) allows us to establish several intuitive properties of the equilibrium. 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. Corollary 1 formalizes these results.

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

1. *For each type $\theta$, prices and quality levels satisfy $p(\theta) \geq p^0(\theta)$ and $q(\theta) \geq q^0(\theta)$.*

2. *For each quality level $q$, the nonlinear pricing schedules satisfy $p(q) \leq p^0(q)$.*

Figure 3 shows the nonlinear pricing schedules for the parameter values above.

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

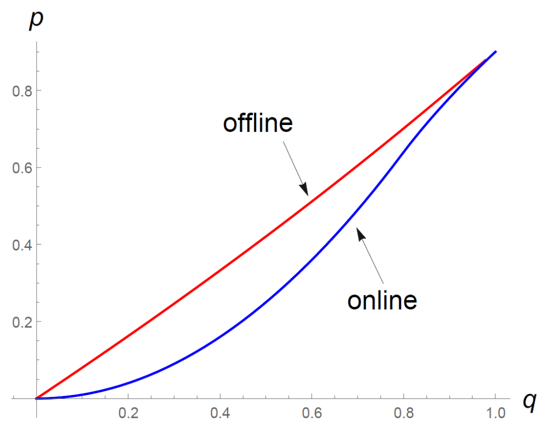


FIGURE 3: Nonlinear tariffs: every offline variety $q$ is sold at a lower price online.

**Proposition 3 (Platform Size)**

1. *$q^0(\theta)$ is decreasing in the fraction of online consumers $\lambda$ for all $\theta$.*
2. *There exists $\bar{\lambda} < 1$ such that $q^0(\theta) \equiv 0$ for all $\lambda \geq \bar{\lambda}$.*

An immediate consequence of this result is that expected consumer surplus online and offline is decreasing in $\lambda$. As the number of firms changes, however, the response of the equilibrium offline quality schedule is more subtle. Figure 4 illustrates.

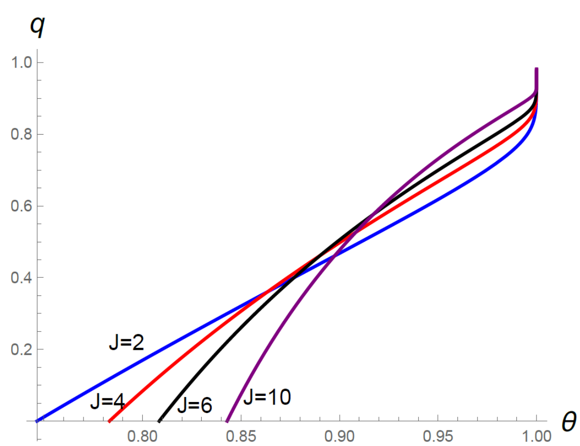


FIGURE 4: Offline Quality Schedule, $\lambda = 1/2, G(m) = m, F(\theta) = Beta(\theta, 1/6, 1/6)$.

## 3.4 Welfare Effects

An immediate consequence of Proposition 2 is 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. Aggregate consumer surplus is given by

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

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

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

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 the total producers' surplus by difference.

We now examine the impact of the size of the platform $\lambda$ and of the number of bidders $J$ on all parties' surplus levels.

### 3.4.1 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.

### 3.4.2 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 $\lambda$ reduces offline quality (and hence prices), in a way that more than offsets

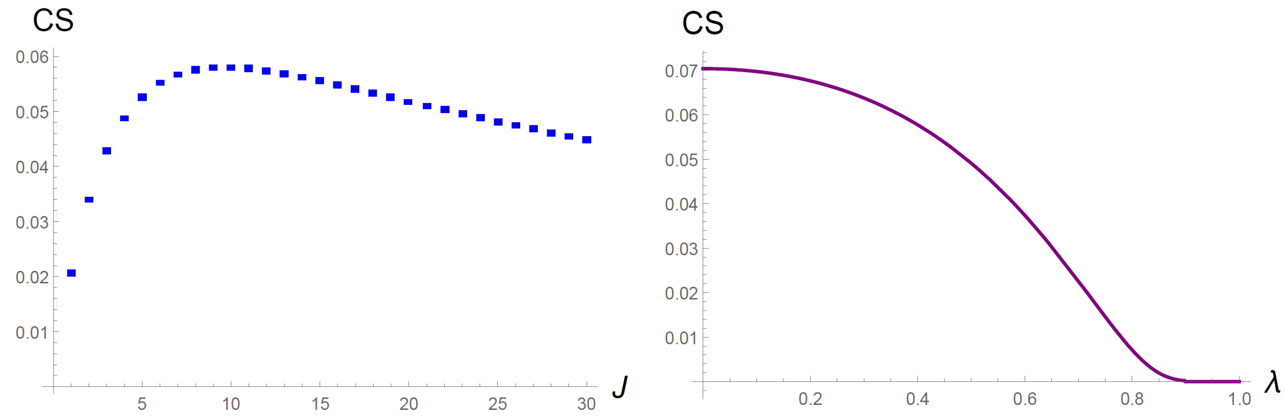


FIGURE 5: Consumer surplus: higher $J$ improves match quality but reduces information rents (left); a higher $\lambda$ reduces quality offline until eventual shutdown (right).

the shift of consumers toward the more lucrative online market. When $\lambda$ is sufficiently high, however, the offline market is inactive ($q^0 \equiv 0$) and producer surplus increases in $\lambda$.

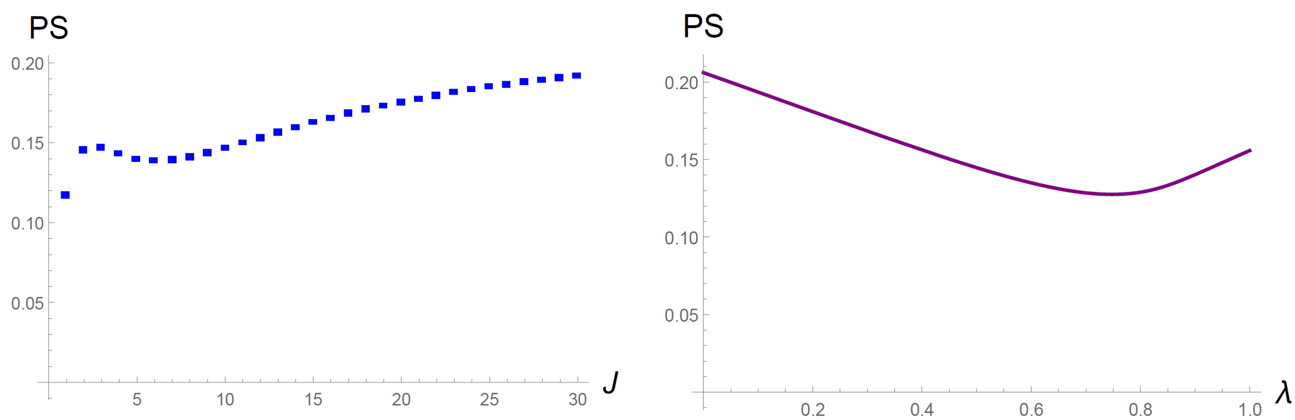


FIGURE 6: Producer surplus as a function of $J$ and $\lambda$.

### 3.4.3 Total Surplus

As consumers move to the platform ($\lambda$ increases), the average realized match value between consumers and producers increases, because more consumers receive the efficient quality level online. At the same time, offline quality decrease. Similar effects occur when more sellers join the platform. Figure 7 shows that total surplus increases both in $\lambda$ and in $J$. The picture is slightly misleading, however, because $TS/\lambda$ decreases, i.e., the total welfare gains are less than proportional to the platform's growth.

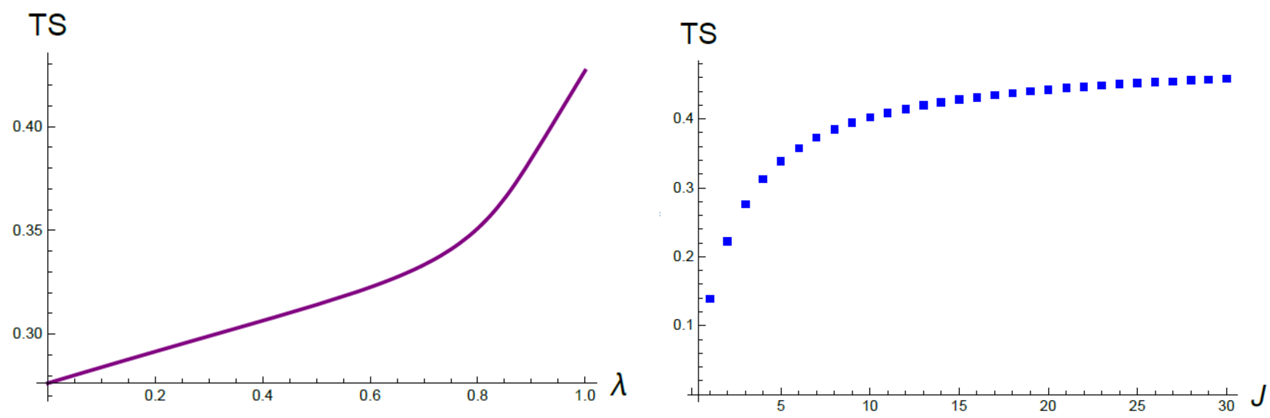


FIGURE 7: Total surplus, for $J = 4$ (left) and $\lambda = 1$ (right).

### 3.4.4 Platform Revenue

As we have shown, for sufficiently high $\lambda$, 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 8 illustrates.

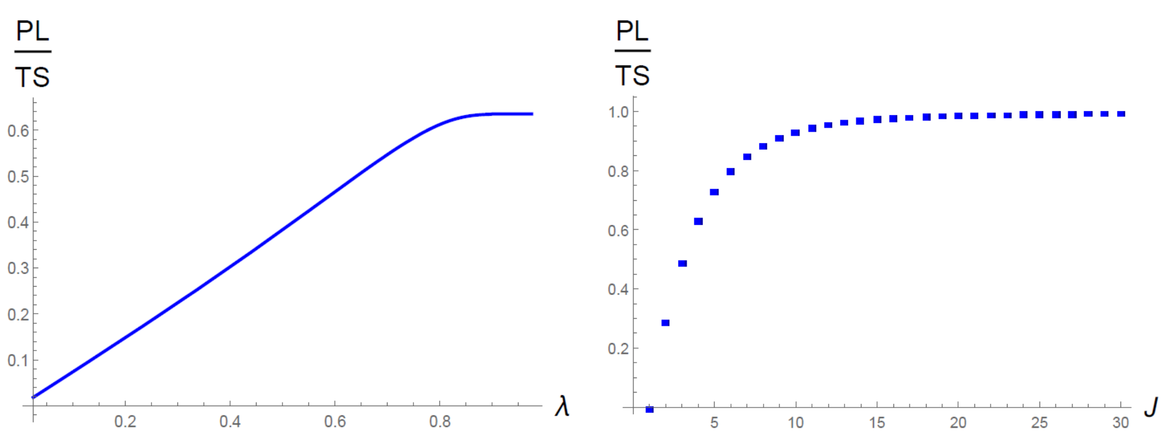


FIGURE 8: Platform revenue as a share of total surplus, $J = 4$ (left) and $\lambda = 1$ (right).

# 4 Information Design

So far, we have imposed no restrictions on platform's ability to share data $\theta$ with bidders. In practice, the amount of data sharing can be limited both by regulation and by the platform's design choices. In this section, we compare two alternative data-governance regimes: (a) the "contextual ads" regime, and (b) the "cohort-based" ads. In the former regime, information is symmetric: the platform limits the amount of information it allows consumers to reveal, and shares all such information with bidders. In the latter information is asymmetric: the platform collects partial information from the consumers but shares only coarse information with the bidders.

## 4.1 Contextual Ads

We now study how the platform can limit the amount of information shared with consumers and bidders online, while maintaining the symmetry of this information. In particular, recall that the consumer's true valuations are distributed according to $F$. We now let the platform design a signal that induces a distribution of posteriors $\tilde{F}$ for both the consumer and the bidders, where

$$F \succ_{mps} \tilde{F} \succ_{mps} G.$$

In a special case of this model, the platform acquires no superior information, and allows bidders to target ads based on input from the consumer only, i.e., $\tilde{F} = G$. Even in this case, the platform holds considerable power, because it can auction 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 equilibrium quality provision is again given by Proposition 2, where we replace the distribution of fundamentals $F$ with the distribution $\tilde{F}$ chosen by the platform. We describe the comparative statics of the equilibrium schedule in the following proposition, where we compare two different choices of $\tilde{F}$, one more informative than the other.

**Proposition 4 (Contextual Ads)**
*Suppose $F_1 \succ F_2$ in the rotation order, and that $F_1 \leq F_2$ for all $\theta$ such that $q^0(\theta; F_2) > 0$. Then the offline quality under the more informative distribution is lower than under the less informative distribution,*

$$q^0(\theta; F_1) \leq q^0(\theta; F_2),$$

*with strict inequality whenever $q^0(\theta; F_2) \in (0, 1)$.*

The welfare effects of information on the platform's reduces to the question of whether a less informative distribution $\tilde{F}$ is preferable 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 $\theta$. 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. The main tension, in terms of total surplus, is that contextual ads are less efficient due to the probability of a mismatch, but they induce smaller quality distortions offline. We now examine the consequences of information precision on consumer and producer surplus, separately on and off the platform.

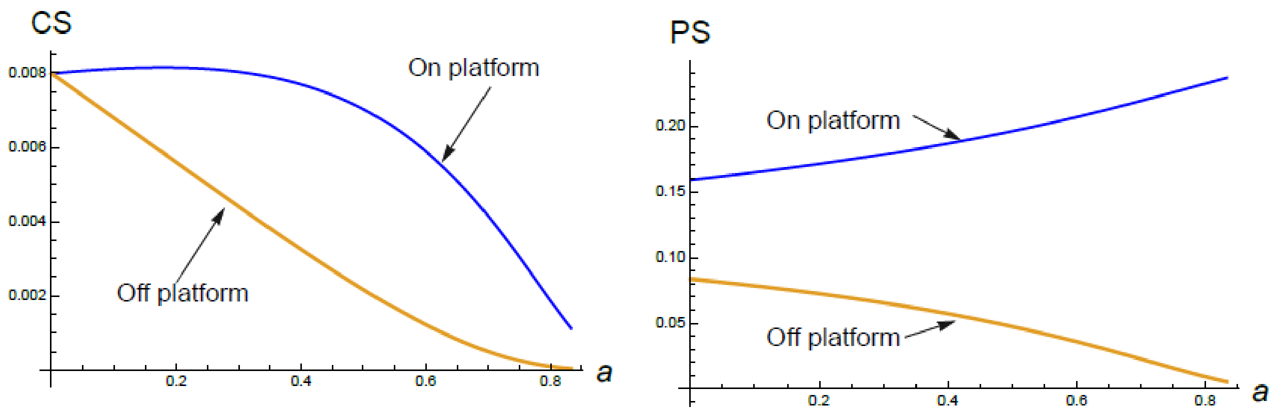


FIGURE 9: Effect of information precision, $\lambda = 4/5, G(m) = m, J = 2$.

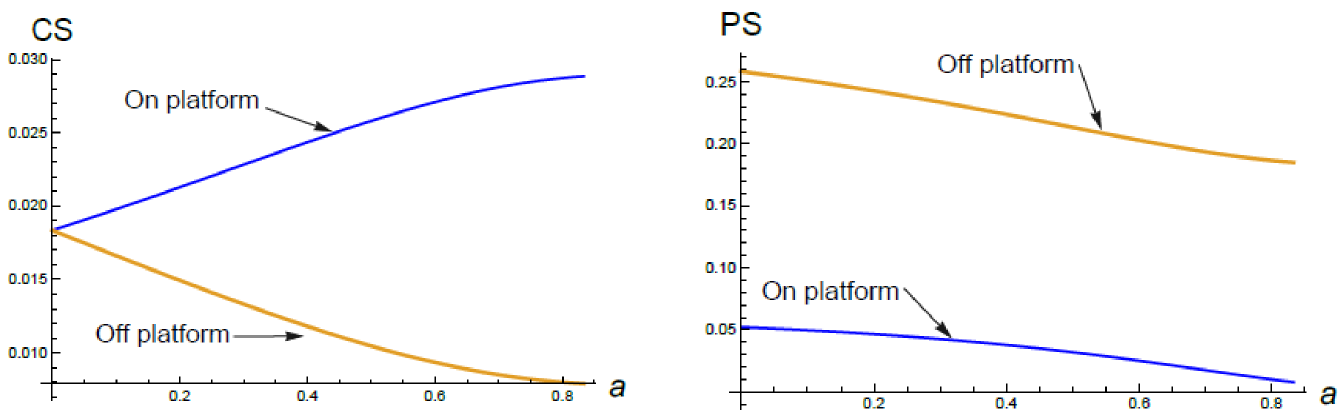


FIGURE 10: Effect of information precision, $\lambda = 4/5, G(m) = m, J = 12$.

## 4.2 Cohort-Based Ads

Now suppose information about each individual $\theta_i$ is shared with the online consumer $i$ but not with the sellers. In contrast, the platform publicly announces the ranking of consumer valuations $\theta_j$ to all bidders. Thus, the platform runs a finite set of auctions—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 the recent Google Privacy Sandbox proposals 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 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 $i \to j$ 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 $\theta$ 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 5 (Coarse Targeting)**
*Assume $F^J \succ_{lr} G^J$ over all $\theta$ 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 $\theta$ and $m$, respectively. Thus, cohort-based ads yield higher quality provision offline, but lower quality online, relative to the baseline model with full disclosure $\theta$. The comparison of offline qualities with the contextual ads case, on the other hand, is more involved.

**Corollary 2 (Quality and Consumer Surplus)**

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

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

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

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

A critical implication of Proposition 5 is that all consumers are now better off, relative to the full information case, as a consequence of the greater quality provision offline. Total surplus can be higher, too, as a consequence of greater offline quality, although online quality is lower.

## 4.3 Data Governance: Comparison

We study different information structures in the two models above. In this subsection, we focus on a parametrized example with Beta distributions, illustrated in Figure 11, that satisfies the assumptions of Proposition 5.

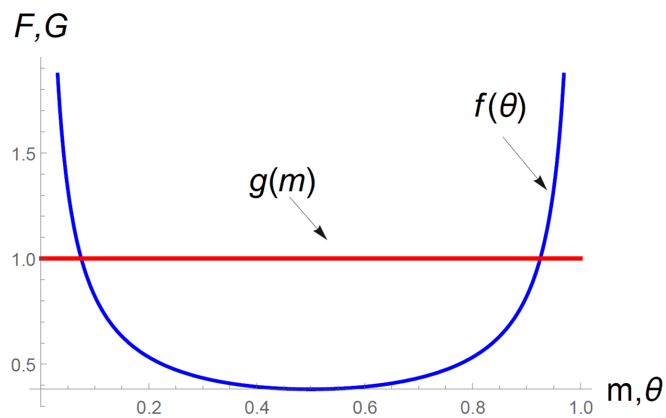


FIGURE 11: Contextual vs. Unrestricted Targeting

**Example 1 (Beta Distribution)**
*Consumer valuations for each brand* $j$ *are given by* $v_j \in \{0, 1\}$ *with equal probability, independently across* $j$ *and . Consumer beliefs* $m_j$ *are uniformly and independently distributed over* $[0, 1]$ *for each* $j$*. The platform reveals to both consumers and bidders a signal that induces beliefs* $\theta_j$ *over valuations that follow a symmetric Beta distribution with parameter* $1 - a$*, where* $a \in [0, 1)$.

For this example, we consider the welfare effects (for consumers, bidders, and the platform) of the precision of the consumers' information, as captured by $F(\theta, 1-a)$.

### 4.3.1 Consumer Surplus

- For low $J$, greater precision hurts consumers—offline schedule gets worse.
- For high $J$, greater precision helps consumers—order statistic already selects highest signal, but match quality still improves, up to a certain point.
- Greater precision is more beneficial for low $\lambda$ (fewer consumers are hurt by offline schedules getting worse).

Figure 12 illustrates this intuition.

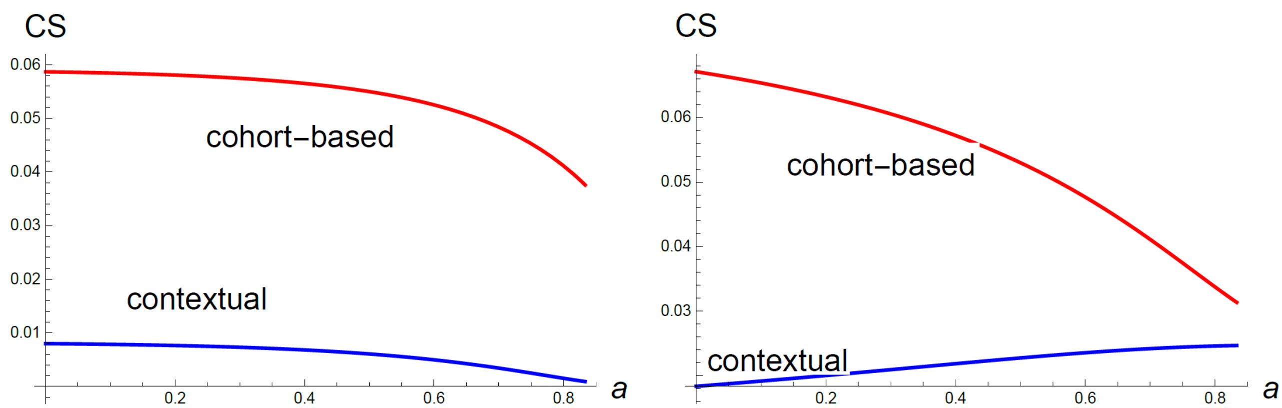


FIGURE 12: Effect of information precision, $\lambda = 4/5$ (both), $J = 2$ (left), $J = 12$ (right).

### 4.3.2 Producer Surplus

- With few competitors, producers benefit from better match quality. The distance between the expectation of the first and second order statistics grows.
- With many competitors, greater precision toughens competition. The distance between the expectation of the first and second order statistics shrinks.

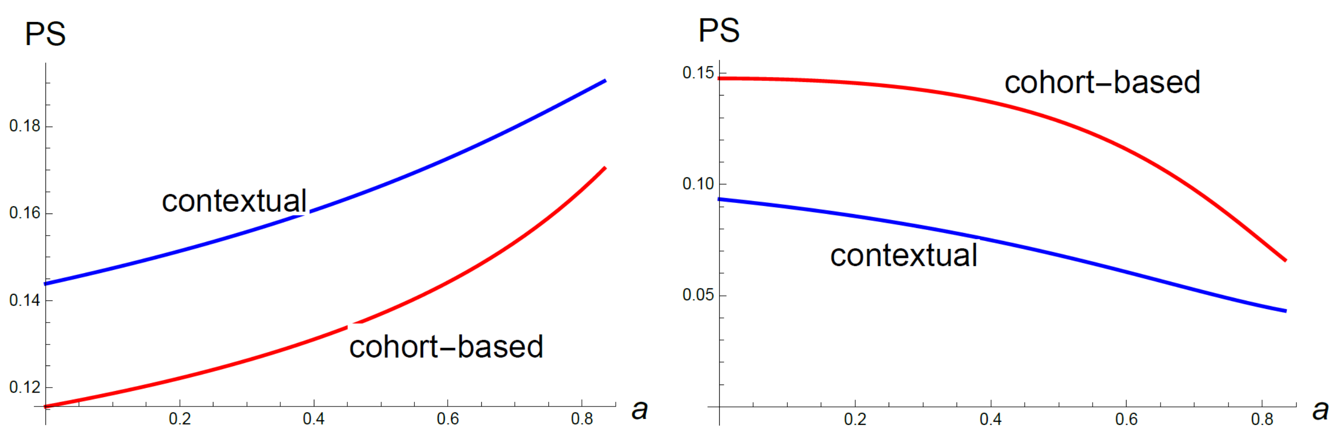


FIGURE 13: Effect of information precision, $\lambda = 4/5$ (both), $J = 2$ (left), $J = 12$ (right).

### 4.3.3 Total Surplus

- Total surplus increases with precision, but not obvious.

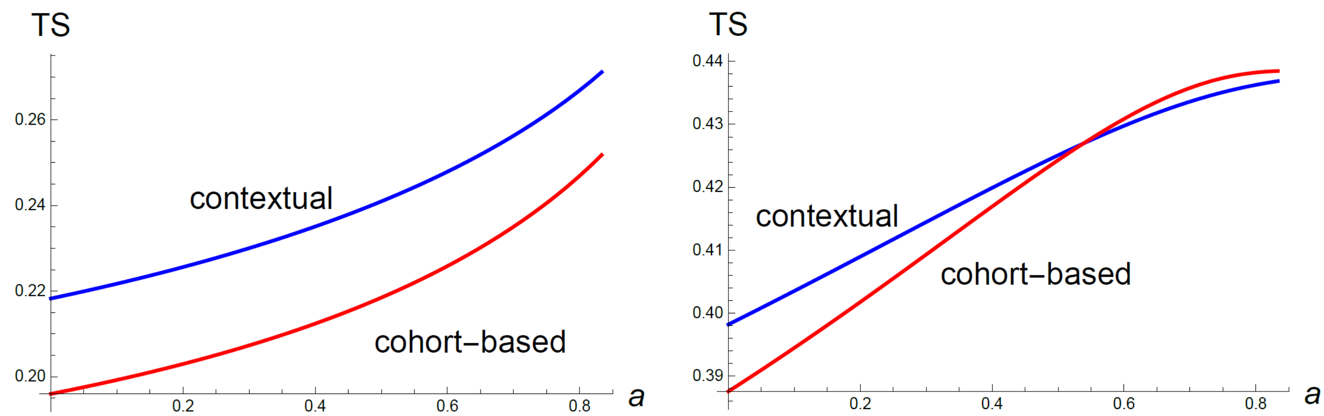


FIGURE 14: Effect of information precision, $\lambda = 4/5$ (both), $J = 2$ (left), $J = 12$ (right).

### 4.3.4 Platform Revenue

- Platform surplus increases with precision, but not obvious.

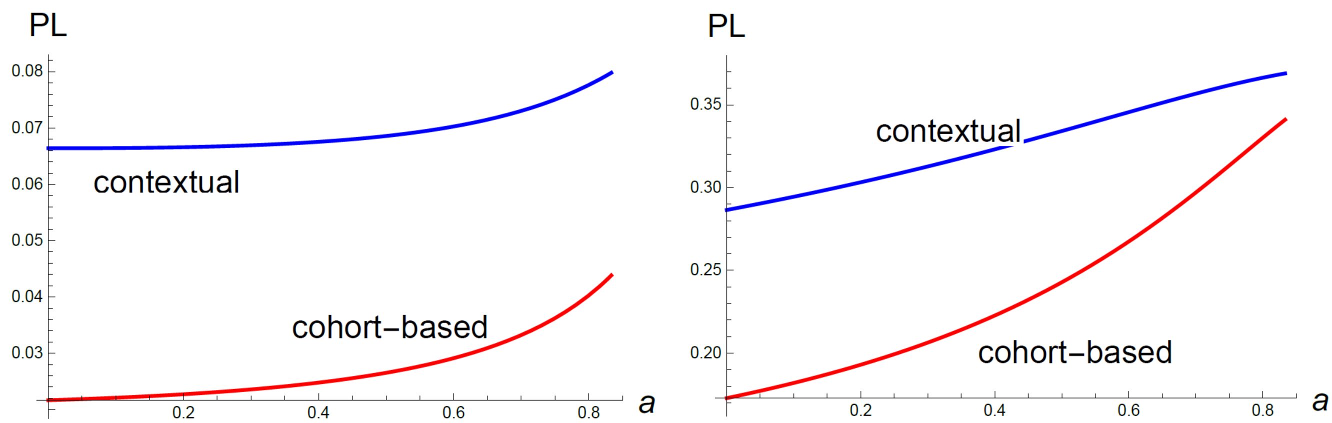


FIGURE 15: Effect of information precision, $\lambda = 4/5$ (both), $J = 2$ (left), $J = 12$ (right).

# 5 Competition Design

So far we have assumed that the platform reveals only the winning bidder's type $\theta_j$ to the buyer, and shows sponsored content only. Equivalently, we could have assumed that: (a) the platform reveals to the consumer her $\theta_j$ for the winning firm; (b) it contextually shows her the winning bidder's personalized offer; (c) it does not reveal values or prices for the losing bidders; (d) searching online beyond the sponsored link is costly (even if vanishingly so).

In our baseline, the consumer infers (correctly, on the path of play) that the auction winner provides her with the highest value, and only considers the winning firm's online and offline offers. In practice, platforms manage both information flows and the degree of competition on the results page.

## 5.1 Organic Links

We now consider an extension to our model, whereby the online buyer can buy from *any* offline producer. This version of our model corresponds to a platform that displays numerous "organic links." The timing of the game is as follows, and is summarized in Figure 16 below.

1. Sellers simultaneously set off-platform menus $\left(q_j^0, p_j^0\right)$.
2. The platform posts all menus and reveals $\theta$ to all online buyers.
3. Sellers place bids $b_j(\theta)$ for each type.
4. The winner of the auction winner offers a single product of quality $q_j(\theta)$ at price $p_j(\theta)$.
5. The consumer can buy on-platform from seller $j$ or off-platform from *any* seller.

In this more intensively competitive model, the offline menus can be used to directly affect the online market shares: any firm could offer more utility to consumers offline, lose the auction for some $\theta$, and still make a sale online. In turn, these incentives affect the other bidders' willingness to pay at the auction.

Under complete information (which we have assumed), all firms can react to any offline deviation by changing the contracts they would offer online to each type $\theta$, and adjusting their bids accordingly. We then characterize the auction outcome, and the offline equilibrium menus.

Fix a profile of offline menus and associates rent functions $\{U_j^0\}_{j=1}^N$. With organic links present, buyer $\theta$'s outside option when online is given by

$$\bar{U}^0(\theta) := \max_j U_j^0(\theta_j).$$

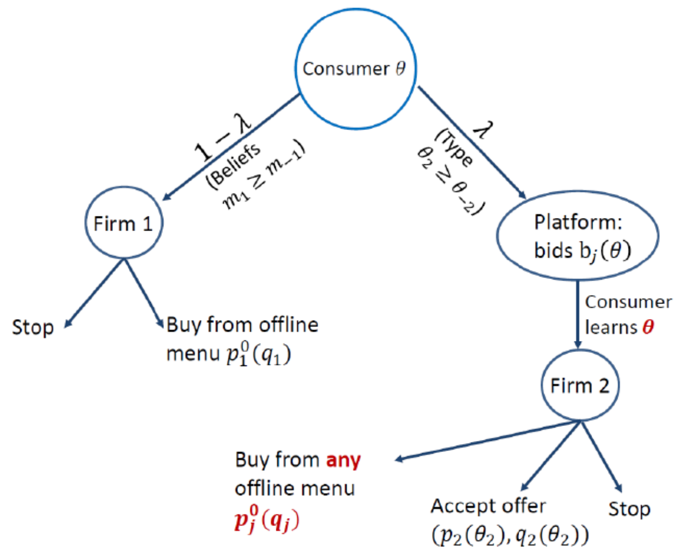


FIGURE 16: Sample search patterns for consumer $i$ online and offline. Any consumer on the platform can choose between a personalized offer and every offline menu.

It follows that every seller $j$ bids

$$b_j(\theta) = \max\{0, \theta_j^2/2 - \bar{U}^0(\theta)\}$$

in the auction for buyer $\theta$. Now let

$$h(U^0(\theta_j)) := \sqrt{2U^0(\theta_j)}$$

denote the lowest *competing* type $\theta_{-j}$ for which the second highest bid is positive.

Therefore, seller $j$'s best response problem is given by

$$\max_{q^0,U^0} \lambda \int_0^1 \int_0^{h(U^0)} \underbrace{\left(\frac{\theta^2}{2} - U^0(\theta)\right)}_{\text{losing bid } = 0} \mathrm{d}F^{J-1}(s)\mathrm{d}F(\theta) + \lambda \int_0^1 \int_{h(U^0)}^{\theta} \underbrace{\frac{1}{2}\left(\theta^2 - s^2\right)}_{\text{losing bid } > 0} \mathrm{d}F^{J-1}(s)\mathrm{d}F(\theta)$$

$$+ (1-\lambda) \int_0^1 \left[\theta q^0(\theta) - \frac{q^0(\theta)^2}{2} - U^0(\theta)\right] G^{J-1}(\theta)\,\mathrm{d}G(\theta).$$

In words, for every offline type $\theta$, seller $j$ knows that, whenever they are the high-valuation bidder, there is a positive probability the second bid is nil. In that case, any rents left offline to type $\theta$ simply reduce the price the firm can charge online. With the complement probability, however, the second highest bid is strictly positive, in which case $U_j^0(\theta)$ reduces all bids by the same amount: the necessary discount on the price to the buyer is offset by a reduction in the price paid to the platform for the sponsored link.

As a result we have the following characterization of the symmetric equilibrium.

**Proposition 6 (Equilibrium with Organic Links)**

1. *There exists a unique symmetric equilibrium.*
2. *In equilibrium, $q^*(\theta) = \theta$ and $q^0(\theta)$ is weakly higher than without organic links.*
3. *The online consumer's utility level is correspondingly higher.*

Figure 17 illustrates the equilibrium quality levels.

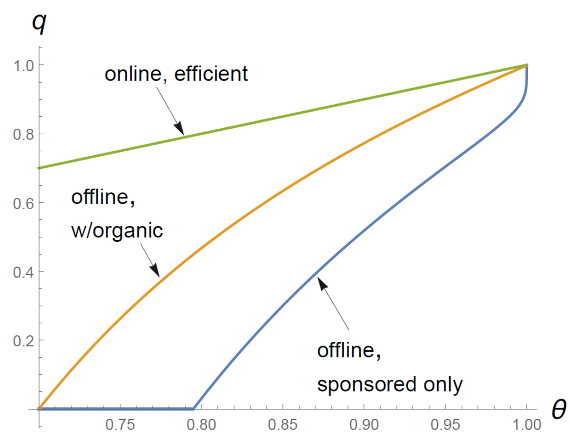


FIGURE 17: Equilibrium quality, $\lambda = 1/2, J = 5, G(m) = m, F(\theta) = Beta(\theta, 1/4, 1/4)$

The intuition for this result is that the winning bidder competes with the best off-platform offer across *all bidders.* Therefore, the ranking of the bidders' willingness to pay for sponsored links is uniquely determined by the total potential surplus. In this setting, raising the off-platform utility for type $\theta_j$ lowers all (strictly positive) bids whenever $\theta_j = \max \theta$. This yields cost savings for rent provision, relative to the baseline model with sponsored links only. In that case, off-platforms rents only decreased on-platform prices.

## 5.2 Multiple Sponsored Links

Our baseline model is one of "perfect steering:" upon winning the second-price auction, the seller competes only with its own offline price-quality schedule. In practice, however, it is standard practice for multiple advertisements—often for related products—to be shown to a single consumer. We now extend our framework to incorporate this feature. Doing so will clarify how data governance can influence how competition between sellers online interact with competition between online and offline offerings. Throughout, we maintain the assumption that the platform knows the consumer's type $\theta$. Of course, a potential explanation for why multiple advertising slots are auctioned off is that the platform is yet uncertain about the consumer's exact preferences.

At the start of the game, each seller $j$ chooses a non-linear price-quality schedule $(p_j^0, q_j^0)$ which it offers to the fraction $1-\lambda \in [0,1]$ of consumers who buy offline. Online, consumers and sellers learn some (potentially noisy) information about $\theta$. Following this, the sellers participate in an ad auction in which sellers bid against each other in order to access the online consumer. In the main text, we considered the case in which a single advertising slot is auctioned off—there, the winner of the ad auction had exclusive access to the online consumer and faced the showrooming constraint only: in order to make the sale online, it had to deliver more utility than the (now informed) consumer can obtain from its offline menu. We now enrich this setting by allowing multiple slots to be auctioned off: the winning sellers now face two sources of competition: as before, they continue to face the showrooming constraint, but now also compete against each other to make the sale online.

There are well-known difficulties in analyzing multi-unit auctions. We will consider third-price auctions in which identical advertising slots are allocated to the highest and second-highest bidder. This is a special case of a sealed bid uniform price multi-unit auction in which (i) each buyer can purchase up to a single unit; and (ii) the price paid is the highest losing bid.

We remain in the full information regime in which both the consumer and sellers learn $\theta$ perfectly online. We show that there is a symmetric equilibrium in which:

1. all firms choose identical offline price/utility-quality schedules $(q^0, U^0)$;

2. sellers $\{(2), \dots (J)\}$ bid 0; and

3. seller (1) bids $b > 0$ and pays 0.

We assume that the allocation rule breaks ties in favour of the seller for which the consumer has higher valuation.[2]

We will tie-break in favour of the seller which is the consumer values most highly. Sellers (1) and (2) are then allocated ad slots and the platform raises no revenues. Both sellers then choose a price and quality to offer the online consumer. The unique equilibria here is analogous to Bertrand competition: Seller (2) offers $q^*_{(2)}(\theta) = \theta_{(2)}$ and price $p^*_{(2)}(\theta) = \theta^2_{(2)}/2$

[2]Since $F$ has a density, there are a.s. no ties in valuations.

which solves the problem

$$\max_{q_{(2)},p_{(2)}} \theta_{(2)} q_{(2)}(\theta) - p_{(2)}$$
$$s.t. \quad p_{(2)} - q_{(2)}(\theta)^2/2 \geq 0$$

i.e., online, (2) gives as much surplus as possible to the consumer without violating its non-negativity constraint on profits. Firm (1) now faces this additional constraint posed by firm 2: for it to sell to the consumer online, it must fulfil

$$\begin{aligned} &\text{(IC - Showrooming)} \quad \theta_{(1)} q_{(1)}(\theta) - p_{(1)} \geq U^0_{(1)}(\theta_{(1)}) \\ &\text{(IC - Competition)} \quad \theta_{(1)} q_{(1)}(\theta) - p_{(1)} \geq \theta_{(2)} q_{(2)}(\theta) - p_{(2)} = \theta^2_{(2)}/2. \end{aligned}$$

(IC - Showrooming) states that the utility delivered to the consumer of type $\theta$ online must exceed what the consumer can obtain offline; (IC - Competition) states that firm (1) must deliver more utility online than firm (2). When $\theta_{(1)} - \theta_{(2)}$ is large, we expect (IC - Showrooming) to bind; conversely, when $\theta_{(1)} - \theta_{(2)}$ is small, we expect (IC - Competition) to bind. Fixing $\theta_{(1)}$, the threshold at which we cross from the showrooming to competition constraint is when $U^0(\theta_{(1)}) = \theta^2_{(2)}/2$, or $\theta_{(2)} = (2U^0(\theta_{(1)})^{1/2} := h(U^0)$.

It is then immediate to show that the equilibrium with two sponsored and no organic links is outcome-equivalent to the one-slot case *with* organic links. Similarly, inspecting the firm's problem under two slots, the value of the dynamic program is plainly lower than under the one-slot case. In fact, under the two slot case, since the online quality remains efficient, and the offline quality is now less inefficient than before and there are more gains from trade. We might think of the full information design with multiple slots as shifting (i) these additional gains from trade to CS; (ii) shifting the reduction in producer surplus to CS; and (iii) shifting all platform revenues to CS.

# 6 Extensions and Robustness

## 6.1 Endogenous Participation

So far, consumers' search behavior (on or off the platform) was assumed exogenous. In order to relax this assumption, we introduce a small cost of using the platform. This additive cost $\varepsilon$ 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 $\varepsilon$.

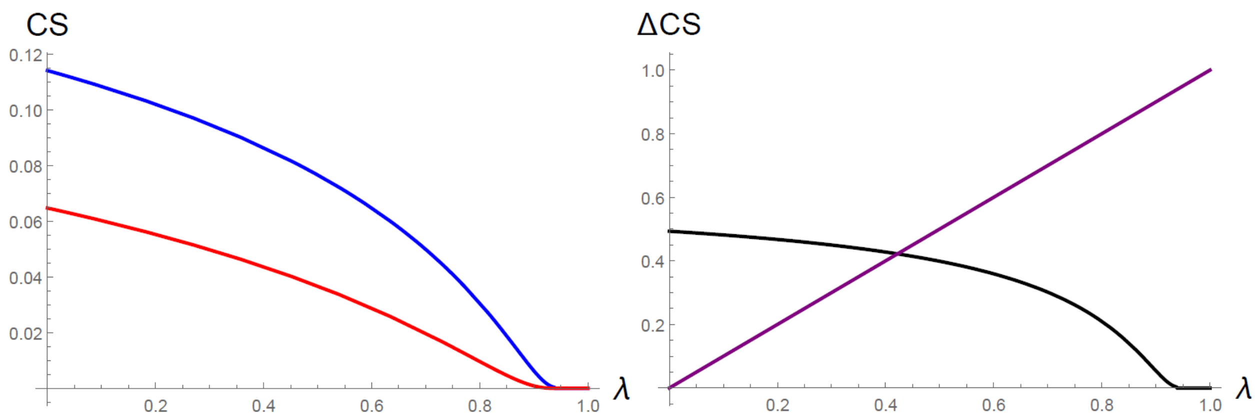


FIGURE 18: Left: On-off platform gap in consumer surplus decreases with $\lambda$. Right: $\varepsilon \sim [0, 1/50]$. Platform participation decisions are strategic substitutes. Unique equilibrium.

Figure 18 (left) shows expected consumer surplus online and offline as a function of platform size $\lambda$. Figure 18 (right) shows the gap in expected consumer utility as a function of $\lambda$, and the critical cost level

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

for which $\lambda$ is the platform size in equilibrium.

Finally, Figure 19 shows that, for small supports of $\varepsilon$, the equilibrium platform size $\lambda^*$ is increasing in the number of sellers $J$.

## 6.2 Retail Platforms

The most immediate interpretation of our model is the ranking provided by the search engine, notably Google sponsored search. However, a ranking mechanism is of course provided by Amazon as well. A more complete description of the transactions occurring on Amazon would have to take account a richer fee structure that is based on sales commissions and sponsored search. A second aspect that enters when considering Amazon is that Amazon itself might act as a seller on the platform. (While the dual role of Amazon is discussed in the context of digital commerce, something akin to the dual role of course occurs in large retailers who offer their own private labels.) A short-cut to analyze self-preferencing would be to say that Amazon has its own product, and could choose to highlight its product when the competing product is not to far. Then steering would have an enormous value, and while it would reduce the value of the consumer, this might be attenuated.

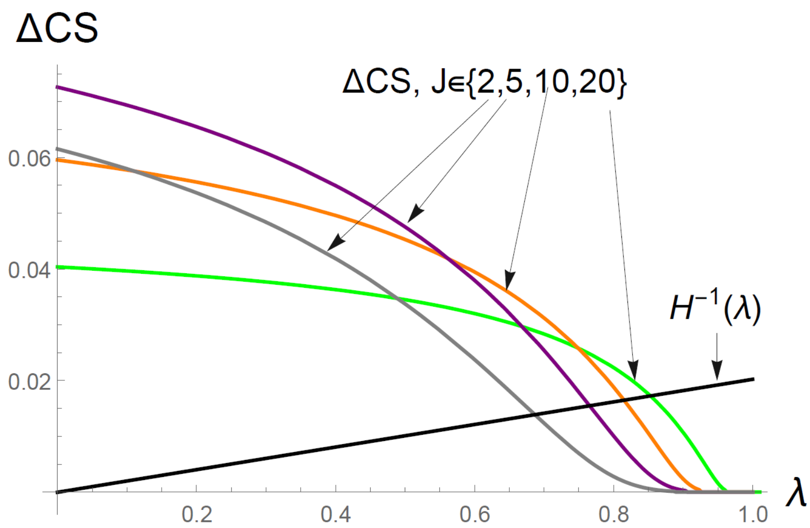


FIGURE 19: $\varepsilon \sim [0, 1/50]$. Equilibrium $\lambda^*$ increases with $J$.

## 6.3 Consideration Sets

The equilibrium construction in the moment is one of "perfect" steering. The constraint on the winning seller in the choice of the consumer is the showrooming constraint rather than the competitor. We could also consider the case where the winning competitor on the platform receives the consumer, and then only faces his offline offer as competition. We could think about extensions where the consumer searches with probability $1 - \alpha$ among competing offers, say there are two sponsored search listing, after having reviewed the top offer. The competition may then depend on the entire type profile. If $\theta_1$ and $\theta_2$ are far apart, then the competition would be weak. If $\theta_1$ and $\theta_2$ are nearby, then the competition would be hard, and the probability of moving to the second ranked allocation would be relevant. Suppose the second firm knows that it is loser, then it could adjust its bidding strategy.

The symmetry in the consideration set (i.e. there is competition with probability one) is of course useful to attain the results in a transparent manner. Suppose now that we maintain two slots, but that there is some stochastic element that precludes competition. Then, we have two dimensions along which there can differentiation, the recognition probability and the information that the firms have about each other. We may wish to consider an intermediate solution. The specific competition that ensures is then a version of Bergemann, Brooks, and Morris (2021), but with quality differentiated products.

# 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.

# 8 Appendix

The appendix collects the remaining proofs for the results presented in the paper.

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

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

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 5.** 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 $\theta$ 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. These are sufficient because the problem is linear in $q$, concave $U$, and additively separable in these two variables. See Jullien (2000) for a similar approach. 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

$$\begin{aligned}
\gamma &= -\frac{\lambda JF^{J-1}f}{\left(1-\lambda\right)JG^{J-1}g+\lambda JF^Jf}\left(1-\left(1-\lambda\right)G^J-\lambda F^J\right)\\
\gamma^0 &= -\frac{\left(1-\lambda\right)JG^{J-1}g}{\left(1-\lambda\right)JG^{J-1}g+\lambda JF^Jf}\left(1-\left(1-\lambda\right)G^J-\lambda F^J\right)
\end{aligned}$$

and differentiating both with respect to $\theta$, 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 $\theta$, 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 $G^J$. ■

**Proof of Proposition 6.** Each seller solves the following problem:

$$\begin{aligned}
\max_{q^0,U^0} \quad & \lambda \int_0^1 \Bigg[ \int_0^{h(U^0)} \underbrace{\left(\frac{\theta^2}{2} - U^0(\theta)\right)}_{\substack{\text{showrooming}\\ \text{binds}}} dF^{J-1}(s) + \int_{h(U^0)}^{\theta} \underbrace{\left(\frac{\theta^2 - s^2}{2}\right)}_{\substack{\text{competition}\\ \text{binds}}} dF^{J-1}(s) \Bigg] dF(\theta) \\
& + (1-\lambda)\int_0^1 \left(\theta q^0(\theta) - \frac{q^0(\theta)^2}{2} - U^0(\theta)\right) G^{J-1} dG(\theta) \\
& \text{s.t.} \quad \partial U^0/\partial\theta = q^0, U^0 \geq 0.
\end{aligned}$$

We will maximize this pointwise. Write the Hamiltonian (where $q^0$ is the co-state, $U^0$ is the state, and $\gamma$ is the multiplier):

$$\begin{aligned}
\mathcal{H}(q^0, U^0, \gamma) := \lambda f \Bigg[ & \underbrace{\int_0^{h(U^0)} \left(\frac{\theta^2}{2} - U^0(\theta)\right) dF^{J-1}(s)}_{\text{(IC - Showrooming) binds}} + \underbrace{\int_{h(U^0)}^{\theta} \left(\frac{\theta^2 - s^2}{2}\right) dF^{J-1}(s)}_{\text{(IC - Competition) binds}} \Bigg] \\
& + (1-\lambda)g[\theta q^0(\theta) - q^0(\theta)^2/2 - U^0(\theta)]G^{J-1}(\theta) \\
& + \gamma q^0(\theta).
\end{aligned}$$

A necessary condition for the pair $(q^0, U^0)$ to be optimal is:

$$\begin{aligned}
\mathcal{H}_{q^0} &= (1-\lambda)gG^{J-1}(\theta - q^0) + \gamma = 0 \\
\mathcal{H}_{U^0} &= \lambda f \left[ h'^0) f_{J-1}(h(U^0)) \left(\frac{\theta^2}{2} - U^0\right) - F^{J-1}(h(U^0)) - h'^0) f_{J-1}(h(U^0)) \left(\frac{\theta^2 - h(U^0)^2}{2}\right) \right] \\
&\quad - (1-\lambda)G^{J-1}g \\
&= -\lambda f F^{J-1}(h(U^0)) - (1-\lambda)G^{J-1}g \\
&= -\dot{\gamma}
\end{aligned}$$

where $f_{J-1}(\theta) := (J-1)F^{J-2}(\theta)f(\theta)$ is the density of the highest draw from $J-1$ independent draws. The second equality in $\mathcal{H}_{U^0}$ follows from observing that $h(U^0)^2/2 = U^0$ so the second order changes do indeed cancel out, formalizing our intuition developed above. We also have our boundary condition: $\gamma(1) = 0$ which, as in Mussa and Rosen (1978), leaves the highest type undistorted. Further, $U^0$ and $q^0$ are linked through standard envelope theorem arguments:

$$U^0(\theta) = U(0) + \int_0^\theta q^0(s)ds.$$

We now compare the optimality conditions as derived for the organic-links case above against the baseline model. We use the subscripts '$1S'$' and '$2S'$' to denote the co-state variables for the baseline and the new case, respectively. Rewrite $\mathcal{H}_{U^0}$ as

$$\begin{aligned}\dot{\gamma}_{2S} &= \lambda f F^{J-1}(\underbrace{h(U^0)}_{\theta_{(2)}}) + (1-\lambda)G^{J-1}g \\ &\leq \lambda f F^{J-1} + (1-\lambda)G^{J-1}g \\ &= \dot{\gamma}_{1S}\end{aligned}$$

since $\theta_{(2)} \leq \theta_{(1)}$. Further observe that $\mathcal{H}_{q^0}$, is the same both on and offline so

$$\begin{aligned}\gamma_{1S} &= \frac{1}{J}(\lambda F^J + (1-\lambda)G^J - 1) \\ &= \gamma_{1S}(1) - \int_\theta^1 \dot{\gamma}_{1S}(s)ds \\ &\leq \gamma_{2S}(1) - \int_\theta^1 \dot{\gamma}_{2S}(s)ds \\ &= \gamma_{2S}\end{aligned}$$

where the inequality follows from noting that our boundary conditions are the same for both problems. In other words, with organic links, there is less cost to giving the consumer more utility offline. Substituting into $\mathcal{H}_{q^0}$, we see that this translates into less distortion of the offline quality schedule:

$$\begin{aligned}q^0_{1S} &= \min\left\{\theta + \frac{\gamma_{1S}}{(1-\lambda)G^{J-1}g}, 0\right\} \\ &\leq \min\left\{\theta + \frac{\gamma_{2S}}{(1-\lambda)G^{J-1}g}, 0\right\} = q^0_{2S}\end{aligned}$$

so by the envelope theorem,

$$\begin{aligned} U^0_{1S} &= U^0_{1S}(0) + \int_0^{\theta} q^0_{1S} \\ &\leq U^0_{2S}(0) + \int_0^{\theta} q^0_{2S} = U^0_{2S}. \end{aligned}$$

Now coupling this with the observation above that $U^*_{2S}$ is the upper envelope of $U^0$ and $\theta^2_{(2)}/2$,

$$U^*_{2S} = \max\left\{U^0_{2S}, \theta^2_{(2)}/2\right\} \geq U^0_{2S} \geq U^0_{1S} = U^*_{1S}$$

where the last equality follows from the main text where in equilibrium, the seller delivers the same utility schedule on and offline. Consumer welfare for each type unambiguously improves on and offline hence

$$CS_{2S} = \lambda \int U^*_{2S}(s) dF^J(s) + (1-\lambda) \int U^0(s) dG^J(s) \geq CS_{1S}.$$

This ends the proof. ■

# 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.

Anderson, S. P., and Ö. Bedre-Defolie (2021): "Hybrid platform model," Discussion Paper 16243, CEPR.

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.

Bergemann, D., B. Brooks, and S. Morris (2021): "Search, information, and prices," *Journal of Political Economy*, 129(8), 2275–2319.

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

Calzolari, G., and V. Denicolo (2015): "Exclusive contracts and market dominance," *American Economic Review*, 105(11), 3321–51.

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.

Fuchs, W., and A. Skrzypacz (2015): "Government interventions in a dynamic market with adverse selection," *Journal of Economic Theory*, 158, 371–406.

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.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93(1), 1–47.

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

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

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*.

Philippon, T., and V. Skreta (2012): "Optimal interventions in markets with adverse selection," *American Economic Review*, 102(1), 1–28.

Tirole, J. (2012): "Overcoming adverse selection: How public intervention can restore market functioning," *American Economic Review*, 102(1), 29–59.

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