Data Sharing and Market Power with Two-Sided Platforms*

Rishabh Kirpalani† and Thomas Philippon‡

First version: November 2019; this version: April 2021

Abstract

We study an economy where consumers and merchants interact on a two-sided platform. The platform gathers consumer data and sells it to merchants. Data sharing increases gains from trade by improving match quality, but also generates two externalities which increase the market power of the platform and lower merchant entry. The *gatekeeper* externality arises from the platform's control over access to consumers. The *copycat* externality derives from its ability to compete with its own merchants. These effects are not internalized by consumers and data sharing can be socially excessive. We study implications for the regulation of digital versus traditional platforms.

*We are especially grateful to German Gutierrez for helping us understand the business model of Amazon, and to Luis Cabral, Steve Davis, Andres Drenik, Janice Eberly, Paolo Martellini, Diego Perez, Christopher Sullivan, Larry White, Rob Seamans, and seminar participants at NBER, University of Chicago, and New York University for stimulating discussions. We are grateful to the Smith Richardson Foundation for a research grant.

†University of Wisconsin-Madison, rishabh.kirpalani@wisc.edu.

‡New York University, CEPR and NBER, tphilipp@stern.nyu.edu

"They [Wall Street analysts] paid less attention to the thing that was attracting so many customers and keeping them coming back for more: the data. For Amazon wasn't just able to give its visitors the books they came to buy, but the books that they didn't even realize they wanted."

Margaret O'Mara, The Code, 2019.

O'Mara (2019)'s description of the rise of Amazon underscores the importance of data and technology to the success of the company. For instance, an important reason for locating the firm in Seattle was the well established presence of Microsoft and the access to thousands of tech engineers, and indeed, Amazon became a technological and logistical powerhouse. Amazon and other platforms collect vast amounts of data on both consumers and sellers using their websites. While the use of such data constitutes a perfectly legitimate business strategy, Amazon has recently come under scrutiny for allegedly using this data unfairly against other sellers participating on its website. In April 2020, the *Wall Street Journal* reported that Amazon employees on the private-label side of its business had used data about individual third-party sellers on its site to create competing products.[1] Similarly, in July 2020, the *Wall Street Journal* reported that Amazon appeared to use its investment and deal-making process [on its VC side] to help develop competing products. Is this business as usual, or does it reflect excessive market power by Amazon? Can data access create anti-competitive effects? Can data collection and usage be excessive? If so, when, and why? These are some of the questions we try to analyze in this paper.

Large internet platforms have changed the way market participants interact. One reason for this is the extraordinary ability of platforms such as Amazon and Google to gather and analyze large amounts of data. Platforms use this data to enable better matching between participants as well as for commercial purposes, including sale to third parties (Gutierrez, 2020b). In this paper, we study the consequences for market participants as well as the welfare implications of data sharing in the context of large platforms. In our model, information shared by buyers enables better matching between consumers and merchants on the platform. However, more information/data endogenously increases the market power of the platform relative to sellers. We ask if data sharing can be excessive and under what conditions. In doing so, we shed light on the qualitative differences between new platforms such as Amazon and Google and more traditional retail platforms.

In our model, consumers and merchants interact on a two-sided platform. Consumers can share information with the platform regarding their tastes for different varieties of a good. The platform then sells this data to merchants. Merchants and consumers interact in a directed search market on the platform. As the information gathered from the consumers becomes more precise, merchants can better predict the varieties desired by the consumers which improves overall match quality. On the other hand more precise information increases the market power of the platform relative to the merchants.

The platform and the merchants bargain over the price of information. As is standard, this price depends on the outside options of both the platform and the merchant. In our model, as information becomes more precise, the outside option of the merchant decreases while that of the platform increases. We call the former the *gatekeeper* effect and the latter the *copycat* effect. The gatekeeper effect arises because platform partially

[1] https://www.wsj.com/articles/amazon-scooped-up-data-from-its-own-sellers-to-launch-competing-products-11587650015. Amazon said it was conducting an internal investigation into the practices described in the story.

controls access to consumers. Consumers have the choice to purchase on the platform or at stores outside the platform, which we call the outside market. Similarly, merchants have the option to sell the good on the platform or at the outside market. As information gets more precise, a larger fraction of consumers choose to purchase on the platform and thus do not show up at the outside market. Therefore, the seller is less likely to be matched outside the platform leading to a lower outside option. This effect is present in all digital platforms.

The copycat effect occurs when platforms also have the option to compete with their own merchants. The platform could use its information to produce the good by itself and sell directly to consumers. As information gets more precise, the value of doing so for the platform increases. While this effect is not present in booking platforms such as Uber or OpenTable, it is particularly relevant for platforms like Amazon which have private-label products.

These two outside option effects combine to increase the price of information the platform can charge as information gets more precise, which lowers the seller's payoff. Thus, whether merchants are made better off as a consequence of increased data sharing depends on if the additional matching efficiency benefits outweigh the increase in the price of information.

We show that as information gets more precise, the latter effect dominates the benefits of better matches which in turn implies that the number of merchants participating on the platform decreases. This lowers consumer welfare. Since consumers are small, they do not internalize the effect of their information disclosure on seller entry via the matching efficiency and the gatekeeper and copycat effects. We show that due to the gatekeeper and copycat externalities, there may be excessive information sharing by consumers compared with that chosen by a social planner. Consequently, regulation that restricts the information shared by consumers can increase overall welfare. A corollary of this result is that if information technology decreases the private cost of information disclosure, there is more likely to be too much information disclosure.

Currently, there are two largely separate pieces of regulation that are being aimed at platforms. The first, motivated by privacy issues, is regulation that explicitly targets data collection. The most prominent recent example of such regulation is GDPR. Second, there are anti-trust suits being filed against large platforms. These explicitly target the anti-competitive behavior of platforms. One message of our paper is that the goals targeted by these two regulatory measures are closely linked. In other words, regulating data collection may increase competitiveness.

A natural question that arises is how platforms like Amazon differ from traditional retail platforms like Walmart and Target. For example, the latter also gather information about consumers tastes via sales, and produce their own in-house private label products. However, the key difference is the amount of data collected and processed by Amazon relative to Walmart, as well as the efficiency of their matching technologies. For example, physical stores can only observe the products actually purchased by consumers while Amazon can observe the consumers browsing history including the items in the cart (but not actually purchased). This enables Amazon to better predict consumers' tastes. As we argue in the paper, it is exactly when the level of information is high that there is likely to be excessively high disclosure. This provides a rationale for greater regulation of online platforms relative to the physical stores.

The paper is organized as follows. We present our baseline model in Section 1. We compare the planner and the decentralized equilibria in Section 2. We study an extension with heterogeneous consumers in Section 3.

Related Literature Katz and Shapiro (1985) provide an early analysis of network externalities in a Cournot model. They characterize equilibria with varying degree of compatibility across firms. Rochet and Tirole (2003) incorporate the insights of network economics and multi-product pricing to study competition among two sided platforms. Caillaud and Jullien (2003) model platforms as matchmakers and study competition with multi-homing, and price discrimination. They consider two-tier price systems with registration and transaction fees. Armstrong (2006) emphasizes the role of consumers who join all platforms and analyzes cross-group externalities. The study of single- versus multi-homing decisions has expanded rapidly since these early contributions, in particular regarding media markets. Ambrus et al. (2016) model consumers who spread their attention between several outlets. Park et al. (2020) model the entry of TV stations in local newspaper markets in the 1950s. Gutierrez (2020a) estimates a large scale model of consumers' choices on Amazon's platform and studies the impact of private labels. These papers abstract from the interaction of information disclosure by consumers and the market power of the platform relative to sellers, which is the central focus of our paper.

A growing literature studies information disclosure by consumers. On the one hand, it is clear that the use of personal data can improve the allocation of online resources. Excessive information sharing can potentially harm consumers, however. Bergemann et al. (2015) analyze how a monopolist can use information about consumers' tastes to engage in third degree price discrimination. Bergemann et al. (2018) study how a data seller should optimally design and sell statistical experiments. Acemoglu et al. (2019) and Bergemann et al. (2019) study environments in which there can be excessive data sharing since consumers do not internalize that their data can reveal information about other consumers. See also Bergemann and Bonatti (2019) for a survey of the literature on data markets.

Our model of taste uncertainty is similar to Bergemann et al. (2015), Bergemann et al. (2018), and Ichihashi (2019). Compared to these papers, our contribution is to analyze information in a standard directed search environment where we can analyze the market power of a platform vis-a-vis its merchants. The key externality in our model operates through the outside options of sellers and the platform which in turn affects the seller's entry decisions.

To model the interaction between buyers and sellers we use the directed search framework developed by Shimer (1996) and Moen (1997) among others. See Wright et al. (2019) for an excellent survey. Directed search determines both trade on the platform as well as the outside market which in turn affects the sellers' outside option.

1 Baseline Model

We consider a model with an exogenous mass $\bar{N}_b$ of consumers (buyers, b, he), an exogenous number of legacy merchants $\bar{N}_o$, an endogenous number of entering merchants N_s^e (sellers, s, she), and a market structure where consumers and merchants can match and trade. There is a single good which comes in several varieties and consumers are initially unsure about which variety suits their needs. Consumers and merchants can interact in either a standard retail market or on a platform (M).

1.1 Equilibrium without a platform

We start by describing the model without the platform and then introduce the platform in the next subsection. This helps us understand exactly when the platform will be used by consumers and merchants. We think of the model without the platform as a standard retail market and refer to this market as the *outside market*, indexed by 'o' in the remainder of the paper. We model this outside market as a standard directed search environment with free entry. We assume that there is a mass $\bar{N}_o$ of existing merchants in the outside market, referred to as legacy merchants. Each consumer wants to buy a fixed quantity (normalized to 1) of a good that comes in several varieties $i \in \mathcal{I} = [1, .., I]$. Each consumer has exactly one preferred variety which we refer to as the consumer's *taste*. Tastes are i.i.d. across varieties and across consumers, and all varieties are equally likely to be the preferred one ex-ante. Formally, consumption of variety i delivers utility u_i with $\max_{i \in \mathcal{I}} \{u_i\} = u > 0$, and $u_j = 0$ otherwise.The economy has an information technology with allows consumers to share data about their tastes with merchants. This technology depends on the the information processing capacity of merchants. Formally, the consumer chooses the precision δ of a signal $\sigma \in \mathcal{I}$ about his taste. This signal is observed by the merchants. If $u_i = u$, the signal realization is $\sigma = i$ with probability δ, i.e., $\Pr(\sigma = i \mid u_i = u) = \delta$. Since we assume a uniform prior, Bayes' law implies

$$\Pr(u_i = u \mid \sigma = i) = \frac{\Pr(\sigma = i \mid u_i = u) \Pr(u_i = u)}{\Pr(\sigma = i)} = \frac{\delta \times {}^1\!/I}{\delta \times {}^1\!/I + (1 - {}^1\!/I)(1 - \delta)\frac{1}{I-1}} = \delta.$$

The signal precision δ is therefore also the posterior probability that $u_\sigma = u$. We assume that $\delta \in \left[{}^1\!/I, \bar{\delta}_o\right]$ where $\bar{\delta}_o \leq 1$. The parameter $\bar{\delta}_o$ captures the information processing capacity of the merchants. A larger value of $\bar{\delta}_o$ implies that merchants are able to process more data which in turn can help predict consumers tastes more accurately.

All consumers and merchants are ex-ante identical. To simplify our notation, we present the model in terms of a representative buyer and a representative merchant.

The timing of the model is as follows. There is a single period with three stages:

1. In stage 1, consumers choose a disclosure policy $\delta \in \left[{}^1\!/I, \bar{\delta}_o\right]$.

2. In stage 2, new merchants decide whether to pay the entry cost κ. Let N_s^e be the mass of entering merchants.

3. In stage 3, consumers $N_{b,o}$ and merchants $N_{s,o} = \bar{N}_o + N_s^e$ match in a directed search environment.

We solve the model by backward induction, starting from stage 2. Each merchant must choose a single variety to produce and has z units to sell of that variety.We normalize the marginal cost of production to zero. The matching function is assumed to be constant elasticitiy and the number of matches is $\bar{\alpha}_o \bar{N}_{b,o}^{\gamma} \left(z\bar{N}_o + zN_s^e\right)^{1-\gamma}$, where $N_{b,o} = \bar{N}_b$ in equilibrium. Let $n_o \equiv \left(z\bar{N}_o + zN_s^e\right)/\bar{N}_b$ denote market tightness. The probability that a merchant meets a buyer is given by $\alpha_o(n_o) = \bar{\alpha}_o n_o^{-\gamma}$. The Cobb-Douglass constant elasticity functional form simplifies the exposition but is not crucial. The important feature is that α_o is decreasing and convex.[2]

Merchants post prices and consumers direct their search after observing all the prices. Thus, there are potentially many sub-markets characterized by their price and tightness, (p, n). Following the search literature, we consider a market utility approach in which merchants maximize their payoffs by selecting (p, n) subject to participation by consumers.[3] To sell the good the merchant must clear two hurdles: first, she must match with a buyer – which happens with probability $\alpha_o(n_o)$; second, she must produce the correct variety of the good. Since $\delta \geq 1/I$, it is optimal for these merchants to produce the variety $j = s$ after observing signal $\sigma = s$. To characterize the equilibrium, we first consider the outcomes for a given δ and then discuss the consumer's choice of δ.

The program of the merchant for each unit is

$$v_{s,o} = \max_{n,p} \alpha_o(n)\, \delta p$$

subject to

$$V_{b,o} = n\alpha_o(n)\, \delta(u - p).$$

This problem is equivalent to

$$\max_n \alpha_o(n)\, \delta u - \frac{V_{b,o}}{n}. \tag{1}$$

The first order condition is $\alpha_o'(n)\, \delta u + V_b/n^2 = 0$, which implies $p_o = \left(1 + \frac{n_o \alpha_o'}{\alpha_o}\right) u$. Given our Cobb-Douglass matching function we get $p_o = (1 - \gamma)\, u$. The value for consumers is then

$$V_{b,o} = \gamma u \delta \bar{\alpha}_o n_o^{1-\gamma}.$$

The value function per unit for merchants is $v_{s,o} = (1 - \gamma)\, \bar{\delta}_o u \bar{\alpha}_o n_o^{-\gamma}$. Since each merchant has z units to sell, the total value is $V_{s,o} = z v_{s,o}$ or

$$V_{s,o} = z(1 - \gamma)\, u \delta \bar{\alpha}_o n_o^{-\gamma}. \tag{2}$$

The free entry condition requires

$$V_{s,o} \leq \kappa. \tag{3}$$

The equilibrium in this model is fully characterized by the market tightness n_o. Assuming positive entry, the equilibrium tightness solves

$$n_o\left(\bar{\delta}_o\right) = \left((1 - \gamma)\, \frac{z}{\kappa} u \delta \bar{\alpha}_o\right)^{\frac{1}{\gamma}}. \tag{4}$$

[2]We also require – in all our matching functions – that the number of matches satisfies $\alpha N_b^{\gamma} z N_s^{1-\gamma} \leq \min(N_b, zN_s)$. If we denote by $n = zN_s/N_b$ the market tightness, we need to ensure that $\alpha n^{-\gamma} < 1$ and $\alpha n^{1-\gamma} < 1$, which of course is the same as checking that the probabilities of matching are less than one.

[3]Formally, "money" is a divisible good x that can be produced by both consumers and merchants at cost $C(x) = x$ and yields utility $U(x) = x$, and p is the price of the traded good in terms of "money".

The number of entrants is therefore

$$N_s^e = \max\left(0; z^{\frac{1-\gamma}{\gamma}} \bar{N}_b \left(\left(1-\gamma\right)\frac{u\bar{\alpha}_o\delta}{\kappa}\right)^{\frac{1}{\gamma}} - \bar{N}_o\right). \tag{5}$$

Notice that the number of entrants and thus the market tightness is increasing in the disclosure choice δ

Consider now the choice of δ by consumers. For simplicity we assume that there are no exogenous disclosure costs to consumers and so they solve

$$\max_{\delta\in\left[1/I,\bar{\delta}_o\right]} \gamma u\delta\bar{\alpha}_o n_o^{1-\gamma}.$$

Since the objective function is strictly increasing in δ, we have that $\delta = \bar{\delta}_o$.

A well known result is that the directed search environment is efficient. To see this, consider the planner problem:

$$\max_{N_s^e} \bar{\alpha}_o \bar{N}_b^\gamma z^{1-\gamma}\left(\bar{N}_o + N_s^e\right)^{1-\gamma}\bar{\delta}_o u - \kappa N_s^e.$$

The first order condition of the planner's problem is

$$\left(1-\gamma\right)\bar{\alpha}_o z^{1-\gamma}\left(\frac{\bar{N}_b}{\bar{N}_o + N_s^e}\right)^\gamma \bar{\delta}_o u = \kappa$$

which implies that N_s^e is given by (5), which coincides with the equilibrium condition. As we will show this efficiency result will no longer hold when we introduce the platform.

1.2 Directed Search on the Platform

We now introduce a platform where consumers and merchants can match. The platform M combines two technologies: a matching technology and an information technology. If N_b consumers and N_s merchants are active on the platform, the number of matches is $\bar{\alpha}N_b^\gamma\left(zN_s\right)^{1-\gamma}$. As before, we assume the existence of an information technology which allows consumers to share data about their tastes. We assume that as the information processing technology is improved, individual merchants can no longer process this information themselves. This generates a role for a specialized platform which has the capability of processing large amounts of information. Thus, in order to access this information, merchants must choose to use the platform. To do so, they need to pay a fee to the platform. In exchange for the fee m, the merchant gets access to the matching technology of the platform and to the the consumer "data" (σ, δ). We assume that the processing technology on the platform is given by $\bar{\delta}$ and so the consumers can choose a disclosure level $\delta \in \left[1/I, \bar{\delta}\right]$.

The timing of the model with the platform is:

1. Consumers decide to participate on either the platform or the outside market. If they participate on the platform they choose a disclosure policy $\delta \in \left[1/I, \bar{\delta}\right]$ while if they participate in the outside market, they choose $\delta \in \left[1/I, \bar{\delta}_o\right]$.

2. Merchants decide whether to pay the entry cost κ. Let N_s^e be the number of entrants.

3. The merchants and the platform bargain. If they agree, the merchant pays m and gets the chance to trade on the platform and receive information. If not, the merchant can sell in the outside market.

4. Directed search and matching take place in two venues

   (a) Platform consumers and merchants match on the platform. The merchants receive the consumer's data (σ, δ) via the platform.

   (b) Outside market consumers and merchants match in the outside market. In this case the merchants directly observe (σ, δ)

Suppose that N_s merchants have paid the fee m to purchase the data (σ, δ) and participate in the directed search market. As before, since $\delta \geq 1/I$, it is optimal for these merchants to produce the variety $j = s$ after observing signal $\sigma = s$. Conditional on (N_b, N_s, δ) the platform is a conventional directed search market. Therefore the price is $p = (1-\gamma)\,u$, and the value for the merchant is

$$V_s = z\,(1-\gamma)\,\delta u \bar{\alpha} n^{-\gamma}, \tag{6}$$

where market tightness is $n \equiv zN_s/N_b$, while the value for the buyer is

$$V_b\,(\delta, n) = \gamma \delta u \bar{\alpha} n^{1-\gamma}. \tag{7}$$

Two endogenous variables play a crucial role in our model: the precision of disclosure δ, and the number of merchants N_s. Disclosure is chosen by consumers while the number of merchants is determined by free entry. Disclosure creates gains from trade and therefore increases both V_s and V_b. The number of merchants, on the other hand, increases V_b and decreases V_s.

1.3 Coexistence of the Platform and Outside Market

Suppose that we start from an economy with only the traditional retail market and then introduce the platform. As in Section 1.1, we assume that there is an exogenous mass $\bar{N}_o$ of legacy merchants in outside market. We provide a condition for the platform to be *active*, i.e., attract new entrants. In other words, under this condition, consumers and merchants will choose to participate on the platform. This condition depends on the disclosure choice of consumers, δ, on the platform as well as the technological advantages of the platform captured by $\bar{\delta}$ and $\bar{\alpha}$. Consider the outside market equilibrium with free entry, characterized by market tightness $n_o = \left((1-\gamma)\,\frac{z}{\kappa} u \bar{\delta}_o \bar{\alpha}_o\right)^{\frac{1}{\gamma}}$. Suppose a mass of consumers and a mass of merchants migrate to the platform keeping n_o constant. The outside market would not change. On the platform, if consumers choose δ, merchants would then get $\hat{V}_s = \bar{\alpha}\delta / \left(\bar{\alpha}_o \bar{\delta}_o\right)\kappa$, which is more than κ if and only if $\bar{\alpha}\delta > \bar{\alpha}_o \bar{\delta}_o$.

Lemma 1. *If $\bar{\delta}_o > \delta\bar{\alpha}/\bar{\alpha}_o$ the platform is inactive and the equilibrium is the one in Section 1.1. If $\bar{\delta}_o < \delta\bar{\alpha}/\bar{\alpha}_o$ the platform attracts all the new entrants and only legacy merchants remain in the outside market.*

Proof. First, suppose we have an equilibrium in which both the platform and outside option are active. By this we mean that there are new entrants on both the platform and in the outside market. Then it must be that $V_s = V_{s,o} = \kappa$ and $V_b = V_{b,o}$. These two equations imply that $n = n_o$ and thus it must be that

$$\bar{\delta}_o \bar{\alpha}_o = \delta \bar{\alpha}.$$

Therefore, if this equality does not hold then only one of the two can be active (i.e. attract new entrants) in equilibrium. Suppose now that $\delta \bar{\alpha} < \bar{\alpha}_o \bar{\delta}_o$. Suppose for contradiction that only the platform is active. Since consumers are always indifferent we must have that

$$V_b = \delta \gamma u \bar{\alpha} n^{1-\gamma} = V_{b,o} = \bar{\alpha}_o \bar{\delta}_o \gamma u n_o^{1-\gamma}$$

which implies that

$$n = \left(\frac{\bar{\alpha} \delta}{\bar{\alpha}_o \bar{\delta}_o} \right)^{-\frac{1}{1-\gamma}} n_o.$$

Since merchants can freely enter the outside market, it must be that

$$z (1 - \gamma) u \bar{\delta}_o \bar{\alpha}_o n_o^{-\gamma} \leq \kappa$$

but then

$$z (1 - \gamma) u \delta \bar{\alpha} n^{-\gamma} < \left(\frac{\delta \bar{\alpha}}{\bar{\delta}_o \bar{\alpha}_o} \right)^{\frac{\gamma}{1-\gamma}} z (1 - \gamma) u \bar{\alpha}_o \bar{\delta}_o n_o^{-\gamma} < \kappa$$

which is a contradiction. Thus, the platform can never be active. Finally, suppose that $\delta \bar{\alpha} > \bar{\delta}_o \bar{\alpha}_o$. Suppose for contradiction that only the outside market is active. Then it must be that

$$z (1 - \gamma) u \delta \bar{\alpha} n^{-\gamma} \leq \kappa$$

and so

$$z (1 - \gamma) u \bar{\delta}_o \bar{\alpha}_o n_o^{-\gamma} \leq \left(\frac{\bar{\alpha}_o \bar{\delta}_o}{\bar{\alpha} \bar{\delta}} \right)^{\frac{\gamma}{1-\gamma}} z (1 - \gamma) u \delta \bar{\alpha} n^{-\gamma} < \kappa$$

which is a contradiction. $\square$

To be active, the platform must have a technological advantage in either information or matching. When the platform is active all entrants go to the platform, $N_s = N_s^e$ and thus the only merchants operating in the outside market are the legacy merchants, $\bar{N}_o$. Consumers, on the other hand, must be indifferent between searching on the platform or on the outside market. The indifference condition $V_b = V_{b,o}$ then implies

$$\bar{\alpha} \delta n^{1-\gamma} = \bar{\alpha}_o \bar{\delta}_o n_o^{1-\gamma} \tag{8}$$

where $n_o = z \bar{N}_o / N_{b,o}$ is market tightness in the outside market. Merchants must pay a fee m to the platform, therefore the entry condition becomes

$$V_s - m = \kappa. \tag{9}$$

Market clearing for consumers requires

$$N_b + N_{b,o} = \bar{N}_b. \tag{10}$$

To complete the model we need to derive the fee, m, from the bargaining game between the platform and the merchants.

1.4 Platform Pricing

There are two distinct externalities at the core of our paper which derive from the endogenous interaction between the quality of information and the market power of the platform. The first externality arises from the fact that the platform partially controls access to consumers: we call it the *gatekeeper* externality. The second externality arises from the ability of the platform to compete with the merchants by replicating their successful products: we call it the *copycat* externality.

Gatekeeper Pricing Let V_M denote the outside option of the platform, which we take as given for now and endogenize later with the Copycat technology. The outside option of the merchant is to sell in the outside market, with value $V_{s,o}$ defined in (2). It is worth noting for later that $V_{s,o}$ is a function of n_o which depends on δ through the consumer indifference condition (8). Let θ be the bargaining power of the platform vis-a-vis its merchants. The Nash bargaining problem is then:

$$\max_m (m - V_M)^\theta (V_s - m - V_{s,o})^{1-\theta},$$

assuming that $V_s \geq V_M + V_{s,o}$. The negotiated fee is therefore $m = \theta (V_s - V_{s,o}) + (1-\theta) V_M$, and the total payoff for the merchant if she buys access to the platform is $V_s - m = (1-\theta) V_s + \theta V_{s,o} - (1-\theta) V_M$. Thus, the free entry condition is

$$(1-\theta) V_s + \theta V_{s,o} - (1-\theta) V_M = \kappa. \tag{11}$$

The key point here is that, when the platform becomes more efficient either in matching or in information gathering, the value $V_{s,o}$ of selling in the outside market decreases. When $V_{s,o}$ is low, the platform becomes a de facto Gatekeep for access to consumers. This allows the platform to increase the fees it charges to its merchants. The Gatekeeper effect is an important feature of all digital platforms from Uber to OpenTable and Amazon.

Copycat Technology In addition to controlling access to consumers, some platforms also have the option to compete with their own merchants. This feature is not as universal as the gatekeeper feature. For instance, most booking platforms do not act as sellers (e.g., OpenTable, Ticketmaster, etc.). Other platforms, however, have the option to sell their own products. For instance, Amazon has its AmazonBasics private-label brand (see for example Khan (2016) and Mattiolo (2020)). Mattiolo (2020) documents how Amazon uses data to "decide how to price an item, which features to copy or whether to enter a product segment based on its earning potential".

To capture this idea, we assume that the platform can produce the good itself at cost c, and sell it directly to consumers . The platform controls the search environment and thus can ensure that it matches with a consumer and it is subject to the same information requirements. On the other hand, it must incur a production cost, $c > 0$. The interpretation of this cost is that the platform is less efficient than the merchants at producing the good. The expected profit of the platform is $\delta p - c$, and therefore

$$V_M = \max\left\{0; \delta\left(1-\gamma\right)u - c\right\}. \tag{12}$$

The important feature is that V_M is a (weakly) increasing function of information precision δ.

To summarize, the gatekeeper effect is always present and captured by the endogenous outside option $V_{s,o}$. The copycat effect may or may not be present, depending on the cost parameter c. In particular, our model nests the pure gatekeeper case by letting c go to infinity.

1.5 Decentralized Equilibrium

We describe the equilibrium in two steps. First given disclosure level δ, we characterize an equilibrium on the platform (platform equilibrium). Next, we endogenize δ and define an equilibrium for the environment (decentralized equilibrium).

Given any level of disclosure δ chosen by consumers ex-ante, the platform-equilibrium is characterized by the two market tightness variables (n, n_o) which solve the free entry condition (11) and the consumer indifference condition (8).Combining the two conditions we obtain an expression that relates the platform market tightness, n, to the level of information disclosure δ

$$n^{\gamma} = z\bar{\alpha}\left(1-\gamma\right)u\frac{G\left(\delta\right)}{\kappa + \left(1-\theta\right)V_M\left(\delta\right)}, \tag{13}$$

where

$$G\left(\delta\right) \equiv \left(1-\theta\right)\delta + \theta\left(\frac{\bar{\alpha}_o\bar{\delta}_o}{\bar{\alpha}}\right)^{\frac{1}{1-\gamma}}\delta^{-\frac{\gamma}{1-\gamma}}.$$

Equation (13) shows that the effect of δ on n is ambiguous. The denominator is increasing in δ. This captures the copycat effect in the outside value of the platform. When δ is high, the copycat option is valuable ($V_M\left(\delta\right)$ is large), the platform has a stronger bargaining position and so seller entry and market tightness decrease. The function $G\left(\delta\right)$ captures gains from trade and the gatekeeper effect. The first term, $\left(1-\theta\right)\delta$, is increasing in δ since gains from trade encourage entry and increases market tightness. The second term is decreasing in δ and captures the gatekeeper effect, which arises due to the seller's outside option. When information improves, a larger mass of consumers choose to participate on the platform, which depresses the outside market and therefore the outside option of the merchants, $V_{s,o}$. Taking log-derivatives we obtain

$$\gamma\frac{n'}{n} = \frac{G'\left(\delta\right)}{G\left(\delta\right)} - \frac{V_M'\left(\delta\right)}{\frac{\kappa}{1-\theta} + V_M\left(\delta\right)}$$

A key insight of the model is that an increase in δ can lead to an *decrease* in market tightness.

Lemma 2. *Market tightness n is a decreasing function of δ if and only if*

$$\frac{G'(\delta)}{G(\delta)} - \frac{V_M'(\delta)}{\frac{\kappa}{1-\theta} + V_M(\delta)} < 0.$$

The condition in Lemma 2 is likely to hold when the copycat effect $V_M'(\delta)$ is large, or when the gatekeeper effect is large. A sufficient condition for n to be decreasing in $\bar{\delta}$ is $G'(\delta) < 0$ or

$$\frac{\gamma}{1-\gamma}\frac{\theta}{1-\theta}\left(\frac{\bar{\alpha}_o\bar{\delta}_o}{\bar{\alpha}\delta}\right)^{\frac{1}{1-\gamma}} > 1,$$

which holds when either θ (bargaining power of the platform) or γ (importance of consumer access) is high. As we will see, this can potentially justify regulation of data gathering by the platform.

To close the model we need to solve for δ. As in the previous subsection, the program of a buyer is

$$\max_{\delta} V_b(\delta, n),$$

subject to (7). Since we assume that there are no exogenous disclosure costs, the consumers' private choice is $\delta = \bar{\delta}$.

Using the above results, we can now characterize the decentralized equilibrium.

Proposition 1. *The decentralized equilibrium (n, δ) is characterized by $\delta = \bar{\delta}$ and the free-entry market tightness condition (13).*

2 Social Planner

2.1 Planner's Solution

Consider the problem of a social planner who cares about the welfare of consumers and merchants. We ignore the welfare of the platform in this baseline calculation since we did not model the platform entry decision in the baseline model. Since the expected profits of the merchants is pinned down by free entry, the planner's welfare is completely characterized by that of consumers.

We assume that there exists a threshold $\hat{\delta}$ such that for $\delta \leq \hat{\delta}$, merchants can directly process the information themselves while for $\delta > \hat{\delta}$, the platform must be used. As before, $\bar{\delta} > \hat{\delta}$ denotes the technological frontier of information processing that is available to the planner.

The problem for the planner is thus

$$\delta^* = \arg\max_{\delta \leq \bar{\delta}} \bar{N}_b \left[W^o(\delta)\mathbf{1}_{\delta \leq \hat{\delta}} + W^p(\delta)\mathbf{1}_{\delta > \hat{\delta}}\right],$$

where $W^o(\delta) = V_{b,o}(\delta, n_o(\delta))$, $W^p(\delta) = V_b(\delta, n(\delta))$, $n_o(\delta)$ is given by (4), and $n(\delta)$ is given by (13). Thus the marginal value of information from the perspective of the planner is

$$\frac{dW^o}{d\delta} = \gamma u\bar{\alpha}_o n_o^{1-\gamma} + \gamma(1-\gamma)\delta n_o' n_o^{-\gamma}$$

if $\delta \leq \hat{\delta}$ and

$$\frac{dW^p}{d\delta} = \gamma u \bar{\alpha} n^{1-\gamma} + \gamma (1-\gamma) \delta n' n^{-\gamma}$$

if $\delta > \hat{\delta}$. The first terms in the above equations captures the private marginal benefit of greater disclosure. Since we have assumed no exogenous personal costs of disclosure this term is always positive.

The second term captures the effect of disclosure on market tightness and this is not internalized by the consumers when choosing their disclosure rule. Using (4) it is easy to see that this term in the first equation (i.e. when the platform is inactive) is always positive. Thus, conditional on merchants processing the information, the planner always has a greater marginal value of information than private agents. When the platform is active however, we have seen that n' can be less than zero. In this case, the planner has a lower marginal value of information than the private sector.

Proposition 2. *If the data processing technology δ is less than $\hat{\delta}$, then the equilibrium is efficient. If the data processing technology is larger than $\hat{\delta}$ and $n'(\delta) < 0$, the equilibrium is inefficient and the social planner wants to restrict information disclosure.*

From Lemma 2 we know that a sufficient condition is $\frac{\gamma}{1-\gamma} \frac{\theta}{1-\theta} \left(\frac{\bar{\alpha}_o \bar{\delta}_o}{\bar{\alpha} \bar{\delta}} \right)^{\frac{1}{1-\gamma}} > 1$.

2.2 Implications for Regulation and Entry

The most direct implication of our model is that when the platform is active, consumers may not choose the socially optimal degree of information sharing because they ignore the impact of their disclosure on the entry decisions of merchants which in turn affects their probability of being matched. In other words, there is a relationship between consumer data and the degree of competitiveness in product markets which is not internalized by consumers.

What is surprising about our result is that, in our model, disclosure can be too high even though additional information directly increases gains from trade. This stands in contrast to papers which explicitly assume that one consumer's privacy is directly hurt by the disclosure of other consumers' information, as in Acemoglu et al. (2019). The key for the excessive disclosure result in our model lies in the market power of the platform vis-a-vis the merchants. This can happen either because the platform attracts many consumers and depresses the value of the outside market for the merchants (gatekeeper externality) and/or because better disclosure allows the platform to create copycat products that compete with the merchants' products (copycat externality). Excessive disclosure is likely when either θ (bargaining power of the platform) or γ (importance of consumer access) is high. In this case, one way to implement the socially optimal outcome is to limit the data platforms can collect from consumers. In the context of our model, regulators could impose a cap $\bar{\delta}^{sp} = \delta^* < \bar{\delta}$ on the platform's ability to gather information from consumers.

Notice that neither the gatekeeper nor the copycat externalities are individually necessary to obtain our results. Thus, our model can nest platforms like Uber and OpenTable that collect vast amounts of consumer data but cant directly produce and sell the desired goods themselves. The reason for this is that even if V_M is independent δ, the gatekeeper externality is still present and so our main result that there can be too much data sharing in equilibrium can still hold.

Note that so far we have characterized the effect of disclosure on market tightness $n = zN_s/N_b$.When the platform is inactive, tightness (and the number of entering merchants) is always increasing in disclosure. However, when the platform is active, tightness can decrease. In numerical simulations (see Figure 1), we find that N_b can be a non-monotonic function of δ while N_s is increasing in δ for low values of δ and decreasing afterwards. To see why start from (13) which defines the function $n(\delta)$. Using the indifference condition (8) and the market clearing condition $N_b + N_{b,o} = \bar{N}_b$ we have

$$N_b = \bar{N}_b - \left(\frac{\bar{\alpha}_o}{I\delta\bar{\alpha}}\right)^{\frac{1}{1-\gamma}} \frac{z\bar{N}_o}{n}$$

which shows that N_b is increasing in δ as long as $dn/n > -(1-\gamma)/\delta$. This condition holds for most parameter values, but we can find counter-examples. For N_s, on the other hand, we have

$$N_s = \frac{n}{z}N_b = \frac{n}{z}\bar{N}_b - \left(\frac{\bar{\alpha}_o}{I\delta\bar{\alpha}}\right)^{\frac{1}{1-\gamma}} \bar{N}_o$$

The second term is increasing in δ while the first term decreases if δ is large enough. Therefore, when δ is high, we are more likely to be in the region where N_s decreases with δ.

Suppose due to technical progress the information processing capacity increases over time. So long as $\bar{\delta}_o \leq \hat{\delta}$, the equilibrium level of disclosure (equal to $\bar{\delta}_o$) is always efficient. However as $\bar{\delta}_o > \hat{\delta}$ the platform must be used in order to take advantage of this technology. However, as we have shown above, in this region, the equilibrium level of disclosure can be inefficient. We summarize this in the following result.

Proposition 3. *Information technologies (IT) that improve data processing increase the risk of excessive disclosure.*

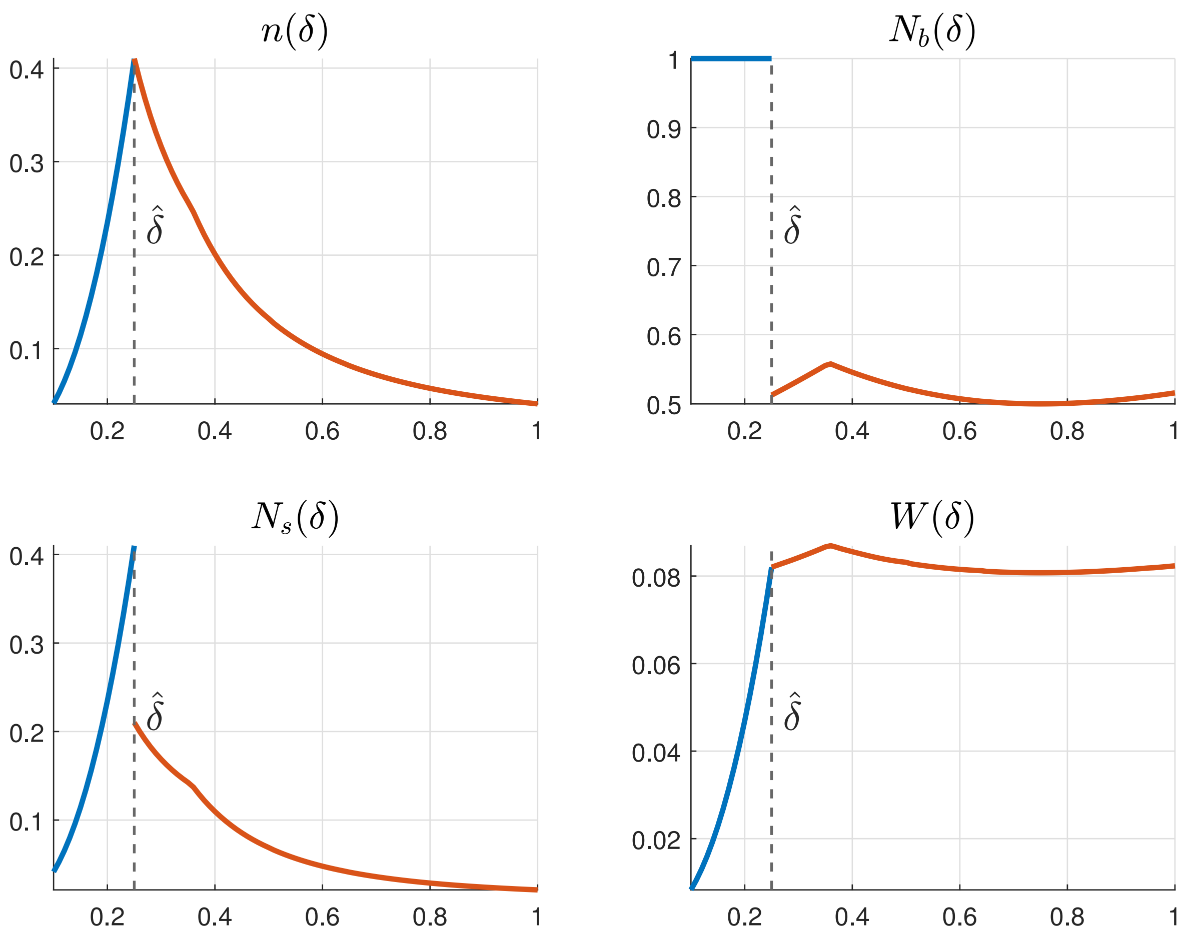
Proposition 3 helps us understand the qualitative difference between a modern platform such as Amazon and traditional retail store. Retail stores are two-sided and operate a matching technology. They typically operate their own brands that compete with their outside merchants. They also gather some information about their customers. What then, is the difference between CVS and Amazon? Why would regulators be worried about the excessive information gathering in one case but not in the other? Some commentators want to prevent platforms from also being able to sell private label products on their site. Why would this policy not apply similarly to traditional retail stores?

Our model provides a precise theoretical argument. Disclosure that requires filling paper work or active action by users is costly. Gathering digital information, on the other hand, is cheap and often passive (e.g., cookies, browsing history). Proposition 3 says that this increases the risk of a negative impact on entry.

3 Heterogeneous Consumers

Our benchmark model assumes that consumers are ex-ante identical. In this section we study the role of heterogeneity among consumers. We assume that consumers, indexed by j, are heterogeneous in their utilities $u_j \in [\underline{u}, \overline{u}]$ and face a fixed entry cost η. Conditional on entry they either participate on the platform

Figure 1: *Numerical Example*



Parameters: $I = 10$, 4, $u = 7$, $\theta = 0.95$, $\bar{\alpha} = 0.2$, $\bar{\alpha}_o = 0.2$, $\kappa = 0.3$

or in traditional retail (outside market). We assume that u_j is observable.[4] Conditional on u_j the directed search market is as before. There exists some threshold type u^* so that all types above enter. The merchants' free entry condition is then

$$z(1-\gamma)U(u^*)\left((1-\theta)\delta\bar{\alpha}n^{-\gamma} + \theta\bar{\alpha}_o\bar{\delta}_o n_o^{-\gamma}\right) = \kappa + (1-\theta)V_M(\delta),$$

where

$$U(u^*) \equiv \mathbb{E}[u \mid u \geq u^*] = \frac{\int_{u^*} uf(u)\,du}{1-F(u^*)}.$$

[4]This assumption is without loss of generality. Even if utilities are privately observed by consumers one can construct a separating equilibrium that is identical to that in this section.

Note that U is an increasing function. When u is observable, the indifference condition holds type by type and so (8) is unchanged. Combining these conditions we obtain

$$n^{\gamma} = z\bar{\alpha}(1-\gamma)U(u^*)\frac{G(\delta)}{\kappa + (1-\theta)V_M(\delta)},$$

with the same function $G(\delta) \equiv (1-\theta)\delta + \theta\left(\frac{\bar{\alpha}_o\bar{\delta}_o}{\bar{\alpha}}\right)^{\frac{1}{1-\gamma}}\delta^{-\frac{\gamma}{1-\gamma}}$ as in the benchmark case. Taking log differentials with respect to δ, we get

$$\gamma\frac{n'}{n} = \varepsilon^*\frac{(u^*)'}{u^*} + \frac{G'(\delta)}{G(\delta)} - \frac{V_M'}{\frac{\kappa}{1-\theta} + V_M}$$

where $\varepsilon^* = \frac{u^*U'(u^*)}{U(u^*)} > 0$ is the elasticity of the conditional expectation.

The marginal type is defined by the free entry condition of consumers

$$\gamma\delta u^*\bar{\alpha}n^{1-\gamma} = \eta,$$

so

$$\frac{(u^*)'}{u^*} = -(1-\gamma)\frac{n'}{n} - \frac{1}{\delta},$$

and combining these conditions we have the following Lemma

Lemma 3. *With free entry by heterogenous consumers and perfect price discrimination, the impact of information disclosure on market tightness is given by*

$$\frac{n'}{n}(\gamma + (1-\gamma)\varepsilon^*) = \frac{G'(\delta)}{G(\delta)} - \frac{V_M'}{\frac{\kappa}{1-\theta} + V_M} - \frac{\varepsilon^*}{\delta}.$$

Note that our baseline model with homogeneous consumers simply corresponds to $\varepsilon^* = 0$, and the sufficient condition in Lemma 2 is unchanged. When consumers are heterogenous, there is an extra factor than can lead to decreasing tightness: as δ increases, the quality of the marginal consumer decreases and thus, the gains from trade decrease. Therefore, with heterogeneous consumers, the externality associated with too much information being shared in equilibrium is more likely to exist.

4 Conclusion

Our model helps explain why excessive data sharing might not have been a problem with traditional retail brick-and-mortar stores but could become one with internet platforms. We show that there can be both under-sharing and over-sharing depending on the information and matching technologies. When matching is efficient and platforms can collect large amounts of information about consumers, then excessive sharing is more likely.

Our model focuses on the case of one good. A natural extension is to introduce several goods and to study the decision by consumers and merchants to single-home or to multi-home.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019). Too much data: Prices and inefficiencies in data markets. Working Paper.

Ambrus, A., E. Calvano, and M. Reisinger (2016, August). Either or both competition: A "two-sided" theory of advertising with overlapping viewerships. *American Economic Journal: Microeconomics 8*(3), 189–222.

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

Bergemann, D. and A. Bonatti (2019). Markets for information: An introduction. *Annual Review of Economics 11*, 85–107.

Bergemann, D., A. Bonatti, and T. Gan (2019). The economics of social data.

Bergemann, D., A. Bonatti, and A. Smolin (2018, January). The design and price of information. *American Economic Review 108*(1), 1–48.

Bergemann, D., B. Brooks, and S. Morris (2015, March). The limits of price discrimination. *American Economic Review 105*(3), 921–57.

Caillaud, B. and B. Jullien (2003). Chicken and egg: Competition among intermediation service providers. *The RAND Journal of Economics 34*(2), 309–328.

Gutierrez, G. (2020a). The seller welfare standard. NYU Stern Working Paper.

Gutierrez, G. (2020b). Ten facts about Amazon. NYU Stern Working Paper.

Ichihashi, S. (2019). Online privacy and information disclosure by consumers. *American Economic Review*.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *The American Economic Review 75*(3), 424–440.

Khan, L. M. (2016). Amazon's antitrust paradox. *Yale lJ 126*, 710.

Mattiolo, D. (2020). Amazon scooped up data from its own sellers to launch competing products; contrary to assertions to congress, employees often consulted sales information on third-party vendors when developing private-label merchandise. *Wall Street Journal (Online)*.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of political Economy 105*(2), 385–411.

O'Mara, M. (2019). *The Code: Silicon Valley and the Remaking of America*. Penguin Press.

Park, K. F., R. Seamans, and F. Zhu (2020). Homing and platform responses to entry: Historical evidence from the u.s. newspaper industry. *Strategic Management Journal n/a*(n/a).

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

Shimer, R. (1996). *Essays in search theory*. Ph. D. thesis, Massachusetts Institute of Technology.

Wright, R., P. Kircher, B. Julien, and V. Guerrieri (2019). Directed search and competitive search: A guided tour. *Journal of Economic Literature*.