# Platform Competition and Information Sharing

Georgios Petropoulos,* Bertin Martens,† Geoffrey Parker,‡ Marshall Van Alstyne§

September 17, 2023

**Abstract**

Digital platforms, empowered by artificial intelligence algorithms, facilitate efficient interactions between consumers and merchants that allow the collection of profiling information which drives innovation and welfare. Private incentives, however, lead to information asymmetries resulting in market failures. This paper develops a product differentiation model of competition between two platforms to study private and social incentives to share information. Sharing information can be welfare-enhancing because it solves the data bottleneck market failure. Our findings imply that there is scope for the introduction of a mandatory information sharing mechanism from big tech to their competitors that help the latter to improve their network value proposition and become more competitive in the market. The price of information in this sharing mechanism matters. We show that price regulation over information sharing like the one applied in the EU jurisdiction increases the incentives of big platforms to collect and analyze more data. It has ambiguous effects on their competitors that depend on the exact relationship between information and network value.

*Keywords: Information sharing, digital platforms, data bottleneck, data portability.*
JEL Classification: D47, D82, K21, L21, L22, L40, L41, L43, L51, L86.

*Massachusetts Institute of Technology, Stanford University and CESifo Network Affiliate. Email: gpetrop@mit.edu.

†Bruegel and Tilburg University. Email: bertin.martens@gmail.com.

‡Dartmouth College. Email: Geoffrey.G.Parker@dartmouth.edu.

§Boston University. Email: mva@bu.edu.

# 1 Introduction

This paper deals with information structures in digital multi-sided markets as well as the implications of information sharing for platform competition and welfare. We explore the implications of a regulatory intervention that enforces a mandatory information sharing mechanism and we study how such a mechanism should be designed to ensure market efficiency and high welfare standards.

Digital platforms fundamentally changed the way information is collected and processed. Traditional offline markets, such as town markets, are organized as multi-sided platforms. Buyers and sellers gather in a physical place and benefit from number-driven network effects: more buyers attract more sellers, and vice versa. However, the platform does not dispose of information collection technology and cannot facilitate matching between users. Platforms cannot collect information on goods and prices that sellers offer nor on buyer preferences or transactions. Users collect their own market information to make their transaction decisions. This decentralized information system is economically inefficient for two reasons. First, because information collection is costly, buyers often collect only part of all available information. With incomplete information, they are more likely to make inefficient decisions. Second, it is socially wasteful because, in the absence of a data-sharing mechanism, each buyer has to collect the same information again.

Online platform markets centralize information and collect a much richer set of market information, including user characteristics and aggregated user interaction and transaction data on all sides of the market. Economies of scale and scope in aggregation and centralization of market information by the platform can, in principle, make markets more transparent and give users access to more complete market information compared to decentralized markets, thereby enabling more efficient decision-making by platform users (Constantinides, Henfridsson, & Parker, 2018; Munger, 2015; Parker, Van Alstyne, & Choudary, 2016). Moreover, economies of scope in the re-use of that market information substantially reduce the social cost of information collection (George, Haas, & Pentland, 2014; Manyika et al., 2011; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012).

The combination of these two factors is sometimes labeled as data-driven network effects (Gregory, Henfridsson, Kaganer, & Kyriakou, 2021, 2022; Prüfer & Schottmüller, 2021; Schaefer & Sapi, 2020; Stucke & Grunes, 2016; Tucker, 2019): the more user data a platform collects, the higher the quality of the service and the network value that it can deliver to its users.

How data creates value and improves firms' market performance has been a topic of attention in multiple disciplines and various contexts. For example, empirical evidence suggest that the employment of data warehouses which store and analyze information about client behaviors and preferences, significantly helps firms to improve their customer services and increase in that way their market performance and profitability (Cooper, Watson, Wixom, & Goodhue, 2000; Wixom & Watson, 2001). Moreover, cross-sectional survey data illustrates that big data enables manufacturing firms to better understand their customer preferences and become more successful (Woerner & Wixom, 2015). Big data also expands the business intelligence possibilities of firms so that they can be more successful Chen, Chiang, and Storey (2012) In fact, benefits from data can be maximized if firms adopt a proper framework for data analysis (Kitchens, Dobolyi, Li, & Abbasi, 2018) and a well-designed data governance structure (Otto, 2011).

Further empirical evidence from other sectors point to the same direction. In the financial industry, data has a multi-billion value for corporations. This value has in fact increased by 25% between 2015 and 2018 (Abis & Veldkamp, 2020). Data appears to be particularly valuable for digital platforms. Big data (including data collected from platform users) helps Amazon to improve the accuracy of its forecasting exercises concerning the number of products, and the number of time periods for which each product is available for sale (Bajari, Chernozhukov, Hortaçsu, & Suzuki, 2019). Furthermore, there are significant economies of scale to data in internet search, or in other words, the quality of search results improves with more data on previous searches (Schaefer, Sapi, & Lorincz, 2018). The same applies for the relationship between data and the quality of content recommendations in online media platforms (Claussen, Peukert, & Sen, 2019).

It is important to emphasize that in the platform context, it is not only the quantity of market data but also the networked quality of the data that helps improve platform services (Arnold, Marcus, Petropoulos, & Schneider, 2018; Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016). It is not the collection of data on each user separately, but the collection of interaction data co-generated by two or more users that generates the data network externality: the social value of interaction data is higher than the sum of separate user data (Duch-Brown, Martens, & Mueller-Langer, 2017; Martens, 2021). The denser the network of interactions, the wider the gap between the social value of networked data and the private value of individual user data. In contrast, decentralized offline markets do not benefit from data-driven network effects.

In practice, however, the centralization of market information by platform operators creates new problems. Online platforms do not share their full market information with their users. Platforms are profit-maximizing companies that retain exclusive control of their market data in order to maximize revenue from the matching service that they offer to users (Evans, 2009; Jones & Tonetti, 2020). They only share narrow information signals with users, through organic rankings and advertising services (Goldfarb & Tucker, 2011). These partial information signals drive a wedge between the interests of users on the same or on different sides of the market. The asymmetric distribution of information between platforms and their users creates bottlenecks for the realization of all data-driven social welfare gains (Parker, Petropoulos, & Van Alstyne, 2022).

This paper's first contribution is that it illustrates such a data bottleneck market failure due to asymmetric information. It builds a simple product differentiation model of platform competition and identifies the channels through which the existence of the data bottleneck market failure limits competition allowing the platform with superior information to extract a disproportionately high share of the value that is created in the ecosystem.

The data bottleneck is a market failure because under great information asymmetry between two competing platforms, the majority of consumers are forced to single-home, increasing the market power of the platform with the data advantage: By attracting the majority of consumers, the platform extracts excessive rents on the supply side by charging a monopoly price.

Data bottlenecks between platforms become particularly onerous when welfare-enhancing network effects result in welfare-reducing monopolistic markets (Crémer, de Montjoye, & Schweitzer, 2019; Furman, Coyle, Fletcher, McAuley, & Marsden, 2019; Scott Morton, 2019). Number-driven network effects motivate users to join the platform with the largest number of users (Dou & Wu, 2021; McIntyre & Srinivasan, 2017; Parker & Van Alstyne, 2005; Rochet & Tirole, 2003). That platform is also likely to collect most market information, offer the strongest data-driven network effects, and leave little to competing platforms. So, exclusive access to large datasets is perceived as an important driver of monopolistic behavior (Cabral et al., 2021). While information is non-rival (Jones & Tonetti, 2020; Lambrecht & Tucker, 2015; Lerner, 2014), information collection is often rival because users will only engage in particular interactions in a single platform. Multi-homing is often discouraged by switching costs and locked in effects, including costs linked to the fragmentation of users' data between platforms. Data-driven network effects amplify traditional number-driven network effects and further entrench platform market positions. More and better data help to improve the quality of algorithms through learning-by-doing within and across users that further entrench market positions for incumbent platforms (Hagiu & Wright, 2020). So, differences in access to data can distort competition between platforms.

The new European regulation, the Digital Markets Act (DMA) recognizes that this problem is particularly acute for 6 large gatekeeper platforms (Alphabet, Amazon, Apple, ByteDance, Meta, Microsoft) and in particular in their provision of 22 core platform products/services that fall into 8 broad categories: search, messaging, social media, intermediation, advertising, video sharing, browser, operating system (European Commission, 2023). It has as an objective to address these information asymmetry concerns on market competition and consumer choice by reducing gatekeeper data exclusivity (European Commission, 2022). This is done by incorporating specific rules with the introduction of data access rights and by facilitating data sharing from gatekeepers to their competitors and other firms in order to create more data-symmetric digital markets. The main rationale is that market information collected by one platform can be useful to improve the services of another similar or complementary platform (Condorelli & Padilla, 2020; Eisenmann, Parker, & Van Alstyne, 2011). Hence, creating an efficient data-sharing mechanism between platforms may overcome the data bottleneck problem.

Specifically, the DMA builds on the the General Data Protection Regulation (GDPR) which establishes the data portability right for individual users over their data: Namely, individuals have the right to port their data located in digital firms, free of charge to other digital undertakings (European Commission, 2016). Article 6 of the DMA includes the following relevant obligations over data access and sharing (European Commission, 2022):

- *Paragraph 9* obliges gatekeepers to ”provide end users and third parties authorised by an end user, at their request and free of charge, with effective portability of data provided by the end user or generated through the activity of the end user in the context of the use of the relevant core platform service...”

- *Paragraph 10* extends the data portability right to business users of big platforms (provided that individuals give their consent when personal data is ported): ”The gatekeeper shall provide business users and third parties authorised by a business user, at their request, free of charge, with effective, high-quality, continuous and real-time access to, and use of, aggregated and non-aggregated data, including personal data, that is provided for or generated in the context of the use of the relevant core platform services or services provided together with, or in support of, the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users.”

- *Paragraph 11* obliges gatekeepers in the core platform service of online search to ”provide to any third-party undertaking providing online search engines, at its request, with access on fair, reasonable and non-discriminatory terms to ranking, query, click and view data in relation to free and paid search generated by end users on its online search engines. Any such query, click and view data that constitutes personal data shall be anonymised.”

So, both individual and business users of gatekeeper platforms have the right to share their data located in the gatekeeper platforms to other digital firms of the online ecosystem. In addition, Google, the only identified gatekeeper with core services in online search, has the obligation to share its market data with its competitors. An interesting element is that data sharing, when it is initiated by the user should take place **free of charge** without allowing the gatekeeper platform to extract rents from sharing the data. In contrast, in search engines, where Google is obliged to share relevant market data with their competitors, the platform is eligible to ask for a positive price over data sharing which should be below the monopoly price and obey some **fair, reasonable and non-discriminatory** (FRAND) terms (whose exact definition is work in progress).

Motivated by the DMA, the second contribution of this paper is to show how information sharing from a gatekeeper platform that has access to a lot of market-relevant data (including users' interaction data) to a smaller competitor platform with data and information disadvantage affect platform competition and welfare. Information sharing helps the competing smaller platform to better understand the preferences of their users and improve their services. It improves the network value it can provide to consumers. This gives rise to multi-homing on the consumer side which implies that the market power associated with the data advantage will decline and the dominant platform will have to reduce its price on the supply side resolving the data bottleneck market failure. As a result, information sharing through increasing platform competition and enabling multi-homing improves both consumer and seller surplus.

The underlying economic mechanism for solving the data bottleneck is analogous to the competitive bottleneck market failure(Armstrong, 2006; Armstrong & Wright, 2007) and how allowing consumers to multi-home can resolve it (Bakos & Halaburda, 2020; Belleflamme & Peitz, 2019b). However, in the case of the data bottleneck, there is an additional dimension of network value. Consumers even if they can in principle multi-home to a competing platform, in equilibrium, they will never do that if the network benefit they get is significantly smaller than the value they get from the gatekeeper platform. So, effective information sharing that increases the network value of the small platform is a necessary ingredient for improving market efficiency and welfare.

The paper then looks at how we should design information sharing mechanisms from a big platform to its smaller competitor, in order to improve competition and consumer welfare. The starting point is to study whether price regulation alone can induce the gatekeeper to share data such that private incentives for data sharing coincide with social ones. The paper provides a full characterization of how private incentives for data sharing depend on how restrictive the respective price regulation is.

We find that if data regulation is only about the price of data sharing, private incentives for data sharing are in principle lower than social ones. Especially when price regulation does not leave sufficient rents to the dominant platform, under a free-of-charge rule over information sharing, similar to the one of the GDPR and DMA. However, we find that even if the gatekeeper has monopoly rights over data, it may still have lower incentives than in the social optimal case. This is because information sharing enables the small platform to become more competitive and capture a larger share of the market. The resulting market loss for the gatekeeper, in most cases, outweighs the benefit it gets by selling information to the small platform when the level of shared information approaches the socially optimal level.

We also adopt a more dynamic perspective, studying how such a mandated information sharing mechanism may affect the incentives of platforms to invest in market data collection and analysis. We show that information sharing reduces gatekeeper's investment to data but when data exhibits increasing returns to network value, sharing information is still welfare improving. An important prediction of our model is that the free-of-charge rule over the price of data sharing does not adversely affect data investments. The negative implications of information sharing on investment come mostly from the increased competitive pressure it incorporates. In fact, the free-of-charge rule is the pricing schedule that maximizes the incentives of the gatekeeper to invest on its data capacity in order to reduce the competitive pressure that data sharing introduces.

Our results have some important practical implications. Regulation over the price of data alone does not suffice. What we need on top of that is to mandate an efficient mechanism of information sharing and oblige gatekeepers with a great data advantage to participate in it and share their collected user interaction and market data with competitors.

The remainder of this paper is organized as follows. Section 2 presents the main model. Section 3 analyzes the model and presents the main results. Section 4 presents the implications of information sharing on platforms' investments on data collection and analysis. Section 5 discusses the research implications and the contribution of our paper. Section 6 presents the limitations of our model and analysis as well as the implications of our work for future research. Section 7 refers to practical aspects of our results and concludes.

# 2 Model

Consider a big gatekeeper platform A and a competing smaller platform B which are located at the extremes of a Hotelling (1929) line, at 0 and 1. The two platforms offer differentiated services to consumers. They match them with sellers of goods and services that join platforms to interact with the demand side. Consumers (C) have a mass of 1 and are distributed uniformly across the line. Consumers who are closer to platform A have preferences that better match platform A's service. The transportation cost incurred by the consumer to arrive at one of the platforms is the parameter that indicates the distance between the platform's service and consumers' preferences. Without a loss of generality, the transportation cost per unit of length for consumers is normalized to 1. Sellers (S) also have a mass of 1 but they do not incur any transportation cost when they join any of the two platforms. Platforms' marginal costs equal 0.

Let $\mu_{Cj}^{k}(q_j) \in (0,1)$ be the per $k = C,S$ agent network value for each consumer at platform $j$. When $k = C$, $\mu_{Cj}^{C}(q_j)$ captures the value a consumer gets from the presence of other consumers on the platform $j$. When $k = S$, $\mu_{Cj}^{S}(q_j)$ captures the value a consumer gets from the presence of sellers on the platform $j$. Sellers also derive network value from their platform interaction with consumers. $\mu_{Sj}^{C}(q_j) = \mu_{Sj}(q_j) \in (0,1)$ denotes the per consumer value a seller gets by joining platform $j$. These network values are functions of the amount of information collected by the platform $j$, $q_j \in [0,\bar{q}]$, where $\bar{q}$ is the finite maximum amount of information that is collected in the market. In other words, $q_j$ is the amount of interaction data, it captures the data collected and analyzed by platform $j$ through the interactions of its users.

Let $\mu_j$ be a more abstract term of network effects on platform $j$ which can take three values: $\mu_j = \{\mu_{Cj}^{C}, \mu_{Cj}^{S}, \mu_{Sj}\}$. Then, $\frac{d\mu_j(\cdot)}{dq_j} > 0, \forall j, q_j$. More information allows platform $j$ to better match agents increasing the value they can get from interacting with each other.

Platform $j$ sets an entry price $p_j$ for each seller who joins the platform. Consumer $i$'s utility from participation in the platform $j$ is

$$\mu_{Cj}^{C} x_j + \mu_{Cj}^{S} s_j - x_{ij},$$

where $x_j$ and $s_j$ is the share of consumers and sellers who join platform $j$, respectively and $x_{ij}$ is the distance of the consumer $i$ from the platform $j$. Consumers differ with respect to their transportation cost (distance) $x_{ij}$ to arrive to platform $j$. Seller's payoff from joining platform $j$ is defined in an analogous way, as

$$\mu_{Sj} x_j - p_j.$$

Depending on the type of the platform, consumers can get a more significant value from same side interactions with other consumers (high $\mu_{Cj}^{C}$) or from their interaction with sellers (high $\mu_{Cj}^{S}$). For example, in a social network platform, we expect that consumers derive more value from interactions with each other than their interactions with advertisers who join the social network. In an e-commerce platform, consumers may benefit from the online reviews of other individuals who share their experience of consuming some products they ordered on the platform, but the more significant value comes from interacting with sellers of goods and services and consuming their products. Sellers of goods and services are primarily interested in interacting with consumers so their value increases with the number of available consumers that can be their potential customers.

Under information asymmetry and without any data-sharing mechanism in place to reduce this asymmetry between the two platforms, we have $q_A > q_B$. This implies that the big platform $A$ can offer a higher network value to its users than platform B, $\mu_A(q_A) > \mu_B(q_B)$.

Let $\bar{x}_A$ and $\bar{x}_B$ be defined as the marginal consumers who would join platform A and B, respectively. We have:

$$\begin{aligned} \mu_{CA}^C \bar{x}_A + \mu_{CA}^S - \bar{x}_A = 0 \quad &\Rightarrow \quad \bar{x}_A(q_A) = \bar{x}_A = \frac{\mu_{CA}^S}{1-\mu_{CA}^C} \\ \mu_{CB}^C(1-\bar{x}_B) + \mu_{CB}^S - (1-\bar{x}_B) = 0 \quad &\Rightarrow \quad \bar{x}_B(q_B) = \bar{x}_B = 1 - \frac{\mu_{CB}^S}{1-\mu_{CB}^C}. \end{aligned} \tag{1}$$

We focus on the interesting case where $\bar{x}_A, \bar{x}_B \in (0,1)$, or equivalently that $\mu_{Cj}^S + \mu_{Cj}^C < 1$, for each $j = A, B$. Note that $\bar{x}_A(q_A)$ and $1 - \bar{x}_B(q_B)$ are increasing in $q_A$ and $q_B$, respectively. Information sharing of amount of data $q \in (0, q_A - q_B]$ from platform A to platform B can in principle allow the smaller platform to provide more efficient services that increase the network value sellers and consumers get in platform $B$, $\mu_B(q_B + q) > \mu_B(q_B)$ if $q$ contains valuable information that is not already captured by $q_B$. So, in such a case, information sharing implies that platform $B$ can compete on a more equal footing with larger platform A for consumers.

Under full information sharing, $q = q_A - q_B$, agents derive equivalent network from both platforms, $\mu_{CA}^k = \mu_{CB}^k$ and $\mu_{SA} = \mu_{SB}$, with $k = C, S$. When, instead, $q < q_A - q_B$, platform $A$ has an advantage in network value, $\mu_{CA}^k > \mu_{CB}^k$ and $\mu_{SA} > \mu_{SB}$. Full information sharing does not only mean that the two platforms are equivalent in network value but also the network values for sellers and consumers have been maximized.

We assume that for $q = q_A - q_B$, it is $\bar{x}_B(q_B + q) = \bar{x}_B(q_A) < \bar{x}_A(q_A)$. As we will see below, this guarantees that under full information sharing, platform $B$'s network value increases to such an extent that there is a potential for multi-homing of consumers in the equilibrium of the Hotelling game.

Platform A makes a take-it-or-leave-it offer and requests a payment $T(q;\alpha)$ from platform B in order to share a fraction $q$ of its data/information. It is $T(q;\alpha) = \alpha(\Pi_B(q+q_B) - \Pi_B(q_B))$, where $\alpha$ is our policy variable, with two extremes: platform $A$ has monopoly power over its information ($\alpha = 1$) and the GDPR's free of charge rule applies ($\alpha = 0$). In other words, with $\alpha = 1$ full bargaining power over the price of information to be shared. In contrast, GDPR assigns all the bargaining power to the recipient of information.

# 3 Analysis

Based on the network value offered by each platform, we have two different cases each of which gives different equilibria when there is no information sharing:

- Case 1: $\bar{x}_A(q_A) \leq \bar{x}_B(q_B)$. Consumers either single-home or do not visit any platform. There is no multi-homing.
- Case 2: $\bar{x}_A(q_A) > \bar{x}_B(q_B)$. There is the potential for multihoming for consumers in $(\bar{x}_A, \bar{x}_B)$. The rest of the consumers in $[0, \bar{x}_B]$ and $[\bar{x}_A, 1]$ either single-home or do not visit any platform.

In both cases, sellers can multi-home if they wish. Since they incur zero transportation cost, they will join a platform if, in equilibrium, the value they derive from it is greater or equal the price they have to pay in order to access it.

Our analysis includes two steps: We first look at the equilibrium pricing strategies of the Hotelling platform game with and without information sharing of amount $q$ as well as the implications of each equilibrium for market demand, supply and payoffs of each market participant. Then, we proceed by studying social and private incentives for information sharing. We derive the optimal choice of $q$ for platform $A$ and how it compares with social planner's preferred information sharing rule for different values of the policy variable $\alpha$ and the different pricing equilibria identified in the first step.

## 3.1 Equilibrium pricing strategies at the Hotelling platform competition game

Starting with Case 1, note that when there is no information sharing, the share of consumers who joins platform A increases with the difference $q_A - q_B$. Since, $q_A > q_B$, a larger share of consumers is captured by platform A. In equilibrium, platform A finds optimal to set a price $p_A^n = \mu_{SA}\bar{x}_A$, while platform B sets price $p_B^n = \mu_{SB}(q_B)(1-\bar{x}_B)$. To see that this is a pricing equilibrium, we can show that there is not any incentive for each platform $j$ to unilaterally deviate from this pricing behavior. If platform A (B) sets its price above $p_A^n$ ($p_B^n$), then no seller has incentives to join platform A (B) because the network value they derive from platform A (B) is lower than the price they have to pay to join the network. This implies that no consumers will have incentives to join platform A (B) since now valuable interactions with sellers will be missing. So, platform A (B) is worse off when charging a price to the sellers that is higher than $p_A^n$ ($p_B^n$). When, instead, platform A (B) charges a lower price than $p_A^n$ ($p_B^n$), then its profit is reduced since it does not fully extracts the rents from the sellers. Since there is not any profitable deviation for any of the two platforms, pricing behavior $(p_A^n, p_B^n)$ constitutes the equilibrium of the Hotelling game.

Given this pricing equilibrium, the share of consumers $[0, \bar{x}_A]$ visits platform A, the share $[\bar{x}_B, 1]$ visits platform B. Consumers in $(\bar{x}_A, \bar{x}_B)$ do not visit any platform. Sellers visit both platforms but derive zero surplus.[1]

This is what we call the data bottleneck market failure. By being able to capture the largest share of the single-homing side of consumers, platform A can extract all the surplus from sellers who still wish to join A to interact with its numerous consumers. In this data bottleneck equilibrium, sellers have only weak incentives to multi-home when $q_A - q_B$ is large because, by selling through platform A only, they can already interact with the largest share of consumers $\bar{x}_A$.

Network effects make it easier for the data bottleneck equilibrium to arise. Strong network

[1] The underlying assumption here is that when sellers are indifferent between single-homing and multi-homing, they choose to multi-home.

externalities help platform $A$ to maximize the excessive rents it extracts for two reasons: First, the higher this network value for sellers ($\mu_{SA}$) is, the higher are the rents the platform A can extract from the sellers. Second, the share of consumers that prefer to join platform A increases with network effects $\mu^C_{CA}$ and $\mu^S_{CA}$ and so does the price platform A charges the sellers (due to cross-side network externalities).

We now study what happens when information $q$ from platform A to platform B takes place. First, note that without any information sharing, the two platforms capture different parts of the demand. They collect different information about the characteristics of consumers and their interactions with the agents of each platform ecosystem. Consequently, $q$ contains different information from $q_B$ and the network value offered to consumers by platform B increases to $\mu_B(q_B+q)$ when it received new valuable information $q$. For a sufficient amount $q$, it will be $\bar{x}_A > \bar{x}_B(q_B+q)$ and there will be the potential of multi-homing for consumers.

In such a case, the two platforms "play" a Bertrand pricing game to attract sellers to their market. Their monopoly power is reduced to the portion of consumers that single-home. In equilibrium, platform A sets $p^I_A = \mu_{SA}\bar{x}_B(q_B+q)$ which is declining in $q$ and platform B sets $p^I_B = \mu_{SB}(q_B+q)(1-\bar{x}_A)$. So, the share of consumers $(\bar{x}_B(q_B+q), \bar{x}_A)$ multi-homes, while the shares $(0, \bar{x}_B(q_B+q))$ and $(\bar{x}_A, 1)$ single-home (see Figure 1).

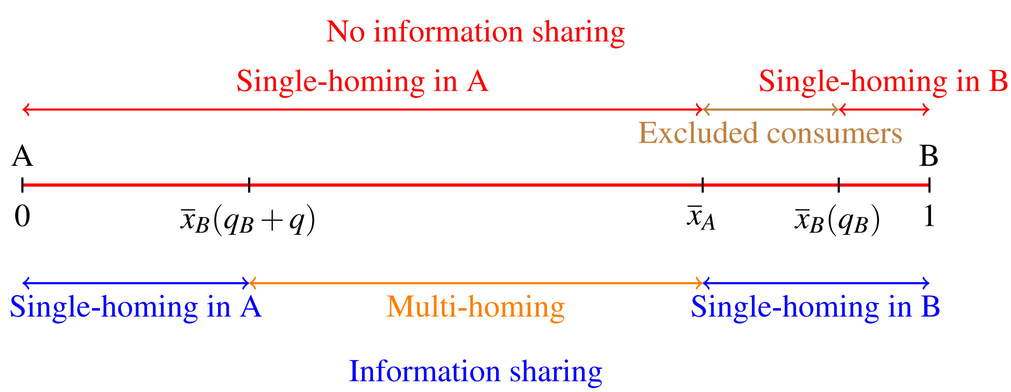


**Figure 1:** Platform competition across the Hotelling line with and without information sharing, when no information sharing implies that some consumers only single-home. The shares of

consumers that single-home and multi-home with and without information sharing are depicted.

Note that under information sharing no platform has incentives to increase its price above the equilibrium level $(p_A^I, p_B^I)$. If one platform sets a higher price sellers will only join its competitor platform (since accessing one platform does not only give them access to consumers that single-home on that platform, but also to the consumers that multi-home). Due to multi-homing, sellers now extract positive surplus and in equilibrium, they all multi-home (given their zero transportation cost) and the data bottleneck market inefficiency is reduced or even resolved if the degree of multi-homing is sufficiently high (or equivalently if platforms become sufficiently more symmetric in information). Platform A now "sees" part of its rents being extracted by the sellers and realizes a lower payoff, since $p_A^I < p_A^n$. This extracted surplus is proportional to the share of consumers that multi-home or in other words to the amount $q$ it is shared. Hence, the data bottleneck market inefficiency is declining in $q$. At the same time, platform B realizes higher rents under information sharing and consumer single-homing since $p_B^I > p_B^n$.

While consumers in $(0, \bar{x}_B(q_B + q))$ and $(\bar{x}_B(q_B), 1)$ have the same welfare as when there is no information sharing, consumers within $(\bar{x}_B(q_B + q), \bar{x}_B(q_B))$ generate higher surplus under information sharing. The share of consumers in $(\bar{x}_A, \bar{x}_B(q_B))$ is not excluded from the market anymore. Consumers in this segment have now a positive surplus. Consumers in $(\bar{x}_B(q_B + q), \bar{x}_A)$ now get a (higher) positive surplus (from both platforms) because they multi-home. So, overall, information sharing and the resulting rise in $\mu_{CB}$ increases both consumer and seller surplus, in comparison to the case where there is no information sharing. The higher the amount $q$ shared, the greater will be the consumer and seller welfare gains and the lower will be the rents extracted by the big platform A.

If, under Case 1, $q$ is sufficiently small such that, $\bar{x}_A \leq \bar{x}_B(q_B + q)$, then there is not any multihoming under information sharing. Equilibrium price by platform $A$ equals $p_A^n$ since under such a small $q$, platform $A$ still enjoys a significant information advantage and it has an exclusive relationship to a share $(0, \bar{x}_A)$ of consumers. This is translated as a monopoly power on the sellers' side. Platform $A$ can still extract all the surplus of the sellers under information sharing because $q$ is small. Sellers more generally have zero surplus regardless of whether information sharing takes place or not. This implies that the data bottleneck market failure is not resolved by information sharing when the amount of information sharing, $q$ is small. Nevertheless, information sharing still improves average consumer welfare since it reduces the share of consumers that are excluded from the market ($\bar{x}_B(q_B + q)$ is declining in $q$ moving to values closer to $\bar{x}_A$). The higher the $q$ is, the smaller the share of excluded consumers will be.

We continue by studying the welfare implications of information sharing in Case 2. The key difference with the previous case is that now there is a potential for multi-homing even when there is no information sharing. Platforms have monopoly power over the segments of demand on which there is no potential for multi-homing. The equilibrium platforms' prices on the sellers' side are $p_A^{n2} = \mu_{SA}\bar{x}_B(q_B)$ and $p_B^{n2} = \mu_{SB}(q_B)(1 - \bar{x}_A)$. We can apply exactly the same approach as above to see why these pricing strategies constitute an equilibrium.[2] This equilibrium implies that consumers in $(0, \bar{x}_B(q_B))$ only visit platform A, consumers in $(\bar{x}_A, 1)$ only visit platform B, and the rest of the consumers multi-home.

Information sharing $q$ again increases the value of platform B for users to level $\mu_B(q_B + q)$ for which we now have $\bar{x}_A > \bar{x}_B(q_B) > \bar{x}_B(q_B + q)$. In the equilibrium, $p_A^{I2} = \mu_{SA}\bar{x}_B(q_B + q) < p_A^{n2}$ and $p_B^{I2} = \mu_{SB}(q_B + q)(1 - \bar{x}_A) < p_B^{n2}$.

Information that the platform $A$ collects, $q_A$, now shares some common information set with $q_B$ due to the segment of demand that multi-home in both platforms. But, there is still a significant segment of demand that without information sharing only visits platform $A$. So, $q$ includes valuable information about consumers that initially is not captured by $q_B$ and therefore can significantly increase platform $B$'s network value.[3]

In equilibrium, information sharing expands the market share of consumers that multi-home

[2] No platform has incentives to unilaterally deviate from this equilibrium.
[3] Platform $B$ will only have incentives to pay a non-negative price, $T \geq 0$ for information that is valuable to it.

(see Figure 2) and consequently, it increases average consumer welfare (as there is a larger share of consumers that visit both platforms and derive extra surplus). The expansion of multi-homing over the Hotelling line is proportional to the amount $q$ of data shared.

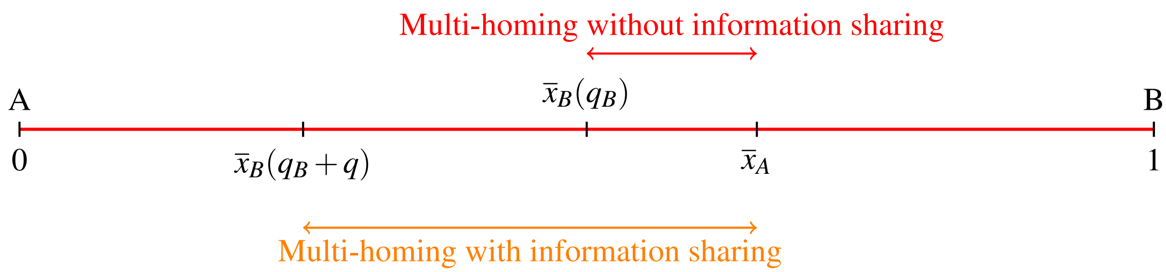


**Figure 2:** Platform competition across the Hotelling line with and without information sharing, when no information sharing implies that at least some consumers multi-home. Information sharing increases the share of consumers that multi-home.

Sellers are again better off since equilibrium platform prices are lower under information sharing. This is because information sharing constrains the ability of both platforms in this case to extract rents from sellers mitigating the data bottleneck market failure.

We summarize the welfare implications of information sharing for all the cases discussed above in the following proposition.

**Proposition 1.** *Let two platforms A and B compete in a Hotelling product differentiation line while being asymmetric in terms of the amount of information they have about their users, $q_A > q_B$. Then, average consumer welfare is strictly increasing in the amount $q$ of information sharing from platform A to platform B. Sellers' profitability is strictly increasing in $q$ if and only if $q$ is sufficiently high such that there is some consumer multi-homing in the equilibrium, $\bar{x}_A > \bar{x}_B(q_B+q)$.*

The direct implication of this proposition is the following corollary.

**Corollary 1.** *Information sharing of amount $q$ from platform A to platform B reduces the data bottleneck market failure if and only if $\bar{x}_A > \bar{x}_B(q_B+q)$.*

## 3.2 Social and private incentives for information sharing

A social planner, who primarily cares about average consumer welfare and sellers' payoff would ideally choose full information sharing.

**Lemma 1.** *Let two platforms A and B compete in a Hotelling product differentiation line, while being asymmetric in terms of the amount of information they have about their users, $q_A > q_B$. The amount $q$ of information sharing from platform A to platform B that maximizes average consumer welfare is $q^{SP} = q_A - q_B$.*

At $q = q^{SP}$ we arrive at a state where platforms A and B have exactly the same amount of information and offer the same network value to their users. Under this optimal social planner's rule, platform $B$ receives all valuable information that was before exclusive to platform A. This leads to an increased average consumer welfare and sellers' payoff, since $\bar{x}_A > \bar{x}_B(q_B + q^{SP})$. No platform has any information advantage and therefore the data bottleneck market failure is not observed in equilibrium.

How do social incentives for information sharing compare with private ones? The amount $T(q;\alpha)$ that platform $A$ requests from platform $B$ for sharing information $q$ with it (take-it-or-leave-it offer) and the implications of information sharing for platform $A$'s direct market profitability are the two factors that determine the incentives of platform $A$ to share amount of information $q$. Policy parameter $\alpha$ defines the ability of platform $A$ to appropriate the information rents when it shares amount $q$.

We now study whether platform $A$ has incentives to share information, and if yes, how much information (optimal $q$) it selects to share in equilibrium for each of the two relevant cases, Case 1 and Case 2.

According to the analysis above, under Case 1, the profit of platform $A$ without any information sharing is $\Pi_A^n = \mu_{SA}\bar{x}_A$. When, platform $A$ shares information $q$, then,

$$\Pi_A^I(q;\alpha)\begin{cases} = \mu_{SA}\bar{x}_A + \alpha\left(\mu_{SB}(q_B+q)(1-\bar{x}_B(q_B+q)) - \mu_{SB}(q_B)(1-\bar{x}_B(q_B))\right), \text{if} \quad \bar{x}_A \leq \bar{x}_B(q_B+q), \\ = \mu_{SA}\bar{x}_B(q_B+q) + \alpha\left(\mu_{SB}(q_B+q)(1-\bar{x}_A) - \mu_{SB}(q_B)(1-\bar{x}_B(q_B))\right), \text{if} \quad \bar{x}_A > \bar{x}_B(q_B+q). \end{cases} \tag{2}$$

The first right-hand term in each of these two expressions refers to the direct channel of market profitability and the second term of each expression refers to the information sharing payment platform A receives from platform $B$.

When $\bar{x}_A \leq \bar{x}_B(q_B+q)$, the direct channel of market profitability is the same both with and without information sharing ($\mu_{SA}\bar{x}_A$). By increasing $q$ in the domain $q \in [0, q^*]$, where $q^*$ is defined by $\bar{x}_A = \bar{x}_B(q_B+q^*)$, platform $A$ does not see any change in its direct market profitability. So, increasing $q$ in this domain is only associated with the payment $T(q \leq q^*; \alpha)$ platform A receives. It is $T(q \leq q^*; \alpha) = \alpha\left(\mu_{SB}(q_B+q)(1-\bar{x}_B(q_B+q)) - \mu_{SB}(q_B)(1-\bar{x}_B(q_B))\right)$. Since $\mu_{SB}(q_B+q)$ is strictly increasing in $q$ and $\bar{x}_B(q_B+q)$ is strictly declining in $q$, we conclude that $T(q \leq q^*; \alpha)$ is strictly increasing in $q$ if and only if $\alpha > 0$. Consequently, $\Pi_A^I(q \leq q^*; \alpha > 0)$ is also strictly increasing in $q$. Platform $A$ never finds optimal to select amount $q < q^*$. If, instead, $\alpha = 0$, platform $A$'s payoff does not depend on $q$ when $q \in [0, q^*]$.

Does the platform $A$ have any incentives to increase the amount of information it shares to levels $q \in (q^*, q_A - q_B]$ when $\bar{x}_A > \bar{x}_B(q_B+q)$?

To answer this question, we need to study how the profit function (2) does depend on the amount $q$ shared with platform B and whether there exists a $q^e \in [q^*, q_A - q_B]$ that maximizes (2). To ensure that the maximization problem is well-defined, we assume that:

$$\mu_{SA}\frac{d^2\bar{x}_B(q_B+q)}{dq^2} + \alpha(1-\bar{x}_A)\frac{d^2\mu_{SB}(q_B+q)}{dq^2} < 0, \tag{3}$$

$\forall\alpha \in [0,1]$ and $q \in [0, q_A - q_B]$.

Note that in this case, when $\alpha > 0$, we have two opposing effects. On the one side, for such higher values of $q$, platform B improves its network value proposition to the extent that it becomes more competitive. Platform $A$ due to the increased market competition sees its direct market profitability drop as the rents it extracts from the sellers decline. The loss from the direct channel of market profitability amounts to $\mu_{SA}(\bar{x}_B(q_B+q)-\bar{x}_A)$. This direct market loss becomes higher with $q$. On the other side, the price of information sharing becomes $T(q \in (q^*, q_A - q_B]; \alpha) = \alpha(\mu_{SB}(q_B+q)(1-\bar{x}_A) - \mu_{SB}(q_B)(1-\bar{x}_B(q_B)))$ which is increasing in $q$. Which of the two effects dominate depends on the sign of the following expression (derived from the first-order-condition of platform $A$'s profit function 2 with respect to $q$):

$$FOC_A(q=q^*;\alpha) = \underbrace{\mu_{SA}\frac{d\bar{x}_B(q_B+q)}{dq}|_{q=q^*}}_{\text{Direct market loss (-)}} + \underbrace{\alpha(1-\bar{x}_A)\frac{d\mu_{SB}(q_B+q)}{dq}|_{q=q^*}}_{\text{Information sharing benefit (+)}}. \tag{4}$$

For $\alpha = 0$, there is only the direct market loss effect $\forall q \in (q^*, q_A - q_B]$ since platform $A$ does not receive any benefit for sharing information. So, it does not have any incentive to raise $q$ above $q^*$. We have multiple equilibria. Any value of $q \in [0, q^*]$ is an equilibrium. We adopt the following equilibrium refinement in order to make meaningful welfare comparisons and investigate in more depth specific market strategies: If platform $A$ is indifferent between different amounts $q$, in equilibrium (that is, there are multiple equilibria over the level of $q$), it selects the one that is preferred by the social planner. So, under $\alpha = 0$, which corresponds to the GDPR's free-of-charge rule, the equilibrium becomes: $q^{GDPR} = q^*$. At this relatively low level of information sharing, the data bottleneck market failure emerges in equilibrium since sellers' surplus is still fully extracted.

Let now $\alpha > 0$ and $q = q^*$. The threshold value $\hat{\alpha}_1$ is defined from expression 4 as $FOC_A(q = q^*; \alpha = \hat{\alpha}_1) = 0$. In other words, this is the value of $\alpha$ below which the platform $A$ does not have any incentive to marginally raise its information sharing above $q^*$. Solving this equation[4] we find

[4]Recall that $\bar{x}_A = \bar{x}_B(q_B+q^*) \Rightarrow 1-\bar{x}_A = \frac{\mu^S_{CB}(q_B+q^*)}{1-\mu^C_{CB}(q_B+q^*)}$.

that

$$\hat{\alpha}_1 = \mu_{SA} \frac{\frac{d}{dq}\left(\ln \frac{\mu_{CB}^S}{1-\mu_{CB}^C}\right)|_{q=q^*}}{\frac{d\mu_{SB}}{dq}|_{q=q^*}}. \tag{5}$$

Given the definition of $\hat{\alpha}_1$, we conclude that for values of $\alpha \in (0, \min[\hat{\alpha}_1, 1]]$, platform $A$ does not have any incentive to raise its information sharing above $q^*$. If $\min\{\hat{\alpha}_1, 1\} = 1$, the GDPR equilibrium of $q^e = q^*$ applies $\forall \alpha \in [0,1]$, so there is not any form of price regulation that can induce platform $A$ to choose any $q > q^*$. It does not matter whether platform $A$ has any power over the rents it can extract from platform $B$ through information sharing or not. It will always select to share amount $q^*$.

Following the definition of $\hat{\alpha}_1$ by expression (5), $\min\{\hat{\alpha}_1, 1\} = 1$ implies that information asymmetry prior sharing is very high and consumers are relatively not ”insensitive” in increases in the network value of platform B:

- Platform A is a gatekeeper, namely, it is very dominant in the market. It attracts the vast majority of users and collects a very large amount of information, $q_A$. As a result, the network value of its sellers $\mu_{SA}$ is very high and $1 - \bar{x}_A$ is very small.

- A marginal increase in $q$ in the neighborhood of $q = q^*$ sufficiently increases the same-side and cross-side network value for consumers that join platform $B$ in comparison to the respective increase to the network value of platform $B$'s sellers. In other words, consumers are not locked in on platform A, but, they can also derive a significant value from platform B if information sharing (towards platform B) takes place.

This holds for any price restriction imposed over information sharing. Price regulation on information does not have any effect on the equilibrium. This is because direct market losses of information sharing due to increased competition exceed the payment $T(q;\alpha)$ for $q > q^*$. Even under $\alpha = 1$, payment $T(q;\alpha)$ is not sufficient to fully compensate platform $A$ for its direct market loss when $q > q^*$. The data bottleneck market failure is still present in the $q^e = q^*$ equilibrium.

If instead, information asymmetry is not very high and consumers are more locked in on platform A, $\min\{\hat{\alpha}_1, 1\} = \hat{\alpha}_1$ at $q = q^*$. Then, for $\alpha \in (0, \hat{\alpha}_1]$, platform A finds and optimal to choose in equilibrium $q^e = q^*$. But, now, for $\alpha \in (\hat{\alpha}_1, 1]$, platform $A$ has incentives to raise its information sharing to levels $q > q^*$ because information sharing incorporates a significantly high payment $T(q = q^*; \alpha \in (\hat{\alpha}, 1])$ to cover direct market losses ($1 - \bar{x}_A$ is higher and $\mu_{SA}$ is lower than in the case of a gatekeeper platform).

So, price regulation can be important, in this case, for the private incentives for information sharing. If regulation leaves small rents to platform A, then, the platform does not have incentives to share a lot of information and the equilibrium is the same as in the case of the gatekeeper, $q^e = q^*$. But, if instead, platform A gets sufficient information rents from sharing, the equilibrium will be $q^e > q^*$.

In this latter case of $\alpha \in (\hat{\alpha}_1, 1]$, we investigate whether the social planner's preferred amount $q^{SP} = q_A - q_B$ can be sustained in equilibrium. To do that, we study the first order condition 4 when $q \to q_A - q_B$. Given assumption (3), a necessary condition for private and social incentives for information sharing to coincide is $FOC_A(q = q_A - q_B; \alpha = 1 \geq 0)$, where:

$$FOC_A(q \to q_A - q_B; \alpha = 1) = \underbrace{\frac{1 - \mu_{CA}^S - \mu_{CA}^C}{1 - \mu_{CB}^C} \frac{d\mu_{SB}}{dq}|_{q \to q_A - q_B}}_{\text{Information sharing benefit (+)}}$$
$$\underbrace{- \frac{\mu_{SA}}{1 - \mu_{CB}^C} \left( \frac{d\mu_{CB}^S}{dq}|_{q \to q_A - q_B} + \frac{\mu_{CB}^S}{1 - \mu_{CB}^C} \frac{d\mu_{CB}^C}{dq}|_{q \to q_A - q_B} \right)}_{\text{Direct market loss (-)}}. \tag{6}$$

$FOC_A(q \to q_A - q_B; \alpha = 1)$ becomes negative i) under high information asymmetry such that platform $A$, prior to information sharing, moves towards a gatekeeper status by capturing a large part of the market and by having access to a large amount of market information. This means that $\bar{x}_A$ is relatively close to 1 such that $1 - \mu_{CA}^S - \mu_{CA}^C$ is now sufficiently small and $\mu_{SA}$ sufficiently high so that information payment cannot counterbalance the direct market losses even if $\alpha = 1$; ii) a marginal decline of $q$ below $q_A - q_B$ leads to a sharper decline in the network value (same-side and/or cross-side) that consumers get from platform B, relative to the respective decline for the network value of sellers. So, consumers are more "sensitive" to changes in the network value of platform B under information sharing and less locked in on platform A.

When $\alpha \in (\hat{\alpha}_1, 1]$ and $FOC_A(q = q_A - q_B; \alpha = 1) < 0$, platform $A$ chooses in equilibrium information sharing $q^e \in (q^*, q^{SP})$. Under condition (3), the equilibrium amount of information sharing is strictly increasing in $\alpha$. Platform A has more incentives to share information if it can extract higher rents from platform B. The socially desirable level of information sharing is not achieved. However, sellers now derive positive rents from their participation in the market because equilibrium information sharing is sufficient to reduce to some extent the data bottleneck market failure.

Can private and social incentives coincide? This can be the case if information asymmetry is only moderate and prior to information sharing platform $A$ misses valuable information about an important segment of the market (e.g., information from a non-trivial share of consumers) such that $\bar{x}_A$ is more distant from 1 and $1 - \mu_{CA}^S - \mu_{CA}^C$ is sufficiently high. In addition, this occurs when consumers are sufficiently locked in on platform $A$. Then, $FOC_A(q = q_A - q_B; \alpha = 1) \geq 0$. So, there exists an $\alpha^{cr} \in [\hat{\alpha}_1, 1]$ such that $FOC_A(q = q_A - q_B; \alpha \geq \alpha^{cr}) \geq 0$. This means that private incentives for information sharing coincide with social ones for $\alpha \in [\alpha^{cr}, 1]$. In such an equilibrium, it is $q^e = q^{SP}$ and the data bottleneck market failure disappears.

We have fully characterized the information sharing equilibria in the case that prior to information sharing we have $\bar{x}_A(q_A) \leq \bar{x}_B(q_B)$ (Case 1). We now move forward with the case where $\bar{x}_A(q_A) > \bar{x}_B(q_B)$ (Case 2).

The profit of platform $A$ without any information sharing is $\Pi_A^{n2} = \mu_{SA}\bar{x}_B(q_B)$. When, platform

$A$ shares information $q$, then,

$$\Pi_A^{i2}(q;\alpha) = \mu_{SA}\bar{x}_B(q_B+q) + \alpha\left(\mu_{SB}(q_B+q) - \mu_{SB}(q_B)\right)(1-\bar{x}_A). \tag{7}$$

For $\alpha = 0$, platform $A$ does not have any incentive to share information since $\bar{x}_B(q_B+q) < \bar{x}_B(q_B)$, $\forall q > 0$. The expression (4) still applies. Let $\hat{\alpha}_2$ be defined (similarly to $\hat{\alpha}_1$) such that $FOC_A(q=0;\alpha=\hat{\alpha}_2)=0$. So, when, $\alpha \in (0,\min\{\hat{\alpha}_2,1\}]$, platform $A$ does not have any incentives to share information, so, in equilibrium, $q_2^e = q_2^{GDPR} = 0$. When, $\min\{\hat{\alpha}_2,1\} = \hat{\alpha}_2$, the equilibrium is very similar to the one derived under Case 1. Namely, for $\alpha \in (0,\hat{\alpha}_2]$, it is $q_2^e = 0$. For $\alpha \in (\hat{\alpha}_2,1]$ and when $FOC_A(q=q_A-q_B;\alpha \in (\hat{\alpha}_2,1])) < 0$, platform $A$, in equilibrium, chooses information sharing $q_2^e \in (0,q^{SP})$. This amount $q_2^e$ is strictly increasing in $\alpha$. If instead, $FOC_A(q=q_A-q_B;\alpha = 1]) \geq 0$, there exists an $\alpha_2^{cr} \in [\hat{\alpha}_2,1]$ such that $FOC_A(q=q_A-q_B;\alpha \geq \alpha_2^{cr}) > 0$. So, for $\alpha \in [\alpha_2^{cr},1]$, private and social incentives for information sharing coincide, with $q_2^e = q^{SP}$. The derivation and intuition follow exactly the same steps illustrated above for Case 1.

So, to sum up,

**Proposition 2.** *Let* $\min[\hat{\alpha}_i,1]=1$, *where* $i=1,2$. *Then, the optimal amount of information sharing selected by platform A is:*

- $q^e = q^* \in (0,q^{SP})$, *when* $\bar{x}_A < \bar{x}_B(q_B)$,
- $q^e = 0$, *when* $\bar{x}_A > \bar{x}_B(q_B)$.

Following the discussion above, Proposition 2 refers to the situation where there is a great information asymmetry with the presence of a gatekeeper platform $A$. Private incentives for information sharing are minimum. In equilibrium information sharing, $q^e$, no consumer is excluded. If the market is not initially fully covered ($\bar{x}_A < \bar{x}_B(q_B)$), then, platform A has incentives to share information up to the point the market becomes fully covered ($\bar{x}_A = \bar{x}_B(q_B+q^*)$). Sharing information in this domain does not affect platform $A$'s profitability. The market expansion of platform

$B$ due to the additional information it receives has only to do with serving new consumers that were excluded prior to information sharing. But, the amount of information $q$ shared is still too modest to allow platform $B$ to improve its network value to a level at which it can "steal" consumers from platform $A$ and resolve the data bottleneck market failure.

Note that gatekeeper platform $A$ does not have any incentives to supply a large amount of information to platform $B$ because it dominates the market and any information sharing would result to direct market losses that exceed information sharing payment, $T(q;\alpha)$ (which is now small because of the great asymmetry, since $\bar{x}_A$ is close to 1).

At first sight, it may seem surprising that the equilibrium information sharing of Proposition 2 does not depend on $\alpha$.

**Corollary 2.** *A strict information price regulation such that $\alpha \in [0, \hat{\alpha}_i]$, where $i = 1,2$, does not reduce private incentives for information sharing if and only if information asymmetry is so high that* $\min\{\hat{\alpha}_i, 1\} = 1$.

This is true especially if consumers are relatively elastic with respect to increases in the network value of platform $B$. The degree of strictness of information price regulation does not affect the equilibrium amount of information sharing selected by the gatekeeper. Platform $A$ chooses to share the same amount of information regardless of the value of $\alpha$. So, in this case, the GDPR's free-of-charge rule ($\alpha = 0$) can effectively achieve a redistribution of information rents towards the small platform $B$ without affecting the equilibrium $q^e$. Nevertheless, price regulation is not sufficient to induce the adoption of socially optimal levels of information sharing.

If instead, information asymmetry is significant, but not excessively high and consumers are relatively less elastic to changes in network value of platform $B$, we have $\min[\hat{\alpha}_i, 1] = \hat{\alpha}_i$. The following equilibrium applies:

**Proposition 3.** *Let $\min[\hat{\alpha}_i, 1] = \hat{\alpha}_i$, where $i = 1,2$. Then, the optimal amount of information sharing selected by platform A depends on the range of values of key parameters as follows:*

- *When* $\alpha \in [0, \hat{\alpha}_i]$*, the equilibrium amount of information sharing is*

  - $q^e = q^* \in (0, q^{SP})$*, when* $\bar{x}_A < \bar{x}_B(q_B)$*,*

  - $q^e = 0$*, when* $\bar{x}_A > \bar{x}_B(q_B)$*.*

- *When* $\alpha \in (\hat{\alpha}_i, 1]$*, the equilibrium* $q^e$ *is:*

  - $q^e \in (q^*, q^{SP})$*, when* $FOC_A(q = q_A - q_B; \alpha \in (\hat{\alpha}_1, 1]) < 0$ *and* $\bar{x}_A < \bar{x}_B(q_B)$*,*

  - $q^e \in (0, q^{SP})$*, when* $FOC_A(q = q_A - q_B; \alpha \in (\hat{\alpha}_2, 1]) < 0$ *and* $\bar{x}_A > \bar{x}_B(q_B)$*,*

  - $q^e = q^{SP}$*, when* $FOC_A(q = q_A - q_B; \alpha \in (\hat{\alpha}_i, 1]) \geq 0$*.*

Now, information price regulation can have important implications for the equilibrium information sharing. When information asymmetry is sufficiently low ($FOC_A(q = q_A - q_B; \alpha \in (\hat{\alpha}_i, 1]) \geq 0$), social and private incentives for information sharing coincide if and only if price regulation $\alpha$ leaves sufficient information rents to platform $A$.

Platform $A$ now does not have a gatekeeper status and it is not so dominant as in the case of Proposition 2. That increases information sharing payment it receives from platform $B$ which is proportional to the share of consumers that do not join platform $A$ (share $1 - \hat{x}_A$ is relatively high). This payment is sufficient to cover platform A's direct market losses that arise from information sharing, especially when consumers are not so elastic to the network value of platform $B$. Hence, platform $A$ chooses in equilibrium, $q = q^{SP}$ and a significant share of consumers multi-homes between the two platforms. Even if, in this case, platform $B$ develops an equivalent to platform $A$'s value proposition, most consumers do not switch exclusively to platform $B$ but choose to visit both platforms, in equilibrium. So, due to the increased multi-homing, platform $A$ does not experience large direct market losses under $q = q^{SP}$ (while it still receives a relatively high monetary benefit from platform $B$ as an information sharing payment).

If the network value elasticity of consumers is higher (but still moderate) or the information asymmetry is significant, then private incentives for information sharing are lower than social ones. Direct market losses become now significant for platform $A$ which does not find optimal to raise $q$ at the social optimal levels. The exact equilibrium depends on the payment $T(q;\alpha)$ and in particular on the parameter $\alpha$. Under a strict information price regulation such that $\alpha \in [0, \hat{\alpha}]$, platform $A$ cannot compensate its direct market losses from the information payment and therefore it has low incentives to share information. The GDPR price rule falls under this category as it eliminates the rents platform $A$ can extract from information sharing.

For less restrictive information price regulation ($\alpha > \hat{\alpha}$), platform $A$ finds it optimal to select to share an intermediate amount of information $q^e$ that is larger than the amount that corresponds to the GDPR rule, but smaller than the social optimum, $q^{SP} > q^e > q^{GDPR}$.

**Corollary 3.** *For a less restrictive information price regulation, $\alpha \in (\hat{\alpha}_i, 1]$, such that, in equilibrium, $q^{SP} > q^e > q^{GDPR}$, amount $q^e$ is strictly increasing in $\alpha$.*

# 4 Data investments and the price of information

Our analysis so far shows that under great information asymmetry, with the presence of a gatekeeper platform which dominates the market, private incentives for information sharing are smaller than the social ones. So, there is a scope for a regulatory intervention through a mechanism that mandates interaction data sharing from the gatekeeper to its competitor. In our static framework we showed that this mechanism should enforce the socially desirable level of information sharing, $q^{SP} = q_A - q_B$.

We now adopt a more dynamic perspective. In such a mandated scheme we do not only consider the market implications of information sharing $q$ on market participants but also how does such a mechanism affect the incentives of the gatekeeper platform to collect and analyze data.

Each platform $j$ facilitates interactions that can produce amount of data up to $\overline{q}_j$, with $\overline{q}_A > \overline{q}_B$. Platforms choose of how much of this data, $q_j \in [0, \overline{q}_j]$ will collect and analyze.

Let $C_j(q_j)$ be the convex cost function of data collection and analysis of amount $q_j$ by platform

$j = A, B$.[5]

The timing of our game is as follows:

1. The regulator implements a mandated sharing scheme. This scheme obliges the sharing of amount $q$ from the gatekeeper platform to its competitor in exchange of price $T(q;\alpha)$.

2. Platform $A$ and platform $B$, after observing the new mandated data sharing regulation, choose the optimal amount of information they collect and analyze, $q_A^i$ and $q_B^i$, respectively.

3. The two platforms choose their pricing schedules for the sellers who consider to join them.

Price $T(q;\alpha)$ is defined as in the main model. We solve the model by backward induction focusing on information sharing $q > q^*$ under Case 1 and $q > 0$ under Case 2, since otherwise, there is not any scope for a regulatory intervention (private incentives of platform $A$ are sufficient to reach $q = q^*$ and $q = 0$ in each of the two cases, respectively). Stage 3 resembles the analysis and in particular of Subsection 3.1. Given those price equilibria, the objective functions of platform $A$ and $B$ in stage 2 are:

$$
\begin{aligned}
\Pi_A^i(q_A, q;\alpha) &= \mu_{SA}(q_A)\bar{x}_B(q_B+q) + \alpha\left(\mu_{SB}(q_B+q) - \mu_{SB}(q_B)\right)(1-\bar{x}_A(q_A)) - C(q_A), \\
\Pi_B^i(q_B, q;\alpha) &= \mu_{SB}(q_B+q)(1-\bar{x}_A(q_A)) + (1-\alpha)\left(\mu_{SB}(q_B+q) - \mu_{SB}(q_B)\right)(1-\bar{x}_A(q_A)) - C(q_B+q).
\end{aligned}
\tag{8}
$$

The optimal amounts of data collected and analysed, $q_A^i$ and $q_B^i$ are given by the first order conditions of objective functions (8). They depend both on the amount $q$ and restrictions over the information rent $\alpha$ that is accrued by the gatekeeper platform A. It is $q_A^i = \text{argmax}_{q_A}\{\Pi_A^i(q_A, q;\alpha)\}$

[5] The convexity of the cost function, in combination with condition (3) ensures that the maximization problem over platforms' profits is well-defined.

and $q_B^i = \text{argmax}_{q_B}\{\Pi_B^i(q_B, q; \alpha)\}$. Optimal amounts satisfy the following first-order-conditions:

$$\begin{aligned}
q_A^i \quad &: \quad \frac{d\mu_{SA}(q_A)}{dq_A}|_{q_A=q_A^i}\bar{x}_B(q_B+q) - \alpha\left(\mu_{SB}(q_B+q) - \mu_{SB}(q_B)\right)\frac{d\bar{x}_A(q_A)}{dq_A}|_{q_A=q_A^i} = \frac{dC(q_A)}{dq_A}|_{q_A=q_A^i}, \\
q_B^i \quad &: \quad \left(\frac{d\mu_{SB}(q_B+q)}{dq_B}|_{q_B=q_B^i} + (1-\alpha)\frac{d}{dq_B}\left(\mu_{SB}(q_B+q) - \mu_{SB}(q_B)\right)|_{q_B=q_B^i}\right)(1-\bar{x}_A) = \\
&= \quad \frac{dC(q_B+q)}{dq_B}|_{q_B=q_B^i}. \qquad (9)
\end{aligned}$$

From the first-order-condition (9) on $q_A^i$, with the help of the envelope theorem, we see that $q_A^i$ is strictly decreasing in $q$. The gatekeeper platform is obliged to share part of the unique insights it gets from the collected market data and such an information sharing leads to an increased competitive pressure by platform $B$ with significant direct market losses. As a result, the benefit from data collection and analysis drops.

The relationship between $q_A^i$ and $\alpha$ is as follows. If the regulator allows for sufficient information rents to be captured by the gatekeeper, then platform $A$ has less incentives to collect and analyze market data. In contrast, the incentives of platform $A$ to collect information are maximized under the GDPR free-of-charge rule. This is because optimal amount of information, $q_A^i$, is strictly decreasing in $\alpha$.

The reason behind this surprising result is the following: As $alpha$ increases, platform $A$ is eligible to receive higher share of the information rents through the mandated sharing scheme. An increase in $q_A^i$ would result to market expansion since consumers get a higher network value by joining platform A which would result to a shift of $\bar{x}_A$ to the right, closer to 1. However, this shift would limit the benefit of platform $B$ from information sharing. Even if platform $A$ receives a larger share of the "pie" through an a higher policy parameter $\alpha$, it still receives a lower benefit just because the "pie" (as a whole) becomes smaller as $q_A^i$ increases. Hence, the higher is the share of the information sharing assigned to it by the regulator, the more incentives the gatekeeper has to further to reduce $q_A^i$ in order to increase the size of the "pie" and consequently get a higher benefit. If, in contrast, the free-of-charge rule $\alpha = 0$ applies, the gatekeeper does not expect any monetary benefit from the information rents and therefore it has higher incentives to increase its profitability through its market operations, by collecting and analyzing a higher amount of data.

**Proposition 4.** *The GDPR's free-of-charge rule over information sharing, $\alpha = 0$ corresponds to a price $T(q; \alpha = 0)$ that maximizes the incentives of the gatekeeper platform to invest in data collection and analysis, for given amount q.*

The implications of the amount information sharing on optimal level $q_B^i$ depend on whether interaction data (or the amount of market information) exhibits increasing, constant or decreasing returns to scale with respect to network value at $q_B + q$ as well as on the convexity of the cost function $C(\cdot)$. In general, if platform $j$ collects the amount of information $q_j = \hat{q}_j$, then, when $\mu_j''(\cdot)|_{q_j=\hat{q}_j} > 0$ ($\mu_j''(\cdot)|_{q_j=\hat{q}_j} < 0$), data exhibits increasing (decreasing) returns to scale at $q_j = \hat{q}_j$. Constant returns to scale at $q_j = \hat{q}_j$ correspond to $\mu_j''(\cdot)|_{q_j=\hat{q}_j} = 0$.

When interaction data exhibits strong increasing returns to scale in network value the the cost function is only weakly convex at $q_B + q$, then, from the first-order-condition (9) on $q_B^i$ we conclude that $q_B^i$ is strictly increasing in $q$. The potential of getting further insights from platform $A$ and improve its market performance makes platform $B$ eager to collect more information on the interaction of their users. Otherwise, $q_B^i$ is strictly declining in $q$.

The implications of $\alpha$ for $q_B^i$ also depend on the second order effects of information on network value. When information exhibits increasing or constant returns to scale, as $\alpha$ increases, platform $B$ gets less benefits from the information sharing mechanism and finds optimal to respond by choosing a lower amount $q_B^i$. When, instead, data exhibits diminishing returns, an increase in $\alpha$ leads to an increase in $q_B^i$.

The regulator has two key instruments with respect to its mandated information sharing mechanism: the amount of information sharing, $q$ and the share of information sharing benefit that is captured by the the gatekeeper, $\alpha$. The regulatory intervention requires to consider the values of these two instruments. Their values define how the mandated mechanism will affect the incentives of platforms to collect and analyze data. Data collection brings network value which is valuable for consumers and sellers. So, on the one side, the regulator wants to solve the data bottleneck market failure, but on the other hand, it should take into account how information sharing of amount $q$ at price $T(q;\alpha)$ affects amounts $q_A^i$ and $q_B^i$ and the network value of joining each platform.

The welfare implications of information sharing illustrated in the previous section remain valid as long as the following proposition applies.

**Proposition 5.** *An information sharing mechanism from platform A to platform B that increases $\mu_B(q_B^i)$ without decreasing $\mu_A(q_A^i)+\mu_B(q_B^i)$ leads to higher consumer and business user welfare.*

This is a sufficient condition. When information sharing leads to a higher (aggregate) network value, it enables more multi-homing and reduces the data bottleneck. This is easier met when data does exhibit increasing returns to scale (at least in the neighborhood of $q_B$ and along the path towards $q_B+q$).

It is not uncommon to consider a S-shaped relationship between data and the value it creates (Hagiu & Wright, 2020; Parker et al., 2022; Posner & Weyl, 2018; Valavi, Hestness, Ardalani, & Iansiti, 2022). Consider, for example, the data-network value relationship depicted in Figure 3. Any sharing of information $q$ will move, network value $\mu_B$ to the right on the curve and network value $\mu_A$ to the left on the curve. If the curve is relatively flat at the region where $\mu_A$ lies (diminishing returns for the network value of platform $A$), the decline in the network value consumers enjoy on platform A will be modest in comparison to the large increase in $\mu_B$ which lies in the steepest region of the curve (increasing returns for the network value of platform B). So, for such functions of network value, information sharing from platform A to platform B is more likely to be welfare-improving.

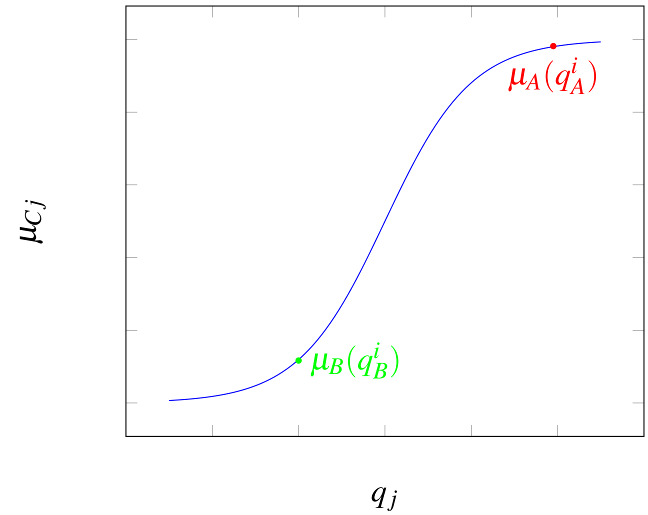


**Figure 3:** Data returns to scale: An illustration.

The regulator now can set $\alpha = 0$. Price $T(q; \alpha = 0)$ has a positive impact on the incentives of platform $A$ to invest on data collection and analysis (Proposition 4). It also provides incentives for platform B to choose a higher amount $q_B^i$ given the increasing returns in that part of the curve.

Then, the socially optimal amount of information sharing $q$ is expected to be significant but lower than $q^{SP}$ of the static case of the previous section. This is because we also need to account for the negative impact of information sharing on $q_A^i$ and potentially on $q_B^i$ (as platform $B$ moves upwards on the curve due to the additional information it receives, after some point, diminishing returns kick in).

The social planner still desires a large scale of information sharing towards platform B, especially if platform A is in the frontier where there is a plateau in the relationship between data and network value while platform B can still benefit a lot from the additional insights it gets through information sharing.

Whether such a plateau exists depends on the position of the gatekeeper platform and the type of application it runs. For example, Bajari et al. (2019) shows that such a plateau exists in their forecasting exercise with Amazon data. If instead, we look at competitors to Google, in online search, it is less likely to find them at a plateau point (Petropoulos, 2016). Yahoo, for example, is more likely to be located at the steep part of the curve Schaefer and Sapi (2020); Schaefer et al. (2018).

## 5 Research implications and contribution to the literature

Under great information asymmetry when a gatekeeper platform is present and collects most of the data available in the market, private incentives for information sharing are much smaller than social ones. Analyzing the pricing strategies by each platform, we illustrate that there exists a data bottleneck market failure. The concentration of market information and the resulting asymmetric information structures help gatekeepers to better monetize their information advantage to the expense of their users. The gatekeeper can extract high information rents from their business users by taking the advantage over its monopoly power on the consumer side. Access to more information makes the gatekeeper able to improve its network value proposition and attract more consumers in its platform. It can then charge a higher price to sellers. This is because sellers have now more incentives to join the platform due to both traditional and data-driven network effects that improve the value proposition of the gatekeeper.

To our knowledge, we are the first to illustrate the possibility of the data bottleneck market failure due to information asymmetry. We present a new mechanism through which market data access and allocation can have important implications for competition in two-sided markets. Literature has either focuses on the data and competition relationship in one-sided firms context (De Corniere & Taylor, 2020; Dubus & Legros, 2022; Farboodi, Mihet, Philippon, & Veldkamp, 2019; Farboodi & Veldkamp, 2021; Hagiu & Wright, 2020; Ichihashi, 2021; Prüfer & Schottmüller, 2021), or on the implications of data within a two-sided platform network, but, without studying the implications of data and information for competition between platform intermediaries (Bergemann & Bonatti, 2023; Bergemann, Bonatti, & Gan, 2022; Bounie, Dubus, & Waelbroeck, 2021).

The presence of the data bottleneck in our analysis has important managerial and welfare implications for market conduct and information sharing. On the managerial side, by looking at the private incentives for information sharing, first, we show that a platform should be sufficiently compensated to be convinced to share its data. In fact, the price of information increases because of the data bottleneck market failure. Second, we illustrate how to approximate the optimal amount of information to be shared. The optimal amount of information sharing should carefully balance the following two countervailing forces: the direct market losses though the increased competitive pressure, on the one side, and the benefit through the data sharing payment, on the other.

Moving to the information recipient side, when small platforms have an additional decision to make, how much they will invest in data collection and analysis, we show that their optimal decisions depend on the price they have to pay for the information they receive through sharing and on whether in their market activities context data exhibits increasing or decreasing returns. If they have to pay a high price for information shared with them, it is in general optimal to invest more on own data analysis only when data exhibits decreasing returns to scale. The opposite holds for increasing data-returns. Furthermore, when a platform is expecting to receive a large amount of information by a competitor through a sharing mechanism, it has incentives to invest more by its own on data collection and analysis under increasing returns, due to the positive synergies between incoming market data through sharing and data collected.

On the welfare side, we show that private incentives for information sharing are lower than the social ones, especially under the presence of gatekeeper platforms and if the social planner primary cares about the users of platforms, consumers and sellers. Hence, there is a scope for a regulatory intervention with the establishment of an effective mandated information sharing mechanism in a way that it enables the increase of the network value of market competitors of the gatekeeper platform. Surprisingly, we find that a strict regulation over the price of information (e.g., similar to the GDPR's free-of-charge rule) can induce the gatekeeper to invest more on data collection and analysis improving the network value the users get from their participation in the market. This implies that FRAND terms for the price of information should be sufficiently restrictive if we want to keep market network value high. The socially optimal amount of data that the gatekeeper will be forced to share depends on data returns to scale and the exact relationship between data and network value. Adopting an S-based relationship between data and network value, we illustrate how information sharing can help competitors of the gatekeeper to improve their network value and improve the welfare of consumers and sellers by resolving the data bottleneck market failure.

# 6 Implications for Society and Practice

Big tech firms have been under regulatory scrutiny due to their great market power in digital markets. A traditional regulatory approach relies on competition policy tools that aim at preventing these dominant firms from abusing their power to the expense of their users and competitors.

In this paper, we show that in addition to these tools, we should also consider policies and ways to make digital markets more symmetric in terms of access to market data and information. Dominant platforms with access to a lot of information about their users and the market can exert monopoly power over the unique market insights they have and keep a lion's share of the efficiency gains they bring to the market through their intermediation services. Data concentration and information asymmetries can have direct implications for market competition (Rubinfeld & Gal, 2017).

We have already seen relevant concerns in recent competition policy cases. In the Facebook case in Germany, part of the issue was the unfair competitive advantage of Facebook against its competitors by being able to combine user data from its subsidiaries and improve its services (Kerber & Zolna, 2022). While evaluating the proposed merger between Google and Fitbit, a company manufacturing wearable devices for the health and fitness, a particular concern was the market advantage Google could get in online advertising ecosystem by accessing Fitbit users' health data, with detrimental effects for consumers (Bourreau et al., 2020). This is the reason, the European Commission approved the merger only under the conditions such that ”Google will not use for Google Ads the health and wellness data collected from wrist-worn wearable devices and other Fitbit devices of users" and that "Google will maintain a technical separation of the relevant Fitbit's user data." (European Commission, 2020).

As we illustrate, more symmetric information structures can solve competition concerns and improve consumer welfare. What we need to achieve this state is in addition to competition policy tools, to develop a proper data policy framework. This framework should be able to mandate information sharing and provide specific goals for the amount of information to be shared and the price over this exchange.

We show that a price cap over information sharing is likely to incentivize big tech to invest more on data collection and analysis. That can in principle increase the network benefit for consumers if it does not generate privacy costs.

Our research findings underline the complementary nature between competition and data policy. In particular, we show that when consumers are locked in a big platform, the social benefits from information sharing will be smaller. Competition policy tools that target to eliminate those locked in effects can make a data policy that enforces information sharing to be more effective.

The exact type of data can have important implications for the characteristics of a data policy that enables information sharing. As discussed, our particular focus should be on data that can increase network value. Since platforms facilitate interactions between their users, it is that interaction data that is important for data policy. This data can either incorporate personal information of individuals or not. Non-personal data typically involve interactions between business users only or it is an aggregated overview of platform interactions, to an extend that the risk of de-anonymization is low (De Montjoye, Radaelli, Singh, & Pentland, 2015). This level of aggregation is achieved at the DMA's obligation for Google to share market data with its competitors in online search (European Commission, 2022, 2023).

Our findings suggest that such mandated data sharing regulations over non-personal data move to the correct direction when a gatekeeper is involved and the level of information asymmetry is high. The DMA adopts a FRAND pricing scheme as a compensation to Google for sharing its non-personal aggregated market data. This FRAND pricing has not been precisely defined yet. Our analysis in Section 4 shows that gatekeeper's incentives to collect and analyze data are not adversely affected if this compensation is small. On the contrary, the lower the price for sharing information the gatekeeper receives, the higher the incentives it has to collect interaction data. Whether the recipients of Google's market data will have more incentives to collect and analyse data depends on data returns to scale. So, we provide some guidance to the regulator on what the implications of the definition of the FRAND pricing will be and how it depends on key parameters of the market. Our research also suggests that there is a scope for extending such information sharing mechanisms beyond online search, also covering other core platform services of our online ecosystem. There is also a scope for such data sharing rules to apply to other jurisdictions. Privacy regulations are an important aspect that shapes data policy and the key parameters for information sharing in the case of personal data. The EU jurisdiction has the most developed privacy regulation, the GDPR. Our research suggest that an important aspect for the information sharing mechanism that should be enforced is that the shared data should enable the competitor platform to improve its network value proposition in order to be able to attract more users and become more competitive. Since platforms are networks that facilitate interactions between their users, it is the sharing of such interaction data that can amplify the network value of the data recipient (Martens, Parker, Petropoulos, & Van Alstyne, 2021).

The GDPR has some flaws that prevent it from enabling an effective information sharing mechanism (Krämer, Senellart, & de Streel, 2020). For example, the language used in this regulation does not clarify which types of data individuals can port and whether their interaction data in the platform is part of the data that can be shared.

The DMA incorporates some improvements with that respect, but it still puts restrictions over the sharing of networked co-generated data. This is because it is a very hard problem to assign property rights over co-generated interaction data. Each individual has the right to only share her own data. But, she does not have the right to share data of someone else. In co-generated data points, the separating line between which data is mine and which data is yours is very blurry. So, enabling an effective information sharing mechanism based on personal data would first require to be able to comply with such privacy regulations and solve the co-generated data property rights problem. Until then, it would be beneficial to focus on mechanisms for information sharing based on non-personal aggregated market data that can help recipients to improve their value proposition and become more competitive, with associated benefits for platform users, consumers and sellers.

# 7 Extensions, limitations and directions for future research

Our product differentiation model captures the main elements of platform competition and illustrates the existence of the data bottleneck market failure and its welfare implications. While we consider consumers as differentiated with each other with respect to their location on the Hotelling line, sellers are homogeneous and a direct implication of this consideration is that in equilibrium they fully multi-home.

One extension of our model would be to differentiate not only across consumers but also across sellers. In this more symmetric treatment of supply and demand side, in equilibrium, we have partial multi-homing in the sellers' side as well. The data bottleneck market failure still emerges even if some of the sellers have a positive surplus due to differentiation (e.g., the ones in the neighborhood of the platform they decide to join). We could also consider that consumers are charged a positive price to join a platform. That would essentially imply that consumers and sellers are to great extent symmetric, in the model. Great information asymmetry between the two platforms and the implications of this asymmetry for the network value each of them offers to their users would result to arrive to a similar equilibrium where the gatekeeper platform captures most of consumers and sellers. The implications of information sharing from the gatekeeper to its competitor, would then result to more competition between the two platforms with more benefits for sellers and consumers (with the share of users that multi-home in each side to be increasing in information sharing).

We could also explore the role of data externalities on the welfare implications of information sharing. For example, when a user participates in a platform and interacts with other users, she may derive less value if she meets the same users in the competing platform. Our model assumed that we have deep markets. For a given user and her preferences, there are multiple other users to interact with. So, it is unlikely to meet the same user if she visits the competing platform. If we move beyond this consideration, then a user that generates a lot of interaction points with multiple users in one platform, will have less incentives to multi-home and visit the competing platform if she believes that she will interact there with the same users. This is formally modelled by Bakos and Halaburda (2020). Bringing their insights to our context, this would me equivalent to an increased locked in effect of users to one platform (e.g., the gatekeeper). As we discussed, such locked in effects provide further incentives for the big platform to share information with its competitor because user locked in constraints the ability of information sharing to lead to more market competition. So, big platform's direct market losses would only be modest. Since it will still receive the data payment from its market rival, it finds it more attractive to share more of its data. Information sharing is still welfare improving but locked in effects limit the benefits it introduces.

In our model we make the assumption that platforms charge participation fees rather than per-transaction fees, as it is usually the case, in practice. This assumption is in agreement with existing literature that explores network effects and platform competition (Armstrong, 2006; Belleflamme & Peitz, 2010, 2019a; Doganoglu & Wright, 2006; Hagiu & Spulber, 2013; Niculescu, Wu, & Xu, 2018; Tan, Anderson Jr, & Parker, 2020). Note that the way we model platform competition is compatible with considering per-transaction fees, instead. Sellers' choice to join a platform depend on how much network value they can derive from their participation and the price they have to pay in order to join. Whether this price takes the form of a participation fee or a per-transaction fee does not matter for sellers' choice.

Per-transaction fees would make a difference if we were studying how users interact within the platform and how the fee paid by sellers affect their offerings within the platform. An interesting avenue for future research is to evaluate how information structures and sharing would affect market equilibria and welfare adopting such a more "microscopic" approach of the platform market. More information across the platform participants would intuitively lead to more efficient transactions. However, platforms might have private incentives to limit the share of valuable market information to their users. For example, if a platform also supplies its own products in the market as a vertically integrated firm, it might have incentives not to share market information with the third-party sellers to the benefit of its own seller.

A second line of research which is a promising avenue for influential contributions has to do with designing mechanisms for data sharing at a privacy-preserving way. Being able to provide some solutions with practical insights for the sharing of personal data can boost the benefits for platform users provided that privacy protection remains strong. Some proposals are already on the table whose main arguments point toward using data and information to make digital markets more efficient (Acemoglu, Makhdoumi, Malekian, & Ozdaglar, 2022; Agarwal, Dahleh, & Sarkar, 2019; Arrieta-Ibarra, Goff, Jiménez-Hernández, Lanier, & Weyl, 2018; Bergemann et al., 2023; Hardjono & Pentland, 2019; Koutroumpis, Leiponen, & Thomas, 2020; Van Alstyne, Petropoulos, Parker, & Martens, 2021).

Last but not least, our paper illustrates that data returns to network value can have important implications for how information sharing affects platform strategies and welfare. Better understanding the second order effects of data on network value is very important for creating efficient information structures and should be a priority for empirical research.

**Acknowledgments:** Our paper has benefited from inspiring discussions with Seth Benzell, Dirk Bergemann, Hemant Bhargava, Alessandro Bonatti, Erik Brynjolfsson, Luis Cabral, Rebecca Christie, Maria Demertzis, Erika Douglas, Nestor Duch-Brown, Michal Gal, Avi Goldfarb, Justus

Haucap, Nina Huang, Ginger Jin, Jan Krämer, Mike Palazzolo, Hani Safadi, Sameer Mehta, Jordan Suchow, Alex Pentland, Nicolas Petit, Maciej Sobolewski, Catherine Tucker, Tommaso Valletti, Reinhilde Veugelers, Alexander White as well as participants in various conferences and other presentations. Georgios Petropoulos gratefully acknowledges financial support from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant No. 799093.

# References

Abis, S., & Veldkamp, L. (2020). The changing economics of knowledge production. *Available at SSRN 3570130*.

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, *14*(4), 218–256.

Agarwal, A., Dahleh, M., & Sarkar, T. (2019). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 acm conference on economics and computation* (pp. 701–726).

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

Armstrong, M., & Wright, J. (2007). Two-sided markets, competitive bottlenecks and exclusive contracts. *Economic Theory*, *32*(2), 353–380.

Arnold, R., Marcus, J. S., Petropoulos, G., & Schneider, A. (2018). Is data the new oil? diminishing returns to scale.

Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., & Weyl, E. G. (2018). Should we treat data as labor? moving beyond "free". In *aea papers and proceedings* (Vol. 108, pp. 38–42).

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational issues of big data and analytics in networked business. *MIS quarterly*, *40*(4), 807–818.

Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. In *Aea papers and proceedings* (Vol. 109, pp. 33–37).

Bakos, Y., & Halaburda, H. (2020). Platform competition with multihoming on both sides: Subsidize or not? *Management Science*, *66*(12), 5599–5607.

Belleflamme, P., & Peitz, M. (2010). Platform competition and seller investment incentives. *European Economic Review*, *54*(8), 1059–1076.

Belleflamme, P., & Peitz, M. (2019a). Managing competition on a two-sided platform. *Journal of Economics & Management Strategy*, *28*(1), 5–22.

Belleflamme, P., & Peitz, M. (2019b). Platform competition: Who benefits from multihoming? *International Journal of Industrial Organization*, *64*, 1–26.

Bergemann, D., & Bonatti, A. (2023). Data, competition, and digital platforms. *arXiv preprint arXiv:2304.07653*.

Bergemann, D., Bonatti, A., & Gan, T. (2022). The economics of social data. *The RAND Journal of Economics*, *53*(2), 263–296.

Bergemann, D., Crémer, J., Dinielli, D., Groh, C.-C., Heidhues, P., Schäfer, M., ... Sullivan, M. (2023). Market design for personal data.

Bounie, D., Dubus, A., & Waelbroeck, P. (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*(2), 283–313.

Bourreau, M., Caffarra, C., Chen, Z., Choe, C., Crawford, G. S., Duso, T., ... others (2020). *Google/fitbit will monetise health data and harm consumers*. Centre for Economic Policy Research.

Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T. M., & Van Alstyne, M. W. (2021). The eu digital markets act: A report from a panel of economic experts. *Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T., and Van Alstyne, M., The EU Digital Markets Act,* *Publications Office of the European Union, Luxembourg*.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165–1188.

Claussen, J., Peukert, C., & Sen, A. (2019). The editor vs. the algorithm: Targeting, data and externalities in online news. *Data and Externalities in Online News (June 5, 2019)*.

Condorelli, D., & Padilla, J. (2020). Harnessing platform envelopment in the digital world. *Journal of Competition Law & Economics*, *16*(2), 143–187.

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). *Introduction—platforms and infrastructures in the digital age.* INFORMS.

Cooper, B. L., Watson, H. J., Wixom, B. H., & Goodhue, D. L. (2000). Data warehousing supports corporate strategy at first american corporation. *MIS quarterly*, 547–567.

Crémer, J., de Montjoye, Y.-A., & Schweitzer, H. (2019). Competition policy for the digital era. *Report for the European Commission*.

De Corniere, A., & Taylor, G. (2020). Data and competition: a general framework with applications to mergers. *Market Structure, and Privacy Policy, CEPR Discussion Papers*(14446).

De Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, *347*(6221), 536–539.

Doganoglu, T., & Wright, J. (2006). Multihoming and compatibility. *International Journal of Industrial Organization*, *24*, 45–67.

Dou, Y., & Wu, D. (2021). Platform competition under network effects: Piggybacking and optimal subsidization. *Information Systems Research*, *32*(3), 820–835.

Dubus, A., & Legros, P. (2022). The sale of data: Learning synergies before m&as. *Available at SSRN 4322714*.

Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2017). The economics of ownership, access and trade in digital data.

Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform envelopment. *Strategic manage-*

*ment journal*, *32*(12), 1270–1285.

European Commission. (2016). *Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)* (Tech. Rep.). European Commission. Retrieved from https://gdpr-info.eu/ (OJ L 119, 4.5.2016, p. 1–88)

European Commission. (2020). *Mergers: Commission clears acquisition of fitbit by google, subject to conditions* (Tech. Rep.). Retrieved from https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_20_2484/IP_20_2484_EN.pdf (Press Release)

European Commission. (2022). *Regulation (eu) 2022/1925 of the european parliament and of the council of 14 september 2022 on contestable and fair markets in the digital sector and amending directives (eu) 2019/1937 and (eu) 2020/1828 (digital markets act)* (Tech. Rep.). European Commission. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R1925 (OJ L 265, 12.10.2022, p. 1–66)

European Commission. (2023). *Digital markets act: Commission designates six gatekeepers* (Tech. Rep.). European Commission. Retrieved from https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4328 (Press Release)

Evans, D. S. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of economic perspectives*, *23*(3), 37–60.

Farboodi, M., Mihet, R., Philippon, T., & Veldkamp, L. (2019). Big data and firm dynamics. In *Aea papers and proceedings* (Vol. 109, pp. 38–42).

Farboodi, M., & Veldkamp, L. (2021). *A model of the data economy* (Tech. Rep.). National Bureau of Economic Research.

Furman, J., Coyle, D., Fletcher, A., McAuley, D., & Marsden, P. (2019). *Unlocking digital competitionreport of the digital competition expert panel* (Tech. Rep.). UK Government.

George, G., Haas, M. R., & Pentland, A. (2014). *Big data and management* (Vol. 57) (No. 2). Academy of Management Briarcliff Manor, NY.

Goldfarb, A., & Tucker, C. E. (2011). Privacy regulation and online advertising. *Management science*, *57*(1), 57–71.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of management review*, *46*(3), 534–551.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2022). Data network effects: Key conditions, shared data, and the data value duality. *Academy of Management Review*, *47*(1), 189–192.

Hagiu, A., & Spulber, D. (2013). First-party content and coordination in two-sided markets. *Management Science*, *59*(4), 933–949.

Hagiu, A., & Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *RAND Journal of Economics*.

Hardjono, T., & Pentland, A. (2019). Data cooperatives: Towards a foundation for decentralized personal data management. *arXiv preprint arXiv:1905.08819*.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, *39*(153), 41–57.

Ichihashi, S. (2021). The economics of data externalities. *Journal of Economic Theory*, *196*, 105316.

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*(9), 2819–2858.

Kerber, W., & Zolna, K. K. (2022). The german facebook case: The law and economics of the relationship between competition and data protection law. *European Journal of Law and Economics*, *54*(2), 217–250.

Kitchens, B., Dobolyi, D., Li, J., & Abbasi, A. (2018). Advanced customer analytics: Strategic value through integration of relationship-oriented big data. *Journal of Management Infor-*

*mation Systems*, *35*(2), 540–574.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660.

Krämer, J., Senellart, P., & de Streel, A. (2020). *Making data portability more effective for the digital economy: Economic implications and regulatory challenges*. Centre on Regulation in Europe asbl (CERRE).

Lambrecht, A., & Tucker, C. E. (2015). Can big data protect a firm from competition? *Available at SSRN 2705530*.

Lerner, A. V. (2014). The role of'big data'in online platform competition. *Available at SSRN 2482780*.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., ... others (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Martens, B. (2021). Data access, consumer interests and social welfare–an economic perspective on data. In *Data access, consumer interests and public welfare* (pp. 69–102).

Martens, B., Parker, G., Petropoulos, G., & Van Alstyne, M. W. (2021). Towards efficient information sharing in network markets.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60–68.

McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic management journal*, *38*(1), 141–160.

Munger, M. (2015). Coase and the 'sharing economy'. In C. Veljanofski (Ed.), *Forever contemporary: the economics of ronald coase* (pp. 187–208). Institute of Economic Affairs.

Niculescu, M. F., Wu, D., & Xu, L. (2018). Strategic intellectual property sharing: Competition on an open technology platform under network effects. *Information Systems Research*, *29*(2), 498–519.

Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, *29*(1), 3.

Parker, G. G., Petropoulos, G., & Van Alstyne, M. W. (2022). Digital platforms and antitrust. *Oxford Handbook of Institutions of Economic Governance and Market Regulation*.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management science*, *51*(10), 1494–1504.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company.

Petropoulos, G. (2016). Search engines, big data and network effects. *Bruegel Blog Post*.

Posner, E. A., & Weyl, E. G. (2018). *Radical markets: Uprooting capitalism and democracy for a just society*. Princeton, NJ: Princeton University Press.

Prüfer, J., & Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, *69*(4), 967–1008.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, *1*(4), 990–1029.

Rubinfeld, D. L., & Gal, M. S. (2017). Access barriers to big data. *Ariz. L. Rev.*, *59*, 339.

Schaefer, M., & Sapi, G. (2020). Learning from data and network effects: The example of internet search.

Schaefer, M., Sapi, G., & Lorincz, S. (2018). The effect of big data on recommendation quality. *The Example of Internet Search*.

Scott Morton, F. (2019). *Final report of the subcommittee on market structure and antitrust* (Tech. Rep.). Stigler Committee on Digital Platforms.

Stucke, M. E., & Grunes, A. P. (2016). Introduction: big data and competition policy. *Big Data and Competition Policy, Oxford University Press (2016)*.

Tan, B., Anderson Jr, E. G., & Parker, G. G. (2020). Platform pricing and investment to drive third-party value creation in two-sided networks. *Information Systems Research*, *31*(1), 217–239.

Tucker, C. (2019). Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility. *Review of industrial Organization*, *54*(4), 683–694.

Valavi, E., Hestness, J., Ardalani, N., & Iansiti, M. (2022). Time and the value of data. *arXiv preprint arXiv:2203.09118*.

Van Alstyne, M. W., Petropoulos, G., Parker, G., & Martens, B. (2021). ’in situ’data rights. *Communications of the ACM*, *64*(12), 34–35.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS quarterly*, 17–41.

Woerner, S. L., & Wixom, B. H. (2015). Big data: extending the business strategy toolbox. *Journal of information technology*, *30*(1), 60–62.