MIS Quarterly

RESEARCH ARTICLE

Regulating Digital Platform Ecosystems Through Data sharing and Data Siloing: Consequences for Innovation and Welfare[1]

Jan Krämer
University of Passau, Chair of Internet and Telecommunications Business
Passau, GERMANY {jan.kraemer@uni-passau.de}

Shiva Shekhar
Tilburg University, Tilburg School of Economics and Management (TiSEM)
Tilburg, THE NETHERLANDS {s.shekhar_1@tilburguniversity.edu}

Digital platform ecosystems thrive on their ability to acquire and leverage user data across multiple data-driven services. This enables dominant platforms to harness insights obtained from their primary markets, where user data is collected, thus gaining a competitive advantage in secondary markets, where they exploit this data. While data cross-use brings about efficiencies, policymakers worldwide have expressed concerns about the economic power and the potential distortion of competition and innovation incentives associated with it. To address these concerns, two distinct and targeted policy interventions have been suggested: data siloing, which restricts the cross-use of data within platform ecosystems, and mandated data sharing with competitors. Using an analytical model that examines data cross-use in digital platform ecosystems, we analyzed the impact of data siloing and data sharing obligations, and their interaction on competition, innovation, consumer welfare, and overall social welfare. Our findings indicate that an optimal policy involves data sharing without data siloing, whereas the EU's Digital Markets Act currently mandates both types of data cross-use regulation.

Keywords: Data cross-use, digital platform ecosystems, regulation, data sharing, data siloing

Introduction

Digital platforms, characterized by their network effects, data-driven operations, and ability to match supply and demand at an unprecedented scale, have attained unparalleled dominance in online markets. Whether focusing on search, social media, or e-commerce, a few dominant platforms have achieved an extraordinary level of market concentration, posing significant challenges to traditional businesses and raising concerns among policymakers worldwide. The pervasive influence of a few digital platforms has allowed them to create expansive ecosystems spanning multiple interconnected markets. The creation of these ecosystems is often fueled by the cross-use of data: As consumers use the service of a dominant platform, they inevitably create a data footprint (e.g., location data, search and click data) that platforms can use to gain valuable insights for other services as well, enabling them to enhance their offerings and maintain a competitive edge. Thus, as dominant platforms expand their influence, they can capitalize on the wealth of data originating from their already established, dominated primary markets, allowing them to exert market power also in not yet dominated (secondary) markets.

[1] Siva Viswanathan was the accepting senior editor for this paper. Hanna Halaburda served as the associate editor. The authors are listed alphabetically and contributed equally.

Given the expanding digital platform ecosystems and the increasing use of data across platforms, policymakers are faced with the challenge of striking a delicate balance between fostering innovation and competition in digital markets. Data cross-use presents both opportunities and challenges for competition and innovation. On the one hand, it can drive innovation by enabling incumbent platforms to gain efficiencies that ultimately benefit users. On the other hand, data-driven competitive advantages in one market can be exploited in other markets, restricting the ability of rival platforms to compete and hindering their innovation efforts. The lack of competition may also discourage incumbent platforms from innovating as vigorously as they would with stronger competition. This would negatively impact consumers. It is primarily this chilling effect on innovation that has prompted policymakers to introduce regulations aimed at leveling the playing field in emerging platform markets, ensuring platforms' competitiveness even when connected to established platform markets that are already monopolized.

In this paper, we investigate two policy interventions specifically targeted at regulating data cross-use in digital platform ecosystems, *data siloing* and *data sharing*, and ask how these policy interventions impact innovation, competition, and welfare in digital platform ecosystems fueled by the cross-use of data. While data siloing puts restrictions on a dominant platform's ability to cross-use data from one market in another market, data sharing requires a dominant platform to share cross-used data with rival firms. Both of these interventions are included in the Digital Markets Act (DMA), a law enacted in the European Union in November 2022, whose explicit goal is to improve contestability, innovation, and fairness in digital markets.[2]

For example, Google's ability to cross-use data and algorithmic insights from consumers' search (e.g., on locations, local businesses, opening hours) through its dominant search engine (primary market) enabled it to enter the market for digital maps (secondary market) and offer a superior service (Prüfer & Schottmüller, 2021). Similarly, Google recently entered the health insurance market (Vengattil & Humer, 2020), where its ability to cross-use search data (as well as other relevant data, such as from maps) allows it to personalize its insurance contracts and improve risk models much better than standard insurance companies (Landy, 2020). In reverse, this also means that the lack of access to similar data troves leads to a distinct disadvantage, which can be a major impediment to rivals seeking to compete and innovate in platform markets (Zhu & Iansiti, 2012; Sun et al., 2024). In this context, strict data siloing would disallow Google to cross-use data from search in the insurance market, thereby leveling the playing field, whereas data sharing would seek to level the playing field by requiring Google to share cross-used data from its search business with other firms in the insurance market.

Another recent example of platform ecosystem building through data cross-use is Apple's strategy to capitalize on the vast data generated by its mobile devices and iOS platform (primary market) to extend its reach into the realm of mobile advertising (Loveless, 2022). By using new features such as app tracking transparency (enforced from iOS 14.5) to prevent other ad tech providers from accessing such user-level data in the name of privacy and through blocking trackers and cookies by default using Private Browsing (starting with macOS Sonoma), Apple has been able to gain access to superior data on consumer behavior from its iOS (primary market), which enables it to offer better-targeted advertising solutions on Apple devices than other ad tech providers.

The DMA regulates how dominant platforms (called "gatekeepers" in the DMA) can leverage user data obtained through their regulated services (called "core platform services") within their ecosystems. Search engines and operating systems—i.e., primary services, in the examples above—are such core platform services, and, among other firms, Google and Apple serve as gatekeepers under the DMA. Specifically, Article 5(2) of the DMA explicitly forbids gatekeepers to "cross-use personal data from the relevant core platform service in other services provided separately by the gatekeeper" unless users explicitly consent to this. This corresponds to *data siloing*. Although data siloing arguably levels the playing field in the secondary market, it comes at the cost of reducing the value generation (efficiencies) associated with data cross-use between markets (Krämer & Schnurr, 2022) and may thus limit innovation in digital markets.

Further, Article 6(11) of the DMA demands that "the gatekeeper shall provide to any third-party undertaking providing online search engines ... access ... to ranking, query, click, and view data ... generated by end users on its online search engines." This represents an example of *mandated data sharing*. Several scholars have cited mandated data sharing as a promising policy intervention (e.g., Argenton & Prüfer, 2012; Krämer & Schnurr, 2022; Prüfer & Schottmüller, 2021; Parker et al., 2021), as it immediately exploits the nonrivalry of data, which is the main economic property that makes data inputs distinct from physical inputs. However, mandated data sharing may also have a chilling effect on incumbents' incentives to

[2] Other jurisdictions, such as the U.K. and the U.S., have proposed similar pieces of regulation, but have not yet formally adopted them. The U.K. is in the process of establishing a Digital Markets Unit (DMU), which in the future could mandate data siloing or data sharing as a possible remedy for platforms with "strategic market status" (Furman et al., 2019; CMA, 2020). In the U.S., the American Innovation and Choice Online Act (H.R. 3816) was enacted by the U.S. Congress in June 2021 and contains similar provisions. These regulations, like the DMA, are considered necessary to complement antitrust law, which has been found to be too slow and not suited to the context of digital markets (Abrahamson, 2014; Graef et al., 2015; Gilbert, 2023).

innovate, as it leads to a loss of exclusivity over the data. At the same time, increased competition due to data sharing can spur innovation. Taken together, the effects of regulated data sharing on innovation are ex ante ambiguous.[3]

Despite the practical relevance of data siloing and data sharing for the regulation of data-centric markets and online platforms, the impact of two approaches to innovation incentives in digital markets has not yet been studied in detail. Using a game-theoretic approach, we account for the potential for various degrees to which data siloing or data sharing may or can be imposed in practice due to deliberate regulatory considerations or technical and legal limitations.[4] Further, we consider the possibility of the two data policies interacting with each other: as more data is siloed, less can be shared.

Our results suggest that data siloing and data sharing have mixed effects on innovation and welfare in platform ecosystems and have important implications for market participants and policymakers. Our findings show that both types of data cross-use regulation decrease the innovation efforts of the incumbent, which also leads to lower demand, and hence less data availability in the primary market, which ultimately limits the source of value generation. In reverse, stricter data sharing regulation increases the data-induced efficiency and innovation efforts of rival platforms in the secondary market. Data siloing also has the ability to increase innovation efforts by rivals, but only if the level of data sharing is low. When the level of data sharing is high, stricter data siloing decreases the innovation efforts of the rival, leading to overall lower innovation and welfare in both markets. Therefore, no regulation (i.e., no siloing and no sharing) may yield better outcomes than some regulation— which would appear to level the playing field (e.g., strict siloing, but low level of data sharing) and may result from a political compromise or be guided by ease of implementation. Our results indicate that the most efficient market outcomes can be achieved with high degrees of data sharing but no data siloing. These results are robust across numerous model extensions. Our results have immediate application for the ensuing regulation of data-driven digital platform ecosystems, and we derive concrete policy proposals that can guide policymakers to derive data cross-use regulations that truly lead to more innovation and greater consumer welfare.

Related Literature

Our paper contributes to the nascent but growing literature on the regulation of digital platforms and data-related remedies that is situated at the intersection of information systems, economics, strategy and innovation research (Easley et al., 2018; Hagiu & Wright, 2023; Prüfer & Schottmüller, 2021; Parker et al., 2021; Tucker, 2019; Krämer & Schnurr, 2022; Cennamo et al., 2023) and relates to several streams of literature.

First, data cross-use represents a form of *demand-side* economies of scope (Gawer, 2014; Henten & Windekilde, 2022). Traditionally, economies of scope refer to *supply-side* synergies that arise from cost savings when producing multiple products (Teece, 1980; Panzar & Willig, 1981). In the context of platforms, Gawer (2014) extended the concept to economies of scope in innovation to denote cost savings from joint innovation. This concept is still driven by a supply-side logic and stems from the observation that manufacturing platforms used within supply chains are increasingly being shared across firms, ultimately leading to "innovation ecosystems" (Adner & Kapoor, 2010). Economies of scope in demand, on the other hand, arise from increased value to users when more products or services are added (Gawer, 2014; Henten & Windekilde, 2022). In the case of data cross-use, insights gained from the users of Service A are used to increase the value of Service B, representing a case of *data-driven* economies of scope in demand.

Second, data cross-use has some commonalities with economies of scale (Stigler, 1958; Wilson, 1975) and organizational learning effects (Levitt & March, 1988) but is distinct from those concepts. Supply-side economies of scale occur because fixed costs of production or information acquisition lead to a decline in average unit costs within the same market as demand increases. Learning effects are similar to supply-side economies of scale, but unit costs fall along a learning curve over time (Cabral & Riordan, 1994; Argote, 2013) rather than with demand. Demand-side economies of scale are network effects that affect users' value *within* a (multisided) market (Katz & Shapiro, 1985; Parker & Van

[3] In theory, antitrust law could be another route to mandate data sharing (if data is deemed an "essential facility") or to require data siloing (e.g., as a remedy in a merger proceeding), but the legal burden for doing so is high, and there are only a few isolated examples of this in the past. For instance, Twitter was obliged to share data about tweets with PeopleBrowser (Graef et al., 2015), and the merger between Facebook and WhatsApp was cleared only under the condition of data siloing (European Commission, 2014).

[4] In practice, data siloing is easy to circumvent and difficult to monitor for regulators. For example, although strict data siloing served as a remedy for the merger between Facebook and WhatsApp, Facebook was later fined because it violated the remedy and combined user data from WhatsApp with other services (European Commission, 2017). Furthermore, regulators may only demand data siloing by default but allow users to opt in to the combination of their data from the various services that they are using so that the data of those users can nevertheless be cross-used. This is also the case for Article 5(2) of the Digital Markets Act. Similarly, there are limitations to data sharing. It may not be technically feasible to share all data in real time, and privacy regulations limit the depth of user data that can be shared. Regulators can also restrict data sharing to certain types of data (e.g., search and query data), although other types of data would also be available to cross-use.

Alstyne, 2005).[5] While economies of scale and learning effects are prevalent in digital services due to near-zero marginal costs and algorithmic improvements over time, this is not the focus of our study. Instead, we are interested in how the level of demand in one market affects demand and innovation in *another market*, even when the services or products provided are seemingly unrelated. Markets are linked by the value created by the cross-use of data. More demand in one market leads to more generation of user data (as consumers inevitably leave a digital footprint when using the service), which in turn can be cross-used to derive demand-enhancing insights in another market.

An important feature of our analysis is that due to the nonrivalry of data, demand-side economies of scope from data cross-use can potentially be shared with rivals. In this respect, our analysis shares features with the literature on data portability and knowledge sharing between firms, but, as argued below, insights from those contexts are not readily transferable.

Knowledge sharing *within* firm boundaries is a well-studied area (for a review, see, e.g., Small & Sage, 2006). However, the literature on knowledge sharing *across* firm boundaries is much less developed (Ritala et al., 2015). External knowledge sharing is often characterized by knowledge trading (Barachini, 2009), where firms exchange knowledge with each other. Firms generally do so only when they expect to benefit individually and can improve their own competitive position or innovation capabilities (Brusoni et al., 2001; Han et al., 2012; Gupta & Polonsky, 2014). In contrast, regulated data sharing, as studied here, is a one-way data exchange and does not require reciprocity. Indeed, in our setting, the dominant firm would not voluntarily share data with its rival, as it would undermine its own competitive position. Hence, data sharing occurs only because it is mandated by a regulator with the intention of improving the strategic position and innovation incentives of the rival firm with which data is shared, ideally without undermining the innovation incentives of the data provider. Knowledge sharing can also occur in a more uncontrolled manner, e.g., as suggested by innovation diffusion theory (Rogers, 1995), or through unintended information leakage (Baughn et al., 1997; Ritala et al., 2015), or through labor mobility and the socialization of workers (Menon, 2018). Our focus is different, as we study data sharing and data siloing as deliberate market interventions controlled and enforced by policymakers in order to maximize overall market efficiency and consumer welfare.

Data portability is another policy intervention that policymakers have adopted to regulate data-rich platforms. However, data portability is different from data siloing and data sharing as conceptualized here. Data portability has been introduced as a result of privacy regulations and enables individual users to access personal data that they have provided to a data controller, e.g., an online platform. Data portability also allows users to instruct the data controller to transfer personal data directly to a third party. Such data portability provisions exist, for example, in the EU's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Real-time data portability is also included in the Digital Markets Act (DMA) under Article 6(9), i.e., as a provision distinct from and in addition to the data sharing and data siloing provisions that we study here. Data portability differs from data sharing, as considered here because data sharing allows competitors to access data from a broad (representative) set of users and that data is not limited to personal data (e.g., a search engine may be obligated to share data on search results rather than just data provided by users). In contrast, data portability allows individual users to access and share only their own personal data (e.g., only the keywords entered into a search engine), on a per-user request basis. The purpose of data portability is to make data-driven markets more competitive by facilitating the switching between providers in the presence of data-induced lock-in effects (Krämer, 2021). Consequently, the existing theoretical literature on data portability focuses on how data portability affects competition within a given market (Wohlfarth, 2019; Krämer & Stüdlein, 2019; Lam & Liu, 2020) but not on how portability affects innovation or competition across markets, which is the focus of our study.

Our study is closely related to recent studies that consider mandatory data sharing as a means to promote competition and innovation in digital markets, especially in the context of search engines (Argenton & Prüfer, 2012; Prüfer & Schottmüller, 2021; Schaefer & Sapi, 2023). Prüfer and Schottmüller (2021) considered the innovation incentives of competing search engines and found that data-driven network effects can lead to a situation of market dominance by one firm (i.e., the market has tipped), leading to a reduction in innovation incentives for all market participants. They argue that data sharing (within the same market) can restore a level playing field that restores innovation incentives. In our paper, we instead consider how both data sharing and data siloing, as well as the interaction between them, affect innovation incentives in the primary (tipped) market as well as in a secondary (untipped) market. The importance of data access for competition and innovation and the demand-enhancing effect of better data access are widely recognized in both the theoretical (Gawer, 2014; Krämer & Schnurr, 2022; Gregory et al., 2021; Henten & Windekilde, 2022; Kwark et al., 2017; Lam & Liu, 2023) and empirical

[5] There is some controversy in the literature regarding the relationship between indirect network effects and demand-side economies of scope. For example, Gawer (2014) considers demand-side economies of scope to be a type of indirect network effect because demand on market side A can affect value or demand on market side B. However, we distinguish between these concepts because we argue that data cross-use creates demand-side economies of scope across two seemingly unrelated markets, and not just between market sides A nd B belonging to the same (multisided) market.

literature (Schaefer & Sapi, 2023; Klein et al., 2022; Sun et al., 2024). For example, Schaefer and Sapi (2023) used real search engine query logs to empirically investigate the quality improvements from more fine-grained search data. To date, however, the discussion on data sharing has largely focused on competition and innovation in a single platform market, whereas the focus of our study is on the leveraging of data across markets. Moreover, the interplay between data sharing and data siloing has so far been ignored.

Our paper more generally complements the literature on data-driven network effects (Rubinfeld & Gal, 2017; Gregory et al., 2021, 2022; Hagiu & Wright, 2023) that arise from a firm's ability to engage in within-user and across-user learning from data. Gregory et al. (2021, 2022) highlighted the practical relevance and importance of such data-driven network effects for digital markets spurred by advances in artificial intelligence and theorized how the interaction between data availability and AI capabilities leads to enhanced value for users. Similarly, Haftor et al. (2021) considered both within-user and across-user data-driven network effects and showed how these can create consumer lock-in while generating value and efficiency. Hagiu and Wright (2023) considered data-enabled learning as a quality-enhancing effect within a market and found that the degree of competition between rival firms depends largely on the availability of data and the shape of the learning curve. We consider both within-user learning and across-user learning but study their impact on competition and innovation across markets within a platform ecosystem. We contribute to the literature by showing that there is a tension between innovation (enhanced user value) and competition in the cross-market context.

Finally, our study also relates to prior research on cross-market promotions and bundling in marketing research (Goić et al., 2011; Prasad et al., 2015; Yan et al., 2022). For example, Goić et al. (2011) considered a setting in which consumers who purchase a good in a primary market receive price discounts that can be redeemed in a secondary market. Our paper shares with them the insight that cross-market leverage can increase efficiency for firms and consumers. However, we show that the mechanism through which this occurs is different, as we consider a cross-market network effect, where the demand-enhancing effect depends not only on one's own consumption in the primary market but foremost on how many other users have consumed the primary good. Furthermore, our study focuses on the impact of data-related policy interventions and their impact on innovation and competition, while prior literature has focused on the impact of cross-market promotions on the sales and profits of the focal firm.

Model

Players and Environment

We consider a game-theoretic model with two platforms $i \in \{1,2\}$ and two markets $m \in \{A, B\}$: (1) a primary (monopolized) market denoted by $m = A$ and (2) a secondary (competitive) market denoted by $m = B$. For expositional clarity, we assume a Cournot-type competition, whereby platforms strategically set optimal output or demand levels, q_i, along an inverse demand curve $P(Q)$ with $Q = \sum_i q_i$.[6] Hence, $P(Q)$ represents the "price" that can be taken from consumers at a certain level of demand. Therefore, "price" should be understood broadly as a means by the platforms to extract consumer surplus. Indeed, our model insights are not restricted to a setting in which platforms demand an explicit price from consumers. Consumer surplus extraction can also occur in other ways such as through advertisements (which incur ad nuisance costs to consumers) or by creating and selling consumer profiles (which incur privacy costs to consumers). However, a higher "price" or degree of surplus extraction always corresponds to a lower demand (as consumers use the service less when it demands a higher price or shows more ads) so that inverse demand curves are downward sloping. Figure 1 visualizes these downward-sloping inverse demand curves.

Consumer Demand and Platform Profit in the Primary, Monopolized Market A

We consider the following parsimonious inverse demand function in Market A in our main model: $P_A(q_A, v_A) = 1 + v_A - 3q_A$, where v_A signifies the innovation efforts that Platform 1 exerts in Market A.[7] While the precise functional form of the demand function is not material for our results, the main assumption here is that a higher innovation effort, v, shifts the demand curve upwards (e.g., as innovation leads to a better service quality such as better algorithms or more service features), leading to a higher demand at any given price, as visualized in Figure 1a. Let $I(v)$ be a platform's service-specific cost of innovation for exerting effort level v. We assume a standard convex cost function, given by $I(v) = v^2/2$, which means that it becomes increasingly more costly to innovate to increase service quality—i.e., $I'(v) > 0$ and $I''(v) > 0$.[8] Thus, Platform 1's profit in Market A is given by $\Pi_A = P_A q_A - I(v_A)$.[9]

[6] Our insights are robust to other ways of modeling competition, such as through Bertrand-type or Hotelling-type price competition. Details are available upon request.

[7] A microfoundation for this specific demand function is provided in Appendix B, but our insights apply more generally for downward-sloping inverse demand functions.

[8] Our insights do not depend on this specific functional form of the investment function.

[9] After defining a function, we suppress the arguments of the function for easier reading. We again employ the arguments of the functions when needed for clarity.

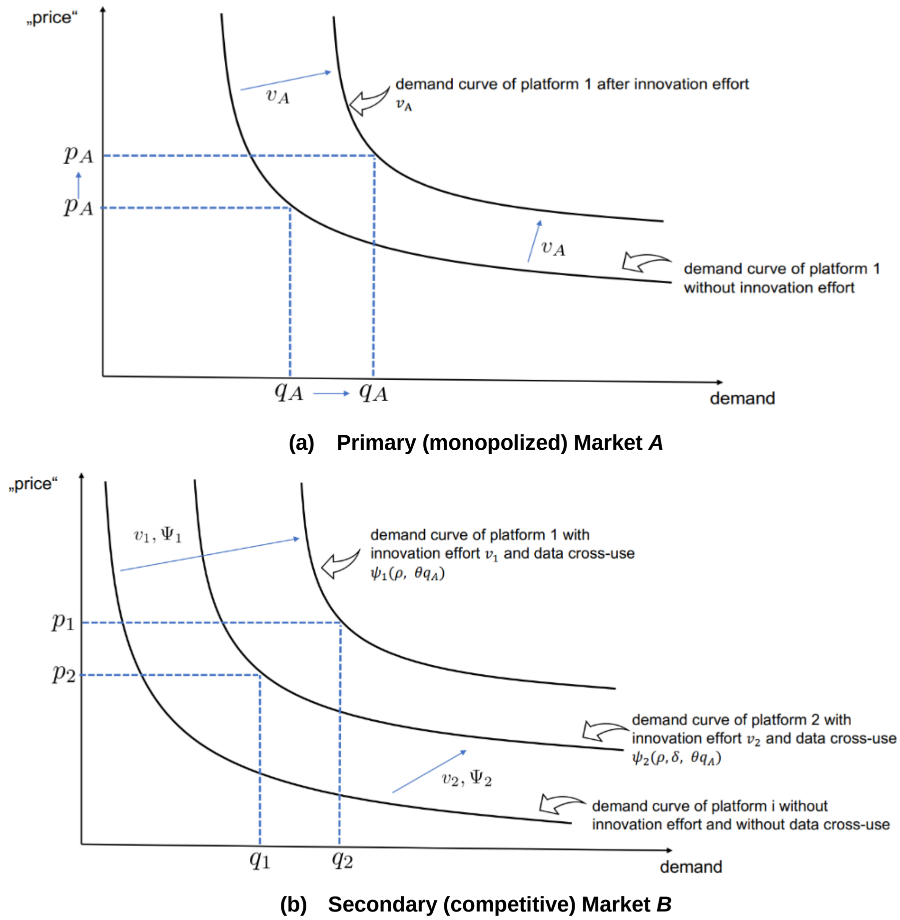


Figure 1. Visualization of the Model Mechanism Through Which Innovation Efforts and Data Cross-Use Affect Market

Table 1. Table of Notations

Variable	Interpretation
Ψ_i	Demand-enhancing effect of data cross-use in Market B for platform $i \in \{1,2\}$.
ρ	The level of mandated data siloing, $\rho \in [0,1]$.
δ	The level of mandated data sharing, $\delta \in [0,1]$.
θ	Degree to which user data from Market A can be cross-used in B.
P_A	Inverse demand function of Platform 1 in Market A.
P_i	Inverse demand function of platform $i \in \{1,2\}$ in Market B.
$I(v)$	Investment cost for innovation effort level v.
Π_i	Profit of platform $i \in \{1,2\}$.
CS_m	Consumer Surplus in market $m \in \{A, B\}$.
Decision variables*:*	
q_A	Consumer demand in Market A.
q_i	Consumer demand of platform i in Market B.
v_A	Innovation effort by Platform 1 in Market A.
v_i	Innovation effort by platform i in Market B for $i \in \{1,2\}$.

Data Cross-Use, Consumer Demand, and Platform Profit in Secondary, Competitive Market B

Our main interest lies in considering the implications of data cross-use between the primary Market A and the secondary Market B on platforms' innovation efforts and on (consumer) welfare, and how such data cross-use should be regulated in order to maximize (consumer) welfare. Specifically, we assume that Platform 1 can leverage its consumer data obtained in Market A to enhance consumer demand in Market B, which provides it with a competitive advantage.

Data Cross-Use Regulation

Data cross-use by Platform 1 in Market B could be regulated in two ways to derive a more level playing field for competition and innovation in Market B:

Data siloing: This refers to a policy whereby the data-rich incumbent is prohibited from exploiting (consumer) data gathered in its primary (monopolized) market to enhance its operations in a secondary (competitive) market. This implies that different business entities of the regulated platform providing services in different markets are proscribed from cross-using data or insights derived thereof in their respective services. We denote the degree of data siloing by $\rho \in [0,1]$ where $\rho = 0$ represents no limitations on data cross-use, and $\rho = 1$ corresponds to a strict data siloing regime where no data can be cross-used.

Data sharing: This policy allows the regulated platform to cross-use the data acquired in the primary market in the secondary market as well but requires the platform to share this data with rivals. We denote the degree of data sharing by $\delta \in [0,1]$, where $\delta = 0$ represents no obligation for data sharing and $\delta = 1$ corresponds to a strict data sharing regime where all data available to Platform 1 in Market B must also be shared with its rival in Market B.

No Regulation

We denote a certain data cross-use regulation by (ρ, δ). This includes the case where $\rho = \delta = 0$, which represents a lenient policy where platforms are not regulated. That is, Platform 1 can make use of all of its data from Market A in Market B (no data siloing) and does not have to share data with the rival platform (no data sharing).

Data Generation and Impact of Cross-Use on Demand

Consumers typically leave a data footprint when using a digital service. For example, by using a search engine, consumers generate query and click data that reveals what they are searching for. Likewise, users of a navigation software inevitably reveal data about their destination and geolocation. Thus, the generation of user data and the usage of a service are often inevitably intertwined. Consequently, the more consumers use the service in Market A, the more that user data is generated in Market A. We therefore proxy the consumer data generated in Market A by the demand in Market A, i.e., q_A. Further, in the main model, we assume that the data generated in Market A is useful to gather insights on consumers more generally in Market B, even if an individual consumer is not directly identifiable in Market B and irrespective of whether that consumer also buys from Platform 1 in Market B (cross-user learning). For example, search and query data from search (Market A) can be useful to identify the health status of a certain region or a certain user population, which would be important information for a health insurance service in Market B. Similarly, data from a navigation software collecting insights on individual routes and traffic (Market A) could be useful for determining the placement of charging stations for electric vehicles (Market B).[10] We parameterize the degree to which data from Market A is relevant for cross-use in Market B by $\theta \in (0,1)$. At $\theta = 0$ there is no value for cross-use (e.g., because the services provided are very different and no inferences can be made from Market A to increase demand in Market B), whereas at $\theta = 1$ the data generated in Market A is also highly useful for enhancing demand in Market B. We model the demand-enhancing potential of data from Market A through cross-use in Market B by means of Platform 1's and 2's inverse demand functions in Market B in the following way:[11] $P_1(v_1, \Psi_1, q_1, q_2) = 1 + v_1 + \Psi_1 - 3(q_1 + q_2)$ and $P_2(v_2, \Psi_2, q_1, q_2) = 1 + v_2 + \Psi_2 - 3(q_1 + q_2)$. Notice that we use the same parsimonious inverse demand function as in Market A, but with two notable differences. First, as we consider a Cournot-type competition model, platform i's inverse demand depends not only on its own strategic demand level, q_i, but also on that of the rival platform, since the two platforms are in competition. Second, and most importantly for our analysis, platform i can shift its inverse demand curve upwards not only by exerting innovation

[10] In Appendix D, we show that our insights are robust to the case where Platform 1 can identify consumers and can offer a cross-market benefit to those consumers who buy from it in both markets (within-user learning).

[11] A microfoundation for these demand expressions can be found in Appendix B.

efforts (v_i), but also by leveraging data cross-use. We denote the magnitude of this data cross-use effect on demand by Ψ_i and note that it is platform specific and depends on the degree to which data from Market A can be repurposed in Market B (parameterized by θ) and on the regulatory policy regime with a given data sharing (δ) and data siloing (ρ) level. Next, we describe the properties of Ψ_i in more detail.

Under a policy regime ($\rho,\ \delta$) imposed on Platform 1, which has immediate access to the data generated in Market A, $\Psi_1(\rho, \theta q_A)$ is a function of the level of cross-use-relevant data generated in Market A, and the degree of data siloing, ρ. For Platform 2, which does not have immediate access to data generated in Market A, $\Psi_2(\delta, \rho, \theta q_A)$ additionally depends on the degree of data sharing, δ. Due to the intermarket linkage via data cross-use, consumers have to anticipate the value generated from demand in Market A when considering which service to use in Market B. In any equilibrium, this anticipated value must be correct.[12]

In favor of clarity, the main results of our paper are shown with a specific functional form for Ψ_i with $\Psi_1(\rho, \theta q_A) = (1-\rho)\theta q_A$ and $\Psi_2(\delta, \rho, \theta q_A) = \delta \Psi_1(\rho, \theta q_A)$. In Appendix F we demonstrate that our insights are robust for more general functional forms for Ψ_i. Importantly, as the amount of data siloing (ρ) increases, both Ψ_1 and Ψ_2 fall, because data siloing limits the amount of data that can be cross-used from Market A in Market B. Furthermore, more regulated data siloing also implies that Platform 1 has less data available for sharing, as it can only share data (originating from Market A) with Platform 2 in Market B if it has access to this data itself in Market B. As a consequence, data siloing and data sharing interact. This is captured by the multiplicative form $(1-\rho)\theta$ embodied in Ψ_i. At the limit $\rho = 1$, no data can be cross-used and $\Psi_1 = \Psi_2 = 0$, yielding the same inverse demand functions in Market B for the two platforms, and thus leveling the playing field for competition and innovation. Instead, when $\rho < 1$ and the level of mandated data sharing (δ) increases, only the inverse demand of Platform 2 increases while the inverse demand of Platform 1 stays the same. This implies that an increase in data sharing also levels the playing field in Market B. Thus, both data siloing and data sharing are different policy means to achieving the same end. But as we will show, they have vastly different implications on innovation and consumer welfare.

Figure 1 visualizes how innovation efforts and data cross-use affect demand and market outcomes. Innovation efforts by Platform 1 in the primary Market A lead to an increased demand in that market, which makes more data available for cross-use and thus also yields a demand-enhancing effect via Ψ_1 in the secondary Market B. The magnitude of that effect is moderated by the policy regime (ρ, δ) and the potential for data cross-use (θ). In case data has to be shared with the rival Platform 2, innovation by Platform 1 in Market A also benefits Platform 2 in Market B via Ψ_2. In addition, each platform can enhance its demand further through innovation efforts in Market B. But as we will show later, innovation incentives in markets A and B are also influenced by the policy regime (ρ, δ) through the intermarket linkages via data cross-use, and competition in Market B.

Platform profits: Platform profits are the sum of revenues in each market that the platform operates in, minus the service-specific innovation costs:

$$\max_{v_A, v_1, q_A, q_1} \Pi_1 \triangleq \underbrace{P_A q_A - I(v_A)}_{\text{Profit in Market } A} + \underbrace{P_1 q_1 - I(v_1)}_{\text{Profit in Market } B}, \text{ and } \max_{v_2, q_2} \Pi_2 \triangleq \underbrace{P_2 q_2 - I(v_2)}_{\text{Profit in Market } B}.$$

Timing and equilibrium concept: In stage $t = 1$, platforms choose their demand-enhancing innovation levels v_A, v_1 and v_2.[13] In stage $t = 2$ observing the innovation levels, consumers anticipate the demand in Market A and the corresponding demand-enhancing effect of data cross-use in Market B.[14] In stage $t = 3$, Platform 1 and Platform 2 simultaneously determine their demand level q_A, q_1 and q_2, respectively. We employ the subgame perfect equilibrium concept. Table 1 summarizes our notation.

The Impact of Data Siloing and Data Sharing on Market Outcomes and Innovation

We solve the game by backwards induction and relegate the proofs to Appendix A. In the following, we highlight the main strategic trade-offs that emerge from the equilibrium analysis.

Demand-Setting Stage

First, we establish the intermarket demand linkage and then discuss how it depends on the policy regime, given innovation effort levels.

[12] As is typical for models with value derived from a network, we assume that the value generated from the network is correctly anticipated in equilibrium. In order to ease notation, we do not differentiate between the actual and anticipated value associated with the network. In the proofs (Appendix A), the model is solved in detail while retaining this distinction.

[13] Innovation effort is chosen before demands because the investment in innovation is a longer term decision than demand and prices.

[14] In Appendix C, we show that our results hold when consumers can observe the demand in Market A.

Lemma 1 (Impact of data cross-use regulation on demand): ***(a)*** *A higher degree of data cross-use (θ) or a higher innovation level in Market A (v_A) always increases Platform 1's demand in Market B, but increases Platform 2's demand in Market B only when coupled with a sufficient level of data sharing ($\delta > 1/2$).* ***(b)*** *More data siloing (larger ρ) leads to lower demand in Market B for Platform 1 but reduces Platform 2's demand in Market B only when the level of data sharing is high ($\delta > 1/2$).* ***(c)*** *More data sharing (larger δ) leads to lower demand for Platform 1 in Market B but a higher demand for Platform 2.* ***(d)*** *More data siloing reduces the total demand in Market B, whereas more data sharing increases the total demand in Market B.*

The lemma shows that demands in Market B are impacted by innovation efforts in Market A and the ability for data cross-use. With a higher innovation effort in Market A (v_A), which results in a larger demand (and hence also more data) in Market A, or a larger relevance of the data cross-use (θ), Platform 1 indeed obtains a larger demand in Market B. This is a direct consequence of the benefits of data cross-use from Market A to Market B.

In contrast, Platform 2 benefits from a higher v_A only if a certain level of data sharing between Platform 1 and Platform 2 is imposed, i.e., $\delta > 1/2$. Intuitively, this means that data sharing regulation must be sufficiently strict in order to benefit the rival Platform 2. A similar argument applies for the effect of an increase in θ on the demand of Platform 2.

Moreover, notice that as the level of data sharing or data siloing increases, Platform 1 finds it profitable to reduce the demand it serves. This result follows directly from the fact that stricter regulation curbs Platform 1's data advantage in Market B and thus lowers its dominance as well. An increase in data siloing levels the playing field for Platform 2, and thus increases its demand, when at the same time data sharing is low. However, an increase in data siloing leads to lower demand of Platform 2 when the data sharing regulation is sufficiently strict i.e., $\delta \geq 1/2$. This is because in this regulatory constellation, Platform 2 would benefit more from data sharing than from data siloing (which limits the amount of data that can be shared) in Market B. Interestingly, the rise in consumer demand served by Platform 2 from increased data sharing is higher than the fall in consumer demand served by Platform 1. This suggests that consumers benefit from data sharing in Market B, given a fixed level of innovation effort in Market A. However, a key feature of our model is that the level of innovation in Market A is not fixed, but chosen strategically by the regulated platform, and is therefore affected by regulation.

Innovation-Setting Stage

In Appendix A, we show that an increase in innovation in Market A encourages Platform 1 to innovate more in Market B as well. In that sense, Platform 1's innovation in Market A is a strategic complement to the innovation of Platform 1 in Market B. Instead, for Platform 2, an increase in innovation by Platform 1 in Market A increases its innovation incentive only for sufficiently high levels of data sharing: $\delta > 9/14$ (see proof of Proposition 1). It is therefore of particular relevance to understand the innovation incentive of Platform 1 in Market A. Differentiating the profit of Platform 1 with respect to v_A yields the following first order condition, highlighting Platform 1's incentives to invest in innovation in Market A:

$$\frac{\partial \Pi_1}{\partial v_A} = \underbrace{\frac{\partial P_A}{\partial v_A} q_A}_{\substack{\text{direct effect in} \\ \text{primary market } (+)}} - \underbrace{\frac{\partial I(v_A)}{\partial v_A}}_{\substack{\text{innovation} \\ \text{cost effect } (-)}} + \underbrace{\left(\underbrace{\underbrace{\frac{\partial P_1}{\partial \Psi_1}}_{(+)} \underbrace{\frac{\partial \Psi_1}{\partial \mathrm{q_A}}}_{(+)}}_{\substack{\text{data cross-use effect} \\ \text{in secondary market}}} + \underbrace{\underbrace{\frac{\partial P_1}{\partial \Psi_1}}_{(-)} \underbrace{\frac{\partial \Psi_1}{\partial \mathrm{q_A}}}_{(+)}}_{\substack{\text{data sharing} \\ \text{effect}}} \right) \underbrace{\frac{\partial \mathrm{q_A}}{\partial v_A} q_1}_{\substack{\text{total demand effect} \\ \text{in primary market } (+)}}}_{\text{Impact of data cross-use on innovation incentives in Market } A} = 0 \quad (2)$$

Innovation incentives in Market A can be decomposed into multiple effects. The sign of the effect denotes whether the effect translates into a positive or a negative innovation incentive. First, there exists a (positive) direct effect and a (negative) innovation cost effect. These effects arise also in the absence of data cross-use and reveal that without data cross-use Platform 1 would face a classic monopolistic optimization problem in Market A. These expressions govern Platform 1's innovation incentives that would arise if the two markets were not linked through data cross-use. However, the complex third term arises only when Platform 1 has the ability for data cross-use in Market B and reveals how Platform 1's innovation incentives in Market A can be positively or negatively affected through data cross-use and regulation. The first term in the parenthesis represents the (positive) *data cross-use effect* in the secondary market: It arises from the fact that a unit increase in innovation effort in Market A increases the marginal revenue of Platform 1 in Market B. This is because increased innovation effort in Market A enhances consumer value which increases demand in Market A, which leads to the generation of more data, which again enhances the demand of Platform 1 in Market B as well. This effect increases Platform 1's incentive to innovate in Market A. Note that a stricter data siloing regulation (higher ρ) will dampen these innovation incentives and lower innovation in Market A. This is because as ρ increases, the marginal gain in Market B from data collection in Market A falls—i.e., $\frac{\partial^2 \Psi_1}{\partial q_A \partial \rho} = -\theta < 0$.

The second (negative) term in the parenthesis, denoted as *data sharing effect*, exists only with mandated data sharing, i.e., if $\delta > 0$. This effect arises from the fact that as innovation in Market A increases, under mandated data sharing, demand of the rival platform also increases, which lowers the marginal revenue of Platform 1 in Market B. This discourages Platform 1 from investing in Market A. Note that as data sharing levels increase (as δ increases), the data availability for Platform 2 increases, which increases the demand of Platform 2, i.e., $\frac{\partial^2 q_2(\cdot)}{\partial q_A \partial \delta} = \frac{2\theta(1-\rho)}{9} \geq 0$. Thus, with more data sharing, the data sharing effect becomes stronger and dampens the incentive of Platform 1 to innovate in Market A further. The sum of the two opposing effects determines whether Platform 1's innovation level in Market A rises or falls vis-á-vis without data cross-use.

Proposition 1 (Innovation): *The innovation effort by Platform 1 in Markets A and B is reduced with stricter data sharing (larger δ) or with stricter data siloing (larger ρ). The innovation effort of Platform 2 always increases with data sharing but increases with data siloing only if the level of data sharing is low ($\delta < 9/14$).*

From the above, it is immediate that without regulation ($\rho = 0, \delta = 0$), the innovation effort reducing data sharing effect is absent and the data cross-use effect is at the highest level, implying that Platform 1's innovation incentives in Market A are the highest in this case.

Insight 1 (Impact on innovation in the primary market): *The level of innovation in the regulated primary market is higher under a data sharing regime ($\rho < 1, \delta > 0$) than under a strict data siloing regime ($\rho = 1$), but always the highest without regulation ($\rho = 0, \delta = 0$).*

We now turn to the discussion of the impact of regulation on innovation incentives in the secondary market. With stricter data siloing or stricter data sharing obligations, innovation incentives by Platform 1 in Market B are also reduced. This result is a combination of two reinforcing effects. First, an increase in data sharing or data siloing increases competition in Market B, which means that Platform 1's benefits from a demand-enhancing innovation are lowered, as more of the increased profits from increased demand are dissipated away in the competitive process. In turn, this lowers the incentive to innovate by Platform 1 in Market B. Second, an increase in data sharing or data siloing also reduces the innovation levels in Market A, which further lowers data cross-use and, in consequence, the incentive to innovate in Market B.

This reduced innovation incentive of Platform 1 affects Platform 2's innovation levels in a more nuanced way. As data sharing increases, Platform 2 has an overall increased innovation incentive resulting from a positive direct and a negative indirect effect. The direct effect arises as follows. An increase in data sharing increases the data cross-use ability of Platform 2, which, ceteris paribus, enhances the innovation incentives of Platform 2. The indirect effect arises from reduced innovation incentives of Platform 1 in Markets A and B. While a reduction in innovation by Platform 1 in Market B positively impacts innovation incentive of Platform 2, a reduction in innovation efforts in Market A may lower the incentive to innovate when data sharing levels are sufficiently high. However, in sum, the positive direct effect dominates and Platform 2 innovates more with data sharing.

Interestingly, an increase in data siloing increases the innovation level of Platform 2 in Market B only when the imposed level of data sharing is relatively low ($\delta < 9/14$). When the level of data sharing regulation is high, Platform 2 significantly benefits from data collected in Market A. Recall that data siloing lowers the amount of data collected as well as the possibility to cross-use the collected data. This twofold reinforcing negative effect of data siloing on data generation and cross-use also reduces Platform 2's innovation incentives.

Insight 2 (Impact on innovation in the secondary market): *For the regulated Platform 1, the level of innovation in the secondary market is higher under a data sharing regime ($\rho < 1, \delta > 0$) than under a strict data siloing regime ($\rho = 1$). However, innovation by Platform 1 is always the highest without regulation ($\rho = 0, \delta = 0$).*

Impact on Welfare

We now consider the impact of data cross-use regulation on (1) consumer surplus, and (2) total welfare. The detailed formal analysis is relegated to Appendix A.

Consumer Surplus

Under Cournot-type competition, the consumer surplus can be derived as a function of demand and is given by $CS_A = \frac{3}{2}(q_A^\star)^2$ and $CS_B = \frac{3(q_1^\star + q_2^\star)^2}{2}$ in Markets A and B, respectively (see Appendix A). The star denotes the equilibrium values. Lemma 1, which considers the impact of regulation on demand but holding innovation efforts constant, already hints at the nuanced effects of a data cross-use regulation on consumer surplus, since the demand of Platform 1 and Platform 2 can be affected in opposing ways. Furthermore, Proposition 1 asserts that the platform's innovation efforts are also affected by the policy regime.

Proposition 2 (Consumer Surplus): *A stricter data siloing regulation (larger ρ) decreases consumer surplus in Markets A and B. A stricter data sharing regulation (larger δ), lowers consumer surplus in Market A, but increases consumer surplus in Market B. The total consumer surplus increases with more data sharing (a larger δ).*

The intuition for this proposition is as follows. A stricter data siloing regime reduces innovation efforts in Market A which also lowers consumer surplus in Market A. In Market B, stricter data siloing affects consumer welfare as follows. First, a direct consumer welfare loss from a decrease in the ability to cross-use data in Market B as well as the welfare loss from a decrease in innovation in Market A, which leads to lower demand and less data being generated in Market A that can be cross-used in Market B. This negative effect on consumer surplus is further reinforced by a reduction in innovation efforts by Platform 1 in Market B. In contrast, Platform 2 increases its innovation with data siloing, but only when the mandated level of data sharing is low. Otherwise, the negative effect on consumer surplus is further reinforced by reduced innovation of Platform 2. The total effect of more data siloing on consumer surplus is unambiguously negative. Thus, a stricter data siloing regime may hurt consumers in both markets.

The impact of an increase in data sharing on consumer surplus is likewise driven by several underlying effects: In Market A, an increase in data sharing lowers innovation efforts which reduces consumer surplus. In Market B, more data sharing has two opposing effects on consumer surplus. First, a direct positive effect arises from increased innovation by Platform 2. Next to increasing demand, this also makes Platform 2 a fiercer competitor in Market B. Second, data sharing negatively impacts consumer surplus through lower innovation efforts by Platform 1 in both Market A and Market B. In Market B, since the positive effect of increased competition and higher innovation effort by Platform 2 dominates, data sharing increases consumer surplus in Market B. Whether more data sharing increases consumer surplus in the aggregate across Markets A and B depends on the strength of the underlying effects and the relative size of each market.[15] The positive impact of data sharing on consumer surplus in Market B can dominate the negative impact on Market A if, at the same time, the level of data siloing is relatively low. This is visualized in Figure 2a, which compares the consumer surplus for all possible degrees of data sharing regulation (δ) and data siloing regulation (ρ) relative to the benchmark of no regulation ($\rho = 0, \delta = 0$), located in the southwest corner. This suggests that the consumer surplus loss in Market A from reduced innovation may be outweighed by the consumer surplus gains in Market B, rendering a data sharing regulation policy potentially beneficial only if coupled with a lenient data siloing regime.

[15] In Appendix E, we consider a model variant where the market size of Markets A and B differ.

Total Welfare

Next, we consider total consumer welfare, which is the sum of producers' surplus (i.e., the platforms' profits) and consumer surplus.

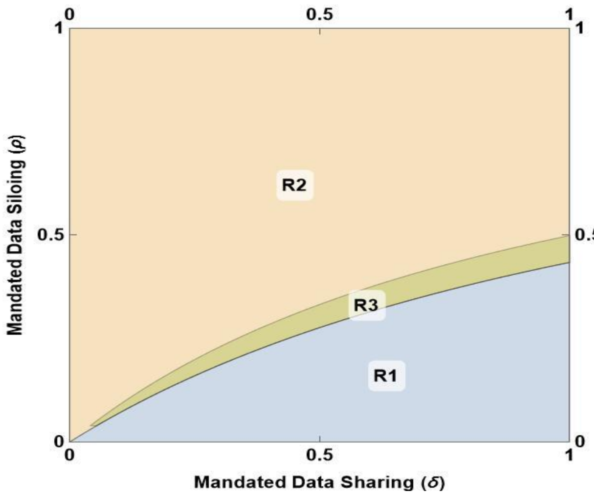
Proposition 3 (Producers' Surplus): *Platform 1's profit decreases with stricter data sharing (larger δ) or stricter data siloing (larger ρ). Platform 2's profit increases with stricter data sharing, but increases with stricter data siloing only if the level of data sharing is low ($\delta < 9/14$).*

Proposition 3 highlights that data cross-use regulation can indeed level the playing field and make the platforms' profits more similar, as both data sharing as well as data siloing tend to have opposite effects on the profits of Platform 1 (the data provider) and Platform 2 (the data recipient). However, Proposition 3 also points again to a possible negative interaction of a data siloing regulation and a data sharing regulation. Based on these insights, we can now derive the impact of the two types of regulations on total welfare.

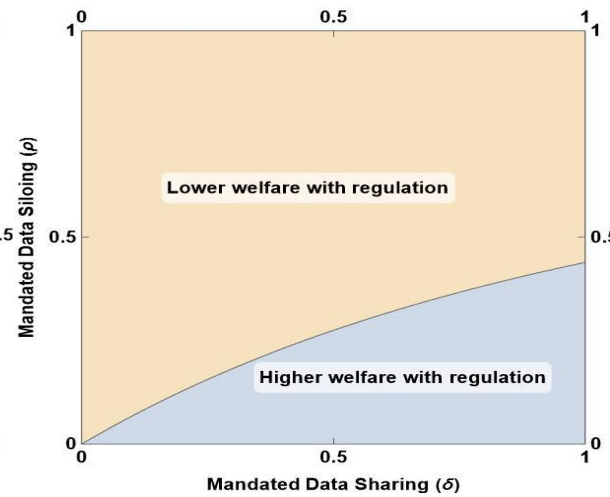
Proposition 4 (Total Welfare): *A stricter data siloing regulation (larger ρ) decreases total welfare, whereas a stricter data sharing regulation (larger δ) increases total welfare.*

Proposition 4 highlights that, although both data siloing and data sharing have the possibility to level the playing field for competition in markets that are connected through data cross-use, a stricter data siloing regime, which limits the extent of the data cross-use, is likely to be harmful overall as it reduces total welfare. In contrast, a stricter data sharing regime unambiguously increases total welfare, everything else being equal. Consequently, with regard to total welfare, data siloing regulation and data sharing regulation negatively interact with each other. Whether total welfare is increased or decreased by some data policy ($\rho > 0, \delta > 0$) relative to the benchmark of no regulation ($\rho = 0, \delta = 0$) depends on parameters. This is visualized by Figure 2b.

Insight 3 (Impact on Welfare): *In comparison to no regulation ($\rho = 0, \delta = 0$), data sharing regulation ($\rho < 1, \delta > 0$) increases consumer welfare or total welfare only if the level of data siloing is low ($\rho < \rho^{CS}$ or $\rho < \rho^{W}$). [See also Figure 2.] A strict data siloing regime ($\rho = 1$) always yields a lower (consumer) welfare than no regulation.*



(a) Consumer Surplus



(b) Total Welfare

Note: Region R1: Higher CS with regulation, R2: Lower CS with regulation, R3: Lower CS with regulation (but higher CS in Market *B*). Figures derived for θ=1/10.

Figure 2. The Impact of Various Degrees of a Data Siloing Regulation (ρ > 0) and a Data Sharing Regulation (δ > 0) in Comparison to No Regulation ($\rho = \delta = 0$)

Taken together, we can immediately derive that the optimal data cross-use regulation, which maximizes (consumer) welfare, does not involve any data siloing, but maximizes data sharing.

Insight 4 (Optimal Data Cross-Use Regulation): *Total welfare and consumer surplus are maximized when no data is siloed and data is fully shared with the rival Platform 2—i.e.,* $\rho = 0$ *and* $\delta = 1$.

Model Robustness and Extensions

The propositions and insights on the impact of data cross-use regulation on innovation and welfare have been derived for a parsimonious base model. As we show next, our insights are robust and continue to hold qualitatively for a range of model extensions.

Alternative Timing

The base model assumes that demand levels are chosen simultaneously in Markets A and B. However, since Market A is considered the primary market, suppose strategic output choices are first made in the primary Market A, and then in the secondary Market B. In Appendix C, we show that our results are robust to this alternative timing. This is because the main strategic trade-offs—that is, the leveraging effect of data from Market A into Market B and its implications on the innovation strategy of Platform 1—remain unaffected by the timing of the game. Thus, we derive the qualitatively same results as in the base model.

Personalized Data Cross-Use

In the base model, data collected for some users in the primary Market A was assumed to shift the demand curve for platform i for $i \in \{1,2\}$ in the secondary Market B, irrespective of whether the same users are also present in the secondary market. Thus, the data gathered in Market A was assumed to bear some general insights about consumer preferences in Market B. Our assumption in the base model is therefore consistent with an *across-user learning* scenario (Hagiu & Wright, 2023). In Appendix D we consider a model variant, where the data collected by Platform 1 in Market A on some consumer is valuable only for the platform in Market B that serves the same consumer on which the data was gathered. This would correspond to a *within-user learning* scenario (Hagiu & Wright, 2023). To model this, we considered a generalized Hotelling-type competition framework (cf. Fudenberg & Tirole, 2000) in which consumers are heterogeneous along two dimensions: (1) their innate preference of the service of Platform 2 over the service of Platform 1 in Market B, denoted by X, with a high X

value denoting a strong preference for Platform 2's service; (2) consumers' privacy preferences, denoted by Y. A high Y value indicates high privacy costs for revealing data when consuming service A. Along both dimensions consumers are independently and identically distributed (i.i.d.) according to a uniform distribution on the unit interval. A consumer of type (X, Y) benefits from data cross-use only if the consumer has used the service in Market A and the consumer's data is available to the platform that services the consumer in Market B. Figure 3 visualizes the ensuing demand system along the two consumer dimensions, X and Y, and illustrates that four demand segments can exist, depending on whether a consumer decides to use the services of Platform 1 or 2 in Market B and whether or not the consumer benefits from data cross-use: Consumers with $Y > Y_A$ do not consume the service of Platform 1 in Market A due to high privacy costs and hence they do not benefit from data cross-use (NS), whereas consumers with $Y \leq Y_A$ use the service of Platform 1 in Market A and could benefit from data cross-use (S). Of the consumers for which data from Market A exists (does not exist), those with $X < X_S$ ($X < X_{NS}$) use the service of Platform 1 in Market B. As shown in the visualization in Figure 3, we found that more consumers tend to choose the service of Platform 1 in Market B whenever data cross-use regulation is not strict ($\rho < 1$, $\delta < 1$), so that $X_{NS} < X_S$. Hence, Platform 1 retains some competitive advantage over those consumers on which it already has data.

In Appendix D we show that the main results of our base model continue to hold. Although only some consumers benefit from data cross-use, this does not qualitatively change the strategic trade-offs as long as there is at least some overlap in the consumer base across the two markets.

Different Market Sizes for the Primary Market and Secondary Market

In the base model, we assume that Markets A and B are of equal size. In Appendix E, we study an extension, where the size of Market A relative to Market B can be varied through parameter $\alpha \in (0,1)$. For a large α, Market A is the larger market, and for a small α, Market B is larger. For markets of equal size ($\alpha = 1/2$) we obtained the same results as in the base model. We found that when Market A is small, and thus less relevant for Platform 1's strategy and consumer welfare, Platform 1 has less incentive to innovate in Market A, which also lowers demand and the data available for cross-use. The only qualitative change to our previous results is that the intermediate region in Figure 2a—where data cross-use regulation yields an overall lower consumer surplus than no regulation but yet improves consumer surplus in Market B—gradually disappears as α becomes smaller. This is because for a small α, since Market A does not contribute as much to consumer welfare, the consumer surplus outcome of Market B dominates the overall consumer surplus outcome. In contrast, if Market A is large (α is large), it becomes more likely that data cross-use regulation will negatively affect consumer surplus, as the negative impact on Market A dominates (see Figure E1 in the Appendix).

Model with General Functional Form for Data Cross-Use Benefit

In the base model we derived our insights based on a specific functional form, Ψ_i, i.e., how data cross-use impacts demand. In Appendix F we show that our results hold when considering more general functional forms of Ψ_i.

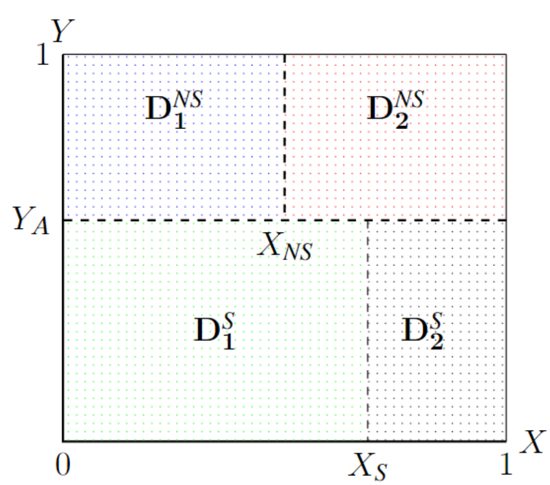


Figure 3. Consumer Demand Under Personalized Data Cross-Use

Conclusion and Policy Implications

The cross-use of data for different services is prevalent within digital platform ecosystems and allows dominant platforms to leverage insights from their primary markets to gain a competitive edge in secondary markets. Although data cross-use yields efficiencies and quality improvements for data-driven services, policymakers around the world are concerned with the economic power and distortion of competition and innovation incentives that may accompany data cross-use and market leverage. It has been suggested that data siloing (i.e., disallowing the cross-use of data within the platform ecosystem) and mandated data sharing with rivals are distinct and targeted policy interventions to restore competition and innovation in data-driven digital markets (e.g., Argenton & Prüfer, 2012; Krämer & Schnurr, 2022; Prüfer & Schottmüller, 2021; Parker et al., 2021). Both policies already apply to digital gatekeepers under the European Union's Digital Markets Act (DMA)—the world's first comprehensive regulation for digital platforms.

We present an analytical model on data cross-use in digital platform ecosystems, which allowed us to analyze the impact of data siloing and data sharing obligations, as well as their interaction, on competition, innovation and, ultimately, consumer surplus and total welfare. Based on our results, we draw several important insights for the regulation of data cross-use in digital platform ecosystems that can inform current and future regulation of digital markets.

First, under a robust set of assumptions, we show that the optimal policy is to mandate as much data sharing as possible, but not to mandate data siloing (Propositions 2 to 4, Insight 4). Although both policies level the playing field with respect to competition, they have vastly different impacts on innovation (Proposition 1, Insights 1 and 2) and consumer welfare (Propositions 2 to 4, Insight 3). Our analysis shows that a strict data siloing obligation is a problematic remedy that tends to reduce (consumer) welfare and leads to the least innovation in the primary market—i.e., even below the level without regulation (Insight 3). Arguably, data siloing is also easier to implement and to administer than a data sharing obligation, which would make it a seemingly attractive remedy for policymakers. However, our results suggest that the ease of implementation should not be a guiding principle for policymakers; rather, a hands-off approach with respect to data siloing may even be better after all. If policymakers want to nevertheless impose data siloing obligations, e.g., due to political pressure, a more lenient data siloing provision is preferable. This could be achieved, for example, by allowing data cross-use by default unless consumers actively opt out. However, under the EU's Digital Markets Act, data siloing is the default and consumers must actively opt in to data cross-use, representing a rather strict data siloing regime.

Further, although optimal in theory, mandated data sharing can never be fully achieved in practice due to technical and legal limitations. Digital platforms collect consumer data at a fine granular and individual level, such as individual clicks and search queries. In order to establish a level playing field, this data would need to be shared instantaneously with rivals, and in full detail. This is practically infeasible in most applications due to the sheer amount of data that would need to be shared (and the potential for many recipients) and due to privacy regulation. However, our results suggest that this should not discourage policymakers from considering mandated data sharing and fostering research on how data can be efficiently shared in a privacy-preserving manner. For example, technological measures (such as federated learning and differential privacy) and institutional measures (such as third-party data trusts, or in-situ rights to run one's own algorithms on the regulated platform's data infrastructure) could be taken to overcome some of the privacy concerns (for a more detailed discussion see Krämer & Schnurr, 2022; Parker et al., 2021).

Second, although more mandated data sharing is optimal from a consumer welfare and total welfare perspective when considering the entire ecosystem, this remedy is not without its trade-offs. While mandated data sharing increases the level of competition in the secondary market and stimulates innovation incentives by the rival platform (Proposition 1), which leads to an increase in welfare, mandated data sharing lowers innovation incentives by the regulated platform, both in the primary (regulated) market (Proposition 1, Insight 1) and in the secondary market (Insight 2). Depending on the size and importance of the primary market, this trade-off and its implications for aggregate (consumer) welfare could play out differently in practice. Typically, the primary market is large and of societal importance, as it otherwise would not be subject to regulation. In such cases, the negative effects of data sharing regulation become more pronounced due to a relatively strong reduction in innovation incentives in the primary market and the ensuing effects on data generation, which is the foundation for data cross-use and the associated efficiencies. Thus, when imposing mandated data sharing, policymakers need to carefully monitor the effects in the primary and secondary markets separately, noting that opposing trends are to be expected. If the negative effects in the primary markets are particularly strong, the data sharing policy may need to be reevaluated (e.g., by limiting the scale of scope of data to be shared), as policymakers may otherwise run the risk of killing the goose that lays the golden eggs.

Third, our results highlight that data sharing regulation and data siloing regulation interact: They cannot be considered in isolation, and one is not a substitute for the other. With more data siloing, less data becomes available for sharing. While data sharing preserves and shares the efficiencies generated through data cross-use, data siloing limits such data efficiencies. We found that in comparison to no regulation, data sharing regulation increases (consumer) welfare only if the level of data siloing is low (Insight 3). Further, more data siloing can increase the innovation effort of the rival, but only if the level of data sharing is low (Proposition 1). If the level of data sharing is high, an increase in data siloing has the opposite effect. Thus, policymakers will need to consider both types of data cross-use regulation jointly in order to account for this interaction. It seems that policymakers are currently willing to impose much stricter data siloing provisions than data sharing provisions. For example, the Digital Markets Act (European Commission, 2020) foresees data sharing only for regulated search engines, whereas data siloing is imposed on all regulated platforms. In fact, our results suggest that the opposite would be more favorable for increasing consumer welfare, i.e., imposing a less strict data siloing regime in tandem with more comprehensive data sharing obligations.

Limitations and Future Research

Finally, we point to some limitations and avenues for future research. Like any theoretical model, our model can be criticized based on its assumptions and the focus of the analysis. Indeed, our model could be extended in a number of ways. First, future research could incorporate data-driven network effects (Gregory et al., 2021) not only across markets but also within each market. This would amplify the benefits of attaining a larger market share and thus lead to a higher level of innovation in the primary market as well as a stronger degree of competition in the secondary market. However, this would not change our insights qualitatively. With mandated data siloing or data sharing, the incentives to innovate in the primary market would still be lower, and the degree of competition in the secondary market would still be higher than without these regulatory interventions.

Second, future research could also allow for competition in the primary market with Platform 1 being the dominant intermarket player. This would also not change our insights qualitatively because it would not fundamentally change the trade-offs considered. However, the degree of competition in the primary market would moderate the impact of the regulation on innovation incentives in the primary market. The stronger the competition in the primary market, the less Platform 1 would reduce its level of innovation following a regulation. In consequence, the negative impact of data sharing or data siloing on welfare in the primary market would be dampened. However, we note that economic regulation can only be imposed if a platform possesses significant market power, which we proxy here through a monopoly position in the primary market. Thus, in practice, a scenario in which Platform 1 would face significant competition in its primary market would most certainly preclude the possibility of imposing data sharing or data siloing regulation in the first place.[16]

Third, future research could also seek to develop a more dynamic framework to study the successive entry of a dominant platform in more and more related markets. Our static, two-stage two-market model could provide a useful starting point for this.

Acknowledgments

This research won the Kauffmann Best Paper Award at the International Conference on Information Systems in 2021. Jan Krämer acknowledges funding by Deutsche Forschungsgemeinschaft (Project 409474477).

References

Abrahamson, Z. (2014). Essential data. *Yale Law Journal, 124*, 867-881.

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306-333. https://doi.org/10.1002/smj.821

Argenton, C., & Prüfer, J. (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics*, 8(1), 73-105. https://doi.org/10.1093/joclec/nhr018

Argote, L. (2013). *Organizational learning: Creating, retaining and transferring knowledge*. Springer Science & Business Media.

Barachini, F. (2009). Cultural and social issues for knowledge sharing. *Journal of Knowledge Management*, *13*(1), 98-110. https://doi.org/10.1108/13673270910931198

Baughn, C. C., Denekamp, J. G., Stevens, J. H., & Osborn, R. N. (1997). Protecting intellectual capital in international alliances. *Journal of World Business*, *32*(2), 103-117. https://doi.org/10.1016/S1090-9516(97)90002-X

Brusoni, S., Prencipe, A., & Pavitt, K. (2001). Knowledge specialization, organizational coupling, and the boundaries of the firm: why do firms know more than they make? *Administrative Science Quarterly*, *46*(4), 597-621. https://doi.org/10.2307/3094825

Cabral, L. M., & Riordan, M. H. (1994). The learning curve, market dominance, and predatory pricing. *Econometrica: Journal of the Econometric Society*, 1115-1140. https://doi.org/10.2307/2951509

Carroni, E., Madio, L., & Shekhar, S. (2024). Superstar exclusivity in two-sided markets. *Management Science*, *70*(2), 991-1011. https://doi.org/10.1287/mnsc.2023.4720

Cennamo, C., Kretschmer, T., Constantinides, P., Alaimo, C., & Santaló, J. (2023). Digital platforms regulation: An innovation-

[16] For example, under the current Digital Markets Act, platforms need to serve as a "gatekeeper" for a given "core platform service" (e.g., online search) and must enjoy an "entrenched and durable position" in order to be subjected to any regulation (see Article 3 of the DMA).

centric view of the EU's Digital Markets Act. *Journal of European Competition Law & Practice*, *14*(1), 44-51. https://doi.org/10.1093/jeclap/lpac043

CMA. (2020). *Online platforms and digital advertising final report. Appendix S: The relationship between large digital platforms and publishers.* Available at https://assets.publishing.service.gov.uk/media/5efb22fbd3bf7f768fdcdfae/Appendix_S_-_the_relationship_between_large_digital_platforms_and_publishers.pdf.

Easley, R. F., Guo, H., & Krämer, J. (2018). Research commentary—From net neutrality to data neutrality: A techno-economic framework and research agenda. *Information Systems Research*, 29(2), 253-272. https://doi.org/10.1287/isre.2017.0740

European Commission. (2014). *Commission Decision of 03/10/2014 declaring a concentration to be compatible with the common market (Case No COMP/M.7217—FACEBOOK/WHATSAPP) according to Council Regulation (EC) No 139/2004.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014M7217

European Commission. (2017). *Mergers: Commission fines Facebook 110 million euros for providing misleading information about WhatsApp takeover*. https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1369.

European Commission. (2020). *Digital Markets Act: Ensuring fair and open digital market.* https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

Fudenberg, D., & Tirole, J. (2000). Customer poaching and brand switching. *RAND Journal of Economics*, *31*(4), 634-57. https://doi.org/10.2307/2696352

Furman, J., Coyle, D., Fletcher, A., McAuley, D., & Marsden, P. (2019). *Unlocking digital competition: Report of the digital competition expert panel.* HM Treasury. https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239-1249. https://doi.org/10.1016/j.respol.2014.03.006

Gilbert, R. J. (2023). Antitrust reform: An economic perspective. *Annual Review of Economics*, *15*(1), 151-175. https://doi.org/10.1146/annurev-economics- 082222-070822

Goić, M., Jerath, K., & Srinivasan, K. (2011). Cross-market discounts. *Marketing Science*, *30*(1), 134-148. https://doi.org/10.1287/mksc.1100.0603

Graef, I., Wahyuningtyas, S. Y., & Valcke, P. (2015). Assessing data access issues in online platforms. *Telecommunications Policy*, *39*(5), 375-387. https://doi.org/10.1016/j.telpol.2014.12.001

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. https://doi.org/10.5465/amr.2019.0178

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. https://doi.org/10.5465/amr.2021.0111

Gupta, S., & Polonsky, M. (2014). Inter-firm learning and knowledge-sharing in multinational networks: An outsourced organization's perspective. *Journal of Business Research*, *67*(4), 615-622. https://doi.org/10.1016/j.jbusres.2013.02.043

Haftor, D. M., Climent, R. C., & Lundström, J. E. (2021). How machine learning activates data network effects in business models: Theory advancement through an industrial case of promoting ecological sustainability. *Journal of Business Research*, *131*, 196-205. https://doi.org/10.1016/j.jbusres.2021.04.015

Hagiu, A., & Wright, J. (2023). Data-enabled learning, network effects, and competitive advantage. *The RAND Journal of Economics*, *54*(4), 638-667. https://doi.org/10.1111/1756-2171.12453

Han, K., Oh, W., Im, K. S., Chang, R. M., Oh, H., & Pinsonneault, A. (2012). Value cocreation and wealth spillover in open innovation alliances. *MIS Quarterly*, *36*(1), 291-315. https://doi.org/10.2307/41410418

Henten, A., & Windekilde, I. (2022). Demand-side economies of scope in big tech business modelling and strategy. *Systems*, *10*(6), 246. https://doi.org/10.3390/systems10060246.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*(153), 41-57. https://doi.org/10.2307/2224214

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review*, *75*(3), 424-440. Available at: https://www.jstor.org/stable/1814809

Klein, T. J., Kurmangaliyeva, M., Prüfer, J., & Prüfer, P. (2022). *How important are user-generated data for search result quality? Experimental evidence* (TILEC discussion paper). SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4229292

Krämer, J. (2021). Personal data portability in the platform economy: economic implications and policy recommendations. *Journal of Competition Law & Economics*, *17*(2), 263-308. https://doi.org/10.1093/joclec/nhaa030

Krämer, J., & Schnurr, D. (2022). Big data and digital markets contestability: Theory of harm and data access remedies. *Journal of Competition Law & Economics*, *18*(2), 255-322. https://doi.org/10.1093/joclec/nhab015

Krämer, J., & Stüdlein, N. (2019). Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the General Data Protection Regulation. *Economics Letters*, *181*, 99-103. https://doi.org/10.1016/j.econlet.2019.05.015

Kwark, Y., Chen, J., & Raghunathan, S. (2017). Platform or wholesale? A strategic tool for online retailers to benefit from third-party information. *MIS Quarterly*, *41*(3), 763-A17. https://doi.org/10.25300/MISQ/2017/41.3.05

Lam, W. M. W., & Liu, X. (2020). Does data portability facilitate entry? *International Journal of Industrial Organization*, *69*, Article 102564. https://doi.org/10.1016/j.ijindorg.2019.102564

Lam, W. M. W., & Liu, X. (2023). Dancing with rivals: How does platform's information usage benefit independent sellers? *European Journal of Operational Research*, *309*(1), 421-431. https://doi.org/10.1016/j.ejor.2022.12.026

Landy, H. (2020). Here's how Google, Amazon, Facebook, and Apple are making a play for health insurance market. Fierce Healthcare. https://www.fiercehealthcare.com/tech/here-s-how-google-amazon-facebook-and-apple-are-making-a-play-for-health-insurance-market

Levitt, B., & March, J. (1988). Organizational learning. *Annual Review of Sociology*, *14*(1), 319-338. https://doi.org/10.1146/annurev.so.14.080188.001535

Loveless, J. (2022). How does Apple's app tracking transparency framework affect advertisers? *Forbes* https://www.forbes.com/sites/forbesbusinesscouncil/2022/08/22/how-does-apples-app-tracking-transparency-framework-affect-advertisers/?sh=1ebac84177c6

Menon, N. M. (2018). Information spillover and semi-collaborative networks in insurer fraud detection. *MIS Quarterly*, *42*(2), 407-426. https://doi.org/10.25300/MISQ/2018/14433

Panzar, J. C., & Willig, R. D. (1981). Economies of scope. *The American Economic Review*, *71*(2), 268-272. http://www.jstor.org/stable/1815729.

Parker, G., Petropoulos, G., & Van Alstyne, M. (2021). Platform mergers and antitrust. *Industrial and Corporate Change*, *30*(5), 1307-1336. https://doi.org/10.1093/icc/dtab048

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, *51*(10), 1494-1504. https://doi.org/10.1287/mnsc.1050.0400

Prasad, A., Venkatesh, R., & Mahajan, V. (2015). Product bundling or reserved product pricing? Price discrimination with myopic and strategic consumers. *International Journal of Research in Marketing*, *32*(1), 1-8. https://doi.org/10.1016/j.ijresmar.2014.06.004

Prüfer, J., & Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, *69*(4), 967-1008. https://doi.org/10.1111/joie.12259

Ritala, P., Olander, H., Michailova, S., & Husted, K. (2015). Knowledge sharing, knowledge leaking and relative innovation performance: An empirical study. *Technovation*, *35*, 22-31. https://doi.org/10.1016/j.technovation.2014.07.011

Rogers, E. M. (1995). *Diffusion of innovation*. Free Press.

Rubinfeld, D. L., & Gal, M. S. (2017). Access barriers to big data. *Arizona Law Review* 59, 339-382. https://arizonalawreview.org/access-barriers-to-big-data/

Schaefer, M., & Sapi, G. (2023). Complementarities in learning from data: Insights from general search. *Information Economics and Policy*, *65*, Article 101063. https://doi.org/10.1016/j.infoecopol.2023.101063

Small, C. T., & Sage, A. P. (2006). Knowledge management and knowledge sharing: A review. *Information Knowledge Systems Management*, *5*(3), 153-169. Available at https://www.jstor.org/stable/724882

Sun, T., Yuan, Z., Li, C., Zhang, K., & Xu, J. (2024). The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. *Management Science*, 70(4), 2645-2660. https://doi.org/10.1287/mnsc.2023.4828

Stigler, G. J. (1958). The economies of scale. *The Journal of Law and Economics*, *1*, 54-71.

Teece, D. J. (1980). Economies of scope and the scope of the enterprise. *Journal of Economic Behavior & Organization*, *1*(3), 223-247. https://doi.org/10.1016/0167-2681(80)90002-5

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. https://doi.org/10.1007/s11151-019-09693-7

Vengattil, M., & Humer, C. (2020). *Alphabet's Verily targets employer health insurance with Swiss Re partnership*. Reuters. https://www.reuters.com/article/us-alphabet-verily-idUSKBN25L1Q9.

Wilson, R. (1975). Informational economies of scale. *The Bell Journal of Economics*, *6*(1), 184-195. https://doi.org/10.2307/3003221

Wohlfarth, M. (2019). Data portability on the internet: an economic analysis. *Business & Information Systems Engineering*, *61*, 551-574. https://doi.org/10.1007/s12599-019-00580-9

Yan, K., Hua, G., Cheng, T. C. E., Wang, S., & Dong, J. X. (2022). Joint promotion of cross-market retailers: models and analysis. *International Journal of Production Research*, *60*(11), 3397-3418. https://doi.org/10.1080/00207543.2021.1922946

Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, *33*(1), 88-106. https://doi.org/10.1002/smj.941

About the Authors

Jan Krämer is a is a professor of information systems at the University of Passau, Germany, where he holds the Chair for Internet & Telecommunications Business. He is also the director of the Research Training Group 2720 on Digital Platform Ecosystems at the University of Passau, which is funded by the German Research Foundation, and academic co-director at the Centre of Regulation in Europe (CERRE), a Brussels-based academic thinktank. He has a *Diplom* degree in business engineering and management, and a Ph.D. in economics, both from the Karlsruhe Institute of Technology; he also engaged in longer research visits and study periods at the Massachusetts Institute of Technology and Boston University. His research interests include the economic regulation of telecommunications and internet markets, as well as digital ecosystems and data-driven business models. His work has been published in journals like *Management Science*, *Information Systems Research*, and *Journal of Industrial Economics*. He frequently serves as an independent academic expert for governmental institutions such as the European Commission. https://orcid.org/0000-0002-5866-7065

Shiva Shekhar is an associate professor of information systems in the Information Systems and Operations Management (ISOM) Department at the Tilburg School of Economics and Management (TiSEM) at Tilburg University. He has a doctorate in economics from the Dusseldorf Institute for Competition Economics (DICE), Duesseldorf, Germany. His research interests include the economics of information systems, platform design, platform regulation, and privacy-related topics. He publishes in both economics and management journals. He was previously a consultant for Compass Lexecon and now externally consults for them. https://orcid.org/0000-0002-6225-5821

Appendix A

Proofs

Proof of Lemma 1: Differentiating the profit of Platform 1 with respect to q_A and q_1 and the profit of Platform 2 with respect to q_2, yields the following first order conditions

$$\frac{\partial \Pi_1}{\partial q_A} = P_A + q_A \frac{\partial P_A}{\partial q_A} = 0, \quad \frac{\partial \Pi_i}{\partial q_i} = P_i + q_i \frac{\partial P_i}{\partial q_i} = 0, \text{ for } i \in \{1,2\}.$$

Solving the above first order conditions yields the equilibrium demand in Market A and the demand in Market B as $q_A(v_A) = \frac{1+v_A}{6}$, $q_1(q_A^e, v_1, v_2) = \frac{1+2v_1-v_2+\theta(2-\delta)(1-\rho)q_A^e}{9}$, $q_2(q_A^e, v_2, v_1) = \frac{1+2v_2-v_1-\theta(1-\rho)(1-2\delta)q_A^e}{9}$. In equilibrium, the consumers' anticipated value needs to be correct. Thus, we solve the demands while imposing $q_A = q_A^e$ where q_A^e is the anticipated demand in Market A. Substituting this in the demand functions yields:

$$\begin{aligned} q_1(v_1, v_2, v_A) &= \frac{6(1+2v_1-v_2)+\theta(2-\delta)(1-\rho)(1+v_A)}{54}, \\ q_2(v_2, v_1, v_A) &= \frac{6(1+2v_2-v_1)-\theta(1-2\delta)(1-\rho)(1+v_A)}{54}, \end{aligned}$$

with $q_1 + q_2 = \frac{12+6(v_2+v_1)+\theta(1+\delta)(1-\rho)(1+v_A)}{54}$. From the above, it follows that total output and q_1 rise in v_A and θ. Differentiating q_2 with respect to v_A and with respect to θ yields $\frac{\partial q_2}{\partial v_A} = -\frac{\theta(1-2\delta)(1-\rho)}{54}$ and $\frac{\partial q_2}{\partial \theta} = -\frac{(1-2\delta)(1-\rho)(1+v_A)}{54}$. The sign of the above two expressions is positive for $\delta \geq 1/2$ and negative otherwise. Similarly, it also follows that as δ increases q_1 falls while q_2 increases. Total demand in Market B rises with an increase in δ. Likewise, it follows that as ρ increases q_1 and total market demand falls. Differentiating q_2 with respect to ρ yields $\frac{\partial q_2}{\partial \rho} = \frac{\theta(1-2\delta)(1+v_A)}{54}$. This comparative static is positive for $\delta < 1/2$ and negative otherwise.

Proof of Proposition 1: Differentiating the profit of Platform 1 and Platform 2 with respect to v_1 and v_2, respectively and employing the envelope theorem yields the following first order conditions:

$$\frac{\partial \Pi_i}{\partial v_i} = \left(\underbrace{\frac{\partial P_i}{\partial v_i}}_{\text{Direct effect } (+)} + \underbrace{\frac{\partial P_i}{\partial q_{-i}} \frac{\partial q_{-i}}{\partial v_i}}_{\text{Indirect effect } (+)} \right) q_i - \frac{\partial I(v_i)}{\partial v_i} = 0 \text{ for } i \in \{1,2\}. \tag{6}$$

Solving the above system of equations yields the platforms' innovation efforts in Market B as a function of the innovation effort in Market A as follows: $v_1^{BR}(v_A) = \frac{2(30+\theta(1-\rho)(1+v_A)(14-9\delta))}{345}$ and $v_2^{BR}(v_A) = \frac{2(30-\theta(1-\rho)(1+v_A)(9-14\delta))}{345}$. Differentiating $v_1^{BR}(v_A)$ with respect to v_A, it is straightforward to observe that as v_A increases, Platform 1 sets higher innovation levels also in Market B—i.e., $\frac{\partial v_1^{BR}(v_A)}{\partial v_A} = \frac{2(\theta(1-\rho)(14-9\delta))}{345} > 0$. This confirms that as v_A increases Platform 1 increases investments in Market B. Similarly, differentiating $v_2^{BR}(v_A)$ with respect to v_A yields $\frac{\partial v_2^{BR}(v_A)}{\partial v_A} = -\frac{2\theta(1-\rho)(9-14\delta)}{345}$. This confirms that as v_A increases Platform 2 increases investments in Market B if and only if $\delta > 9/14$.

Solving Equations (2) and (6) simultaneously, yields the following innovation levels in Market A and Market B as $v_A^\star = \frac{\left(345+(2-\delta)\theta(1-\rho)(30+\theta(14-9\delta)(1-\rho))\right)}{(1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)}$, $v_1^\star = \frac{12(25+\theta(14-9\delta)(1-\rho))}{(1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)}$, and

$$v_2^\star = v_1^\star - \frac{\left(4(1-\delta)(1-\rho)\theta(69+\theta(1-\rho)(2-\delta))\right)}{(1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)}.$$

Differentiating the innovation level by Platform 1 in Market A with respect to δ yields:

$$\frac{\partial v_A^\star}{\partial \delta} = -\frac{90\theta(1-\rho)\overbrace{(575 + 46\theta(1-\rho)(16-9\delta) + 3(1-\rho)^2\theta^2(2-\delta)^2)}^{\mathcal{A}_1}}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)^2} < 0.$$

The sign of the above expression depends on the sign of expression denoted by $\mathcal{A}_1 = (575 + 46\theta(1-\rho)(16-9\delta) + 3(1-\rho)^2\theta^2(2-\delta)^2)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_1}{\partial \delta} = -6\theta(1-\rho)\big(69 + \theta(2-\delta)(1-\rho)\big) < 0$. Further, we note that the value of $\mathcal{A}_1$ at $\delta = 1$ given by $\mathcal{A}_1|_{\delta=1} = 575 + \theta(1-\rho)\big(322 + 3\theta(1-\rho)\big) > 0$. Thus, we can state that $\mathcal{A}_1$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial v_A^\star}{\partial \delta} < 0$.

Differentiating the innovation level by Platform 1 in Market B with respect to δ yields:

$$\frac{\partial v_1^\star}{\partial \delta} = -\frac{12\theta(1-\rho)\overbrace{(15525 + 50\theta(1-\rho)(16-9\delta) + (1-\rho)^2\theta^2(14-9\delta)^2)}^{\mathcal{A}_2}}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^2} < 0.$$

The sign of the above expression depends on the sign of expression denoted by $\mathcal{A}_2 = (15525 + 50\theta(1-\rho)(16-9\delta) + (1-\rho)^2\theta^2(14-9\delta)^2)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_2}{\partial \delta} = -18\theta(1-\rho)\big(25 + \theta(1-\rho)(14-9\delta)\big) < 0$. Further, we note that the value of $\mathcal{A}_2$ at $\delta = 1$ given by $\mathcal{A}_2|_{\delta=1} = 25\left(621 + \theta(1-\rho)\big(14 + \theta(1-\rho)\big)\right) > 0$. Thus, we can state that $\mathcal{A}_2$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial v_1^\star}{\partial \delta} < 0$.
Differentiating the innovation level by Platform 2 in B with respect to δ yields:

$$\frac{\partial v_2^\star}{\partial \delta} = \frac{4\theta(1-\rho)\overbrace{(72450 + 75\theta(1-\rho)(37-28\delta) - 6\theta^2(1-\rho)^2\big(52 + 9\delta(9-7\delta)\big) - 5\theta^3(2-\delta)^2(1-\rho)^3)}^{\mathcal{A}_3}}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^2} > 0$$

The sign of the above expression depends on the sign of expression denoted by $\mathcal{A}_3 = \big(72450 + 75\theta(1-\rho)(37-28\delta) - 6\theta^2(1-\rho)^2\big(52 + 9\delta(9-7\delta)\big) - 5\theta^3(2-\delta)^2(1-\rho)^3\big)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_3}{\partial \delta} = -2\theta(1-\rho)\big(1050 + 27\theta(1-\rho)(9-14\delta) - 5\theta^2(1-\rho)^2(2-\delta)\big) < 0$. Further, we note that the value of $\mathcal{A}_3$ at $\delta = 1$ given by $\mathcal{A}_3|_{\delta=1} = 72450 + 5\theta(1-\rho)\left(135 + \theta\big(84 + \theta(1-\rho)\big)\right) > 0$. Thus, we can state that $\mathcal{A}_3$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial v_2^\star}{\partial \delta} > 0$.

Differentiating the innovation level by Platform 1 in Market A with respect to ρ yields:

$$\frac{\partial v_A^\star}{\partial \rho} = -\frac{30\theta(2-\delta)\overbrace{\Big(1725 + \theta(1-\rho)(14-9\delta)\big(138 + \theta(2-\delta)(1-\rho)\big)\Big)}^{\mathcal{A}_4}}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^2} < 0.$$

The sign of the above expression depends on the sign of expression denoted by $\mathcal{A}_4 = \Big(1725 + \theta(1-\rho)(14-9\delta)\big(138 + \theta(2-\delta)(1-\rho)\big)\Big)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_4}{\partial \delta} = -2\theta(1-\rho)\big(621 + \theta(16-9\delta)(1-\rho)\big) < 0$. Further, we note that the value of $\mathcal{A}_4$ at $\delta = 1$ given by $\mathcal{A}_4|_{\delta=1} = 1725 + 5\theta(1-\rho)\big(138 + \theta(1-\rho)\big) > 0$. Thus, we can state that $\mathcal{A}_4$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial v_A^\star}{\partial \rho} < 0$.

Differentiating the innovation level by Platform 1 in Market B with respect to ρ yields:

$$\frac{\partial v_1^\star}{\partial \rho} = -\frac{12\theta(14-9\delta)\overbrace{\Big(1725 + \theta(2-\delta)(1-\rho)\big(50 + \theta(14-9\delta)(1-\rho)\big)\Big)}^{\mathcal{A}_5}}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^2} < 0.$$

The sign of the above expression depends on the sign of expression denoted by $\mathcal{A}_5 = \left(1725 + \theta(2-\delta)(1-\rho)\big(50 + \theta(14-9\delta)(1-\rho)\big)\right)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_5}{\partial \delta} = -2\theta(1-\rho)\big(25 + \theta(1-\rho)(16-9\delta)\big) < 0$. Further, we note that the value of $\mathcal{A}_5$ at $\delta = 1$ given by $\mathcal{A}_5|_{\delta=1} = 1725 + 5\theta(1-\rho)\big(10 + \theta(1-\rho)\big) > 0$. Thus, we can state that $\mathcal{A}_5$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial v_1^\star}{\partial \rho} < 0$.

Differentiating the innovation level by Platform 2 in B with respect to ρ yields:

$$\frac{\partial v_2^\star}{\partial \rho} = \frac{12\theta(9-14\delta)\overbrace{\big(1725 + (2-\delta)\theta(1-\rho)\big(50 + \theta(14-9\delta)(1-\rho)\big)\big)}^{\mathcal{A}_5}}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^2}.$$

Recall that $\mathcal{A}_5$ has been established as being positive. Therefore, the sign of the above comparative static expression depends on the sign of $(9-14\delta)$, which is negative for $\delta > 9/14$ and positive otherwise. Thus, it follows that $\frac{\partial v_2^\star}{\partial \rho}$ is positive when $\delta < 9/14$ and negative otherwise.

Proof of Proposition 2: The consumer surplus in Market A is $CS_A = \int_{r_A^\star}^1 \frac{(r + v_A^\star - P_A^\star)}{3} dr = \frac{3(q_A^\star)^2}{2}$, where $P_A^\star = 1 + v_A^\star - 3q_A^\star$, $r_A^\star \triangleq P_A^\star - v_A^\star$ and $q_A^\star = \frac{5\big(69 + \theta(2-\delta)(1-\rho)\big)}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)}$. Similarly, the consumer surplus in Market B can be obtained as $CS_B = \int_{\Phi^\star}^1 \frac{(r + v_1^\star + \Psi_1^\star - P_1^\star)}{3} dr = \frac{3(q_1^\star + q_2^\star)^2}{2}$, where $\Psi_1^\star = \theta(1-\rho)q_A^\star$, $\Psi_2^\star = \theta(1-\rho)(1-\delta)q_A^\star$, $P_i^\star = 1 + v_i^\star + \Psi_i^\star - 3(q_1^\star + q_2^\star)$, $\Phi^\star \triangleq P_1^\star - v_1^\star - \Psi_1^\star$ and $q_1^\star + q_2^\star = \frac{450 + 45(1+\delta)\theta(1-\rho) - 3(2-\delta)(1-\delta)\theta^2(1-\rho)^2}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)}$. Notice that a sufficient statistic for the consumer surplus is the total demand in each market. Differentiating with respect to ρ yields:

$$\frac{\partial q_A^\star}{\partial \rho} = -\frac{5(2-\delta)\theta\overbrace{\big(1725 + \theta(1-\rho)(14-9\delta)\big(138 + \theta(2-\delta)(1-\rho)\big)\big)}^{\mathcal{A}_4}}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)^2} < 0$$

$$\frac{\partial(q_1^\star + q_2^\star)}{\partial \rho} = -\frac{45(1+\delta)\theta\overbrace{\big(1725 + \theta(1-\rho)(2-\delta)\big(50 + \theta(1-\rho)(14-9\delta)\big)\big)}^{\mathcal{A}_5}}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)^2} < 0.$$

The above two comparative statics are negative because the expressions denoted by $\mathcal{A}_4$ and $\mathcal{A}_5$ have been shown as being positive in the relevant range. From the above, it follows that the total consumer surplus in both Markets A and B falls with an increase in ρ. Differentiating the total demand in each market with respect to δ yields

$$\frac{\partial q_A^\star}{\partial \delta} = -\frac{15(1-\rho)\theta\overbrace{\big(575 + 46\theta(1-\rho)(16-9\delta) + 3(1-\rho)^2\theta^2(2-\delta)^2\big)}^{\mathcal{A}_1}}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)^2} < 0,$$

$$\frac{\partial(q_1^\star + q_2^\star)}{\partial \delta} = \frac{15(1-\rho)\theta\overbrace{\big(5175 + 75\theta(1-\rho)(1-2\delta) - 9\theta^2(1-\rho)^2\big(20 - 3\delta(2+\delta)\big) - \theta^3(1-\rho)^3(2-\delta)^2\big)}^{\mathcal{A}_6}}{(1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)^2} > 0.$$

The sign of $\frac{\partial q_A^\star}{\partial \delta}$ is always negative as we have previously established that $\mathcal{A}_1$ is positive. The sign of $\frac{\partial(q_1^\star + q_2^\star)}{\partial \delta}$ depends on the sign of expression denoted by $\mathcal{A}_6 = \big(5175 + 75\theta(1-\rho)(1-2\delta) - 9\theta^2(1-\rho)^2\big(20 - 3\delta(2+\delta)\big) - \theta^3(1-\rho)^3(2-\delta)^2\big)$ in the numerator. Differentiating this expression with respect to δ yields $\frac{\partial \mathcal{A}_6}{\partial \delta} = -\theta(1-\rho)(150 - 54(1+\delta)\theta(1-\rho) - 2\theta^2(2-\delta)(1-\rho)^2) < 0$. Further, we note that the value of $\mathcal{A}_6$ at $\delta = 1$ given by $\mathcal{A}_6|_{\delta=1} = 5175 - \theta(1-\rho)\left(75 - \theta(1-\rho)\big(99 + \theta(1-\rho)\big)\right) > 0$. Thus, we can state that $\mathcal{A}_6$ is positive in our relevant parameter region. Using this observation, we confirm that $\frac{\partial(q_1^\star + q_2^\star)}{\partial \delta} > 0$. Differentiating the total consumer surplus across markets yields:

$$\frac{\partial(CS_A^\star + CS_B^\star)}{\partial \delta} = \frac{\kappa}{\big((1725 - (2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\big)^3} > 0,$$

where $\kappa \triangleq 45\theta(\rho-1)(3(\delta-2)^3(\delta-1)\theta^5(\rho-1)^5+9(\delta-2)(\delta(\delta(9\delta+14)-83)+50)\theta^4(\rho-1)^4+15(\delta(2\delta(56\delta+51)-141)-698)\theta^3(\rho-1)^3+5(2\delta(1323\delta-7946)+24035)\theta^2(\rho-1)^2+5(62216\delta+1391)\theta(\rho-1)-2130375)>0$. In the following, we show that $\kappa>0$. Differentiating κ with respect to δ four times, we observe that $\frac{\partial^4\kappa}{\partial\delta^4}=-3240\theta^5\big(27-\theta(1-\rho)\big)(1-\rho)^5<0$. Next differentiating κ with respect to δ thrice and considering its value at $\delta=1$, we note that $\frac{\partial^3\kappa}{\partial\delta^3}|_{\delta=1}=810\theta^4\Big(560-3\theta(1-\rho)\big(32+\theta(1-\rho)\big)\Big)(1-\rho)^4>0$. Next, differentiating κ with respect to δ twice and considering its value at $\delta=1$, we note that $\frac{\partial^2\kappa}{\partial\delta^2}|_{\delta=1}=-810\theta^3\Big(1470-\theta(1-\rho)\Big(730+\theta(1-\rho)\big(69+\theta(1-\rho)\big)\Big)\Big)(1-\rho)^3<0$. The above inequality follows from $0\le\rho\le1$, $0\le\delta\le1$ and $0<\theta<1$. This implies that the slope of the first derivative of κ with respect to δ is always negative. Differentiating κ with respect to δ and observing its sign at $\delta=1$ yields $\frac{\partial\kappa}{\partial\delta}|_{\delta=1}=45\theta^2(1-\rho)^2\bigg(311080+\theta(1-\rho)\bigg(53000-3\theta\Big(1995+\theta\big(54+\theta(1-\rho)\big)\Big)\bigg)\bigg)>0$. The above inequality is always positive. Finally, it remains to be shown that at $\kappa>0$ at $\delta=0$, which holds as $\kappa|_{\delta=0}=45\theta(1-\rho)\gamma>0$ where $\gamma=\bigg(2130375+\theta(1-\rho)\bigg(6955-\theta(1-\rho)\bigg(120175+6\theta(1-\rho)\Big(1745-2\theta\big(75+2\theta(1-\rho)\big)\Big)\bigg)\bigg)\bigg)>0$.

Proof of Proposition 3: The equilibrium profit of Platform 1 and Platform 2 is given as $\Pi_1^\star=P_1^\star q_1^\star+P_A^\star q_A^\star-\frac{(v_A^\star)^2}{2}-\frac{(v_1^\star)^2}{2}$ and $\Pi_2^\star=P_2^\star q_2^\star-\frac{(v_2^\star)^2}{2}$. Differentiating the profit of Platform 1 given by $\Pi_1^\star$ with respect to δ yields $\frac{\partial\Pi_1^\star}{\partial\delta}=-\frac{18\theta(1-\rho)\zeta}{\left(1725-\theta^2(1-\rho)^2(28-32\delta+9\delta^2)\right)^3}<0$, where $\zeta\triangleq7374375-15(\delta-2)^3(9\delta-14)\theta^4(\rho-1)^4-2(\delta(3\delta(859\delta-6162)+39836)-26952)\theta^3(\rho-1)^3+100(14\delta-9)(27\delta-46)\theta^2(\rho-1)^2+50(55933\delta-87318)\theta(\rho-1)$. Differentiating the profit of Platform 1 given by $\Pi_1^\star$ with respect to ρ yields $\frac{\partial\Pi_1^\star}{\partial\rho}=-\frac{6\theta\sigma}{\left(1725-\theta^2(1-\rho)^2(28-32\delta+9\delta^2)\right)^3}<0$, where $\sigma\triangleq405\delta^5\theta^4(\rho-1)^4-18\delta^4\theta^3(\rho-1)^3(205\theta(\rho-1)-859)+40\delta^3\theta^2(\rho-1)^2(\theta(\rho-1)(335\theta(\rho-1)-3223)-2835)-6\delta^2\theta(\rho-1)(2\theta(\rho-1)(4\theta(\rho-1)(505\theta(\rho-1)-8101)-39675)+1398325)+3\delta(8\theta(\rho-1)(2\theta(\rho-1)(\theta(\rho-1)(455\theta(\rho-1)-10558)-12750)+1091475)-7374375)-2(14\theta(\rho-1)-25)(2\theta(\rho-1)(14\theta(\rho-1)(10\theta(\rho-1)-291)-11325)+688275)$. In the following, we show that ζ and σ are always positive using continuity and the sign of its slope.

To show that $\zeta>0$, we differentiate ζ thrice with respect to δ thrice, which yields $\frac{\partial^3\zeta}{\partial\delta^3}=36\theta^3\big(859+10\theta(17-9\delta)(1-\rho)\big)(1-\rho)^3>0$. Next differentiating ζ with respect to δ twice and considering its value at $\delta=0$, we note that $\frac{\partial^2\zeta}{\partial\delta^2}|_{\delta=0}=72\theta^2\Big(1050-\theta(1-\rho)\big(1027+80\theta(1-\rho)\big)\Big)(1-\rho)^2$. The above inequality is positive for $0<\theta<\min\{\frac{1}{160}\sqrt{1390729}\sqrt{\frac{1}{(\rho-1)^2}}+\frac{1027}{160(\rho-1)},1\}$. Instead the above inequality is negative for $\theta>\frac{1}{160}\sqrt{1390729}\sqrt{\frac{1}{(\rho-1)^2}}+\frac{1027}{160(\rho-1)}$. Next, differentiating ζ with respect to δ and it is then sufficient to show that at $\delta=1$ and at $\delta=0$ the sign stays the same. We find that $\frac{\partial\zeta}{\partial\delta}|_{\delta=0}=-2\theta(1-\rho)(1398325+2\theta(1-\rho)\Big(22175-2\theta(1-\rho)\big(9959+450\theta(1-\rho)\big)\Big)<0$ and $\frac{\partial\zeta}{\partial\delta}|_{\delta=1}=-10\theta(1-\rho)\Big(279665+\theta(1-\rho)\Big(1310-\theta(1-\rho)\big(2119+36\theta(1-\rho)\big)\Big)\Big)<0$. The inequalities in the above cases are always negative, thus confirming that the first derivative of ζ with respect to δ is always negative. Now, it is sufficient to show that at $\delta=1$, ζ is positive which is given as $\zeta|_{\delta=1}=7374375+25\theta(1-\rho)\Big(62770-\theta(1-\rho)\Big(380-\theta(1-\rho)\big(242+3\theta(1-\rho)\big)\Big)\Big)>0$.

To show that $\sigma>0$, differentiate σ four times with respect to δ, which yields $\frac{\partial^4\sigma}{\partial\delta^4}=-216\theta^3(1-\rho)^3\big(1718+5\theta(1-\rho)(82-45\delta)\big)<0$. Next differentiating σ with respect to δ thrice and considering its value at $\delta=0$, we note that $\frac{\partial^3\sigma}{\partial\delta^3}|_{\delta=0}=-240\theta^2(1-\rho)^2\Big(2835-\theta\big(3223+335\theta(1-\rho)\big)\Big)<0$. The above inequality is negative for $\theta<\frac{\sqrt{14186629}-3223}{670(1-\rho)}$ and positive otherwise. Differentiating σ with respect to δ twice and considering its value at $\delta=1$ and at $\delta=0$ is respectively given as $\frac{\partial^2\sigma}{\partial\delta^2}|_{\delta=1}=60\theta(1-\rho)\Big(279665+\theta(1-\rho)\Big(4530-\theta(1-\rho)\big(3162+71\theta(1-\rho)\big)\Big)\Big)>0$, and $\frac{\partial^2\sigma}{\partial\delta^2}|_{\delta=0}=12\theta\Big(1398325+2\theta(1-\rho)\Big(39675-4\theta(1-\rho)\big(8101+505\theta(1-\rho)\big)\Big)\Big)>0$. This confirms that $\frac{\partial^2\sigma}{\partial\delta^2}$ is always positive. Next, differentiating σ with respect to δ and considering its value at $\delta=1$, we note that $\frac{\partial\sigma}{\partial\delta}|_{\delta=1}=-75\Big(294975+\theta(1-\rho)\Big(125540-\theta^2\big(720+11\theta(1-\rho)\big)\Big)\Big)<0$. This confirms that $\frac{\partial\sigma}{\partial\delta}$ is always negative. Now, it is sufficient to show that at $\delta = 1$ we find $\sigma > 0$ which is given as $\sigma|_{\delta=1} = 125\left(98325 + \theta(1-\rho)\left(20790 - \theta(1-\rho)\left(180 + \theta\big(86 + \theta(1-\rho)\big)\right)\right)\right) > 0$. This proves that $\sigma > 0$ for all $0 \le \rho \le 1$, $0 \le \delta \le 1$ and $0 < \theta < 1$.

We can proceed in the same manner to show the effect of data cross-use regulation on the profits of Platform 2. Differentiating the profit of Platform 2, given by $\Pi_2^\star$, with respect to δ yields $\frac{\partial \Pi_2^\star}{\partial \delta} = \frac{\chi}{\left(1725-\theta^2(1-\rho)^2(28-32\delta+9\delta^2)\right)^3} > 0$, where $\chi \triangleq (38\theta(\rho-1)\big((\delta-2)(\delta-1)\theta^2(\rho-1)^2 + 3(14\delta-9)\theta(\rho-1) - 75\big)(5(\delta-2)^2\theta^3(\rho-1)^3 + 6(9\delta(7\delta-9)-52)\theta^2(\rho-1)^2 + 75(28\delta-37)\theta(\rho-1) + 72450) > 0$. Employing the same steps as above, we can show that this expression is positive. The explicit long proof is available upon request. Similarly, differentiating the profit of Platform 2 with respect to ρ yields $\frac{\partial \Pi_2^\star}{\partial \rho} = \frac{\mu}{\left(1725-\theta^2(1-\rho)^2(28-32\delta+9\delta^2)\right)^3} < 0$, where $\mu \triangleq 114(14\delta-9)\theta\left((\delta-2)\theta(\rho-1)\big((9\delta-14)\theta(\rho-1)+50\big)+1725\right)\big((\delta-2)(\delta-1)\theta^2(\rho-1)^2 + 3(14\delta-9)\theta(\rho-1) - 75\big)$ with $\mu > 0$ if and only if $\delta < 9/14$ and negative otherwise. Employing the steps as above, we can show that this expression is positive. The explicit long proof is available upon request.

Proof of Proposition 4: Total welfare in a market is the sum of consumer surplus and platform profits in both markets and is given as $TW^\star = CS_A^\star + CS_B^\star + \Pi_1^\star + \Pi_2^\star = \frac{W}{2\left((1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\right)^2}$, where $W \triangleq 3(26708\delta^2 - 31346\delta + 17063)\theta^2(\rho-1)^2 - (\delta-2)^2(2\delta(8\delta-61)+131)\theta^4(\rho-1)^4 + 6(\delta-2)(\delta(577\delta-554)-73)\theta^3(\rho-1)^3 - 10350(19\delta+22)\theta(\rho-1) + 1987200$. Differentiating the total welfare with respect to δ and ρ is respectively given as $\frac{\partial TW^\star}{\partial \delta} = \frac{(1-\rho)\theta W_\delta}{2\left((1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\right)^3} > 0$ and $\frac{\partial TW^\star}{\partial \rho} = -\frac{\theta W_\rho}{\left((1725-(2-\delta)(14-9\delta)(1-\rho)^2\theta^2)\right)^3} < 0$, where $W_\delta \triangleq 339221250 - 6(\delta(221972\delta^2 - 262746\delta - 331879) + 183778)\theta^3(\rho-1)^3 - 650(\delta-2)^3(\delta-1)\theta^5(\rho-1)^5 - 6(\delta-2)(\delta(\delta(5193\delta-1894)-24311)+19162)\theta^4(\rho-1)^4 - 10350(6\delta(203\delta-600)+2975)\theta^2(\rho-1)^2 - 10350(33620\delta-27961)\theta(\rho-1) > 0$ and $W_\rho \triangleq \big((\delta-2)(9\delta-14)\theta^2(\rho-1)^2 - 1725\big)\big(3(26708\delta^2 - 31346\delta + 17063)\theta(\rho-1) - 2(\delta-2)^2(2\delta(8\delta-61)+131)\theta^3(\rho-1)^3 + 9(\delta-2)(\delta(577\delta-554)-73)\theta^2(\rho-1)^2 - 5175(19\delta+22)\big) - 2(\delta-2)(9\delta-14)\theta(\rho-1)(3(26708\delta^2 - 31346\delta + 17063)\theta^2(\rho-1)^2 - (\delta-2)^2(2\delta(8\delta-61)+131)\theta^4(\rho-1)^4 + 6(\delta-2)(\delta(577\delta-554)-73)\theta^3(\rho-1)^3 - 10350(19\delta+22)\theta(\rho-1) + 1987200) > 0$. To show these inequalities, we need to show that $W_\delta > 0$ and $W_\rho > 0$. To show that $W_\delta > 0$, we employ continuity and the sign of its slope. Differentiating W_δ with respect to δ four times yields $\frac{\partial^4 W_\delta}{\partial \delta^4} = -48\theta^4(1-\rho)^4\big(15579 - 325\theta(1-\rho)\big) < 0$. The above inequality follows directly from $0 \le \rho \le 1$, $0 \le \delta \le 1$ and $0 < \theta < 1$. Next, differentiating W_δ with respect to δ thrice and then considering its sign at $\delta = 1$ we find $\frac{\partial^3 W_\delta}{\partial \delta^3}|_{\delta=1} = 36\theta^3(1-\rho)^3\left(221972 - \theta(1-\rho)\big(8492 + 325\theta(1-\rho)\big)\right) > 0$. The above inequality follows from $0 \le \rho \le 1$ and $0 < \theta < 1$. Differentiating W_δ with respect to δ twice and considering its sign at $\delta = 1$, we observe $\frac{\partial^2 W_\delta}{\partial \delta^2}|_{\delta=1} = -60\theta^2(1-\rho)^2(420210 - 80634\theta(1-\rho) - 5241\theta^2(1-\rho)^2 - 65\theta^3(1-\rho)^3) < 0$. The above inequality follows from $0 \le \rho \le 1$ and $0 < \theta < 1$. This implies that the slope of the first derivative of W_δ with respect to δ is always negative.

Differentiating W_δ with respect to δ and then observing its sign at $\delta = 1$ yields $\frac{\partial W_\delta}{\partial \delta}|_{\delta=1} = 10\theta(1-\rho)(34796700 + 1204740\theta(1-\rho) - 114788\theta^2(1-\rho)^2 - 6402\theta^3(1-\rho)^3 - 65\theta^4(1-\rho)^4) > 0$. The above inequality is always positive as the last three terms in the bracket which are negative are significantly smaller than the positive term. Further, these negative terms are also multiplied by our parameters which further make them less negative as $0 \le \rho \le 1$ and $0 < \theta < 1$. It remains to show that at $\delta = 0$ we find $W_\delta > 0$ which is given as $W_\delta|_{\delta=0} = 339221250 - 289396350\theta(1-\rho) - 30791250\theta^2(1-\rho)^2 + 11026668\theta^3(1-\rho)^3 + 229944\theta^4(1-\rho)^4 + 5200\theta^5(1-\rho)^5$. Note that there are only two terms that negatively influence the above expression, and these are significantly smaller than the positive terms. As W_δ is always rising in δ, we then confirm that W_δ is unambiguously positive for any $\delta \in [0,1]$.

To show that W_ρ>0. differentiate W_ρ with respect to δ four times, which yields $\frac{\partial^4 W_\rho}{\partial \delta^4} = 72\theta^3(1-\rho)^3\big(221972 + \theta(1-\rho)(33836 - 25965\delta)\big) > 0$. The above inequality follows directly from $0 \le \rho \le 1$, $0 \le \delta \le 1$ and $0 < \theta < 1$. Next, differentiating W_ρ with respect to δ thrice and considering its sign at $\delta = 1$ yields $\frac{\partial^3 W_\delta}{\partial \delta^3}|_{\delta=1} = -18\theta^2(1-\rho)^2(2101050 + 34982\theta(1-\rho) - 3287\theta^2(1-\rho)^2) < 0$. The above inequality follows from $0 \le \rho \le 1$ and $0 < \theta < 1$. Differentiating W_ρ with respect to δ twice and considering its sign at $\delta = 1$, we observe $\frac{\partial^2 W_\rho}{\partial \delta^2}|_{\delta=1} = 30\theta(1-\rho)(11598900 + 82800\theta(1-\rho) - 63553\theta^2(1-\rho)^2 - 1933\theta^3(1-\rho)^3) > 0$. The above inequality follows from $0 \le \rho \le 1$ and $0 < \theta < 1$. This implies that the slope of the first derivative of W_ρ with respect to δ is always positive and W_ρ is convex in δ. As W_ρ is convex, we derive δ_{min} as the value of δ that minimizes W_ρ. Substituting $\delta = \delta_{min}$, into W_ρ, we find that the minimum value of W_ρ is always positive. This confirms that $W_\rho > 0$ for all $0 \le \rho \le 1$, $0 \le \delta \le 1$ and $0 < \theta < 1$.

Appendix B

Microfoundation of the Demand in the Baseline Model

Consumer demand in Market A: Consumers have a basic valuation r with the support $[-2,1]$ which follows the Uniform distribution, i.e., $r \sim \mathcal{U}[-2,1]$.[17] The utility of a consumer of type r in Market A that buys the service of Platform 1 is given as $U_A(r) = r + v_A - P_A$, where v_A is the value associated with innovation effort of Platform 1 in Market A, and P_A is the price of the service. Consumers employ the services of the Platform 1 in Market A only when they obtain positive utility from doing so.[18] This condition pins down the demand for the Platform 1's service in Market A as $U_A \geq 0 \Longrightarrow r > \hat{r}_A(P_A, v_A) \triangleq P_A - v_A$. Thus, the mass of consumers affiliating with the service of Platform 1 in Market A is $q_A \triangleq 1 - \frac{(\hat{r}_A(P_A,v_A)+2)}{3} = \frac{1+v_A-P_A}{3}$. Rearranging yields: $P_A = 1 + v_A - 3q_A$. Observe that inverse demand in Market A is rising as the value of the service offered by Platform 1 rises and is downward sloping in q_A.

Consumer demand in Market B: In Market B, as in Market A, we model consumers as having a basic valuation r with the support $[-2,1]$ which follows the uniform distribution, i.e., $r \sim \mathcal{U}[-2,1]$. Thus, the utility of a consumer of type r that buys from Platform 1 or from Platform 2 is given as $U_1(r) = r + v_1 + \Psi_1 - P_1$ and $U_2(r) = r + v_2 + \Psi_2 - P_2$. where v_i, Ψ_i and P_i for $i \in \{1,2\}$ are the values associated with innovation efforts, data cross-use levels and price of each platform's service in the Market B. Consumers will buy the service of the platform that provides them the highest net utility. Under the above specification, platforms 1 and 2 will have positive demand only if quality-adjusted price at each platform $\Phi \triangleq P_1 - v_1 - \Psi_1 = P_2 - v_2 - \Psi_2$ is identical implying that the "no arbitrage" condition holds.[19] Furthermore, we assume the value of a consumer's outside option is zero, such that consumers with $U_i < 0$ will not choose any platform. This implies total demand is constituted only by those consumers for whom $r > \Phi$.[20] Hence, total demand for the service in the market is $q_1 + q_2 \triangleq 1 - \frac{(\Phi+2)}{3}$, being the total demand in Market B. Rearranging and inverting the above total demand for each platform i yields the following inverse demand expression at the two platforms as $P_1(v_1, \Psi_1, q_1, q_2) \triangleq 1 + v_1 + \Psi_1 - 3(q_1 + q_2)$, $P_2(v_2, \Psi_2, q_1, q_2) \triangleq 1 + v_2 + \Psi_2 - 3(q_1 + q_2)$. with $\frac{\partial P_i}{\partial q_j} < 0$ for any $j \in \{1,2\}$, $\frac{\partial P_i}{\partial v_i} \geq 0$.

[17] As in Katz and Shapiro (1985), the support also includes negative basic valuation, as some consumers may not use the service, even at a zero price, unless the service quality exceeds a certain threshold. This support assures that there is no corner solution and that the axioms of probability are satisfied when platforms innovate in either market.

[18] Note that we assume that the outside option of consumers provides zero utility. This assumption does not affect our results as any positive (and small enough) but fixed outside option is sufficient to provide qualitatively similar results.

[19] Any consumer of type r should be indifferent between buying from R_1 or R_2 i.e., $U_1(r) = U_2(r)$. This gives us the desired no arbitrage condition.

[20] We obtain this condition from the inequality $U_i(r) \geq 0$ for both platforms.

Appendix C

Alternative Timing

Here we present a variation of our model where the strategic output decisions in the two markets occur sequentially. Specifically, we assume that decisions are first made with respect to the (primary) Market A and then with respect to the (secondary) Market B. The timing is as follows:

Stage 1: Platforms choose their level of innovation v_A, v_1, v_2.

Stage 2: Platform Firm 1 sets demand levels in Market A q_A. *Stage 3:* Platforms 1 and 2 choose their demand levels in Market B. The demand and profits are realized.

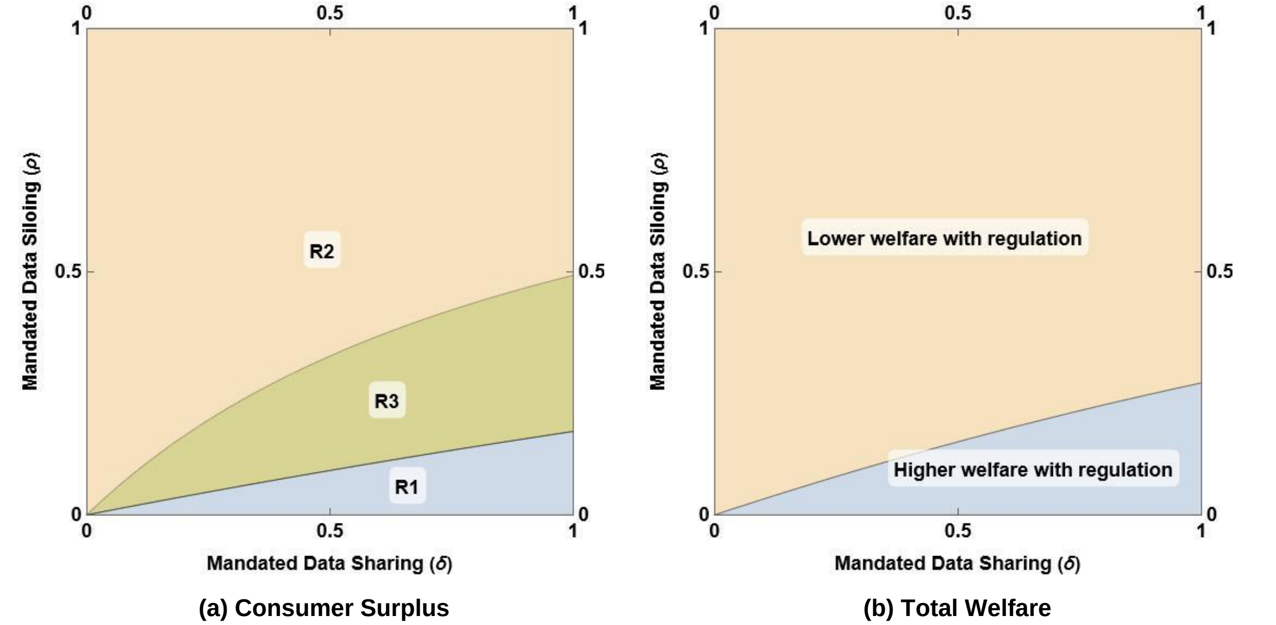
In Stage 3, platforms determine the level of demands in Market B to maximize profits given demand in Market A. Platforms 1 and 2 respectively choose q_1 and q_2 in Market B to maximize profits. Taking the appropriate first order conditions and solving yields the equilibrium demand in Market B as $q_1(v_1, v_2, q_A) = \frac{1+2v_1-v_2+\theta q_A(2-\delta)(1-\rho)}{9}$, $q_2(v_2, v_1, q_A) = \frac{1-v_1+2v_2-\theta q_A(1-2\delta)(1-\rho)}{9}$, with $q_1 + q_2 = \frac{2+v_1+v_2+\theta q_A(1-\rho)(1+\delta)}{9}$. Substituting these demands in profit of the two platforms yields $\Pi_1 = P_A q_A - I(v_A) + P_1(v_1, \Psi_1, q_1, q_2)q_1 - I(v_1)$, $\Pi_2 = P_2(v_2, \Psi_2, q_1, q_2)q_1 - I(v_2)$.
In Stage 2, Platform 1 sets the demand level in Market A to maximize profits. Differentiating the profit of Platform 1 with respect to q_A and solving yields the optimal demand level as $q_A = \frac{27(1+v_A)+2\theta(2-\delta)(1-\rho)(1+2v_1-v_2)}{(162-2\theta^2(2-\delta)^2(1-\rho)^2)}$. Substituting this optimal demand level in Market A into the inverse demand of Platform 1 in Market A yields $P_A(v_A, v_1, v_2) = 1 + v_A - \frac{3\left(27(1+v_A)+2(1+2v_1-v_2)\theta(2-\delta)(1-\rho)\right)}{(162-2\theta^2(2-\delta)^2(1-\rho)^2)}$. Substituting this optimal demand level in Market A into the demands of Platform 1 and Platform 2 in Market B yields demands as $q_1(v_1, v_2, v_A)$, $q_2(v_2, v_1, v_A)$. The associated inverse demands in Market B are $P_1(v_1, v_2, v_A)$ and $P_2(v_2, v_1, v_A)$. In Stage 1, each platform sets innovation effort levels to maximize their own profits. The profit of Platform 1 and 2 as function of innovation levels are given as $\Pi_1(v_1, v_2, v_A) = P_A q_A - I(v_A) + P_1 q_1 - I(v_1)$ and $\Pi_2(v_2, v_1, v_A) = P_2 q_2 - I(v_2)$. Differentiating the profit of Platform 1 with respect to v_1 and v_A and the profit of Platform 2 with respect to v_2 and solving simultaneously, we obtain the equilibrium innovation effort levels as follows:

$$v_1^\star = \frac{4(675 + 27\theta(14-9\delta)(1-\rho) - 5\theta^2(2-\delta)(4-3\delta)(1-\rho)^2 - (2-\delta)\theta^3(1-\rho)^3(10-3\delta(4-\delta)))}{(15525 - (2-\delta)\theta^2(1186-771\delta)(1-\rho)^2 + 2(2-\delta)^2\theta^4(1-\rho)^4\left(10-3\delta(4-\delta)\right)},$$

$$v_2^\star = \frac{2(54-(2-\delta)\theta(1-\rho)^2)(25-\theta(9-14\delta)(1-\rho) - 2(2-\delta)(1-\delta)\theta^2(1-\rho)^2)}{(15525 - (2-\delta)\theta^2(1186-771\delta)(1-\rho)^2 + 2(2-\delta)^2\theta^4(1-\rho)^4\left(10-3\delta(4-\delta)\right)}, \text{ and}$$

$$v_A^\star = \frac{3105 + \theta(2-\delta)(1-\rho)(270 - \theta(1-\rho)(86-57\delta) - 2\theta^2(2-\delta)(4-3\delta)(1-\rho)^2}{(15525 - (2-\delta)\theta^2(1186-771\delta)(1-\rho)^2 + 2(2-\delta)^2\theta^4(1-\rho)^4\left(10-3\delta(4-\delta)\right)}.$$

Substituting these equilibrium values into the profit functions yields the equilibrium profit obtained by each platform as $\Pi_1^\star \triangleq \Pi_1(v_1^\star, v_2^\star, v_A^\star)$, $\Pi_2^\star \triangleq \Pi_1(v_2^\star, v_1^\star, v_A^\star)$. The consumer surplus in Markets A and B is respectively given by $CS_A = \frac{3(q_A^\star)^2}{2}$ and $CS_B = \frac{3(q_1^\star+q_2^\star)^2}{2}$. The total welfare is defined as the sum of the platforms' profits and total consumer surplus and given as $W(\rho, \delta) = CS_A + CS_B + \Pi_1^\star + \Pi_2^\star$. Plotting the main results of our paper, we show that both consumer surplus and total welfare can be lower under regulation than without regulation, and yield qualitatively the same results as in the base model (compare Figure C1 to Figure 2).



Note: Region R1: Higher CS with regulation, R2: Lower CS with regulation, R3: Lower CS with regulation (but higher CS in Market *B*). Figures derived for $\theta = 1/10$.

Figure C1. The Impact of Various Degrees of Data Siloing Regulation ($\rho > 0$) and Data Sharing Regulation ($\delta > 0$) in Comparison to No Regulation ($\rho = \delta = 0$)

Appendix D

Personalized Data Cross-Use

In this extension, we show that our results hold when platforms interact with the same consumers in both markets. Collecting data in Market A allows Platform 1 to offer better value for services in Market B through personalization. Contrary to the base model, we now assume that Markets A and B are populated by the same set of consumers and that their total mass is normalized to one. As in the base model, consumers in each market demand at most one unit of the service. Consumers in our setting are modeled as independently and identically distributed (i.i.d.) along two dimensions on the unit interval—i.e., (1) consumers' relative preference for services of Platform 2 over 1 in Market B denoted by X, and (2) consumers' privacy cost Y when consuming Service A. Consumers' preference for services in Market B denoted by X is distributed according to a cumulative distribution function F with density f. Along the dimension of privacy costs Y consumers are distributed according to the cumulative distribution function G with density g. Since consumers are i.i.d. across these two dimensions, the joint distribution can be written as $P(X \leq x, Y \leq y) = F(x)G(y)$ with density being $f(x)g(y)$. For the purposes of this extension, we impose the following restrictions.

Assumption 1: *We assume that the distributions F and G follow a Uniform distribution—i.e., $F \sim \mathcal{U}[0,1]$ and $G \sim \mathcal{U}[0,1]$.*

In Market A, the utility of consumer of type Y, who buys the service, is $U_A(Y, v_A, p_A) = v + v_A - p_A - Y$. Consumers employ the services of the Platform 1 in Market A only when they obtain positive utility from doing so which yields $U_A \geq 0 \Longrightarrow Y \leq Y_A(v_A, p_A) \triangleq v + v_A - p_A$. In Market B, we adopt the Hotelling model of competition (Hotelling, 1929). Platforms 1 and 2 are located at the end points of the unit interval at coordinates 0 and 1, respectively, and compete by setting prices. Platform i's price is p_i $(i = 1,2)$. A consumer denoted by its type $X \in [0,1]$ that buys from Platform 1 in Market A and in Market B benefits from personalization in Market B and its utility is denoted as $U_1^S(X, p_1)$. Instead when a consumer does not buy from Platform 1 in Market A, there is no data collected by Platform 1 and thus also no personalization benefits. The utility of these consumers is given as $U_1^{NS}(X, p_1)$. The utilities of these two types of consumers that purchase from Platform 1 in Market B are $U_1^S(X, p_1) = v_1 + (1-\rho)\theta - p_1 - tX$, $U_1^{NS}(X, p_1) = v_1 - p_1 - tX$. Here, $(1-\rho)\theta$ is the value of (cross-market) personalization that Platform 1 can offer consumers whose data it has collected in Market A given data siloing regulation ρ, t is the transportation cost and X is consumers' preference mismatch.[21]

Similarly, a consumer of type $X \in [0,1]$ that buys from Platform 1 in Market A and from Platform 2 in Market B may benefit under the regulation ρ and δ and its utility is denoted as $U_2^S(X, p_2) = v_2 + (1-\rho)\delta\theta - p_2 - t(1-X)$. This is because Platform 1 has data regarding this consumer's preference and may be mandated to share this data with Platform 2. Here, $(1-\rho)\delta\theta$ is the value of personalization that Platform 2 can offer consumers, given data cross-use regulation ρ and δ, whose data it has collected in Market A given data siloing regulation ρ. Instead, when a consumer does not buy from Platform 1 in Market A, there is no data collected by Platform 1 and thus also no personalization benefits. The utilities of this consumer type is $U_2^{NS}(X, p_2) = v_2 - p_2 - t(1-X)$. Demand in Market B for Platform 1 is then characterized by the following two indifference conditions. $U_1^S \geq U_2^S \Longrightarrow X \leq X_S \triangleq \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + \Omega}{2t}$ where $\Omega \triangleq \theta(1-\rho)(1-\delta)$ and $U_1^{NS} \geq U_2^{NS} \Longrightarrow X \leq X_{NS} \triangleq \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2)}{2t}$. Since demands are i.i.d., the mass of consumers who buy Service A and those who buy from Platform 1 and those who buy from Platform 2 in Market B is given as $D_1^S = F(X_S)G(Y_A)$ and $D_2^S = \big(1 - F(X_S)\big)G(Y_A)$. Among the set of consumers who do not buy Service A, the mass of consumers in Market B who buy from Platform 1 and those who buy from Platform 2 is given as $D_1^{NS} = F(X_{NS})\big(1 - G(Y_A)\big)$ and $D_2^{NS} = \big(1 - F(X_{NS})\big)\big(1 - G(Y_A)\big)$. Further, the total mass of consumers who buy Service A denoted by D_A. Figure E2 illustrates the four demand segments.

The profit of Platform 1 and Platform 2 are respectively given by $\Pi_1 \triangleq p_A D_A - I(v_A) + p_1(D_1^S + D_1^{NS}) - I(v_1)$ and $\Pi_2 \triangleq p_2(D_2^S + D_2^{NS}) - I(v_2)$. The timing of the game is as follows:

Stage 1: Platforms choose their level of innovation v_A, v_1, v_2.

Stage 2: Platforms set their prices p_A, p_1, p_2.

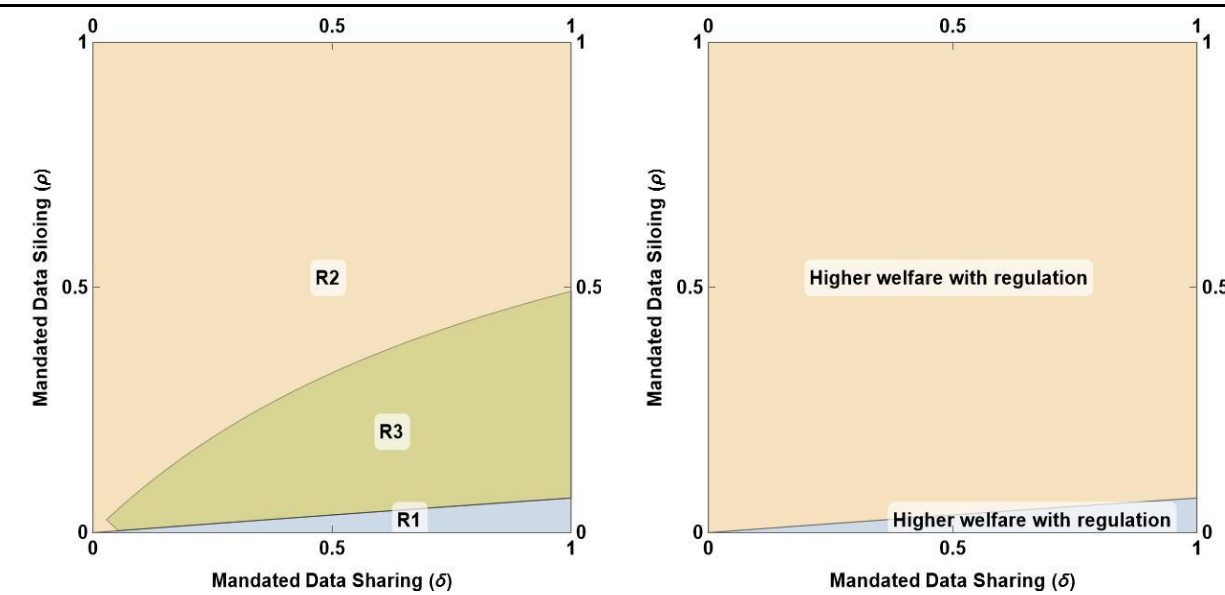
Stage 3: Consumers in Markets A and B buy and profits are realized. The equilibrium concept we employ is subgame perfect Nash equilibrium.

[21] Here, data is like a consumer value enhancing input that Platform 1 owns exclusively and may be obliged by regulation to share with its rival. For details on how a competing platform may share exclusive inputs with rivals see Carroni et al. (2024).

In Stage 2, platforms determine simultaneously the price for their service in each market. Platform 1 chooses the price level p_A, in Market A and the price level p_1 in Market B to maximize profits, while Platform 2 chooses its price p_2. Differentiating the profit of Platform 1 with respect to p_A and p_1 and the profit of Platform 2 with respect to p_2 and solving simultaneously, we get the equilibrium price levels in the markets as a function of innovation levels given as $p_1 = \frac{2t\left(6t+2(v_1-v_2)+\Omega(v+v_A)\right)}{12t-\Omega^2}$, $p_2 = \frac{2t\left(2\left(3t-(v_1-v_2)\right)-\Omega(\Omega+v+v_A)\right)}{12t-\Omega^2}$ and $p_A = \frac{(v+v_A)(6t-\Omega^2)-(3t+v_1-v_2)\Omega}{12t-\Omega^2}$. Substituting these prices into the demands, we get demands as function of investment as D_1^S, D_2^S, D_1^{NS}, D_2^{NS}, and D_A.

In Stage 1, each platform sets innovation levels to maximize profits. Differentiating the profit of Platform 1 with respect to v_1 and v_A and the profit of Platform 2 with respect to v_2 and solving simultaneously, we obtain the equilibrium innovation effort levels denoted as $v_A^\star$, $v_1^\star$ and $v_2^\star$ and are given by $v_A^\star = \frac{\Omega t(8t-\Omega^2)(36t-3\Omega^2-8)+v(12t-\Omega^2)\left(8t(9t-2)-\Omega^2(8t-1)\right)}{(12t-\Omega^2)(8t(9t-2)-\Omega^2(16t-1)+\Omega^4)}$, $v_1^\star = \frac{(8t-\Omega^2)\left(36t^2-v\Omega^3-t\left(8-3\Omega(4v-\Omega)\right)\right)}{(12t-\Omega^2)(8t(9t-2)-\Omega^2(16t-1)+\Omega^4)}$ and $v_2^\star = \frac{8t(4t(9t-2)-12\Omega tv+\Omega^3 v-\Omega^2(13t-1)+\Omega^4)}{(12t-\Omega^2)(8t(9t-2)-\Omega^2(16t-1)+\Omega^4)}$.

The equilibrium price of each platform in Market B and in Market A are given by $p_A^\star = p_A(v_A^\star, v_1^\star, v_2^\star)$, $p_1^\star = p_1(v_1^\star, v_2^\star, v_A^\star)$, and $p_2^\star = p_2(v_2^\star, v_1^\star, v_A^\star)$. Substituting these prices and innovation effort levels, the equilibrium profit obtained by each platform is $\Pi_1^\star \triangleq \Pi_1(v_1^\star, v_2^\star, v_A^\star)$ and $\Pi_2^\star \triangleq \Pi_1(v_2^\star, v_1^\star, v_A^\star)$. The consumer surplus in Market A is given by $CS_A = \int_0^{Y_A^\star}(v + v_A^\star - p_A^\star - Y)\, dY$ where $Y_A^\star \triangleq v + v_A^\star - p_A^\star$ and $\Omega = \theta(1-\delta)(1-\rho)$. Similarly, the consumer surplus in Market B is $CS_B = \int_0^{X_S^\star}\int_0^{Y_A^\star}(v_1^\star + (1-\rho)\theta - p_1^\star - tX)\, dY dX + \int_0^{X_{NS}^\star}\int_{Y_A^\star}^1 (v + v_1^\star - p_1^\star - tX)\, dY dX + \int_{X_S^\star}^1\int_0^{Y_A^\star}\left(v_2^\star + (1-\rho)\delta\theta - p_2^\star - t(1-X)\right) dY dX + \int_{X_{NS}^\star}^1\int_{Y_A^\star}^1\left(v_2^\star - p_2^\star - t(1-X)\right) dY dX$, where $X_S^\star = \frac{1}{2} + \frac{v_1^\star - v_2^\star - (p_1^\star - p_2^\star) + \Omega}{2t}$, $X_{NS}^\star = \frac{1}{2} + \frac{v_1^\star - v_2^\star - (p_1^\star - p_2^\star)}{2t}$. The total welfare is defined as the sum of the platforms' profits and total consumer surplus and given as $W(\rho,\delta) = CS_A + CS_B + \Pi_1^\star + \Pi_2^\star$. Figure D1 visualizes the main results of this extension with respect to consumer surplus and total surplus. We find qualitatively similar results as in the base model (compare Figure D1 to Figure 2).



(a) Consumer Surplus

(b) Total Welfare

Note: Region R1: Higher CS with regulation, R2: Lower CS with regulation, R3: Lower CS with regulation (but higher CS in Market *B*). Figures derived for θ = 1/10, t = 1,v = 0.75, and Δ = 0.2

Figure D1. The Impact of Various Degrees of Data Siloing Regulation ($\rho > 0$) and Data Sharing Regulation ($\delta > 0$) in Comparison to No Regulation ($\rho = \delta = 0$) with Personalized Data Cross-Use

Appendix E

Different Sizes of the Two Markets

We now consider how the relative size of the two markets impacts our result. To this end, we normalize the total mass in the two markets to one and let the size of Market A be $\alpha \in (0,1)$ and the size of Market B be $(1-\alpha)$. This specification implies that the profit from Market A becomes more valuable for Platform 1 as α increases. Incorporating this difference in market size in the profit of Platform 1 and Platform 2 yields $\Pi_1 \triangleq \alpha P_A q_A - I(v_A) + (1-\alpha)P_1 q_1 - I(v_1)$ and $\Pi_2 \triangleq (1-\alpha)P_2 q_2 - I(v_2)$.

In Stage 3, platforms determine simultaneously the level of demands they seek to achieve in each of the two markets to maximize profits considering the belief of consumers regarding the demand in Market A. Notice that given innovation effort in Stage 1, the demand levels of each platform do not change under the timing as specified in the base model. This is because in Stage 2, innovation levels are given and the choice of demands in the different markets is simultaneous. The equilibrium demand in market B and in Market A as function of innovation effort levels are as in the base model (see (sec (4)). The profit of Platform 1 and 2 as a function of innovation effort levels are given as $\Pi_1 \triangleq \alpha P_A q_A + (1-\alpha)P_1 q_1 - I(v_A) - I(v_1)$ and $\Pi_2 \triangleq (1-\alpha)P_2 q_2 - I(v_2)$.

In Stage 1, each platform sets innovation effort levels to maximize profits as presented above. Differentiating the profit of Platform 1 with respect to v_1 and v_A and the profit of Platform 2 with respect to v_2 and solving simultaneously the first order conditions, we obtain the equilibrium innovation effort levels denoted as $v_A^\star$, $v_1^\star$ and $v_2^\star$ as follows:

$$v_A^\star = \frac{(48\alpha^3 + 336\alpha^2 + 345\alpha + (2-\delta)\theta(1-\alpha)(1-\rho)(30 + 4\alpha(6+\theta(1-\rho)) + \theta(14-9\delta)(1-\rho)))}{\hat{Z}},$$

$$v_1^\star = \frac{12(1-\alpha)(30 - 4\alpha^2 + \alpha(19 + 4\theta(1-\rho)) + \theta(14-9\delta)(1-\rho))}{\hat{Z}}, \text{ and}$$

$$v_2^\star = \frac{\begin{array}{c}4(1-\alpha)(90 - 12\text{\textbackslash alpha}^2 - 3\theta(9-14\delta)(1-\rho) - \theta\text{\^{}}2(2-\delta)(1-\delta)(1-\rho)\text{\^{}}2 + \\ \alpha(57 + \theta(1-\rho)(2\theta(1-\rho) + \delta^2\theta(1-\rho) + 3\delta(4-\theta(1-\rho)))))\end{array}}{\hat{Z}},$$

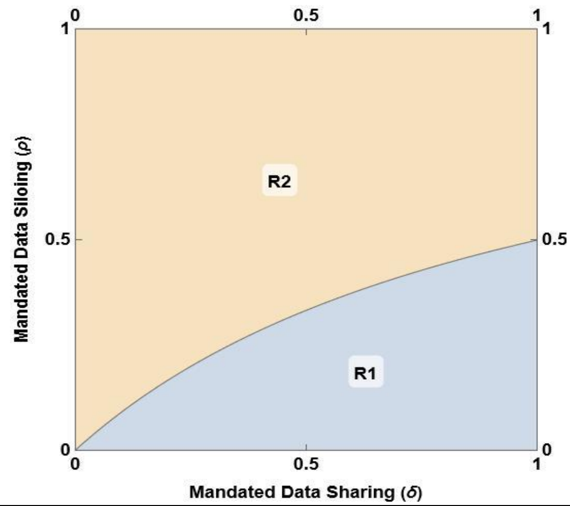
where $\hat{Z} = (2070 - 48\alpha^3 - 4\alpha^2(12 - (2-\delta)\theta^2(1-\rho)^2) + \alpha(1671 + \theta^2(2-\delta)(10-9\delta)(1-\rho)^2) - \theta^2(2-\delta)(14-9\delta)(1-\rho)^2)$.

Substituting these equilibrium innovation levels into the equilibrium market prices in Market B and in Market A yields $P_A^\star \triangleq P_A(v_A^\star)$ $P_1^\star \triangleq P_1(v_1^\star, v_2^\star, v_A^\star)$ and $P_2^\star \triangleq P_2(v_2^\star, v_1^\star, v_A^\star)$.

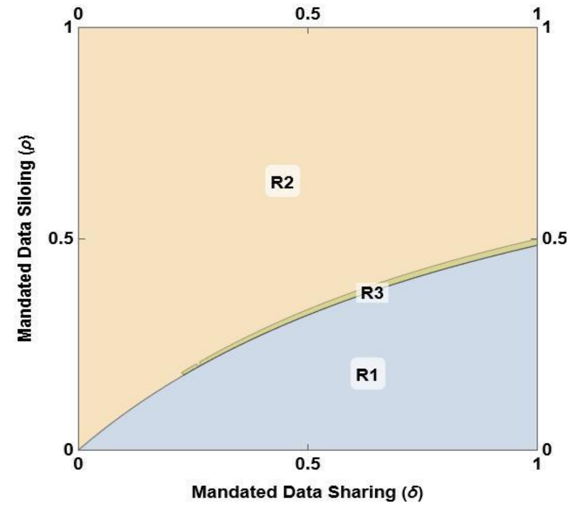
Welfare: The consumer surplus in Market A and B is as in the base model and given by $CS_A = \frac{3(q_A^\star)^2}{2}$ and $CS_B = \frac{3(q_1^\star + q_2^\star)^2}{2}$. Total consumer surplus depends on the relative market sizes of each market and is given as $CS_T = \alpha CS_A + (1-\alpha)CS_B$. Figure E1 visualizes the impact of different market sizes with respect to consumer surplus. We find qualitatively similar results as in the base model (compare Figure E1 with Figure 2). It can also be observed that α affects mainly the size of Region R3 in which the regulation has an opposing effect on Market A and Market B. However, our main policy implications are not affected by this.

The total welfare is defined as the sum of the platforms' profits and total consumer surplus and given as $W(\rho,\delta) = CS_T + \Pi_1^\star + \Pi_2^\star$.

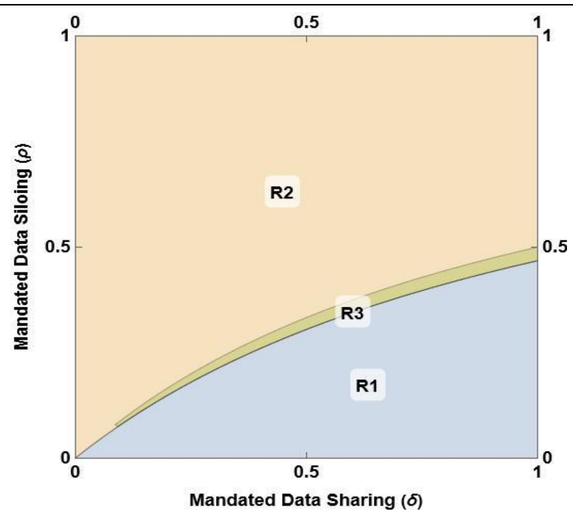
Figure E2 visualizes the impact of different market sizes with respect to total welfare, which shows qualitatively similar results as in the base model (compare Figure E2 with Figure 2).



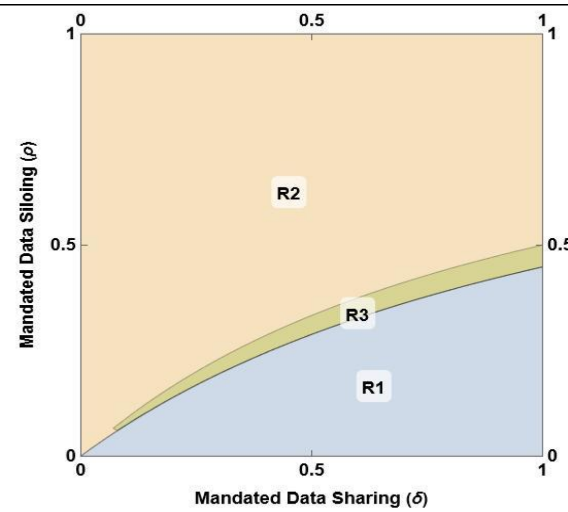
(a) Consumer Surplus for $\alpha = 0.01$



(b) Consumer Surplus for $\alpha = 0.25$



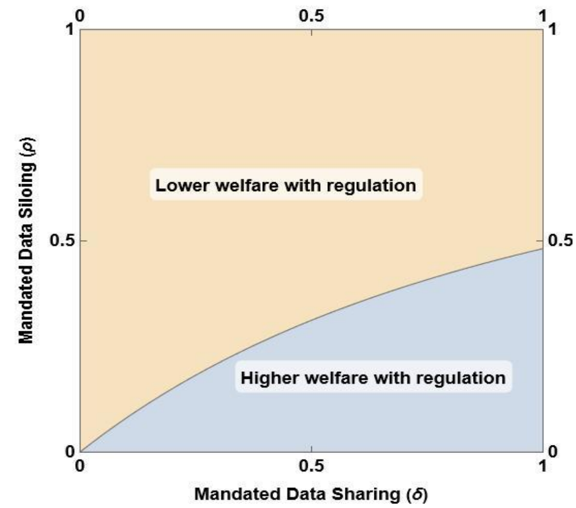
(c) Consumer Surplus for $\alpha = 0.5$



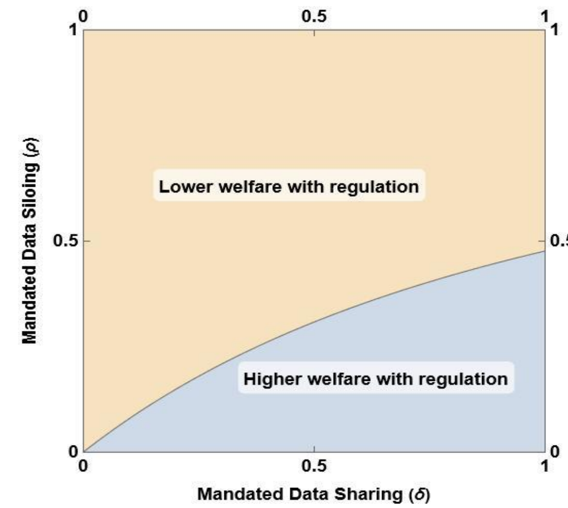
(d) Consumer Surplus for $\alpha = 0.75$

Note: Region R1: Higher CS with regulation, R2: Lower CS with regulation, R3: Lower CS with regulation (but higher CS in Market *B*). Figures derived for $\theta = 1/10$

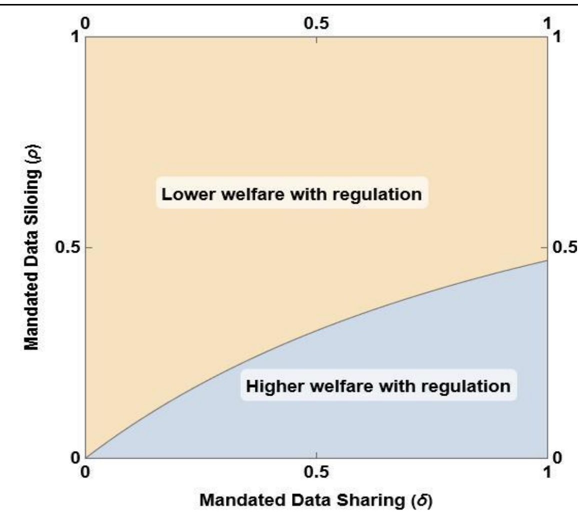
Figure E1. The Impact of Various Degrees of Data Siloing Regulation ($\rho > 0$) and Data Sharing Regulation ($\delta > 0$) in Comparison to No Regulation ($\rho = \delta = 0$)



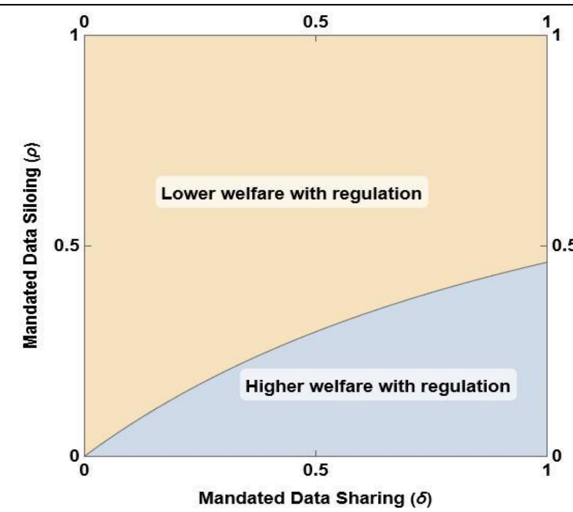
(a) Total Welfare for $\alpha = 0.01$



(b) Total Welfare for $\alpha = 0.25$



(c) Total Welfare for $\alpha = 0.5$



(d) Total Welfare for $\alpha = 0.75$

Note: Region R1: Higher CS with regulation, R2: Lower CS with regulation, R3: Lower CS with regulation (but higher CS in Market B). Figures derived for θ =1/10

Figure E2. The Impact of Various Degrees of Data Siloing Regulation (ρ > 0) and Data Sharing Regulation (δ > 0) in Comparison to No Regulation (ρ = δ = 0)

Appendix F

General Functional Form for Data Cross-Use Benefit

We show that, under the following reasonable assumption on Ψ_i, our results hold also under more general function forms for the data cross-use benefit Ψ_i.

The impact of data cross-use on demand $\Psi_i(\cdot)$ exhibits the following properties:

(a) *The platform that has immediate access to the data can use at least as much of the data as the platform with whom data is shared. As more data becomes available, Ψ_i weakly increases, everything else being equal—i.e.,* $\frac{\partial \Psi_1}{\partial q_A} \geq \frac{\partial \Psi_2}{\partial q_A} \geq 0$ and $\frac{\partial \Psi_1}{\partial \theta} \geq \frac{\partial \Psi_2}{\partial \theta} \geq 0$.

(b) *As data siloing requirements become stricter, less data can be cross-used such that Ψ_i decreases, everything else being equal—i.e.,* $\frac{\partial \Psi_1}{\partial \rho} < \frac{\partial \Psi_2}{\partial \rho} < 0$, $\frac{\partial^2 \Psi_i}{\partial \rho\, \partial q_A} < 0$, *with strict data siloing implying* $\Psi_1(1, \theta q_A) = \Psi_2(\delta, 1, \theta q_A) = 0$.

(c) *As data sharing requirements become stricter, only Ψ_2 of the data receiving platform increases, while there is no direct impact on Ψ_1 of the data providing platform—i.e.,* $\frac{\partial \Psi_1}{\partial \delta} = 0$, $\frac{\partial \Psi_2}{\partial \delta} > 0$, *where* $\Psi_1(\rho, \theta q_A) > \Psi_2(\delta, \rho, \theta q_A)$ *and* $\frac{\partial^2 \Psi_2}{\partial \delta\, \partial q_A} > 0$ *while* $\frac{\partial^2 \Psi_1}{\partial \delta\, \partial q_A} = 0$, *with strict data sharing ($\delta = 1$) implying* $\Psi_1(\rho, q_A) = \Psi_2(1, \rho, q_A)$.

Now, we present generalized versions of Lemma 1 and Proposition 1, which capture the main driving forces behind our (welfare) results and insights.

Lemma 1' (Impact of data cross-use regulation on demand): ***(a)*** *A higher degree of data cross-use (θ) or a higher innovation level in Market A (v_A) always increases Platform 1's demand in Market B, but increases Platform 2's demand in Market B only when coupled with a sufficient level of data sharing ($\delta > \delta_1$).* ***(b)*** *More data siloing (larger ρ) leads to lower demand in Market B for Platform 1 but reduces Platform 2's demand in Market B only when the level of data sharing is high ($\delta > \delta_2$).* ***(c)*** *More data sharing (larger δ) leads to lower demand for Platform 1 in Market B, but to a higher demand for Platform 2.* ***(d)*** *More data siloing reduces the total demand in Market B, whereas more data sharing increases the total demand in Market B.*

Proof of Lemma 1': In Stage 3, the first order conditions are as presented in the main text. Solving this system of equation yields the demands in Markets A and B as a function of anticipated value from demand (Ψ_i) and innovation levels as follows: $q_A(v_A) = \frac{1+v_A}{6}$, $q_1(v_1, v_2, \Psi_1, \Psi_2) = \frac{1+2v_1-v_2+2\Psi_1-\Psi_2}{9}$ and $q_2(v_2, v_1, \Psi_2, \Psi_2) = \frac{1+2v_2-v_1+2\Psi_2-\Psi_1}{9}$. In equilibrium, consumers' anticipated value must be correct $q_A^e = q_A(v_A)$ yields $\Psi_1(\rho, v_A)$ and $\Psi_2(\delta, \rho, v_A)$.

Comparative statics with respect to v_A and θ. Differentiating q_1 with respect to v_A yields $\left(2\frac{\partial \Psi_1}{\partial q_A^e} - \frac{\partial \Psi_2}{\partial q_A^e}\right)\frac{\partial q_A}{\partial v_A} \geq 0$, and by the assumptions presented above, $\frac{\partial \Psi_1}{\partial q_A^e} - \frac{\partial \Psi_2}{\partial q_A^e} \geq 0$. Similarly, differentiating q_1 with respect to θ yields $2\frac{\partial \Psi_1}{\partial \theta} - \frac{\partial \Psi_2}{\partial \theta} \geq 0$. Thus, it follows that Platform 1's demand always increases with an increase in v_A or θ. Instead, Platform 2 benefits from a higher v_A or θ only under a sufficiently high degree of data cross-use for Platform 2, i.e., $\delta > \delta_1$. To see this, note that the demand of Platform 2 increases in v_A when $2\frac{\partial \Psi_2}{\partial v_A} - \frac{\partial \Psi_1}{\partial v_A} > 0$. To show this, we recall two properties from the assumptions presented above: (i) $\frac{\partial^2 \Psi_2}{\partial v_A\, \partial \delta} > 0$ is increasing in δ while $\frac{\partial^2 \Psi_1}{\partial v_A\, \partial \delta} = 0$; (ii) $\Psi_2(\delta = 1, \rho, v_A) = \Psi_1(\rho, v_A)$ and $\Psi_2(\delta = 0, \rho, v_A) = 0$. Noting that $\left(2\frac{\partial \Psi_2}{\partial v_A} - \frac{\partial \Psi_1}{\partial v_A}\right)|_{\delta=0} < 0$ and $\left(2\frac{\partial \Psi_2}{\partial v_A} - \frac{\partial \Psi_1}{\partial v_A}\right)|_{\delta=1} > 0$, by continuity there exists a cutoff δ_1 such that when $\delta > \delta_1$, it must be that $\left(2\frac{\partial \Psi_2}{\partial v_A} - \frac{\partial \Psi_1}{\partial v_A}\right) > 0$ and otherwise $\left(2\frac{\partial \Psi_2}{\partial v_A} - \frac{\partial \Psi_1}{\partial v_A}\right) < 0$. Further, by the assumptions stated above, Ψ_2 is a function of θq_A^e, and q_A^e is increasing in v_A. Hence, as $\theta > 0$ and is multiplied with v_A, a similar analysis can be performed.

Comparative statics with respect to ρ. Differentiating demands in Market B with respect to ρ yields

$$sign\left(\frac{\partial q_1}{\partial \rho}\right) = sign\left(2\frac{\partial \Psi_1}{\partial \rho} - \frac{\partial \Psi_2}{\partial \rho}\right) < 0,\ sign\left(\frac{\partial q_2}{\partial \rho}\right) = sign\left(2\frac{\partial \Psi_2}{\partial \rho} - \frac{\partial \Psi_1}{\partial \rho}\right).$$

The first inequality as presented above is always negative as $\frac{\partial \Psi_1}{\partial \rho} \leq \frac{\partial \Psi_2}{\partial \rho} < 0$. The sign of the second inequality is more nuanced. As before, note that as δ increases, $\frac{\partial^2 \Psi_2}{\partial \rho\, \partial \delta} < 0$. Recalling that at $\delta = 0$, $\Psi_2 = 0$ which implies that $\frac{\partial q_2}{\partial \rho} > 0$. On the contrary, at $\delta = 1$ it must be that $\Psi_2 = \Psi_1$ implying $\frac{\partial q_2}{\partial \rho} < 0$. By continuity, there exists δ_2 such that for $\delta < \delta_2$, we have $\frac{\partial q_2}{\partial \rho} > 0$. *Comparative statics with respect to* δ. Differentiating demands in Market B with respect to ρ, note that $sign\left(\frac{\partial q_1}{\partial \delta}\right) = sign\left(-\frac{\partial \Psi_2}{\partial \delta}\right) < 0$, and $sign\left(\frac{\partial q_2}{\partial \delta}\right) = sign\left(2\frac{\partial \Psi_2}{\partial \delta}\right) > 0$.[22]

Proposition 1' (Innovation): *The innovation effort by Platform 1 in Market A and B is reduced with stricter data sharing (larger δ) or with stricter data siloing (larger ρ). The innovation effort of Platform 2 always increases with data sharing but increases with data siloing only if the level of data sharing is low ($\delta < \delta_3$).*

Proof of Proposition 1': Solving the first order conditions presented in the main text simultaneously, we get innovation levels in Market B as a function of v_A as follows $v_1(v_A) \triangleq \frac{4(5+14\Psi_1-9\Psi_2)}{115}$ and $v_2(v_A) \triangleq \frac{4(5+14\Psi_2-9\Psi_1)}{115}$. The comparative statics with respect to v_A arises from: (i) $\frac{\partial \Psi_1}{\partial q_A^e} \geq \frac{\partial \Psi_2}{\partial q_A^e} \geq 0$ and $\frac{\partial q_A}{\partial v_A} > 0$, and (ii) at $\delta = 0$, $\frac{\partial \Psi_1}{\partial q_A^e} > 0$ and $\frac{\partial \Psi_2}{\partial q_A^e} = 0$, and at $\delta = 1$, $\frac{\partial \Psi_1}{\partial q_A^e} = \frac{\partial \Psi_2}{\partial q_A^e} > 0$. By continuity, $\frac{\partial v_1(v_A)}{\partial v_A} > 0$ always holds. Comparative statics of $v_2(v_A)$ with respect to v_A is more nuanced. Specifically, there exists a critical level of data sharing denoted by δ_3 at which $\frac{\partial v_2(v_A)}{\partial v_A}|_{\delta=\delta_3} = 0$ and above which ($\delta > \delta_3$) we have $\frac{\partial v_2(v_A)}{\partial v_A} > 0$. Substituting $v_1(v_A)$ and $v_2(v_A)$ in the first order condition in the main text and solving characterizes the equilibrium innovation level $v_A^\star$. The equilibrium innovation levels in Market B are $v_1^\star = v_1(v_A^\star) = \frac{4(5+14\Psi_1-9\Psi_2)}{115}$, and $v_2^\star = v_2(v_A^\star) = \frac{4(5+14\Psi_2-9\Psi_1)}{115}$. Employing the implicit function theorem, we obtain the derivative of the equilibrium innovation level $v_A^\star$ with respect to ρ and δ as $\frac{\partial v_A^\star}{\partial \rho} = \frac{6H}{S}\left(2\frac{\partial^2 \Psi_1}{\partial q_A^e\, \partial \rho} - \frac{\partial^2 \Psi_2}{\partial q_A^e\, \partial \rho}\right) < 0$ and $\frac{\partial v_A^\star}{\partial \delta} = -\frac{6H}{S}\left(\frac{\partial^2 \Psi_2}{\partial q_A^e\, \partial \delta}\right) < 0$, with $\left(2\frac{\partial^2 \Psi_1}{\partial q_A^e\, \partial \rho} - \frac{\partial^2 \Psi_2}{\partial q_A^e\, \partial \rho}\right) < 0$, $\left(\frac{\partial^2 \Psi_2}{\partial q_A^e\, \partial \delta}\right) > 0$, $H = (5 + 14\Psi_1 - 9\Psi_2) > 0$ and $S = 1725 - \left(2\frac{\partial \Psi_1}{\partial q_A^e} - \frac{\partial \Psi_2}{\partial q_A^e}\right)\left(14\frac{\partial \Psi_1}{\partial q_A^e} - 9\frac{\partial \Psi_2}{\partial q_A^e}\right) > 0$.

Next, we demonstrate that $v_1^\star$ falls in both δ and ρ. First, the total derivative of Ψ_1 with respect to ρ and δ is given as $\frac{d\Psi_1}{d\rho} = \frac{\partial \Psi_1}{\partial \rho} + \frac{\partial \Psi_1}{\partial v_A}\frac{\partial v_A^\star}{\partial \rho} < 0$, and $\frac{d\Psi_1}{d\delta} = \frac{\partial \Psi_1}{\partial v_A}\frac{\partial v_A^\star}{\partial \delta} < 0$.[23] Second, the total derivative of $\Psi_2(\delta, \rho, v_A^\star)$ with respect to ρ and δ is respectively given as $\frac{d\Psi_2(\delta,\rho,v_A^\star)}{d\rho} = \frac{\partial \Psi_2}{\partial \rho} + \frac{\partial \Psi_2}{\partial v_A}\frac{\partial v_A^\star}{\partial \rho} < 0$ and $\frac{d\Psi_2(\delta,\rho,v_A^\star)}{d\delta} = \frac{\partial \Psi_2}{\partial \delta} + \frac{\partial \Psi_2}{\partial v_A}\frac{\partial v_A^\star}{\partial \delta} > 0$. The second inequality must be positive because intuitively the direct positive effect of an increase in δ must always outweigh any indirect effect. Else, Platform 1 would have incentives to share data with Platform 2 to lower its data advantage, which is unreasonable. Differentiating $v_1^\star$ with respect to ρ and δ yields $\frac{\partial v_1^\star}{\partial \rho} = \frac{4}{115}\left(14\frac{d\Psi_1}{d\rho} - 9\frac{d\Psi_2(\delta,\rho,v_A^\star)}{d\rho}\right) < 0$ and $\frac{\partial v_1^\star}{\partial \delta} = \frac{4}{115}\left(-9\frac{d\Psi_2}{d\delta}\right) < 0$. Differentiating $v_2^\star$ with respect to ρ and δ yields $\frac{\partial v_2^\star}{\partial \rho} = \frac{4}{115}\left(14\frac{d\Psi_2}{d\rho} - 9\frac{d\Psi_1}{d\rho}\right)$ and $\frac{\partial v_2^\star}{\partial \delta} = \frac{4}{115}\left(14\frac{d\Psi_2}{d\delta} - 9\frac{d\Psi_1}{d\delta}\right) > 0$. The second inequality is unambiguously always positive and follows directly from the fact that $\frac{d\Psi_2}{d\delta} > 0$ and that $\frac{d\Psi_1}{d\delta} < 0$. However, the sign of $\frac{\partial v_2^\star}{\partial \rho}$ depends on δ. At $\delta = 0$, we have that $\frac{\partial v_2^\star}{\partial \rho} > 0$. Instead, for $\delta = 1$, we observe that $\frac{\partial v_2^\star}{\partial \rho} < 0$ as $\Psi_1 = \Psi_2$. In addition, it must be that $\frac{\partial^2 v_2^\star}{\partial \rho\, \partial \delta} < 0$. This is because as the data sharing regulation gets stricter (as δ increases), increased data siloing reduces Platform 2's marginal revenues as value from data cross-use is reduced. Suppose that instead if this inequality were positive, Platform 2 would prefer an increase in data siloing as the data sharing regulation gets stricter (as δ increases). However, this would not be reasonable, as the margin (and hence profitability) of Firm 2 increases with the demand-enhancing effect of data sharing, whereas data siloing only limits these benefits. This intuition holds under a wide range of functional forms for Ψ_i. In sum, there exists a δ_3 such that $\frac{\partial v_2^\star}{\partial \rho} > 0$ if and only if $\delta < \delta_3$ and $\frac{\partial v_2^\star}{\partial \rho} \leq 0$ when $\delta \geq \delta_3$.

[22] The above two inequalities arise from Ψ_1 being a constant in δ while Ψ_2 is rising in δ.
[23] The above inequalities follow directly from the assumptions on Ψ_i and that $v_A^\star$ falls in both ρ and δ.