# Data Neutrality, Data Supply, and Market Competition*

Hanming Fang[†] Soo Jin Kim[‡]

April 27, 2026

**Abstract**

We analyze the effects of data neutrality regulations on downstream market competition, the incentive of the platform to produce data, and consumer welfare. In our framework, data neutrality requires that firms seeking access to the platform's data be treated equally, irrespective of whether they are affiliated with the platform. We consider two forms of regulation. Under weak data neutrality, the platform must provide the same amount of data to affiliated and unaffiliated sellers; under strong data neutrality, it must also charge the same price. We show that weak data neutrality can be largely ineffective, as the platform may restore exclusion through discriminatory pricing. Strong data neutrality is more consequential, but it does not necessarily raise welfare. Although it broadens access and intensifies downstream competition, it also reduces the incentive of the platform to refine and produce data. Consequently, data neutrality can reduce the equilibrium amount of data available in the market, and this data-reduction effect can dominate its benefits, which enhance competition. These findings suggest that regulating access to platform data requires balancing fair competition against the incentive to generate valuable data inputs.

**Keywords**: Data Neutrality, Data Intermediary, Data Supply, Data Foreclosure

**JEL Classification Numbers:** D4, L1, L4, L5

*An earlier version of the paper was circulated under the title "*Platform, Data Neutrality, and Market Competition*." We thank Joan Calzada, Yucheng Ding, Itay Fainmesser, Simone Galperti, Wayne Gao, Ricard Gil, Jeremy Greenwood, Kevin He, Po-Hsuan Hsu, Jessie Liu, Andrew Postlewaite, Juuso Toikka, Rakesh Vohra, Christopher Yoo, John Zhang, and participants at Penn Theory Lunch, Asia Pacific Industrial Organization Conference (2021), the 19th International Industrial Organization Conference (2021), China Meeting of Econometric Society (CMES 2021), the Inaugural Penn Economics of Digital Services Research Symposium (2021), Beijing-Humboldt Forum ZEW Session (2021), and Asian Innovation and Entrepreneurship Association (AIEA) Webinar for the valuable discussions and comments. Fang gratefully acknowledges the partial financial support from the Center for Technology, Innovation and Competition (CTIC) and the Warren Center for Network & Data Sciences at the University of Pennsylvania. All remaining errors are our own.

[†]Department of Economics, University of Pennsylvania, 133 S. 36th Street, Philadelphia, PA 19104; and the NBER. Email: hanming.fang@econ.upenn.edu.

[‡]School of Economics, Chung-Ang University, Seoul, Korea. E-mail: soojinkim@cau.ac.kr.

"*Some people say data is the new oil. My view is that it's like oil in one way, in that it has to be refined to be useful. Raw data itself is not really worth that much. You've got to analyse it and process it to be able to come up with information, knowledge and understanding that helps you apply it to real-world problems.*"

"*From an economic point of view, unlike oil, data is non-rival. If I have a barrel of oil and I give it to you, then I lost a barrel and you gained a barrel. But if I have some data or information and I share it with you, then we both have the same information.*" — Hal Varian (2018)

# 1 Introduction

Internet platforms, such as social media and e-commerce websites, act as data intermediaries. They often provide free services to users, but monetize the personal data collected from their utilization of that free service. The personal data collected by platforms can enable online sellers to learn about potential users' specific needs, which facilitates the sellers to customize the products to better meet the consumers' needs and increase sales.[1]

In this paper, we consider an interesting and important problem posed by the potential asymmetry in data availability when the upstream platform has a vertical relationship with some downstream sellers. If the platform, as a data intermediary, is affiliated with a downstream seller, then the affiliated firm may be able to use more platform data, even possibly for free, while its competitors either must pay higher fees or be altogether excluded from data use. For example, Amazon started offering in-house products under the label AmazonBasics and achieved great success in a short period of time.[2] A key factor behind the success of AmazonBasics is the huge amount of data Amazon has—using these data, AmazonBasics can be more effective in designing and directing customers to its private-label brands.[3] Given this preferential treatment given to its affiliated firms, the EU started questioning whether the platform abuses its data-driven advantages in its dual role as a downstream retailer and an upstream marketplace platform.[4] Indeed, Hagiu et al. (2022) examine the effect of regulations on the

[1]There are other uses of personal data by sellers, for example, data can allow the sellers to engage in more effective extraction of consumer surplus via market segmentation (e.g., Yang, 2022); it can also lower consumers' search cost (e.g., Liu et al., 2020).

[2]For example, AmazonBasics batteries have outsold both Energizer and Duracell on Amazon.com. See https://www.nytimes.com/2018/06/23/business/amazon-the-brand-buster.html.

[3]Indeed, Amazon admitted that it uses its internal data to better promote its in-house products. See https://www.cnbc.com/2019/11/19/amazon-uses-aggregated-data-from-sellers-to-build-its-own-products.html. Other unaffiliated competitors can obtain some of Amazon's data through a program called Amazon Retail Analytics Premium. However, it costs a minimum of $100,000 to access the database, which creates an uneven playing field between unaffiliated and affiliated sellers.

[4]See https://www.businessinsider.com/amazon-investigated-by-eu-commissioner-margrethe-vestager-2018-9 and https://www.thelocal.de/20181129/german-competition-watchdog-launches-probe-against-amazon.

platform's role choice, such as whether it acts as a pure marketplace or as a pure retailer.

Instead of examining the effect of a direct ban on the dual roles of a platform, we focus on the source of its market power, i.e., the preferential use of upstream data for the affiliated downstream seller. We examine how flexible and open access to a platform's data by *all* downstream sellers regardless of their vertical affiliation status with the platform, a new hypothetical regulation, which we refer to as *data neutrality*, affects the platform's data production and pricing strategies, the downstream sellers' data acquisition and pricing decisions, and ultimately consumer welfare.

Data neutrality is a relatively new concept, and only a small number of studies have discussed it. In Easley et al. (2018), data neutrality is defined as a concept analogous to net neutrality: data neutrality is considered within the context of an *industry* where there is an agency that collects and distributes market data and downstream sellers have equal rights to access those data. However, we focus on a somewhat narrower definition of data neutrality within the context of a *specific platform*: all firms that want access to the platform's data should be treated equally in terms of data usage. If the platform monetizes data under data neutrality, then any downstream firm, regardless of whether it is affiliated with the platform, should be able to obtain the same amount of data for the same fee; if the platform does not share any user data with others, then even the affiliated firms should not have access to the data under data neutrality.

Several attempts are currently underway to promote open data access. For example, as discussed in Richter and Slowinski (2019), the International Data Spaces Association, a peer-to-peer network, plays a role as a data sharing platform that manages secure data exchange and sharing between business players.[5] Another example is a big data information-sharing program deployed by one of the large Spanish banks, which provides unique data on card transactions of customers to firms (see Galdon-Sanchez et al., 2025). In addition to such private sector efforts, the European Commission proposed the Digital Market Act (DMA) in 2020, which explicitly stipulates that data sharing by platforms should be mandated for fair competitions.[6]

To theoretically examine the effect of open data access as represented by data neutrality regulations on market competition and consumers' welfare, we propose a model with the following players. The monopolistic *platform* as a data intermediary,[7] downstream duopolistic *sellers* who may obtain consumer data from the platform to send targeted ads to attract consumers, and *consumers* whose heterogeneous valuation with respect to the quality of the seller's targeted ads is uniformly distributed on $[0, \bar{\theta}]$. Each consumer obtains higher utility from the

[5]See https://www.internationaldataspaces.org/our-approach/#about-us.

[6]The proposal was signed into law by the European Parliament and the Council of the European Union in September 2022; see https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en and https://ec.europa.eu/commission/presscorner/detail/en/IP_22_6423.

[7]The platform may acquire the data from the services it provides to consumers, such as social media (e.g., Facebook or Tencent) or e-commerce (Amazon or Alibaba), or is simply in the data business.

better-targeted ads from the sellers; thus, the sellers have incentives to acquire data from the platform. We interpret the seller's targeted ads as an improvement of the overall product quality; and it should be understood to also capture the use of data for customized product design. Thus, data acquisition in our context implies an investment in quality: given that two sellers are initially differentiated with respect to targeting quality, sellers decide whether to purchase additional data services from the platform to improve their targeting quality. Additionally, we assume that one of the sellers is affiliated with the platform, whereas the other is not. This asymmetry assumption allows us to explore how the platform's incentive to change its optimal data-related decisions under data neutrality depends on the asymmetric vertical affiliation. The integrated firm–the platform and its affiliated downstream seller–decides the optimal amount of data provision that maximizes its profit.

Specifically, our concern is the platform's incentive to provide exclusive data access to its vertically affiliated seller, which will give it a considerable competitive advantage. Such excludability may result in greater harm to consumers given that data are non-rival goods. Unlike rival goods, the same set of data can be provided to multiple sellers without incurring additional marginal costs. By incorporating these unique features of data as non-rival but excludable goods into the model, we examine two types of data neutrality regulation: *weak* and *strong* data neutrality. Under *weak* data neutrality, the platform should provide the same amount of data to both sellers regardless of its vertical affiliations, although it is still allowed to price discriminate for data. Under *strong* data neutrality, the platform is required not only to provide the same set of data, but also to engage in nondiscriminatory data pricing.

In the absence of data neutrality regulation, we find that the platform will optimally choose to sell the data exclusively to *one* seller only. However, to which seller will the platform sell the data depends on the initial strength of the two sellers: if the affiliated seller's initial targeting quality is higher than that of the unaffiliated seller, then the platform forecloses the unaffiliated seller from data access and grants the affiliated seller exclusive access, and vice versa.

This result is somewhat surprising in two aspects: first, the platform will always sell the data only to one downstream seller and foreclose the other seller from data access; second, the platform may not sell the data exclusively to the affiliated seller. The intuition for this result is as follows. The platform decides which seller or sellers to sell data to based on the following considerations. If selling the data to the affiliated seller, how much additional downstream profits it can generate by having the data advantage, and/or if selling the data to the unaffiliated seller, how much it is willing to pay for the data. The key insight is that selling the data to both sellers reduces the strategic value of the data in the downstream competition, thus their willingness to pay for the data; in the extreme, if both sellers have the same amount of data, the data does not provide much advantage to either seller! This explains why the platform will sell the data exclusively only to one of the downstream sellers. When the affiliated seller has initial strength relative to the unaffiliated seller, giving it exclusive access to the data can further strengthen its competitive advantage; in contrast, if the platform offers a partial or full amount of data to the unaffiliated seller, then the targeting qualities of the two sellers are less differentiated, which leads to more intense price competition in the downstream market. However, such *intense price competition effects* dominate any *data-selling revenue effects* that come from extracting any additional revenues from the unaffiliated seller for data acquisition. Conversely, if the affiliated seller is weaker initially, then the platform is incentivized to provide exclusive data access to the unaffiliated seller to enjoy greater data-selling revenues; any data provided to the weaker affiliated seller will increase the affiliated seller's profits from the downstream competition, but this is dwarfed by the reductions in the willingness to pay for the data by the unaffiliated seller when it no longer has the exclusive access to the data.

Under any type of data neutrality, nondiscriminatory symmetric data acquisition is guaranteed by regulation. However, under *weak* data neutrality, which does not require nondiscriminatory data pricing, the platform can always bypass the regulation and preclude the unaffiliated seller from obtaining data access by charging a price higher than that unaffiliated seller's maximum willingness to pay. If weak data neutrality does not impose an additional regulation on data pricing, such as a price cap that guarantees the unaffiliated seller's data acquisition, it may have *de facto* no impact on the market. Therefore, we consider a stricter regulation, called *strong* data neutrality, that requires nondiscriminatory data pricing in addition to symmetric data provision. Indeed, as long as the marginal cost of data provision is sufficiently low, strong data neutrality guarantees that both sellers obtain the same amount of data at the same price.

However, weak and strong data neutrality that create a level playing field for the downstream sellers do not necessarily enhance welfare because the platform reduces the amount of data under the regulation: we show that the amount of data under no regulation is always greater than that under either weak or strong data neutrality. Due to the *data-reducing effect*, data neutrality does not always enhance consumer surplus despite its *competition-enhancing effect*. We show that, when the initial targeting quality between sellers is less differentiated, the data-reducing effect of data neutrality regulation is more likely to dominate the competition-enhancing effect, thus data neutrality is more likely to become welfare-reducing.

**Related Literature.** Data neutrality, as considered in this paper, is a new hypothetical economic regulation; thus, it has been mentioned in very few previous studies which define data neutrality as a broader form of net neutrality. Easley et al. (2018) consider data neutrality in the context of a software platform, such as a search platform or operating system, in which certain gatekeepers control the flow of information between content providers and consumers. Similarly, Krämer and Schnurr (2018) focus on platform neutrality in general. Several others (Grace and Leskovich, 2015; Pylon Network, 2018; Whitfield, 2017) mention data neutrality in the sense of open and neutral databases for easy and transparent data exchange. However, those papers consider data neutrality in a different context from ours, as we focus on equal data treatment between the platform and sellers (or content providers) and its impacts on the market. Our paper contributes to the emerging literature on data sharing and platform regulation by studying how a nondiscriminatory data-access rule operates in a vertically related market.

The recent literature on open data sharing policy, which shares the core concept with data neutrality, is the most closely related to our study. Martens and Zhao (2021) and Martens et al. (2020) emphasize the effectiveness of mandatory data sharing in terms of economies of scope in data aggregation. Borgogno and Colangelo (2019) note that strong technical and legal support is necessary for the systematic adoption of data sharing. Graef et al. (2019) analyze how the competitive structure, data protection, and consumer laws affect data sharing incentives from a legal perspective. Gu et al. (2022) show that incentives for sharing private data between platforms depend on the nature of the data, such as whether the data are substitutable or complementary. Krämer and Shekhar (2025) show that data-sharing obligations can increase welfare, whereas data-siloing constraints may reduce it. Our study differs from the above-mentioned studies insofar as we consider how a nondiscriminatory data access policy should be designed to be effective as intended by analyzing different types of regulation, called *weak* and *strong* data neutrality. Furthermore, by showing that data neutrality, which creates a level playing field for downstream players, reduces the amount of data the platform wants to offer sellers, we find that a non-discriminatory policy designed to be welfare-enhancing may result in unintended consequences, which has been neglected in the recent literature.

In addition to being broadly related to the literature on the general effect of data availability on downstream market structure or competition (Campbell et al., 2015; Kim, 2024; Goldfarb and Tucker, 2011; Taylor, 2004; Taylor and Wagman, 2014), our work is related to the study of data-related strategies of firms and their impact on relevant market competition or consumers (Bergemann and Bonatti, 2015; Casadesus-Masanell and Hervas-Drane, 2015; De Cornière and Nijs, 2016; Ichihashi, 2020; Ichihashi, 2021; Kirpalani and Philippon, 2020; Montes et al., 2019; Esteves and Carballo-Cruz, 2023). Going one step further, our model takes into account the unique features of data as non-rival but excludable goods. In this regard, two previous papers use a framework setup similar to ours: Jones and Tonetti (2020) find that consumer, instead of firm, ownership generates a more socially optimal data allocation because consumers who balance privacy concerns against gains to data sharing can sell the same set of data as non-rival goods to multiple firms, and Ali et al. (2024) consider the non-rival characteristics of data in the context of data resale and find that a monopolistic data seller with significant market power has less incentive to monetize data unless data reselling is prohibited. We show that, if we fail to take the non-rival feature of data into consideration and apply traditional models for rival goods to data-related markets, we may reach a misleading conclusion, which highlights the need for data-specific regulation.

Given that we model the upstream monopoly data supplier as providing input to downstream firms that compete with each other, our study is also broadly related to the literature on vertical integration and input foreclosure (Katz and Shapiro, 1986; Hart and Tirole, 1990; Kamien et al., 1992; O'Brien and Shaffer, 1992; McAfee and Schwartz, 1994). We distinguish our model from the above-mentioned studies by considering the particular aspects of data as non-rival goods, which results in production cost-reducing efficiency for the upstream platform. Indeed, our model allows us to investigate why we need a data-specific, non-discriminatory regulatory policy.

The remainder of this paper is structured as follows. In Section 2, we describe the model; in Section 3, we analyze the equilibrium outcomes under different data regulatory regimes; in Section 4, we describe the welfare effects of data neutrality regulations; in Section 5, we discuss the results under variations of the model and regulatory remedies; and finally in Section 6 we conclude. All the proofs are relegated to the Appendix.

# 2 A Model

## 2.1 Modelling Data

It is useful to first discuss how the platform aggregates data and how the data is defined in the model by focusing on two important features, namely, (*i*) how the data is valuable for the players of the game and (*ii*) what makes the data different from traditional goods.

**Data aggregation.** Depending on the role of the platform, it can collect different categories of data, such as location, browsing history, and financial data. Given the type of data $k \in \mathcal{K}$, where $\mathcal{K}$ is a finite set of all available data, each data type is denoted as $d_k$, e.g. $d_l$ as location and $d_b$ as browsing history data. The total aggregated amount of data for the platform is denoted as $\mathcal{D} = \cup_{k \in \mathcal{K}} d_k$, which is normalized to one. The platform sells a subset of the total aggregated data, which is denoted as $D \subseteq \mathcal{D}$, to sellers. Our focus is to examine how the platform's data production, i.e., "refinement" in the language of Varian (2018), is affected by data neutrality regulations.

**Value of data.** For downstream sellers, data obtained from the platform can be used for many different purposes. Sellers can use data to predict consumers' willingness to pay, thereby engaging in personalized pricing for the products they sell. Alternatively, data can be used to send better-matched targeted ads to consumers. More data on consumers' current wants or needs can help sellers offer more suitable products, which allows them to raise product prices. In this paper, we focus on the latter use of data and assume that sellers do not engage in price discrimination for products and that consumers benefit from better-matched targeting.

**Data as non-rival but excludable goods.** Data are distinguishable from traditional goods in several ways. The most salient difference is that the platform can sell the same set of data to many downstream sellers, which means that data are non-rival in consumption, as noted in Bourreau and de Streel (2019), Jones and Tonetti (2020), and Ali et al. (2024). As in the case of public goods, the non-rival feature of data can generate inefficiency absent policy intervention if the social value of providing data exceeds the private value. In our model, we also focus on the non-rivalry of data, which makes the analysis different from that for markets with rival goods: given that the platform technically sells the same product, i.e., aggregated data, to many sellers at the same time, the marginal cost of providing data is incurred only once and does not increase in the number of data buyers.

However, data are not pure public goods because they are excludable. In the absence of regulation, the platform as a data intermediary can offer exclusive access to a specific seller. Specifically, our concern is when the platform provides such exclusive data access to its vertically affiliated product seller, which provides a considerable competitive advantage. Given the nonrivalry of data, exclusive data access may cause greater harm to consumers.

By incorporating these unique features of data as nonrival but excludable goods into the model, we first examine how different types of data neutrality regulations, for example, *weak* and *strong* data neutrality, have different impacts on the market compared to other existing vertical regulations for rival goods: under *weak* data neutrality, the amount of data to be sold should be nondiscriminatory across sellers regardless of their vertical affiliation with the platform, whereas under *strong* data neutrality, in addition to the amount of data, data price should be nondiscriminatory.

## 2.2 Players

There are three types of players in the model. The first is a monopolistic platform as a data intermediary that collects raw data about consumers from the services it provides to consumers (e.g., social media or search engines), and then processes the raw data into a data product and monetizes it. The second type is the two downstream sellers, denoted $j \in \{1, 2\}$, who compete for consumers on the platform. They may obtain/purchase the data product from the platform (i.e. the data intermediary) in order to send targeted ads or to better customize products to consumers. One of the downstream sellers, say seller 1 without loss of generality, is affiliated with the platform, while seller 2 is independent. The third type is a continuum of heterogeneous consumers with unit measure who may choose to purchase up to one product from either seller 1 or 2. We assume that consumers are also active users of the platform, and as a result the platform has their raw data.

**Consumers.** There is a unit measure of potential consumers. Consumer $i \in [0, 1]$ receives a heterogeneous valuation $\theta_i$ from a generic product, and we assume that $\theta_i$ is independent and identically drawn from $U[0, \bar{\theta}]$, where $\bar{\theta} \geq 1$.

Sellers' products differ in quality. The initial quality of seller $j$'s product is denoted by $\bar{q}_j > 0$.[8] However, the final quality of seller $j$'s product can be enhanced by better targeting or better customization of the product variety to the consumer's wants, if the seller has access to the data product from the platform. Specifically, we assume that the surplus that a consumer of valuation type $\theta_i$ receives when she purchases from seller $j$ is given by:

$$u_{ij} = \theta_i \bar{q}_j (1 + d_j^{\alpha}) - p_j, \tag{1}$$

where $p_j$ denotes the product price provided by seller $j$; $d_j$ is the amount of data that seller $j$ uses, and

$$q_j(d_j; \bar{q}_j, \alpha) :\equiv \bar{q}_j (1 + d_j^{\alpha}), \tag{2}$$

denotes the quality level of the seller $j$ with the data $d_j$ given the initial targeting technology $\bar{q}_j$, where $\alpha \in [0, 1]$ can be interpreted as a measure of how data contribute to the improvement of product quality through better matching or better customization.

A consumer with unit demand chooses one seller by comparing the net utility gained from the two sellers and with the outside option of not purchasing at all. Note that the market may not be fully covered: if a consumer's valuation $\theta_i$ is too low, she chooses to stay out of the market; thus, to the extent that data can enhance consumers' values from the products, it can also potentially expand the market.

**Downstream sellers.** We denote the market share for seller $j$ in the downstream market as $x_j$, which depends on the price vector $\mathbf{p} = (p_1, p_2)$, data vector $\mathbf{d} = (a_1 d_1, a_2 d_2)$, where $a_j$ is 1 if seller $j$ buys data, or 0 otherwise. Given the data purchasing menus the platform offers to the two downstream sellers, $\langle (d_1, r_1), (d_2, r_2) \rangle$, where $(d_j, r_j)$ are, respectively, the data quantity and per-unit data price the platform offers to seller $j \in 1, 2$, seller $j$ solves the following profit maximization problem:

$$\max_{\{a_j, p_j\}} \pi_j = p_j x_j(\mathbf{p}, \mathbf{d}) - r_j a_j d_j, \tag{3}$$

where we assume that the marginal cost of producing the product is 0. Note that sellers are competing for market share using both the price and the data instruments, because data can increase the quality of the seller's product as in (2).

**Platform.** The platform earns profits from selling the data products to downstream sellers; in addition, as we assumed without loss of generality that the platform is affiliated with seller 1, it also includes seller 1's profits in the downstream market. The asymmetric affiliation assumption permits us to explore how such asymmetry affects the platform's incentives in its

[8] If seller $j$ has many similar varieties of products, one can also interpret $\bar{q}_j$ as the average quality of a randomly selected variety from seller $j$'s lineup of products.

data-related strategy, i.e. data production and provisions to the downstream sellers, and how such incentives are affected by potential data neutrality regulations (see below). The profit maximization problem for the platform is as follows:

$$\max_{\{d_1,d_2,r_1,r_2\}} \pi_p = \underbrace{\sum_{j=1}^{2} r_j a_j d_j - c \times \max\{a_1 d_1, a_2 d_2\}}_{\text{Data sales profit}} + \underbrace{p_1 x_1(\mathbf{p}, \mathbf{d}) - r_1 a_1 d_1,}_{\text{Downstream market profit}}, \tag{4}$$

where $c > 0$ is the marginal cost of producing data products from raw data. The platform's profit includes both the profit from data sales and the downstream market profit from its affiliated seller 1. Note that in (4), $a_j$ is seller $j$'s data purchase choice and is a function of the platform's choice of data purchasing menu $\langle (d_1, r_1), (d_2, r_2) \rangle$; also note that the cost function for data production is $c \times \max\{a_1 d_1, a_2 d_2\}$, which captures the non-rivalry nature of data as described in Section 2.1 in that the total cost of the data production depends on the *maximum* quantity of data purchased by the two sellers. We assume that the marginal cost $c$ is sufficiently small so that the platform has an incentive to provide data to at least its affiliated seller.

**Remark 1.** *Since the platform and seller 1—its affiliated downstream seller—are vertically integrated, the platform could dictate the profit maximization decisions for seller* 1 *by choosing both* $d_1, r_1$ *and in the downstream market,* $p_1$*. To ease the exposition, however, we will instead assume that seller 1 is an independent profit maximizer in its competition against the unaffiliated seller 2; of course, the platform, as the upstream player, can always choose the data purchasing menu* $\langle (d_1, r_1), (d_2, r_2) \rangle$ *such that the optimal choices of seller 1 as an independent seller would also maximize the platform's overall profit as given by (4).*

**Timing.** The timing of the game is as follows:

1. The monopolistic platform first decides on the amount of data $D$ to produce;
2. The monopolistic platform then decides on the data purchase menus for the two downstream sellers: $\langle (d_1, r_1), (d_2, r_2) \rangle$;
3. The sellers decide whether to acquire data ($a_j = 1$) or not ($a_j = 0$) from the platform;
4. Both sellers set the product price simultaneously, which is denoted as $p_j$. Using the obtained consumer data, each seller sends targeted ads or makes customized products to attract consumers;
5. A consumer chooses a seller to buy the product, and the market share of each seller $j$, $x_j$, is determined.

### 2.3 Data Neutrality Regulations

We consider three scenarios with respect to how the data market is regulated. The first scenario is *no data neutrality* regulation whatsoever, which corresponds to the status quo and serves as the benchmark. In the absence of any data regulation, the platform can discriminate against the unaffiliated seller in both data price and quantity in the sense that the platform can provide different amounts of data to the two sellers, i.e., $d_1 \neq d_2$, at different prices, i.e., $r_1 \neq r_2$.

Then we consider two scenarios with some regulations on data neutrality, which we refer to as *weak* and *strong* data neutrality, respectively. A weak data neutrality regulation requires that the platform make the same amount of data available to the unaffiliated seller 2 and the affiliated seller 1, that is, $d_1 = d_2 = D$; data prices to the two sellers, however, are allowed to differ, that is, $r_1 \neq r_2$, except for the restriction that the data price charged to the unaffiliated seller 2, $r_2$, must not exceed seller 2's maximum willingness to pay. A strong data neutrality regulation requires that the platform provides the same set of data (i.e., $d_1 = d_2 = D$ where $D$ is an endogenous choice of the platform) and charges the same unit price (i.e., $r_1 = r_2 = r$) in its data purchase menu.

# 3 Equilibrium

In this section, we analyze the equilibrium of the model under different data neutrality regulations. We first characterize the demand for each seller when the two sellers' data purchases are $(d_1, d_2)$ and choose prices $(p_1, p_2)$. The choice of a consumer of type $\theta_i$ is given by

$$\max\{u_{i1}, u_{i2}, 0\},$$

where $u_{ij}, j \in \{1, 2\}$ is specified by (1), and 0 is the value from not purchasing from either seller. The market share for seller $j$ depends on the relative price-quality ratio, $\frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$, between the two downstream sellers. Lemma 1 characterizes the market share:

**Lemma 1.** *Let $j \neq k$ index the two sellers, and suppose that $\mathbf{d} = (d_j, d_k)$ are such that $q_j(d_j; \bar{q}_j, \alpha) > q_k(d_k; \bar{q}_k, \alpha)$. Then the equilibrium prices in the product market competition between the two sellers, $\mathbf{p} = (p_k, p_j)$, must satisfy $0 < \frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} < \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$; as a result, the market is partially covered, and the demand faced by each seller is given by*

$$x_j(\mathbf{p}, \mathbf{d}) = \frac{1}{\bar{\bar{\theta}}}\left(\bar{\theta} - \theta^H\right); \quad x_k(\mathbf{p}, \mathbf{d}) = \frac{1}{\bar{\bar{\theta}}}\left(\theta^H - \theta^L\right), \tag{5}$$

*where*

$$\theta^H = \frac{p_j - p_k}{q_j(d_j; \bar{q}_j, \alpha) - q_k(d_k; \bar{q}_k, \alpha)}; \quad \theta^L = \frac{p_k}{q_k(d_k; \bar{q}_k, \alpha)}. \tag{6}$$

Note that under the conditions stated in Lemma 1, $\theta^H$ is the marginal valuation type that makes a consumer indifferent between buying from sellers $j$ and $k$, and $\theta^L$ is the marginal valuation type that makes a consumer indifferent between buying from the lower-quality seller (i.e., seller $k$) and not buying. Figure 1 illustrates the market demand described in Lemma 1.

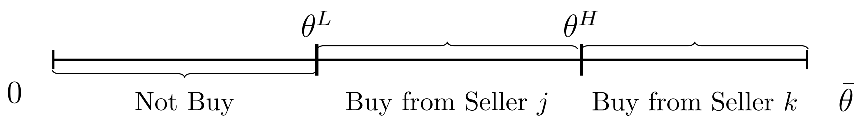


Figure 1: Illustration of Market Demand in Lemma 1.

With the demand function characterized in Lemma 1, the sellers' pricing decisions given data profile $\mathbf{d} = (a_1 d_1, a_2 d_2)$, where $a_j \in \{0, 1\}$ are determined as the solution to the following first-order conditions:

$$\zeta_{p_j} \equiv x_j(\mathbf{p}, \mathbf{d}) + p_j(\mathbf{d}) \frac{\partial x_j(\mathbf{p}, \mathbf{d})}{\partial p_j} = 0, j \in \{1, 2\}. \tag{7}$$

Below, we will analyze the full equilibrium of the model under the three scenarios of data neutrality regulations. To simplify notation, we normalize $\bar{q}_2 = 1$ and denote $\bar{q}_1 = \bar{q}$. In the main analysis, we analyze the case of $\bar{q} > 1$, which implies that the platform is affiliated with an initially strong seller. The case of $\bar{q} < 1$, which implies that the platform is affiliated with an initially weaker seller, is examined in Section 5.1.

## 3.1 No Data Neutrality

In the benchmark case without data neutrality regulations, we can use the system of first-order conditions described in (7) to solve for the equilibrium prices in the downstream market, for a given data purchase profile $\mathbf{d} = (d_1, d_2)$, provided that $q_1(d_1; \bar{q}, \alpha) > q_2(d_2; 1, \alpha)$:[9]

$$p_1^*(\mathbf{d}) = \frac{2\bar{q}\bar{\theta}(1 + d_1^\alpha)[\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)]}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}; \quad p_2^*(\mathbf{d}) = \frac{\bar{\theta}(1 + d_2^\alpha)[\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)]}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}. \tag{8}$$

With the prices given above, the market shares as characterized by (5) simplify as follows (where, by some slight abuse of notation, we write them as functions of $\mathbf{d}$ only):

$$x_1^*(\mathbf{d}) = \frac{2\bar{q}(1 + d_1^\alpha)}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}; \quad x_2^*(\mathbf{d}) = \frac{\bar{q}(1 + d_1^\alpha)}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}. \tag{9}$$

[9]The details of the derivation are collected in Appendix B.

With the equilibrium price and market share characterized by (8) and (9), we can write the equilibrium profit of seller $j$ in the sub-game with the data profile $\mathbf{d} = (d_1, d_2)$ with $d_1 \geq d_2$ as:

$$\pi_j^*(\mathbf{d}) = p_j^*(\mathbf{d}) x_j^*(\mathbf{d}) - r_j d_j. \tag{10}$$

**Data prices.** Given the equilibrium of the downstream market in terms of $\mathbf{d}$, the platform first sets the optimal data prices. First, consider the data price $r_2$ for the unaffiliated seller 2. As is clear from the platform's profit function as described in (4), the platform's profit increases in $r_2$, provided that seller 2 purchases the data. Thus, the platform will set the maximum possible data price by extracting all rents from seller 2. The optimal $r_2$ is therefore the solution to

$$\pi_2^*(\mathbf{d}^{(1,1)}) = \pi_2^*(\mathbf{d}^{(1,0)}), \tag{11}$$

where the superscript denotes $\mathbf{a} = (a_1, a_2)$, e.g., $\mathbf{d}^{(1,1)} = (d_1, d_2)$ and $\mathbf{d}^{(1,0)} = (d_1, 0)$. The optimal $r_2(\mathbf{d})$ is given as follows.

$$r_2(\mathbf{d}) = \frac{\bar{q}^2 \bar{\theta} \left(d_1^\alpha + 1\right)^2 d_2^{\alpha-1} \left\{7(d_2^\alpha + 1) - 16\bar{q}(2 - \bar{q}) + 16\bar{q}\left\{d_1^\alpha \left[\bar{q}(d_1^\alpha + 2) - 2\right] - \left(d_1^\alpha + 1\right) d_2^\alpha\right\}\right\}}{\left[4\bar{q}\left(d_1^\alpha + 1\right) - 1\right]^2 \left[4\bar{q}\left(d_1^\alpha + 1\right) - \left(d_2^\alpha + 1\right)\right]^2}, \tag{12}$$

if

$$\bar{q} \geq \bar{q}^*(\mathbf{d}) \equiv \frac{2\left(d_2^\alpha + 2\right) d_1^\alpha + 2d_2^\alpha + \sqrt{\left(d_1^\alpha + 1\right)^2 \left[\left(4d_2^\alpha + 9\right) d_2^\alpha + 9\right]} + 4}{4\left(d_1^\alpha + 1\right)^2}; \tag{13}$$

and 0, otherwise.

In other words, the willingness of the unaffiliated seller 2 to pay for data from the platform is positive only if the initial targeting quality of seller 1 is sufficiently higher than that of seller 2, i.e., $\bar{q} \geq \bar{q}^*(\mathbf{d})$. Recall that we assume in this section that affiliated seller 1 has an advantage in its initial targeting quality ($\bar{q}_1 = \bar{q} > 1 = \bar{q}_2$), thus seller 1 enjoys an initial competitive advantage. If seller 2 acquires data from the platform, there are two countervailing effects: on the one hand, purchasing data can allow seller 2 to improve its targeting quality to improve its quality, and thus *ceteris paribus,* increase the competitiveness of its product relative to that of the affiliated seller 1; on the other hand, the improved product quality of seller 2 reduces the product differentiation between the two sellers, which leads to more intense price competition. If the affiliated seller's initial targeting quality is not too far ahead of that of the unaffiliated seller, i.e., $\bar{q} < \bar{q}^*(\mathbf{d})$, then the more fierce price competition effect dominates if the unaffiliated seller 2 acquires data from the platform; in contrast, if $\bar{q} \geq \bar{q}^*(\mathbf{d})$, then seller 2 gains more from closing the gap in quality with seller 1.

It can also be shown that $\frac{\partial r_2(\mathbf{d})}{\partial d_2} < 0$. The intuition for this is as follows. From (11), we know that

$$r_2(\mathbf{d}) = \frac{p_2(\mathbf{d}^{(1,1)}) x_2(\mathbf{d}^{(1,1)}) - p_2(\mathbf{d}^{(1,0)}) x_2(\mathbf{d}^{(1,0)})}{d_2} \equiv \frac{\Delta \text{Rev}}{d_2}.$$

We first check that $\frac{\partial \Delta \text{Rev}}{\partial d_2} > 0$. As seller 2 obtains more data, it earns more revenue than in the case of no data availability, due to the *market expansion effect* (see Crawford et al., 2022). A marginal increase in $d_2 > 0$ will lead some consumers who refuse to buy any product before the increase in $d_2$ to enter the market: as shown in Figure 2, the low marginal valuation $\theta^L$ becomes $\widetilde{\theta}^L$ as $d_2$ increases. This effect of market expansion is partly reduced by intense price competition: as $d_2$ increases, $p_1$ falls, which leads some of the customers of seller 2 to switch to seller 1; the *customer switching effect* is captured by the fact that $\theta^H$ becomes $\widetilde{\theta}^H$ as $d_2$ increases, as shown in Figure 2. Given that

$$\frac{\partial \theta^L}{\partial d_2} = -\frac{3\alpha\bar{q}\bar{\theta}\left(d_1^\alpha + 1\right)d_2^{\alpha-1}}{\left(d_2^\alpha - 4\bar{q}d_1^\alpha - 4\bar{q} + 1\right)^2}; \quad \frac{\partial \theta^H}{\partial d_2} = -\frac{2\alpha\bar{q}\bar{\theta}\left(d_1^\alpha + 1\right)d_2^{\alpha-1}}{\left(d_2^\alpha - 4\bar{q}d_1^\alpha - 4\bar{q} + 1\right)^2},$$

the net effect of $d_2$ on $x_2$ is positive, i.e., $\left|\frac{\partial \theta^L}{\partial d_2}\right| > \left|\frac{\partial \theta^H}{\partial d_2}\right|$. However, more $d_2$ results in less differentiated targeting between sellers, which intensifies price competition, placing downward pressure on $r_2$. We find that the negative effect of $d_2$ on $r_2$ outweighs the positive effect; thus, $\frac{\partial r_2(\mathbf{d})}{\partial d_2} = \frac{-\Delta\text{Rev} + \frac{\partial \Delta\text{Rev}}{\partial d_2}d_2}{d_2^2} < 0$.


Figure 2: Changes in Market Share as $d_2$ Increases.

Proposition 1 summarizes this discussion:

**Proposition 1.** *Fix the platform's data quantity offer profile* $\mathbf{d} = (d_1, d_2)$ *with* $d_1 > d_2$. *The unaffiliated seller 2 will purchase the data offer* $d_2$ *at a non-negative price, as specified by (12) only if the initial targeting quality of seller 1 is sufficiently higher than that of seller 2, i.e., only if* $\bar{q} \geq \bar{q}^*(\mathbf{d})$ *where* $\bar{q}^*(\mathbf{d})$ *is defined in (13); moreover,* $\partial r_2(\mathbf{d})/\partial d_2 < 0$.

For the internal data price $r_1(\mathbf{d})$, the platform–as an integrated entity–can choose $r_1(\mathbf{d})$ such that the optimal $d_1$ chosen by seller 1 also maximizes the total profit of the platform. Noting that $d_1 > d_2$, the platform's total profit as specified by (4) can be simplified as:

$$\pi_p(\mathbf{d}) = p_1(\mathbf{d})x_1(\mathbf{d}) + r_2(\mathbf{d})a_2 d_2 - cd_1. \tag{14}$$

Recall that seller 1's downstream profit is given by (10), thus if the platform sets

$$r_1(\mathbf{d}) = c - r_2(\mathbf{d})a_2\frac{d_2}{d_1}, \tag{15}$$

then the optimal amount of data chosen by seller 1 as an independent profit maximizer is the same as that chosen by the platform as a joint profit maximizer.

**Data quantities.** From Proposition 1, we know that the unaffiliated seller will purchase data $d_2$ offered by the platform at a positive price when seller 1 purchases $d_1$ only if $\bar{q} \geq \bar{q}^*(\mathbf{d})$. Given that $\bar{q} > 1$, affiliated seller 1 always acquires data, i.e., $a_1 = 1$. We first focus on the case of $\bar{q} \geq \bar{q}^*(\mathbf{d})$, in which case both sellers buy data from the data purchasing menu $\langle \mathbf{d}, \mathbf{r} \rangle$ offered by the platform, where $\mathbf{r} = (r_1(\mathbf{d}), r_2(\mathbf{d}))$ are given by (15) and (11). Given $\mathbf{r}$, the platform chooses the optimal $d_1$ and $d_2$ by solving the following system of equations (assuming an interior solution):

$$\frac{\partial \pi_p(\mathbf{d})}{\partial d_1} = \underbrace{\frac{\partial p_1(\mathbf{d})}{\partial d_1}}_{(+)} x_1(\mathbf{d}) + p_1(\mathbf{d}) \underbrace{\frac{\partial x_1(\mathbf{d})}{\partial d_1}}_{(-)} + \underbrace{\frac{\partial r_2(\mathbf{d})}{\partial d_1}}_{(+)} d_2 - c = 0; \tag{16}$$

$$\frac{\partial \pi_p(\mathbf{d})}{\partial d_2} = \underbrace{\underbrace{\frac{\partial p_1(\mathbf{d})}{\partial d_2}}_{(-)} x_1(\mathbf{d}) + p_1(\mathbf{d}) \underbrace{\frac{\partial x_1(\mathbf{d})}{\partial d_2}}_{(+)}}_{\text{Term 1: Intense competition effect } (-)} + \underbrace{\underbrace{\frac{\partial r_2(\mathbf{d})}{\partial d_2}}_{(-)} d_2 + r_2(\mathbf{d})}_{\text{Term 2: Data selling revenue effect } (?)} = 0. \tag{17}$$

We first consider the trade-offs captured in the first order condition (17) with respect to $d_2$. The first effect of marginally increasing $d_2$ is referred to as the *intense competition effect*, which captures the effects of the downstream market on the affiliated seller 1 when the unaffiliated seller is offered more data. When seller 2 obtains more data, i.e., when $d_2$ increases, the two sellers' targeting/customization capacities become less differentiated; thus, affiliated seller 1 faces downward pressure on its price. Although such downward pressure on $p_1$ leads some of seller 2's customers to switch to seller 1, the net effects of an increase in $d_2$ on seller 1's downstream revenue can be simplified as:

$$\text{Term 1} = \frac{4\alpha \bar{q}^2 \bar{\theta} \left(d_1^\alpha + 1\right)^2 d_2^{\alpha-1} \left[d_2^\alpha + 1 + 2\bar{q}\left(d_1^\alpha + 1\right)\right]}{\left[d_2^\alpha + 1 - 4\bar{q}\left(d_1^\alpha + 1\right)\right]^3}, \tag{18}$$

can be shown to be negative.[10] The second effect of marginally increasing $d_2$ is referred to as the *data selling revenue effect*, which reflects how much extra data sales revenue the platform can extract from the unaffiliated seller 2 by providing more data. As $d_2$ increases, the platform's data selling revenue increases by the unit data price $r_2(\mathbf{d})$, but it is mitigated by the opposite force that the maximal willingness to pay for unit data $r_2(\mathbf{d})$ is reduced, as we have shown in Proposition 1, i.e. $\partial r_2(\mathbf{d})/\partial d_2 < 0$ if $\bar{q} \geq \bar{q}^*(\mathbf{d})$. In general, the overall data selling revenue effect, which can be simplified as

$$\text{Term 2} = \frac{\alpha \bar{q}^2 \bar{\theta} \left(d_1^\alpha + 1\right)^2 d_2^{\alpha-1} \left[7(d_2^\alpha + 1) - 4\bar{q}(d_1^\alpha + 1)\right]}{\left[d_2^\alpha + 1 - 4\bar{q}(d_1^\alpha + 1)\right]^3}, \tag{19}$$

[10] See Appendix B for details.

has ambiguous signs. Indeed, inspecting (19) reveals that the data selling revenue effect is positive only when

$$\bar{q} > \frac{7(1+d_2^\alpha)}{4(1+d_1^\alpha)}.$$

That is, marginally increasing data sharing leads to higher data selling revenues only if the initial targeting qualities of the two sellers are sufficiently differentiated.

However, combining the two effects as indicated by (18) and (19), we can see that the total effect of increasing $d_2$ on the total profits of the platform is negative:

$$\frac{\partial \pi_p(\mathbf{d})}{\partial d_2} = \frac{\alpha \bar{q}^2 \bar{\theta} \left(d_1^\alpha + 1\right)^2 d_2^{\alpha-1} \left[11(d_2^\alpha + 1) + 4\bar{q}(d_1^\alpha + 1)\right]}{\left[d_2^\alpha + 1 - 4\bar{q}(d_1^\alpha + 1)\right]^3} < 0, \tag{20}$$

where the inequality follows from the fact that the denominator is negative, since $d_2 < d_1$ and $\bar{q} > 1$, and the numerator is positive. Thus, the optimal choice of the platform for the data offering to the unaffiliated seller 2 is 0. Now, if we denote the platform's optimal data quality menu with $\mathbf{d} = (d_1 = D, d_2 = 0)$, then $D$ is determined by the first condition (16):

$$\zeta_N \equiv \left.\frac{\partial \pi_p(\mathbf{d})}{\partial d_1}\right|_{d_1=D, d_2=0} = \frac{4\alpha \bar{q}^2 \bar{\theta} D^{\alpha-1} \left(D^\alpha + 1\right) \left\{\bar{q}\left(D^\alpha + 1\right)\left[4\bar{q}\left(D^\alpha + 1\right) - 3\right] + 2\right\}}{\left[4\bar{q}\left(D^\alpha + 1\right) - 1\right]^3} - c = 0. \tag{21}$$

**Proposition 2.** *In the absence of data neutrality regulations, if affiliated seller 1 has an advantage in the initial targeting quality over unaffiliated seller 2, i.e., if $\bar{q} > 1$, then in equilibrium the platform forecloses the unaffiliated seller from data access and grants the affiliated seller exclusive access. The equilibrium amount of data produced and offered to the affiliated seller under no data neutrality, denoted as $D_N$, solves (21).*

Given $D_N$, we can then calculate the prices and market shares of the two competing downstream sellers by plugging $\mathbf{d} = (D, 0)$ into (8) and (9):

$$p_1^N = \frac{2\bar{\theta}\bar{q}\left(1 + D_N^\alpha\right)\left[\bar{q}(1 + D_N^\alpha) - 1\right]}{4\bar{q}\left(1 + D_N^\alpha\right) - 1}; \quad p_2^N = \frac{\bar{\theta}\left[\bar{q}(1 + D_N^\alpha) - 1\right]}{4\bar{q}\left(1 + D_N^\alpha\right) - 1}; \tag{22}$$

$$x_1^N = \frac{2\bar{q}\left(1 + D_N^\alpha\right)}{4\bar{q}\left(1 + D_N^\alpha\right) - 1}; \quad x_2^N = \frac{\bar{q}(1 + D_N^\alpha)}{4\bar{q}\left(1 + D_N^\alpha\right) - 1}. \tag{23}$$

The platform's total profits are given by:

$$\pi_p^N = \frac{4\bar{q}^2\bar{\theta}\left(1 + D_N^\alpha\right)^2\left[\bar{q}(1 + D_N^\alpha) - 1\right]}{\left[4\bar{q}\left(1 + D_N^\alpha\right) - 1\right]^2} - c \times D_N. \tag{24}$$

Notice that, under no regulation, the platform produces data products, which is a non-rival good, for exclusive use by its affiliated seller. We refer to this phenomenon as *data foreclosure.*

Conditioning on the quantity of data produced, $D_N$, it is *ex post* socially inefficient for the non-rival and productive data products to be exclusively used by one seller only.[11]

### 3.1.1 Discussion

**Data foreclosure.** Why does the platform find it optimal to foreclose the data product from the unaffiliated seller in this model? The reason is that the value of the data in the competition in the downstream market depends on whether the competitors also have access to it. Selling the data to both sellers reduces the strategic value of the data in the downstream competition to either of them, thus their willingness to pay for the data; in the extreme, if both sellers have the same amount of data, the data does not provide much advantage to either seller!

**Irrelevance of Affiliation Status.** In Proposition 2, we show that the platform will provide data exclusively to the affiliated seller 1 when seller 1 enjoys an advantage in the initial targeting quality over the unaffiliated seller 2, i.e., if $\bar{q} > 1$. However, in Section 5.1, we show that when the platform is integrated with an initially weak seller, i.e., $\bar{q} < 1$, the platform will choose to offer exclusive data sales to the unaffiliated seller, because the value of exclusive access to the data to the unaffiliated seller, and thus its willingness to pay for the data, is higher than the lost downstream profits from the affiliated seller. In this sense, we show that the affiliation status is *irrelevant* to the platform's optimal data provision and pricing policies in the absence of data neutrality regulations.

**Competitive Data Intermediaries.** Our data foreclosure result in the absence of data neutrality regulations hinges on the data platform being a monopoly. If the data intermediary market is competitive and downstream sellers can obtain a similar set of data from other platforms, it is likely that the data intermediaries will eschew data foreclosure. We leave the analysis of competitive data intermediaries as a topic for future research. However, we argue that our data foreclosure result for the monopoly data intermediary case is practically relevant. Platforms in different sub-markets, e.g., social media or e-commerce, often collect different sets of data, as shown in a few previous studies (e.g., Maier, 2019; Graef, 2015; Stucke and Grunes, 2016). For example, the data available on Google Maps are mostly related to consumers' location information, while the data on Amazon (in the U.S.) or JD (in China) show consumers' search and purchase histories. Even within the same sub-market, each platform can strategically differentiate its own data product from others (e.g. Gu et al., 2019).

[11] As $\alpha$ and $\bar{\theta}$ are central primitives affecting both data value and downstream competition, we analyze their comparative statics as well; for brevity, the full discussion is relegated to Appendix C.

## 3.2 Weak Data Neutrality

There are already several regulations that encourage data sharing among firms, such as the EU General Data Protection Regulation (GDPR) and the 2020 EU Digital Market Act (DMA). The DMA requires that data sharing implemented by a platform must guarantee fair competition. In this section, we consider two types of hypothetical regulation, which we refer to as *weak* and *strong* data neutrality, respectively.

**Definition 1** (Weak Data Neutrality). *The data platform is required to sell the same amount of data to the downstream sellers regardless of their vertical affiliation status, i.e.* $d_1 = d_2 = D$*; moreover, the maximum data price the platform can charge to the unaffiliated seller 2 is capped at seller 2's maximum willingness to pay.*

Notice that, under weak data neutrality, the data intermediary is still allowed to charge asymmetric prices to the two sellers, i.e., $r_1 \neq r_2$.

It follows from Equations (8)-(9), when we substitute $d_1 = d_2 = D$, that the downstream equilibrium demand and prices of the two sellers are given as follows, where the superscript $W$ denotes weak data neutrality:

$$x_1^W = \frac{2\bar{q}}{4\bar{q}-1}; \quad x_2^W = \frac{\bar{q}}{4\bar{q}-1}; \quad p_1^W = \frac{2\bar{\theta}\bar{q}\left(\bar{q}-1\right)\left(1+D^{\alpha}\right)}{4\bar{q}-1}; \quad p_2^W = \frac{\bar{\theta}\left(\bar{q}-1\right)\left(1+D^{\alpha}\right)}{4\bar{q}-1}, \tag{25}$$

and the data platform's profit is given by

$$\pi_p^W(\mathbf{r}^W|D) = \frac{4\bar{\theta}\bar{q}^2\left(\bar{q}-1\right)\left(1+D^{\alpha}\right)}{(4\bar{q}-1)^2} + (r_2^W - c) \times D. \tag{26}$$

Recall that under weak data neutrality, the data intermediary can still charge differential prices to the affiliated and unaffiliated sellers, thus as before, the platform can set

$$r_1^W = c - r_2^W,$$

to ensure that the level of $D$ that solves seller 1's downstream profit maximization problem (10) coincides with the data level $D$ that solves the platform's joint profit maximization problem (4).

Now, taking the amount of data $D$ as given, the derivative of the platform's profit (4) with respect to $r_2$ is $D$, which is strictly positive. This implies that the platform will charge as high a price $r_2$ as possible to the unaffiliated seller so that it extracts all the extra profits that may accrue to seller 2 from acquiring $D$ amount of data. We denote the equilibrium $r_2$ under weak data neutrality as $r_2^W$, and $r_2^W$ is obtained from (12) by substituting $d_1 = d_2 = D$, which is given by

$$r_2^W(D) = \begin{cases} \dfrac{\bar{q}^2\bar{\theta}D^{\alpha-1}(1+D^\alpha)\{16\bar{q}[(\bar{q}-1)D^\alpha+\bar{q}-2]+7\}}{(4\bar{q}-1)^2[4\bar{q}(1+D^\alpha)-1]^2} & \text{if } \bar{q} \geq \bar{q}_W^*(D); \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

where

$$\bar{q}_W^*(D) \equiv \frac{2D^\alpha + \sqrt{4D^{2\alpha}+9(1+D^\alpha)}+4}{4(1+D^\alpha)}, \tag{28}$$

is obtained from $\bar{q}^*(\mathbf{d})$, which is defined in (13) by substituting $d_1 = d_2 = D$. Equation (27) shows that for any quantity of data $D > 0$ the willingness of the unaffiliated seller 2 to pay for the data is positive only if its initial targeting quality is sufficiently lower than that of the affiliated seller, i.e. $\bar{q} \geq \bar{q}_W^*(D)$. Moreover, we have

$$\frac{\partial \bar{q}_W^*(D)}{\partial D} = -\frac{\alpha D^{\alpha-1}\left(D^\alpha + 4\sqrt{(4D^\alpha+9)D^\alpha+9}+9\right)}{8(D^\alpha+1)^2\sqrt{(4D^\alpha+9)D^\alpha+9}} < 0, \tag{29}$$

which implies that as the quantity of data provision $D$ increases, the differentiation required in the initial targeting qualities between seller 1 and seller 2 is lower in order for seller 2 to have a positive willingness to pay for the data. The intuition for this result is that symmetric data acquisition by both sellers have countervailing effects on the unaffiliated seller: on the one hand, it intensifies the downstream price competition, which hurts its profits; on the other hand, it expands the market, which could increase its profit. When the quantity of data $D$ is low, the price competition effect dominates due to the disadvantage of seller 2's initial targeting quality relative to the affiliated seller 1; however, higher levels of $D$ strengthens the market expansion effect. Figure 3 depicts combinations of $(D, \bar{q})$ in which the unaffiliated seller 2 has a positive willingness to pay for the data provision when $\alpha = 1/2$ and $\bar{\theta} = 3/2$.

Plugging $r_2^W(D)$ as expressed by (27) into the platform's profit function under weak neutrality, as given by (26), and taking the derivative with respect to $D$, we have the following first order condition that characterizes the profit-maximizing data quantity $D_W$ under weak data neutrality:

$$\begin{aligned} \zeta_W \equiv \frac{\partial \pi_p^W(\mathbf{r}^W(D)|D)}{\partial D} &= \frac{\alpha\bar{q}^2\bar{\theta}D^{\alpha-1}}{(1-4\bar{q})^2[4\bar{q}(D^\alpha+1)-1]^3} \\ &\times \left\{(\bar{q}-1)(4\bar{q}+1)[4\bar{q}(D^\alpha+1)-1]^3 - (1-4\bar{q})^2[2\bar{q}(D^\alpha+1)+1]\right\} - c = 0. \end{aligned} \tag{30}$$

**Proposition 3.** *Consider a weak data neutrality regulation where the platform is required to offer the same amount of the data to the affiliated and non-affiliated sellers and that the price for the unaffiliated seller is capped at the unaffiliated seller's maximum willingness to pay. Suppose that the initial targeting quality of the affiliated seller $\bar{q}$ is large enough so that $r_2^W(D)$ as given by (27) is positive, then the equilibrium data quantity $D_W$ solves (29).*

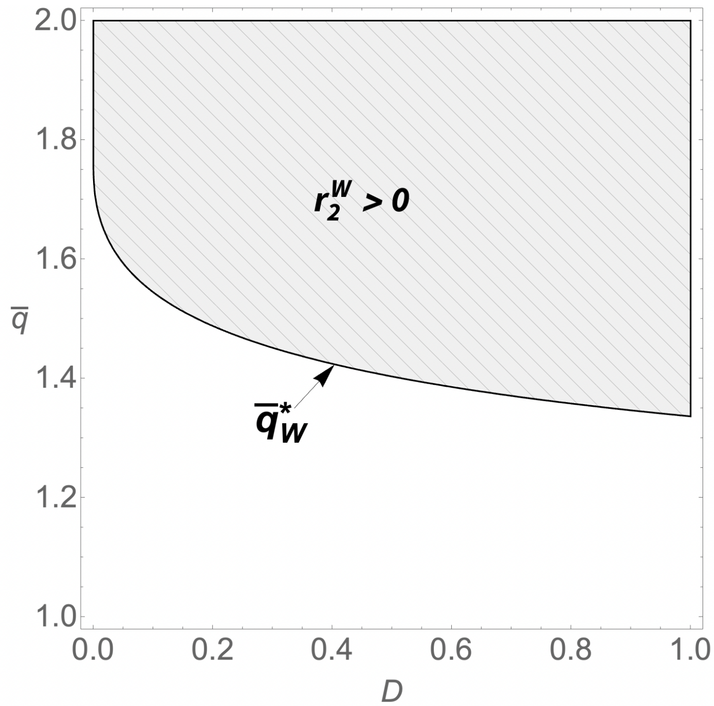


Figure 3: When Seller 2's Willingness to Pay for Data is Positive.

*Notes:* This figure depicts the combinations of $(D, \bar{q})$ in which the unaffiliated seller 2 has a positive willingness to pay for the data provision when $\alpha = 1/2$ and $\bar{\theta} = 3/2$.

As we will show in Section 3.4, the equilibrium amount of data under *weak* data neutrality is less than that under no regulation, even if the platform extracts all rent from the unaffiliated seller in the form of data price. As the platform offers more data to both sellers, i.e., as $D$ increases, the downstream market profit of the affiliated seller increases, but the revenue from the sale of data extracted from the unaffiliated seller decreases. However, if the affiliated seller has the advantage relative to the unaffiliated seller, i.e., when $\bar{q} > 1$, the latter dominates the former; as such, the platform prefers to offer a smaller amount of data under weak data neutrality in order to extract more data revenue from the unaffiliated seller.

**Remark 2.** *As shown in (27), if the initial targeting qualities between the affiliated and unaffiliated sellers are not sufficiently differentiated, i.e., if $\bar{q} < \bar{q}_W^*(D)$, then the unaffiliated seller 2's willingness to pay for the data is zero. In such a case, as long as the data price is non-negative, the equilibrium data acquisition is $\mathbf{d} = (D, 0)$, which is qualitatively the same outcome as that under no data neutrality.*

**Remark 3.** *When $\bar{q} \geq \bar{q}_W^*(D)$, i.e., when seller 2 is willing to pay a positive price to obtain data $D$, Proposition 2 states that the profit-maximizing decision for the platform without regulation is to monopolize the data by making them unavailable to the unaffiliated seller. Thus, if weak data neutrality were only to require that the platform offer the same quantity of data to both sellers, without prohibiting the platform from charging a price exceeding seller 2's maximum willingness* *to pay, the platform can simply set $r_2^W > r_2^W(D)$ where $r_2^W$ is given by (27), and restore the data foreclosure outcome under the no data regulation case. Thus, a weak data neutrality guarantees symmetric data acquisition only when the maximum data price for the unaffiliated seller 2 is also capped at seller 2's maximum willingness to pay.*

## 3.3 Strong Data Neutrality

We now consider a *strong* data neutrality, which requires that the platform provide the same amount of data at the same price to the two downstream sellers:

**Definition 2** (Strong Data Neutrality)**.** *The platform is required to treat the downstream sellers equally in both quantity and price charged for any data products, i.e.,* $d_1 = d_2 = D$ and $r_1^S = r_2^S = r$.

As stated in Section 2, we assume that affiliated seller 1 is an independent profit maximizer, chooses the optimal amount of data $D$ to maximize the downstream profit only, and the platform chooses the optimal price of the data $r$ to maximize the joint profits.

**Remark 4.** *If the platform were in charge of the operations (data acquisition and pricing decisions) of the affiliated seller 1, then it can always obtain the same equilibrium outcome as that in the weak data neutrality regime; that is, the strong and weak data neutrality regulations would have the same effect. To see this, note that the platform could simply set $r_1^S = r_2^S = r_2^W(D)$, as defined in (27). Given that $r_1^S$ is a pure internal transfer, any value of $r_1^S$ would result in the same profit for the platform, which is the sum of profits from the upstream data sales and the downstream market from the affiliated seller 1. Thus, whether a strong data neutrality regulation has more bite than a weak data neutrality depends on the affiliated seller being able to decide on its own operations.*

Under strong data neutrality $\mathbf{d} = (D, D), r_1^S = r_2^S = r$, the platform's profit (4) is

$$\pi_p(\mathbf{d}) = p_1(\mathbf{d})x_1(\mathbf{d}) + (ra_2 - c)D, \tag{31}$$

and the affiliated seller 1's profit (3) is given by

$$\pi_1(\mathbf{d}) = p_1(\mathbf{d})x_1(\mathbf{d}) - rD. \tag{32}$$

Inspecting the objective functions (31) for the platform and (32) for the affiliated seller 1, we have:

**Lemma 2.** *The profit maximizing data quantity $D$ for affiliated seller 1 coincides with that for the platform if the platform sets*

$$r_1^S = r_2^S = r = \frac{c}{1 + a_2}. \tag{33}$$

By comparing the price (33) with seller 2's maximum willingness to pay for data $D$, $r_2^W(D)$ as given by (27), it is clear that when the marginal cost $c$ is sufficiently low, we must have $r < r_2^W(D)$, which implies that seller 2 will accept the symmetric data offer at the price $r = \frac{c}{1+a_2}$, that is, $a_2 = 1$. Then, when the marginal cost of data production $c$ is sufficiently small, we must have:

$$r = \frac{c}{2}.$$

Thus, the profit-maximizing data quantity $D_S$ under strong data neutrality solves the following first order condition:

$$\zeta_S \equiv \frac{\partial \pi_1(\mathbf{d})}{\partial D} = \frac{\partial \pi_p(\mathbf{d})}{\partial D} = \frac{4\alpha(\bar{q}-1)\bar{q}^2\bar{\theta}D^{\alpha-1}}{(4\bar{q}-1)^2} - \frac{c}{2} = 0. \tag{34}$$

**Proposition 4.** *Consider a strong data neutrality regulation where the platform is required to offer the same amount of the data to the affiliated and non-affiliated sellers at the same price. Suppose that the marginal cost of data production $c$ is sufficiently small, then the equilibrium data quantity $D_S$ solves (34).*

## 3.4 Equilibrium Data Quantities

We now characterize how the equilibrium data quantities are affected by the data neutrality regulations by comparing the first order conditions for $D_N$ as in (21), for $D_W$ as in (29), and for $D_S$ as in (34).

First, we compare the equilibrium data quantity under no data neutrality and weak data neutrality. Taking the difference in the first order conditions (21) and (29), we have:

$$\zeta_N - \zeta_W = \frac{\alpha\bar{q}^2\bar{\theta}D^{\alpha-1}\Phi(D|\bar{q},\alpha)}{(4\bar{q}-1)^2\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^3}, \tag{35}$$

where

$$\begin{aligned}\Phi(D|\bar{q},\alpha) &= 11 + 22D^\alpha \\ &+ 4\bar{q}\Big\{16\bar{q}^2\left(D^\alpha+1\right)^3 + 4\bar{q}\left(D^\alpha+1\right)\left[\left(5D^\alpha+7\right)D^\alpha + 9 - \left(15D^\alpha+41\right)D^\alpha\right] - 21\Big\}.\end{aligned}$$

Through algebraic manipulation, it can be shown that $\Phi(D|\bar{q},\alpha) > 0$, which implies that $\zeta_N > \zeta_W$; given that the second order conditions with respect to $D$ are satisfied, this in turn implies that $D_N > D_W$.

Next, we compare the equilibrium data quantity $D_W$ under weak and $D_S$ under strong data neutrality. Again taking the differences in the first order conditions (29) and (34), we have:

$$\zeta_W - \zeta_S = \frac{\alpha \bar{q}^2 \bar{\theta} D^{\alpha-1} \widetilde{\kappa}(D|\bar{q},\alpha)}{(4\bar{q}-1)^2 \left[4\bar{q}\left(D^\alpha+1\right)-1\right]^3} - \frac{c}{2}, \tag{36}$$

where

$$\begin{aligned} \widetilde{\kappa}(D|\bar{q},\alpha) &= 4\bar{q}\left\{(12D^\alpha+31)\,D^\alpha + 16\bar{q}^2\left(D^\alpha+1\right)^3 + 15 - 4\bar{q}\left(D^\alpha+1\right)\left[(4D^\alpha+11)\,D^\alpha+9\right]\right\} \\ &- 7\left(2D^\alpha+1\right). \end{aligned}$$

Whether $\widetilde{\kappa}(D|\bar{q},\alpha)$ can be positive or negative depends on the size of $\bar{q}$; however, we can show that, if $\bar{q} > \bar{q}^*_W(D)$ where $\bar{q}^*_W(D)$ is defined in (28) above which the unaffiliated seller 2's willingness to pay for the data under the weak data neutrality regulation is non-negative, then $D_W > D_S$.

**Proposition 5** (Equilibrium Data Quantities under Data Neutrality Regulations)**.** *The equilibrium quantities produced by the platform under different data neutrality regimes are ranked as follows:* $D_N > D_W \geq D_S$.

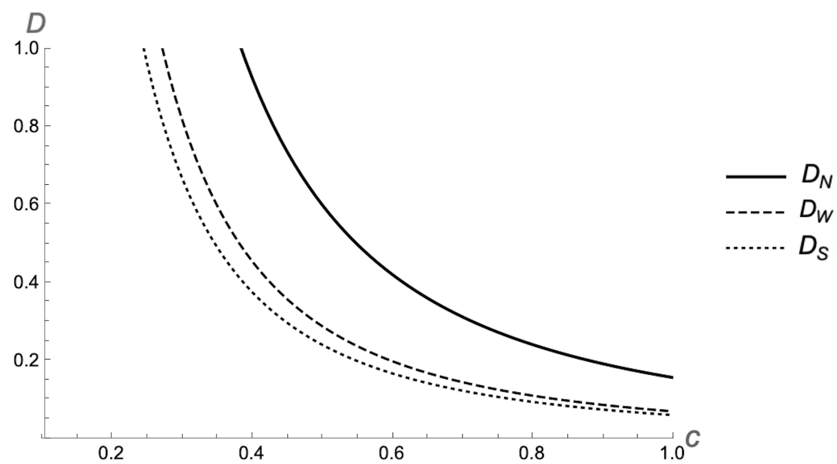


Figure 4: Equilibrium Data Quantities as a Function of $c$ under Different Regulatory Regimes: $\alpha = 1/2$, $\bar{\theta} = 3/2$, and $\bar{q} = 2$.

Figure 4 illustrates the rankings of the equilibrium data quantities under different data regulatory regimes as the marginal cost of data production varies, for fixed values of $\alpha = 1/2, \bar{\theta} = 3/2$ and $\bar{q} = 2$. Under data neutrality, the revenue of the affiliated seller decreases compared to the case without regulation in which the affiliated seller has exclusive data access (Proposition 2). When the platform is required to also sell the data to the unaffiliated seller under data neutrality, it does receive additional data sales revenues from the unaffiliated seller, but this effect is dominated by the profit loss suffered by the affiliated seller, as the affiliated seller faces more intense competition when the unaffiliated seller closes its initial disadvantage in targeting due to the data purchase from the platform. This trade-off drives the result that $D_W < D_N$. Moreover, the platform can extract more marginal rent from the unaffiliated seller 2 under weak data neutrality than under the strong data neutrality regulations, which implies that the platform has a higher incentive to provide data under weak data neutrality than under strong data neutrality.

# 4 Welfare Comparison

In this section, we assess the welfare impact of different data neutrality regulations. On the one hand, when the amount of data produced by the platform is *fixed,* data neutrality regulations would increase social welfare, as granting both sellers access to the data not only intensifies downstream market competition, but also improves the matching values of consumers, thus expanding market coverage; however, as Proposition 5 shows, data neutrality regulations dilute the incentives for the platform to produce data, and ceteris paribus, less data lower the matching value of buyers and could reduce social welfare. As such, the net welfare effect of data neutrality regulations hinges on the trade-offs between these countervailing forces: data neutrality regulation creates a level playing field for downstream sellers, but less data will be produced by the platform.

First, consumer surplus, denoted as $CS$, is derived as follows:

$$CS = \int_{\theta^L}^{\theta^H} [\theta q_2(d_2|\bar{q}_2, \alpha) - p_2] \frac{1}{\bar{\theta}} d\theta + \int_{\theta^H}^{\bar{\theta}} [\theta q_1(d_1|\bar{q}_1, \alpha) - p_1] \frac{1}{\bar{\theta}} d\theta, \tag{37}$$

where $\theta^L = \frac{p_2}{q_2(d_2|\bar{q}_2,\alpha)}$ and $\theta^H = \frac{p_1 - p_2}{q_1(d_1|\bar{q}_1,\alpha) - q_2(d_2|\bar{q}_2,\alpha)}$ denote two marginal consumer types who are indifferent between buying from seller 2 and staying out of the market and indifferent between buying from sellers 1 and 2, respectively.

The total social welfare, which is denoted as $SW$, is defined by the sum of the consumer surplus, the profit of the platform (which includes the profit of the affiliated seller) and the profit of the unaffiliated seller:

$$SW = CS + \pi_p + \pi_2. \tag{38}$$

Given that $q_j(d_j|\bar{q}_j, \alpha) = \bar{q}_j(1 + d_j^\alpha)$ and $\bar{q} \equiv \bar{q}_1 > \bar{q}_2 = 1$, where $\bar{q} \geq \bar{q}^*$, such that seller 2 always accepts the symmetric data offer, i.e., $a_2 = 1$, we compare $CS_k$, where $k \in \{W, S\}$, with $CS_N$, and obtain:

$$CS_k - CS_N = \frac{1}{2}\bar{q}^2\bar{\theta}^2 \left\{ \frac{(4\bar{q}+5)\,(D_k^\alpha + 1)}{(4\bar{q}-1)^2} - \frac{(D_N^\alpha + 1)^2\,[4\bar{q}\,(D_N^\alpha + 1) + 5]}{[4\bar{q}\,(D_N^\alpha + 1) - 1]^2} \right\}. \tag{39}$$

Compared to no regulation, data neutrality makes the two sellers' overall targeting quality more similar and induces more intense price competition, thereby enhancing consumer welfare. This competition-enhancing effect can be captured by the fact that $CS_S = CS_W > CS_N$ if $D_S = D_W = D_N \equiv D$: as long as the regulation does not change the amount of data available in the market, consumers are better off under data neutrality because it leads to fiercer price competition. Due to these, which we will refer to as the *competition-enhancing effect* with a fixed amount of data $D$, the consumer surplus under data neutrality, weak or strong, is greater than in the absence of regulation, as shown in Figure 5.

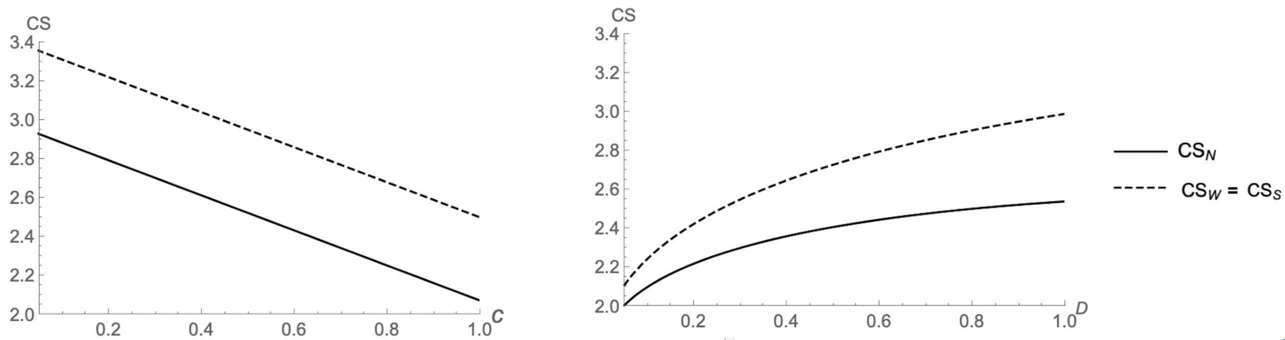


Figure 5: Consumer Surplus under an Exogenous Quantity of Data $D$.

*Notes:* The left panel varies $c$ holding $D = 1$; the right panel varies $D$ holding $c = 1/2$, assuming that $\bar{q} = 2$, $\alpha = 1/2$, and $\bar{\theta} = 3/2$.

In contrast, according to Proposition 5, data neutrality reduces the amount of available data in the market, which deteriorates the overall quality of the targeting, thereby reducing consumer welfare: given $D_N$, as $D_W$ decreases, the gap between the two surplus levels narrows. If the gap between $D_W$ and $D_N$ is sufficiently large, data neutrality may make consumers worse off. Due to these conflicting effects, namely the competition-enhancing effect and the *data-amount-reducing effect*, the net welfare effect is therefore ambiguous. Comparing social welfare levels yields qualitatively identical results.

**Proposition 6.** *Whether data neutrality is welfare-enhancing or welfare-reducing compared to no regulation is ambiguous and depends on the relative size of the competition-enhancing effects and data-amount-reducing effects.*

This result has a direct counterpart in the literature on platform data-sales strategies (e.g. Braulin and Valletti, 2016; Montes et al., 2019). In particular, Braulin and Valletti (2016) conclude that selling data to both downstream firms, which can occur under data neutrality in our context, is the first-best result, but never prevails in equilibrium in their setup. A critical difference is that we endogenize the optimal amount of data to offer by solving the platform's profit maximization problem. As summarized in Proposition 6, if we do not consider the data-amount-reducing effect of data neutrality regulations, a well-intentioned nondiscriminatory regulation can potentially backfire against consumers.

The relative size of the two conflicting effects depends on how the two sellers are initially differentiated in terms of targeting, which is represented by $\bar{q}$. As $\bar{q}$ approaches one, making the two sellers more symmetric in terms of initial targeting quality, the equilibrium data quantity $D$ decreases, countervailing against any additional consumer surplus obtained from the competition-enhancing effect. This downward pressure on the amount of data is greater under data neutrality because of the more intense competitive structure. Proposition 7 summarizes this finding.

**Proposition 7.** *Compared to no regulation, data neutrality is more likely to reduce the amount of data in the market, when the initial targeting quality between the two sellers is less differentiated.*

As the initial targeting qualities of the downstream sellers become similar, the data-amount-reducing effect of data neutrality regulations becomes more severe, potentially reducing consumer surplus and even social welfare. This is summarized in Claim 1:

**Claim 1.** *Data neutrality is more likely to reduce welfare relative to no regulation, when the initial targeting quality between the two sellers is less differentiated*

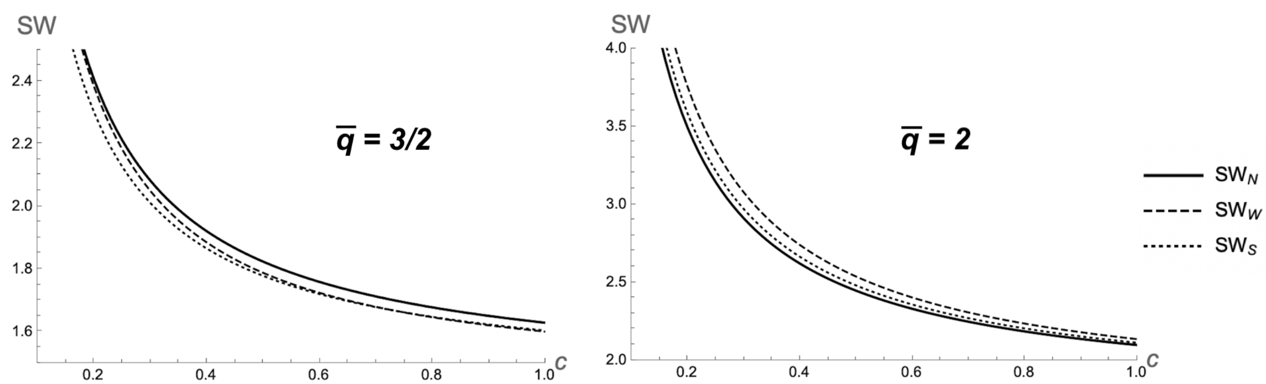


Figure 6: Social Welfare as a Function of $c$ under Different Regulatory Regimes: $\alpha = 1/2$ and $\bar{\theta} = 3/2$.

*Notes:* The left panel: $\bar{q} = 3/2$; the right panel: $\bar{q} = 2$.

Figure 6 provides a graphical example of Claim 1. Claim 1 is similar to the findings in Kirpalani and Philippon (2020). They also show that data sharing, which improves the matching between buyers and sellers, is not unambiguously good, but in their paper the countervailing effect is that better data also strengthens the platform's bargaining power with sellers, which could discourage seller entry and reduce consumer welfare. They thus also argue that regulations that limit information sharing can be potentially welfare-enhancing. Our welfare result complements theirs in that we identify a new mechanism by focusing on the platform's data-selling strategies, i.e., the data-amount-reducing effect, through which welfare can be reduced by data neutrality regulations.

Additionally, according to Proposition 5, $D_W$ is larger than $D_S$, which makes $SW_W$ greater than $SW_S$ because the only difference between $SW_W$ and $SW_S$ arises from differences in the equilibrium amount of data. Thus, as $\bar{q}$ decreases, no regulation produces the highest social welfare, while strong data neutrality produces the lowest.

In summary, when designing data policies, it is important that policymakers take into account the initial competitive structure of the market. If the downstream sellers provide less differentiated products to begin with, e.g., possess similar initial targeting technology in our context, weak or strong data neutrality regulations, which aim to symmetrically improve product qualities for both sellers and to induce more intense downstream competition, may ultimately reduce social welfare because the platform will respond to such data neutrality regulations by reducing the amount of data provided in equilibrium.

# 5 Discussion and Extensions

In this section, we discuss the robustness of our main results and consider several extensions.

## 5.1 Platform's Affiliation with an Initially Weaker Seller: $\bar{q} < 1$

In this section, we consider the case of $\bar{q} < 1$ to see how the equilibrium results are altered if the platform's affiliated seller is initially weaker than the competitor, i.e. $\bar{q} < 1$. To guarantee the interior solution in this case, we assume that $\bar{q} < \frac{1}{1+D^\alpha}$. Note that under the interior solution condition, $q_1(d_1|\bar{q}_1, \alpha) < q_2(d_2|\bar{q}_2, \alpha)$ is also guaranteed. Using Lemma 1, we find the corresponding market share for each seller. The remaining steps are the same as those in the main analysis. The optimal price and market share for the case of $\bar{q} < 1$ are given as follows.

$$\begin{aligned} &\widehat{p}_1(\mathbf{d}) = \frac{\bar{q}\bar{\theta}(1+d_1^\alpha)[(1+d_2^\alpha) - \bar{q}(1+d_1^\alpha)]}{4(1+d_2^\alpha) - \bar{q}(1+d_1^\alpha)}; \quad \widehat{p}_2(\mathbf{d}) = \frac{2\bar{\theta}(1+d_2^\alpha)[(1+d_2^\alpha) - \bar{q}(1+d_1^\alpha)]}{4(1+d_2^\alpha) - \bar{q}(1+d_1^\alpha)}; \\ &\widehat{x}_1(\mathbf{d}) = \frac{1+d_1^\alpha}{4(1+d_2^\alpha) - \bar{q}(1+d_1^\alpha)}; \quad \widehat{x}_2(\mathbf{d}) = \frac{2(1+d_1^\alpha)}{4\bar{q}(1+d_1^\alpha) - (1+d_2^\alpha)}, \end{aligned} \tag{40}$$

where $\widehat{\ }$ denotes the case of integration with an initially weak seller. For the external data price $\widehat{r}_2(\mathbf{d})$, the platform's profit increases in $\widehat{r}_2(\mathbf{d})$ such that it sets the maximum possible data price by extracting all rents from seller 2. The optimal $\widehat{r}_2(\mathbf{d})$ is given as follows.

$$\widehat{r}_2(\mathbf{d}) = \frac{4\bar{\theta}\left\{(d_2^\alpha+1)^2[\bar{q}(d_1^\alpha+1)-4]^2[d_2^\alpha+1-\bar{q}(d_1^\alpha+1)] + [\bar{q}(d_1^\alpha+1)-1][4(d_2^\alpha+1)-\bar{q}(d_1^\alpha+1)]^2\right\}}{d_2[\bar{q}(d_1^\alpha+1)-4]^2[4(d_2^\alpha+1)-\bar{q}(d_1^\alpha+1)]^2}, \tag{41}$$

which is always positive given $\bar{q} < \frac{1}{1+D^\alpha}$.

The platform sets the internal data price $\widehat{r}_1(\mathbf{d})$ to $\frac{\widehat{r}_2(\mathbf{d})-c}{d_2}$ such that the optimal $d_1$ chosen by seller 1 as an independent profit maximizer is the same as that chosen by the platform as a joint profit maximizer. Suppose $d_1 \equiv \beta D$ and $d_2 \equiv D$, where $\beta \in [0,1]$ and $D \in \mathbb{R}^+$. Then, seller 1 and the platform choose the optimal $\beta$ and $D$ by solving the following system of equations:

$$\begin{aligned} \frac{\partial \pi_1(\mathbf{d})}{\partial \beta} &= \frac{\alpha \bar{q}\bar{\theta}\,(D^\alpha+1)^2\,(\beta D)^\alpha\,\{4\,(D^\alpha+1)+11\bar{q}\,[(\beta D)^\alpha+1]\}}{\beta\,\{-4(D^\alpha+1)+\bar{q}[(\beta D)^\alpha+1]\}^3}; \\ \widehat{\zeta}_N \equiv \frac{\partial \pi_p(\mathbf{d})}{\partial D} &= \frac{\bar{\theta}\alpha\widehat{\Lambda}(\mathbf{d})}{D} - c, \end{aligned} \tag{42}$$

where $\widehat{\Lambda}(\mathbf{d})$ is defined in Appendix A. Clearly, $\frac{\partial \pi_1(\mathbf{d})}{\partial \beta} < 0$, given that its denominator is negative; thus, in equilibrium, the affiliated seller 1 will not receive any data. We confirm that $\widehat{\Lambda}(\mathbf{d}^{(0,1)})$ is positive; thus, the optimal $\widehat{d}_2^N = \widehat{D}_N$ is determined by $\widehat{\zeta}_N(\mathbf{d}^{(0,1)}) = 0$. That is, if the platform is affiliated with an initially weak seller, it does not have a data foreclosure incentive, even without any regulation, unlike in the case of affiliation with an initially strong seller. Furthermore, the platform earns more profit when selling data exclusively to the unaffiliated seller than when selling to both sellers non-exclusively. Thus, if $\bar{q} < 1$, the platform sets $r_2$ such that the unaffiliated seller 2 is willing to buy data, which results in $\widehat{\mathbf{a}}^N = (0,1)$.[12]

**Proposition 8.** *Suppose that the platform is affiliated with an initially weaker seller, i.e.,* $\bar{q} < 1$. *In the absence of data neutrality regulation, the platform grants the unaffiliated seller exclusive data access and does not provide data to its affiliated seller. The unaffiliated seller always accepts the exclusive data offer.*

If $\bar{q} < 1$, the overall profit of the platform is smaller when it monopolizes access to data to its affiliated seller, because the revenue from selling the data exclusively to the unaffiliated seller exceeds the additional profit of its affiliated seller in the downstream competition. Due to its initial disadvantage in targeting, the effect of data foreclosure on downstream market dominance is relatively small. On the other hand, when selling data to a high-quality unaffiliated seller, the platform is able to extract more rent; thus, its data selling revenue is sufficiently large. In addition, selling data to both makes the two sellers more symmetric, leading to more intense price competition, which harms the initially weak affiliated seller. This outcome incentivizes the platform to choose the unaffiliated seller as the sole data acquirer.

Under weak data neutrality, the platform is required to provide data non-exclusively, i.e. $d_1 = d_2 = D$. As in (41), the equilibrium data price charged to unaffiliated seller 2 can be derived and denoted as $\widehat{r}_2^W(\mathbf{d})$. Unlike in the case with $\bar{q} > 1$, when the platform is integrated with a weak seller, the unaffiliated strong seller always wants to acquire $D$ under *weak* data neutrality; that is, $\widehat{r}_2^W(\mathbf{d})$ is always positive.

[12] Note that we assume that the marginal cost $c$ is sufficiently small such that at least one seller is provided with data from the platform. If $c$ is sufficiently high, the platform may want to give up selling data entirely, leading to $\widehat{\mathbf{a}}^N = (0,0)$.

**Proposition 9.** *Under weak data neutrality, the unaffiliated seller always accepts the symmetric data offer if it is an initially strong seller.*

Under strong data neutrality, according to the same logic as in Section 3.3, the platform sets $\widehat{r} = \frac{c}{2}$ to make the optimal $D$ from the independent profit maximization of seller 1 the same as that of the joint profit maximization. In each case (no regulation, weak and strong data neutrality), the optimal amount of $D$ is determined by maximizing the corresponding downstream profit of seller 1.

Finally, we compare the first order condition with respect to $D$ under no regulation with that under data neutrality regulations to determine whether the regulation causes a data-amount-reducing effect. In Appendix A, we show that $\widehat{\zeta}_W < \widehat{\zeta}_N$, where $\widehat{\zeta}_N \equiv \frac{\partial \widehat{\pi}_p^N}{\partial D}$ and $\widehat{\zeta}_W \equiv \frac{\partial \widehat{\pi}_p^W}{\partial D}$, which implies that $\widehat{D}_W < \widehat{D}_N$ given that the second order conditions are satisfied. Similarly, we show that $\widehat{\zeta}_S < \widehat{\zeta}_W$, which implies that $\widehat{D}_S < \widehat{D}_W$.

**Proposition 10.** *Data neutrality regulation reduces the amount of data available in the market even when the platform is affiliated with an initially weak seller. Additionally, stricter regulation reduces the amount of data further than weak data neutrality.*

The intuition behind Proposition 10 is the same as that in Proposition 5. Although integration with an initially weak seller alters the platform's incentive to provide data absent regulation, the broad implications from the main analysis can be carried over: data neutrality results in a nonexclusive data access regime, whereas the platform maximizes its profit by offering exclusive data access to one seller. However, such nondiscriminatory regulation does not necessarily improve welfare because the regulation causes the platform to reduce the amount of data it provides.

**Remark 5.** *The implications of data neutrality depend on both the identity and the initial strength of the downstream affiliate of the platform. In the absence of regulation, the platform always prefers exclusive data provision, though not necessarily to its affiliated seller. When the affiliate is initially strong, exclusivity serves a foreclosure purpose: the platform grants the affiliate exclusive access and denies data access to the unaffiliated rival. When the affiliate is initially weak, this foreclosure motive disappears and the platform instead prefers to sell data exclusively to the stronger unaffiliated seller. Under the effective neutrality regimes considered here, by contrast, the data access becomes nonexclusive in equilibrium in either case. The central trade-off is therefore unchanged: neutrality broadens access to data but also weakens the incentive of the platform to supply it. Consequently, even when the regulation forces the platform to provide data to its initially weak affiliate, the equilibrium amount of data supplied is lower than under no regulation.*

## 5.2 Alternative Behavioral Remedy: Allowing Downstream Sellers to Choose the Amount of Data

Both weak and strong data neutrality regulations in the main model allow the platform to choose the amount of data supplied. We now consider an alternative behavioral remedy under which the platform must charge a common data price (i.e., $r_1 = r_2 = r$), while the two downstream sellers are allowed to choose their own data purchases, so that $d_1$ and $d_2$ may differ.

As in the main specification, the platform chooses $r_1(\mathbf{d})$ such that the optimal $d_1$ chosen by seller 1 maximizes the platform's joint profits. Focusing on the case with sufficiently large $\bar{q}$ such that unaffiliated seller 2 buys data (i.e., $a_2 = 1$), the optimal symmetric data price, denoted as $\tilde{r}$, is set at $\frac{c \times \max\{d_1, d_2\}}{d_1 + d_2}$.

We first show that there is no equilibrium with $d_2 > d_1$. Intuitively, if the unaffiliated seller chose more data than the affiliated seller, the common price would be pinned down by $d_2$, which is inconsistent with the incentive of the integrated firm to align the pricing rule with the objective of the affiliated seller. The relevant equilibrium region is therefore $d_1 \geq d_2$, which implies that $\tilde{r} = \frac{cd_1}{d_1 + d_2}$.

Comparing the sellers' first order conditions implies that the affiliated seller indeed chooses more data in equilibrium: $\tilde{D}_1 > \tilde{D}_2$. Thus, even though the policy imposes a common unit price, in equilibrium the two sellers will choose to purchase a different amount of data.

The source of this asymmetry is straightforward. Given that $\frac{\partial \tilde{r}}{\partial d_1} = \frac{cd_2}{(d_1 + d_2)^2} \geq 0$, a higher demand for data from the affiliated seller raises the common data price $\tilde{r}$. The resulting price increase makes the data less attractive for the unaffiliated seller and depresses its equilibrium data purchase. Allowing sellers to choose their own data quantities under a nondiscriminatory pricing rule therefore does not preserve balanced access. Instead, it shifts data usage to the affiliated seller.

Relative to strong data neutrality, this alternative remedy therefore leads to a higher common data price, i.e., $\tilde{r} > r$, and the unaffiliated seller receives less data than under strong data neutrality, i.e., $\tilde{D}_2 < D_S$. At the same time, the affiliated seller purchases more data. The alternative remedy therefore weakens the equalizing effect of nondiscriminatory pricing by allowing the stronger demand of the affiliated seller to feed back into the common price itself.

**Proposition 11.** *Allowing downstream sellers to choose their data quantities under nondiscriminatory data pricing raises the data price and tilts data usage toward the affiliated seller: the unaffiliated seller purchases less data, whereas the affiliated seller purchases more.*

Note that under strong data neutrality, the platform directly chooses the amount of data and restricts data provision because any increase in data must be extended equally to the unaffiliated rival. Under the alternative remedy considered here, that discipline is weakened once sellers are permitted to choose different quantities at a common price. As a result, nondiscriminatory pricing alone is insufficient to provide symmetric data access in equilibrium.

## 5.3 Relation to Net Neutrality

Data neutrality shares a fundamental purpose with other nondiscriminatory regulations, such as net neutrality. Unlike data neutrality, net neutrality in the U.S. has experienced repeated regulatory reversals: the 2018 rollback (Federal Communications Commission, 2018), the 2024 reinstatement (Federal Communications Commission, 2024), and the January 2, 2025 Sixth Circuit decision setting aside the 2024 order (United States Court of Appeals for the Sixth Circuit, 2025).[13] Given that both neutrality regulations are similar in that they aim to create a level playing field in the downstream market (e.g., product sellers for data neutrality and content providers for net neutrality), we compare them to emphasize how our findings concerning data neutrality are different from those for net neutrality.

Under net neutrality, all Internet traffic should be treated equally such that Internet service providers (ISPs) are prohibited from favoring certain content providers (CPs) over others using monetary transfers; analogously, under data neutrality, a platform should not discriminate against certain sellers in terms of data provision (the amount of data to provide and/or data price). Because both concepts impose nondiscriminatory access requirements, the net neutrality literature provides a useful benchmark for our analysis. In particular, studies of net neutrality that combine vertical integration with paid prioritization show that the equilibrium effects depend on the relative efficiency of the integrated downstream firm (Guo et al. (2010); Brito et al. (2014)). Our setting delivers a parallel foreclosure logic absent regulation: when the platform is integrated with a seller that initially has higher targeting quality, it prefers to grant its affiliated seller exclusive data access.

However, the key difference is technological. The network capacity is rival, whereas data are non-rival. When an ISP allocates scarce bandwidth across content providers, prioritizing one provider can impose congestion-related externalities on others, and a neutrality rule that prevents traffic-intensive providers from contributing to capacity expansion may weaken investment incentives and reduce welfare (Peitz and Schuett, 2016; Choi et al., 2018). By contrast, when a platform provides the same data to multiple downstream sellers, doing so does not proportionally raise production cost and does not generate congestion externalities for consumers.

Even so, we show that data neutrality can still reduce welfare. However, the mechanism is different: instead of operating through congestion, it works by weakening the incentive of the platform to refine and provide data. This data-amount-reducing effect is analogous to the investment-reducing effect emphasized in the net neutrality literature, but because data is non-rival, the welfare implication need not be as severe. Indeed, as Claim 1 and Figure 6 show, data neutrality can enhance welfare when downstream sellers are sufficiently differentiated in

[13] However, whether reintroducing the regulation remains necessary is still debatable. For instance, state and local governments in the US, including California, have undertaken action to revive net neutrality: California signed into law S.B. 822 in response to the revocation of net neutrality. Refer to https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB822 for details.

their initial targeting quality.

## 6 Conclusion

This paper studies hypothetical data neutrality regulations that require a platform to treat downstream sellers equally in access to its data.

Under no data neutrality regulation, the platform can discriminate both in the amount of data provided and the price charged to affiliated versus unaffiliated sellers. The main result is that the platform optimally grants *exclusive data access to only one seller*, rather than sharing data with both. When the affiliated seller is initially stronger, the platform forecloses the unaffiliated seller and gives the data exclusively to its affiliate; when the affiliated seller is initially weaker, the platform instead sells the data exclusively to the stronger unaffiliated seller. The intuition is that exclusivity preserves the strategic value of the data in downstream competition, whereas sharing data with both sellers reduces differentiation and intensifies price competition. In this sense, the platform prefers *data foreclosure* under no regulation.

Under weak data neutrality, the platform must offer the same amount of data to both sellers but may still charge different prices. The paper shows that this regime can be *de facto* ineffective. If the platform is still allowed to discriminate on price, it can set the unaffiliated seller's data price high enough to prevent meaningful access, thereby recreating the data foreclosure outcome in the unregulated case. Even when the weak rule is strengthened so that the unaffiliated seller's price is capped at its willingness to pay, the platform responds by reducing the total amount of data it makes available relative to the no-regulation benchmark. So weak neutrality may either do almost nothing, or, if made binding, it broadens access only at the cost of lower data provision.

Under strong data neutrality, the platform must offer the same amount of data at the same price to both sellers. This is the most restrictive regime and does ensure symmetric access when production costs are low enough. But we show that strong neutrality still does not necessarily improve welfare. Although it creates a more level playing field and intensifies downstream competition, it also weakens the incentive of the platform to refine and provide data. As a result, the equilibrium quantity of data under strong neutrality is lower than under no regulation, and generally lower than under weak neutrality as well.

The central welfare result of the paper is that data neutrality is ambiguous in welfare terms. If the amount of data were kept fixed, neutrality would help consumers by intensifying competition and improving matching. However, in equilibrium, neutrality reduces the amount of data the platform chooses to provide. That data-amount-reducing effect can offset or dominate the pro-competitive effect. This makes even strong neutrality potentially welfare-reducing. The paper also shows that neutrality is more likely to reduce welfare when the two sellers are initially less differentiated because in that case the reduction in data provision is especially important relative to the competition benefit. In other words, data neutrality regulation moves the market from a regime of exclusive access with strong platform incentives to produce data toward a regime of broader access but weaker incentives to produce data. The policy question is therefore not just whether access is fair, but whether regulation undermines the supply of the very data input it is trying to equalize.

Data neutrality is attractive because it promises fairer access to an increasingly important competitive input. However, our analysis shows that equal access rules alone are not sufficient to guarantee better market outcomes. In the absence of regulation, the platform prefers exclusive data provision; under weak neutrality, that exclusivity may persist in practice through discriminatory pricing; and under strong neutrality, broader access can come at the cost of lower data production. The central policy lesson is therefore that neutrality can intensify downstream competition while simultaneously discouraging upstream data supply. Whether such regulation improves welfare depends on the balance between these two forces. Therefore, effective policy toward platform data markets must consider not only how data is shared, but also how regulation shapes the platform's incentives to produce and monetize data.

# References


Ali, S. N., Chen-Zion, A., and Lillethun, E. (2024). Reselling Information. *Games and Economic Behavior*, 148:23–43.

Bergemann, D. and Bonatti, A. (2015). Selling Cookies. *American Economic Journal: Microeconomics*, 7(3):259–294.

Borgogno, O. and Colangelo, G. (2019). Data Sharing and Interoperability: Fostering Innovation and Competition through APIs. *Computer Law & Security Review*, 35:105314.

Bourreau, M. and de Streel, A. (2019). Digital Conglomerates and EU Competition Policy. Technical report, CERRE.

Braulin, F. and Valletti, T. (2016). Selling Customer Information to Competing Firms. *Economics Letters*, 149:10–14.

Brito, D., Pereira, P., and Vareda, J. (2014). On the Incentives of an Integrated ISP to Favor Its Own Content. *20th ITS Biennial Conference, Rio de Janeiro 2014: The Net and the Internet-Emerging Markets and Policies 106901, International Telecommunications Society (ITS).*

Campbell, J., Goldfarb, A., and Tucker, C. (2015). Privacy Regulation and Market Structure. *Journal of Economics and Management Strategy*, 24(1):47–73.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015). Competing with Privacy. *Management Science*, 61(1):229–246.

Choi, J., Jeon, D., and Kim, B. (2018). Net Neutrality, Network Capacity, and Innovation at the Edges. *Journal of Industrial Economics*, 66(1):172–204.

Crawford, G., Courthoud, M., Seibel, R., and Zuzek, S. (2022). Amazon Entry on Amazon Marketplace. *CEPR Press Discussion Paper No. 17531.*

De Cornière, A. and Nijs, R. (2016). Online Advertising and Privacy. *RAND Journal of Economics*, 47(1):48–72.

Easley, R., Guo, H., and Krämer, J. (2018). From Network Neutrality to Data Neutrality: A Techno-Economic Framework and Research Agenda. *Information Systems Research*, 29(2):253–272.

Esteves, R.-B. and Carballo-Cruz, F. (2023). Can Data Openness Unlock Competition When an Incumbent Has Exclusive Data Access for Personalized Pricing? *Information Economics and Policy*, 64(C):101046.

Federal Communications Commission (2018). Declaratory Ruling, Report and Order, and Order. *Restoring Internet Freedom*, 17-108.

Federal Communications Commission (2024). Safeguarding and Securing the Open Internet; Restoring Internet Freedom. FCC 24-52, WC Docket Nos. 23-320 and 17-108, released May 7, 2024.

Galdon-Sanchez, J., Gil, R., and Uriz-Uharte, G. (2025). The Value of Information in Competitive Markets: Evidence from Small and Medium-Sized Enterprises. *Journal of Political Economy*, 133(1):252–305.

Goldfarb, A. and Tucker, C. (2011). Privacy Regulation and Online Advertising. *Management Science*, 57(1):57–71.

Grace, R. and Leskovich, R. (2015). Challenges in the Use of Social Media Data for the Next Generation Analyst. *SPIE Digital Library Proceedings 9499*.

Graef, I. (2015). Market Definition and Market Power in Data: The Case of Online Platforms. *World Competition*, 38(4):473–506.

Graef, I., Tombal, T., and Streel, A. (2019). Limits and Enablers of Data Sharing An Analytical Framework for EU Competition, Data Protection and Consumer Law. *TILEC Discussion Paper*, DP 2019-024.

Gu, Y., Madio, L., and Reggiani, C. (2019). Exclusive Data, Price Manipulation and Market Leadership. *CESifo Working Paper 7853*.

Gu, Y., Madio, L., and Reggiani, C. (2022). Data Brokers Co-opetition. *Oxford Economic Papers*, 74:820–839.

Guo, H., Bandyopadhyay, S., Cheng, H., and Yang, Y. (2010). Net Neutrality and Vertical Integration of Content and Broadband Services. *Journal of Management Information Systems*, 27:243–276.

Hagiu, A., Teh, T., and Wright, J. (2022). Should Amazon be Allowed to Sell on Its Own Marketplace? *RAND Journal of Economics*, 53:297–327.

Hart, O. and Tirole, J. (1990). Vertical Integration and Market Foreclosure. *Brookings Papers on Economic Activity: Microeconomics*, pages 205–276.

Ichihashi, S. (2020). Online Privacy and Information Disclosure by Consumers. *American Economic Review*, 110(2):569–595.

Ichihashi, S. (2021). Competing Data Intermediaries. *RAND Journal of Economics*, 52(3):515–537.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, 110(9):2819–2858.

Kamien, M., Oren, S., and Tauman, Y. (1992). Optimal Licensing of Cost-Reducing Innovation. *Journal of Mathematical Economics*, 21:483–508.

Katz, M. and Shapiro, C. (1986). How to License Intangible Property. *The Quarterly Journal of Economics*, 101(3):567–590.

Kim, S. (2024). Privacy, Information Acquisition, and Market Competition. *Working Paper*.

Kirpalani, R. and Philippon, T. (2020). Data Sharing and Market Power with Two-Sided Platforms. *NBER Working Paper 28023*.

Krämer, J. and Schnurr, D. (2018). Is There a Need for Platform Neutrality Regulation in the EU? *Telecommunications Policy*, 42:514–529.

Krämer, J. and Shekhar, S. (2025). Regulating Digital Platform Ecosystems Through Data Sharing and Data Siloing: Consequences for Innovation and Welfare. *MIS Quarterly*, 49(1):123–154.

Liu, Z., Sockin, M., and Xiong, W. (2020). Data Privacy and Temptation. *NBER Working Paper 27653*.

Maier, N. (2019). Closeness of Substitution for ‘Big Data’ in Merger Control. *Journal of European Competition Law & Practice*, 10(4):246–252.

Martens, B., de Streel, A., Graef, I., Tombal, T., and Duch-Brown, N. (2020). Business-to-Business Data Sharing: An Economic and Legal Analysis. *JRC Digital Economy Working Paper*.

Martens, B. and Zhao, B. (2021). Data Access and Regime Competition: A Case Study of Car Data Sharing in China. *Big Data & Society*, 8(2).

McAfee, R. and Schwartz, M. (1994). Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity. *American Economic Review*, 84(1):210–230.

Montes, R., Sand-Zantman, W., and Valletti, T. (2019). The Value of Personal Information in Markets with Endogenous Privacy. *Management Science*, 65(3):1342–1362.

O’Brien, D. and Shaffer, G. (1992). Vertical Control with Bilateral Contracts. *RAND Journal of Economics*, 23(3):299–308.

Peitz, M. and Schuett, F. (2016). Net Neutrality and Inflation of Traffic. *International Journal of Industrial Organization*, 46:16–62.

Pylon Network (2018). The Energy Blockchain Platform. *WhitePaper v.2.0.*

Richter, H. and Slowinski, P. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *International Review of Intellectual Property and Competition Law*, 50:4–29.

Stucke, M. and Grunes, A. (2016). Big Data and Competition Policy. *Oxford University Press.*

Taylor, C. (2004). Consumer Privacy and the Market for Customer Information. *RAND Journal of Economics*, 35(4):631–650.

Taylor, C. and Wagman, L. (2014). Customer Privacy in Oligopolistic Markets: Winners, Losers, and Welfare. *International Journal of Industrial Organization*, 34:80–84.

United States Court of Appeals for the Sixth Circuit (2025). In re: MCP No. 185; Ohio Telecom Association, et al. v. Federal Communications Commission. Nos. 24-7000/3449/3450/3497/3504/3507/3508/3510/3511/3519/3538, decided January 2, 2025.

Whitfield, C. (2017). Data-Informed Policy Innovations in Tennessee: Effective Use of State Data Systems. *State Higher Education Executive Officers Association.*

Yang, K. H. (2022). Selling Consumer Data for Profit: Optimal Market-Segmentation Design and Its Consequences. *American Economic Review*, 112(4):1364–93.

# Appendix

# A Proofs

**Proof of Lemma 1.** Given $q_j(d_j;\bar{q}_j,\alpha) > q_k(d_k;\bar{q}_k,\alpha)$, we can specifically characterize the market share for each seller $j$ depending on the relative size of the hedonic prices of the two sellers, $\frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$, as follows:

(a) If $0 < \frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} < \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$, one can easily verify that consumers whose type-$\theta_i$ exceeds $\theta^H$ will choose seller $j$, and those whose $\theta_i$ is between $\theta^L$ and $\theta^H$ will choose seller $k$, and those whose $\theta_i$ is lower than $\theta^L$ will choose not to purchase at all, where $\theta^H$ and $\theta^L$ are defined in (6). Notice that the condition $0 < \frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} < \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$ implies that $\theta^H > \theta^L$. Since we assume that $\theta$ is uniformly distributed on $[0,\bar{\theta}]$, we conclude that the market is partially covered, and the demand faced by each seller is given by: $x_j = \frac{1}{\bar{\theta}}\left(\bar{\theta} - \theta^H\right)$, $x_k = \frac{1}{\bar{\theta}}\left(\theta^H - \theta^L\right)$, as stated in (5).

(b) If $0 = \frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} < \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}$, we can analogously show that the market will be fully covered, and the demand faced by each seller is given by:

$$x_k = \frac{1}{\bar{\theta}}\left(\frac{p_j}{q_j(d_j;\bar{q}_j,\alpha) - q_k(d_k;\bar{q}_k,\alpha)}\right), \quad x_j = 1 - \frac{1}{\bar{\theta}}\left(\frac{p_j}{q_j(d_j;\bar{q}_j,\alpha) - q_k(d_k;\bar{q}_k,\alpha)}\right).$$

(c) If $\frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} \geq \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)} > 0$, we can show that the market is partially covered, and the demand faced by each seller is given by:

$$x_j = 1 - \frac{1}{\bar{\theta}}\left(\frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)}\right); \quad x_k = 0.$$

(d) If $\frac{p_k}{q_k(d_k;\bar{q}_k,\alpha)} \geq \frac{p_j}{q_j(d_j;\bar{q}_j,\alpha)} = 0$, then the market is fully covered with $x_k = 0$ and $x_j = 1$.

Note, however, cases (b), (c) and (d) are inconsistent with equilibrium because in either case, at least one of the sellers will be receiving zero revenues: in case (b), firm $k$'s revenue is zero because $p_k = 0$; in case (c), the revenue for seller $k$ is zero because $x_k = 0$; and in case (d), both firms will have zero revenue either because $x_k = 0$ or $p_j = 0$. As such, the sellers will have an incentive to deviate to case (a), which results in positive revenues for both sellers. Thus, the only case consistent with equilibrium is case (a). □

**Proof of Proposition 1.** From (12), if the sum of the terms in the curly brackets, $Q(\bar{q}) \equiv 7(d_2^\alpha + 1) - 16\bar{q}(2-\bar{q}) + 16\bar{q}\left\{d_1^\alpha\left[\bar{q}(d_1^\alpha + 2) - 2\right] - (d_1^\alpha + 1)\, d_2^\alpha\right\}$, is positive, $r_2(\mathbf{d})$ is also positive; it is easy to see that $\frac{\partial Q(\bar{q})}{\partial \bar{q}} = 16\,(d_1^\alpha + 1)\left[2\bar{q}d_1^\alpha - d_2^\alpha + 2(\bar{q}-1)\right] > 0$. The solution to $Q(\bar{q}) = 0$ is

$\frac{2\left(d_2^\alpha+2\right)d_1^\alpha+2d_2^\alpha+\sqrt{\left(d_1^\alpha+1\right)^2\left[\left(4d_2^\alpha+9\right)d_2^\alpha+9\right]+4}}{4\left(d_1^\alpha+1\right)^2} \equiv \bar{q}^*(\mathbf{d})$ as in (13). The second claim that $\partial r_2(\mathbf{d})/\partial d_2 < 0$ is proved in the main text. □

**Proof of Proposition 2.** It is easy to show that $\frac{\partial \pi_p(\mathbf{d})}{\partial d_2}$, as in (20), is negative. $\Lambda(\mathbf{d})$ in $\zeta_N$ is given as follows.

$$\begin{aligned}\Lambda(\mathbf{d}) \equiv & \frac{11\left(D^{2\alpha}-1\right)}{(\beta D)^\alpha - 4\bar{q}\left(D^\alpha+1\right)+1} - \frac{\left(D^\alpha+1\right)^2\left[4\bar{q}\left(6D^\alpha-23\right)+11\right]}{\left[(\beta D)^\alpha-4\bar{q}\left(D^\alpha+1\right)+1\right]^2} \\ & + \frac{48\bar{q}(4\bar{q}-1)\left(D^\alpha+1\right)^3}{\left[(\beta D)^\alpha-4\bar{q}\left(D^\alpha+1\right)+1\right]^3} - \frac{2\left(D^\alpha+1\right)D^\alpha\left\{8\bar{q}\left(D^\alpha+1\right)\left[4\bar{q}\left(D^\alpha+1\right)-3\right]+7\right\}}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^3}.\end{aligned} \tag{A.1}$$

From $\frac{\partial \pi_p(\mathbf{d})}{\partial d_2} < 0$, $\beta = 0$, which results in $\zeta_N$, as in (21). Additionally, the second order condition is satisfied, as follows: given that $\frac{\partial \zeta_N}{\partial D} = \frac{4\alpha\bar{q}^2\bar{\theta}D^{\alpha-2}\widetilde{\Lambda}}{[4\bar{q}(D^\alpha+1)-1]^4}$, where $\widetilde{\Lambda} \equiv 2(1-\alpha)+2D^\alpha(1-2\alpha)+16\bar{q}^2(1-\alpha)(1+D^\alpha)^3[1-\bar{q}(1+D^\alpha)]-\bar{q}(1+D^\alpha)[11(1-\alpha)+(11-\alpha)D^\alpha]$, the supremum of $\widetilde{\Lambda}$, which is attained as $\bar{q}$ approaches one, $D=0$, and $\alpha=1$, is zero, which implies that $\frac{\partial \zeta_N}{\partial D}$ is always negative for $\bar{q} > 1$.

Additionally, we show that $\bar{q}^{Rev} - \bar{q}^* = \frac{3+5d_2^\alpha-\sqrt{9d_2^\alpha+4d_2^{2\alpha}+9}}{4(d_1^\alpha+1)}$ is positive as follows: $\frac{3+5d_2^\alpha-\sqrt{9d_2^\alpha+4d_2^{2\alpha}+9}}{4(d_1^\alpha+1)} > \frac{3+5d_2^\alpha-\sqrt{(2d_2^\alpha+3)^2}}{4(d_1^\alpha+1)} = \frac{3d_2^\alpha}{4(d_1^\alpha+1)} > 0.$

Finally, we show the market expansion and the customer-switching effects, as in Figure 2, analytically. Given $\mathbf{d} = (D, \beta D)$, where $\beta \in [0,1]$, we check that $\frac{\partial \theta^L}{\partial \beta} = -\frac{3\alpha\bar{q}\bar{\theta}(D^\alpha+1)(\beta D)^\alpha}{\beta[(\beta D)^\alpha-4\bar{q}(D^\alpha+1)+1]^2}$ and $\frac{\partial \theta^H}{\partial \beta} = -\frac{2\alpha\bar{q}\bar{\theta}(D^\alpha+1)(\beta D)^\alpha}{\beta[(\beta D)^\alpha-4\bar{q}(D^\alpha+1)+1]^2}$, which are negative. □

**Proof of Proposition 3.** $\zeta_W$ is derived from $\frac{\partial \pi_p^W(D)}{\partial D}$, where $\pi_p^W(D)$ is given in (26) and $r_2$ is given by $r_2^W$, as in (27). □

**Proof of Proposition 4.** This is proven in the main analysis. □

**Proof of Proposition 5.** The second order conditions with respect to $D$ under no data neutrality and *weak* data neutrality are, respectively, given as follows.

$$\frac{\partial \zeta_N}{\partial D} = \frac{4\alpha\bar{q}^2\bar{\theta}D^{\alpha-2}\Psi}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^4}; \quad \frac{\partial \zeta_W}{\partial D} = \frac{\bar{q}\bar{\theta}\alpha D^{\alpha-2}\widetilde{\Psi}}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^4}, \tag{A.2}$$

where

$$\Psi \equiv 2D^\alpha(1-2\alpha)+2(1-\alpha)-16(1-\alpha)\bar{q}^2\left(D^\alpha+1\right)^3\left[\bar{q}(D^\alpha+1)-1\right]-\bar{q}\left(D^\alpha+1\right)\left[(11-\alpha)D^\alpha+11(1-\alpha)\right],$$

and

$$\begin{aligned}\widetilde{\Psi} \equiv\ & (\alpha-1)(\bar{q}-1)(4\bar{q}+1)\left[1-4\bar{q}\left(D^\alpha+1\right)\right]^4+(4\bar{q}-1)^2\{\alpha+2\bar{q}\{D^\alpha+\alpha\left(6D^\alpha-1\right)\\ +\ & 4\bar{q}\left(D^\alpha+1\right)\left[D^\alpha+\alpha\left(D^\alpha-1\right)+1\right]+1\}-1\}.\end{aligned}$$

First, we can show that $\Psi<0$. The maximum possible value of $\Psi$, which is attained when $D=0$, is $(1-\alpha)[2+16\bar{q}^2(1-\bar{q})-11\bar{q}]$, which is always negative, given $\bar{q}>1$. This suggests that the second order condition is satisfied because $\frac{\partial \zeta_N}{\partial D}<0$. The second order condition under *weak* data neutrality is satisfied if $\widetilde{\Psi}<0$.

Assuming that the second order conditions are satisfied, we can rank the equilibrium amount of data under different cases by comparing the corresponding first order conditions. From (35), it is sufficient to show that $\Phi(D|\bar{q},\alpha)>0$ to show that $\zeta_N>\zeta_W$. We first show that $\frac{\partial\Phi(D|\bar{q},\alpha)}{\partial\bar{q}}=4\left\{48\bar{q}^2\left(D^\alpha+1\right)^3+8\bar{q}\left(D^\alpha+1\right)\left[\left(5D^\alpha+7\right)D^\alpha+9\right]-\left(15D^\alpha+41\right)D^\alpha-21\right\}$ is positive: the infimum of the terms in the curly brackets is obtained as $\bar{q}$ approaches one, whose value is $\left.\frac{\partial\Phi(D|\bar{q},\alpha)}{\partial\bar{q}}\right|_{\bar{q}=1}=4\{48(D^\alpha+1)^3+8(D^\alpha+1)[(5D^\alpha+7)D^\alpha+9]-(15D^\alpha+41)D^\alpha-21\}$. The infimum of $\Phi(D|\bar{q},\alpha)$ increases in $D$: $\frac{\partial \left.\frac{\partial\Phi(D|\bar{q},\alpha)}{\partial\bar{q}}\right|_{\bar{q}=1}}{\partial D}=12\alpha D^{\alpha-1}\left[2\left(44D^\alpha+75\right)D^\alpha+77\right]>0$. Thus, the infimum of $\Phi(D|\bar{q},\alpha)$, which is attained at $\bar{q}=1$ and $D=0$, is 135, implying that $\Phi(D|\bar{q},\alpha)>0$ since its infimum remains positive. This suggests that $\Phi(D|\bar{q},\alpha)$ is always positive, which proves $\zeta_N>\zeta_W$. According to the conditions in (21) and (29), $\zeta_N(D_N)=0$ and $\zeta_W(D_W)=0$. Given $\zeta_N>\zeta_W$, we have $\zeta_W(D_N)<\zeta_N(D_N)=0$. According to the second order conditions, both $\zeta_N(D)$ and $\zeta_W(D)$ are nonincreasing in $D$; therefore, $D_W<D_N$.

Next, the second order condition with respect to $D$ under *strong* data neutrality is satisfied as follows.

$$\frac{\partial\zeta_S}{\partial D}=\frac{-4\alpha(1-\alpha)(\bar{q}-1)\bar{q}^2\bar{\theta}D^{\alpha-2}}{(4\bar{q}-1)^2}<0. \tag{A.3}$$

It is sufficient to show that $\kappa(D|\bar{q},\alpha)$ is positive to show that $\zeta_N>\zeta_S$. From $\kappa(D|\bar{q},\alpha)\equiv 2D^\alpha+1+\bar{q}\left[32\bar{q}^2\left(1+D^\alpha\right)^3-4\bar{q}\left(4-D^\alpha\right)\left(1+D^\alpha\right)D^\alpha-\left(10+3D^\alpha\right)D^\alpha-6\right]$,

$$\frac{\partial\kappa(D|\bar{q},\alpha)}{\partial\bar{q}}=96\bar{q}^2\left(D^\alpha+1\right)^3+8\bar{q}\left(D^\alpha-4\right)\left(D^\alpha+1\right)D^\alpha-\left(3D^\alpha+10\right)D^\alpha-6$$

can be shown as positive: the minimum possible value of the first term, attained when $D=0$, is $96\bar{q}^2$, whereas that of the last three negative terms, attained when $D=1$, is $-48\bar{q}-19$. Thus, $\frac{\partial\kappa(D|\bar{q},\alpha)}{\partial\bar{q}}>96\bar{q}^2-48\bar{q}-19>96-19-48=29>0$. Since the minimum possible value is positive, $\frac{\partial\kappa(D|\bar{q},\alpha)}{\partial\bar{q}}$ is positive. Then, the infimum of $\kappa(D|\bar{q},\alpha)$, attained as $\bar{q}$ approaches one, is $9\left(D^\alpha+1\right)\left[\left(4D^\alpha+5\right)D^\alpha+3\right]>0$. Since the infimum is still positive, $\kappa(D|\bar{q},\alpha)$ is always positive. Then, $\zeta_N-\zeta_S$ is positive as long as $c$ is sufficiently small.

From the comparison of $\zeta_W$ with $\zeta_S$, there exists $\bar{q}^S$ that satisfies $\widetilde{\kappa}(D|\bar{q}^S, \alpha) = 0$. Specifically, $\bar{q}^S = \frac{1}{12(D^\alpha+1)^3}\left\{9 + 15D^{2\alpha} + 4D^{3\alpha} + 20D^\alpha + \widetilde{\widetilde{\kappa}}(D|\alpha) + \frac{4(D^\alpha+1)^2\{\{[(4D^\alpha+13)D^\alpha+16]D^\alpha+15\}D^\alpha+9\}}{\widetilde{\widetilde{\kappa}}(D|\alpha)}\right\}$. We graphically show that $\bar{q}^S$ is always smaller than $\bar{q}^*$, which suggests that $D_W > D_S$. □

**Proof of Proposition 6.** The consumer surplus and social welfare levels under different cases are given as follows:

$$\begin{aligned}
CS_N &= \frac{\bar{q}^2\bar{\theta}^2\left(D_N^\alpha+1\right)^2\left[4\bar{q}\left(D_N^\alpha+1\right)+5\right]}{2\left[4\bar{q}\left(D_N^\alpha+1\right)-1\right]^2}; \quad CS_k = \frac{\bar{q}^2(4\bar{q}+5)\bar{\theta}^2\left(D_k^\alpha+1\right)}{2(4\bar{q}-1)^2};\\
SW_N &= \frac{1}{2\left[4\bar{q}\left(D_N^\alpha+1\right)-1\right]^2}\Big\{\bar{q}^2\left(D_N^\alpha+1\right)^2\left[\bar{\theta}(5\bar{\theta}-6)-32cD_N\right]-2\bar{q}\left(D_N^\alpha+1\right)\left(\bar{\theta}-8cD_N\right)\\
&\quad+4\bar{q}^3\bar{\theta}(\bar{\theta}+2)\left(D_N^\alpha+1\right)^3-2cD_N\Big\};\\
SW_k &= \frac{\bar{q}\left[16c(1-2\bar{q})D_k+\bar{q}(4\bar{q}+5)\bar{\theta}^2\left(D_k^\alpha+1\right)+2(\bar{q}-1)(4\bar{q}+1)\bar{\theta}\left(D_k^\alpha+1\right)\right]-2cD_k}{2(4\bar{q}-1)^2}.
\end{aligned} \tag{A.4}$$

We show that the difference between consumer surplus and total welfare with and without data neutrality is always positive for $D_N = D_k$.

$$\begin{aligned}
CS_k - CS_N &= \frac{\bar{q}^2\bar{\theta}^2D^\alpha\left(D^\alpha+1\right)\left[112\bar{q}^2\left(D^\alpha+1\right)-4\bar{q}\left(D^\alpha+2\right)-5\right]}{2(4\bar{q}-1)^2\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^2};\\
SW_k - SW_N &= \frac{\bar{q}^2\bar{\theta}D^\alpha\left(D^\alpha+1\right)\left[16\bar{q}^2(7\bar{\theta}-2)\left(D^\alpha+1\right)+22-4\bar{q}(\bar{\theta}+10)\left(D^\alpha+2\right)-5\bar{\theta}\right]}{2(4\bar{q}-1)^2\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^2}.
\end{aligned} \tag{A.5}$$

For the consumer surplus comparison, it is easy to show that the square bracket in the numerator is always positive for $\bar{q} > 1$, which indicates $CS_k > CS_N$. Similarly, the square brackets in the numerator in the social welfare comparison increase in $\bar{\theta}$ and $D$; thus, the infimum of them, which is attained as $\bar{\theta}$ approaches one and $D$ approaches zero, is zero. Since the sum of terms in the square bracket is always positive, $SW_k > SW_N$. These results indicate that, absent the data-amount-reducing effect, data neutrality is welfare enhancing due to the competition-enhancing effect.

To illustrate the *data-amount-reducing effects*, we show that welfare increases in the amount of data.

$$\begin{aligned}
\frac{\partial SW_N}{\partial D} &= \frac{\alpha\bar{q}\bar{\theta}^2D_N^{\alpha-1}\left(D_N^\alpha+1\right)\left[\bar{q}\left(D_N^\alpha+1\right)\widetilde{\Phi}+1\right]}{\left[4\bar{q}\left(D_N^\alpha+1\right)-1\right]^3}-c;\\
\frac{\partial SW_k}{\partial D} &= \frac{\alpha\bar{q}\bar{\theta}D_k^{\alpha-1}\widetilde{\widetilde{\Phi}}}{2(4\bar{q}-1)^2}-c,
\end{aligned} \tag{A.6}$$

where $\widetilde{\Phi} \equiv 2\bar{q}(\bar{\theta}+2)\left(D_N^\alpha+1\right)\left[4\bar{q}\left(D_N^\alpha+1\right)-3\right]+5(2-\bar{\theta})$ and $\widetilde{\widetilde{\Phi}} \equiv \bar{q}[4\bar{q}(\bar{\theta}+2)+5\bar{\theta}-6]-2$. The infimum of $\widetilde{\Phi}$, which is attained as $\bar{q}$ and $\bar{\theta}$ approach one, is $6\left(D_N^\alpha+1\right)\left[4\left(D_N^\alpha+1\right)-3\right]+5>0$;

thus, $\frac{\partial SW_N}{\partial D} > 0$ as long as $c$ is sufficiently low. Similarly, the infimum of $\tilde{\tilde{\Phi}}$, which is attained as $\bar{q}$ approaches one, is $9\bar{\theta} > 0$, which shows $\frac{\partial SW_k}{\partial D} > 0$. The comparative statics for consumer surplus levels are qualitatively the same. Given that less data availability reduces welfare, the data-amount-reducing effects of data neutrality regulation offset the competition-enhancing effects. □

**Proof of Proposition 7.** We first show that the equilibrium amounts of data with and without data neutrality increase in $\bar{q}$, which suggests that less initial targeting quality differentiation leads to a reduced amount of data in both cases. By taking the derivatives of $\zeta_N$ and $\zeta_W$ with respect to $\bar{q}$, we obtain the following result.

$$\begin{aligned} \frac{\partial \zeta_N}{\partial \bar{q}} &= \frac{4\alpha\bar{q}\bar{\theta}D^{\alpha-1}\left(D^\alpha+1\right)\left\{\bar{q}\left(D^\alpha+1\right)\left[16\bar{q}^2\left(D^\alpha+1\right)+1\right]-4\right\}}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^4}; \\ \frac{\partial \zeta_W}{\partial \bar{q}} &= \alpha\bar{\theta}D^{\alpha-1}\left\{\frac{2\bar{q}\left(8\bar{q}^2-6\bar{q}+5\right)+1}{(4\bar{q}-1)^3}+\frac{4\bar{q}\left(D^\alpha+1\right)\left[2\bar{q}\left(D^\alpha+1\right)+3\right]+1}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^4}\right\}. \end{aligned} \tag{A.7}$$

First, $\frac{\partial \zeta_N}{\partial \bar{q}}$ is positive given that the infimum of the terms in the curly brackets, which is attained as $\bar{q}$ approaches one and $D = 0$, is 13. Similarly, it is easy to see that $\frac{\partial \zeta_W}{\partial \bar{q}}$ is also positive. By the implicit function theorem, the sign of $\frac{\partial D}{\partial \bar{q}}$ is the same as that of $\frac{\partial \zeta}{\partial \bar{q}}$. Additionally, by comparing the two cross partial derivatives in (A.7) with one another, we obtain the following result:

$$\frac{\partial \zeta_W}{\partial \bar{q}} - \frac{\partial \zeta_N}{\partial \bar{q}} = \frac{\alpha\bar{\theta}D^{\alpha-1}}{4}\left\{\frac{36}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^4}+\frac{38}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^3}+\frac{7}{\left[4\bar{q}\left(D^\alpha+1\right)-1\right]^2}+\frac{28\bar{q}+5}{(4\bar{q}-1)^3}\right\},$$

which is easy to prove as positive; this means that the downward pressure of the lower $\bar{q}$ on the amount of data is greater under data neutrality. The comparison with strong data neutrality is similar and is omitted here. □

**Proof of Proposition 8.** It is easy to show that $\frac{\partial \pi_1(\mathbf{d})}{\partial \beta}$ is negative from (42). $\widehat{\Lambda}(\mathbf{d}^{(0,1)})$ in $\widehat{\zeta}_N$ is given as follows.

$$\widehat{\Lambda}(\mathbf{d}^{(0,1)}) \equiv \frac{\bar{\theta}\alpha D^{\alpha-1}\left\{4\left(D^\alpha+1\right)^2\left[4(D^\alpha+1)-3\bar{q}\right]+\bar{q}^3+10\bar{q}^2\left(D^\alpha+1\right)\right\}}{\left[4(D^\alpha+1)-\bar{q}\right]^3}, \tag{A.8}$$

which is always positive because $\bar{q} < \frac{4(D^\alpha+1)}{3}$. Since $\widehat{\Lambda}(\mathbf{d}^{(0,1)})$ is positive, we can find the interior solution for $\widehat{D}_N$ satisfying $\widehat{\zeta}_N(\mathbf{d}^{(0,1)}) = 0$. The set of equilibria under no data neutrality is given as follows.

$$
\begin{aligned}
\widehat{x}_1^N &= \frac{1+\widehat{D}_N^\alpha}{4\left(1+\widehat{D}_N^\alpha\right)-\bar{q}}; \quad \widehat{x}_2^N = \frac{2(1+\widehat{D}_N^\alpha)}{4\left(1+\widehat{D}_N^\alpha\right)-\bar{q}};\\
\widehat{p}_1^N &= \frac{\bar{q}\bar{\theta}\left[1+\widehat{D}_N^\alpha-\bar{q}\right]}{4\left(1+\widehat{D}_N^\alpha\right)-\bar{q}}; \quad \widehat{p}_2^N = \frac{2\bar{\theta}\left(1+\widehat{D}_N^\alpha\right)\left[\bar{q}-(1+\widehat{D}_N^\alpha)\right]}{\bar{q}-4\left(1+\widehat{D}_N^\alpha\right)};\\
\widehat{\pi}_p^N &= \frac{\bar{q}\bar{\theta}\left(1+\widehat{D}_N^\alpha\right)\left(1+\widehat{D}_N^\alpha-\bar{q}\right)}{\left[4\left(1+\widehat{D}_N^\alpha\right)-\bar{q}\right]^2} - (\widehat{r}_2^N - c)\times \widehat{D}_N,
\end{aligned}
\tag{A.9}
$$

where $\widehat{r}_2^N$ is in (41). □

**Proof of Proposition 9.** Under weak data neutrality, i.e., $\mathbf{d} = (D, D)$, the set of equilibria is given as follows.

$$
\begin{aligned}
\widehat{x}_1^W &= \frac{1}{4-\bar{q}}; \quad \widehat{x}_2^W = \frac{2}{4-\bar{q}};\\
\widehat{p}_1^W &= \frac{\bar{\theta}\bar{q}\,(1-\bar{q})\,(1+D^\alpha)}{4-\bar{q}}; \quad \widehat{p}_2^W = \frac{2\bar{\theta}\,(1-\bar{q})\,(1+D^\alpha)}{4-\bar{q}};\\
\widehat{r}_2^W &= \frac{4\bar{\theta}\left\{\frac{[\bar{q}(D^\alpha+1)-1]}{[\bar{q}(D^\alpha+1)-4]^2} - \frac{(\bar{q}-1)(D^\alpha+1)}{(\bar{q}-4)^2}\right\}}{D}.
\end{aligned}
\tag{A.10}
$$

We show that $\widehat{r}_2^W$ is always positive. The terms in the curly brackets can be simplified as $\frac{(4-\bar{q})^2[\bar{q}(D^\alpha+1)-1]+(1-\bar{q})(D^\alpha+1)[\bar{q}(D^\alpha+1)-4]^2}{[\bar{q}-4(D^\alpha+1)]^2[\bar{q}(D^\alpha+1)-4]^2}$, and the minimum possible value of the numerator here, attained when $D=0$, is zero. Thus, $\widehat{r}_2^W$ is always positive or equal to zero. □

**Proof of Proposition 10.** The platform's profit under *weak* data neutrality is given by

$$
\widehat{\pi}_p^W(\mathbf{r}|D) = \frac{\bar{\theta}\,(1-\bar{q})\,(1+D^\alpha)}{(4-\bar{q})^2} + (\widehat{r}_2^W - c)\times D. \tag{A.11}
$$

Note that by setting $\widehat{r}_1^W = c - \widehat{r}_2^W$, the optimal amount of data chosen by seller 1 coincides with that chosen by the platform. The optimal amount of data under *weak* data neutrality is set by the following first order condition.

$$
\widehat{\zeta}_W \equiv \alpha\bar{\theta}D^{\alpha-1}\left\{\frac{4\bar{q}\left[\bar{q}(D^\alpha+1)+2\right]}{\left[(4-\bar{q}(D^\alpha+1)\right]^3} + \frac{(1-\bar{q})(4+\bar{q})}{(\bar{q}-4)^2}\right\} - c = 0. \tag{A.12}
$$

By comparing $\widehat{\zeta}_N$ with $\widehat{\zeta}_W$, we obtain the following result.

$$\widehat{\zeta}_N - \widehat{\zeta}_W = \frac{\alpha \bar{q} \bar{\theta} D^{\alpha-1} \Delta}{(4-\bar{q})^2 \left[4(D^\alpha+1)-\bar{q}\right]^3}, \tag{A.13}$$

where $\Delta \equiv 12(4-\bar{q})(4+5\bar{q})D^{2\alpha} + 16(4+5\bar{q})D^{3\alpha} + 2(4-\bar{q})^2(6+11\bar{q})D^\alpha + (4-\bar{q})^2(4+11\bar{q})$. It is easy to see that $\Delta$ is always positive for $\bar{q} < 1$, which guarantees that $\widehat{\zeta}_N > \widehat{\zeta}_W$. Given that the second order conditions are satisfied, this results in $\widehat{D}_N > \widehat{D}_W$.

Next, the optimal amount of data under strong data neutrality is determined by the following first order condition.

$$\widehat{\zeta}_S \equiv \frac{\alpha(1-\bar{q})\bar{q}\bar{\theta} D^{\alpha-1}}{(4-\bar{q})^2} - \frac{c}{2} = 0. \tag{A.14}$$

By comparing $\widehat{\zeta}_W$ with $\widehat{\zeta}_S$, we obtain the following result.

$$\widehat{\zeta}_W - \widehat{\zeta}_S = \frac{4\bar{\theta}\alpha D^{\alpha-1}\widehat{\Delta}}{(4-\bar{q})^2 \left[4-\bar{q}(D^\alpha+1)\right]^3} - \frac{c}{2}, \tag{A.15}$$

where $\widehat{\Delta} \equiv (1-\bar{q})\left[4-\bar{q}(D^\alpha+1)\right]^3 + (4-\bar{q})^2\bar{q}\left[\bar{q}(D^\alpha+1)+2\right]$. It is easy to see that $\widehat{\Delta} > 0$; thus, as long as the marginal cost $c$ is sufficiently low, $\widehat{\zeta}_W > \widehat{\zeta}_S$, which results in $\widehat{D}_W > \widehat{D}_S$. $\square$

**Proof of Proposition 11.** First, we show that $d_1 \geq d_2$. For simplicity, suppose that $d_1 = \beta D$ and $d_2 = D$, where $\beta \in [0,1]$. Denoting the first order condition with respect to $d_j$ as $\zeta_{d_j}$, we want to show that $\zeta_{d_1} - \zeta_{d_2}$ is always positive, suggesting that $d_1 > d_2$, which contradicts $\mathbf{d} = (\beta D, D)$. Given that $\zeta_{d_1} - \zeta_{d_2}$ is a continuous function of all parameters and variables whose domains are closed and bounded (i.e., compact), its sign can be graphically mapped over the relevant set of parameters: we can graphically show that $\zeta_{d_1} - \zeta_{d_2}$ decreases in $D$ and $\alpha$, but increases in $\bar{q}$ and $\bar{\theta}$. Then, its infimum, attained at $D = 1$, $\alpha = 1$, $\bar{\theta} = 1$, and $\bar{q}$ approaching one, is $\Phi(\beta) \equiv \frac{(\beta+1)\left\{\beta(\beta+1)[14-4(\beta+1)]+4\beta\left[4(\beta+1)^2-6(\beta+1)+8\right]\right\}}{\beta[4(\beta+1)-2]^3} - \frac{2\beta}{(\beta+1)^2}$, which decreases in $\beta$. Thus, the minimum possible value of $\Phi(\beta)$, attained at $\beta = 1$, is $\frac{1}{18}$, which is positive, implying that $\zeta_{d_1} > \zeta_{d_2}$.

Using $\mathbf{d} = (D, \beta D)$, i.e., $d_1 \geq d_2$, we can graphically show that $\zeta_{d_1} - \zeta_{d_2}$ decreases in $D$ and $\alpha$, but increases in $c$, $\beta$, $\bar{q}$, and $\bar{\theta}$. Then, the infimum, attained at $D = 1$, $\alpha = 1$, $c = 0$, $\bar{q}$ approaching one, and $\bar{\theta} = 1$, is $\widetilde{\Phi}(\beta) \equiv \frac{2\left\{2\beta[7(\beta+1)-8]+4\beta\left[2(\beta+1)^2-6(\beta+1)+16\right]\right\}}{(7-\beta)^3\beta}$, which increases in $\beta$. Thus, the minimum possible value of $\widetilde{\Phi}(\beta)$, attained at $\beta = 0$, is $\frac{92}{343}$, which is positive, implying that $\zeta_{d_1} > \zeta_{d_2}$. Given that the second order conditions are satisfied, this suggests that $\tilde{D}_1 \equiv D = d_1 > d_2 = \beta D \equiv \tilde{D}_2$, as shown in Figure A.1. Finally, $\mathbf{d} = (D, \beta D)$ results in $\tilde{r} = \frac{c}{1+\beta} \geq r = \frac{c}{2}$. $\square$

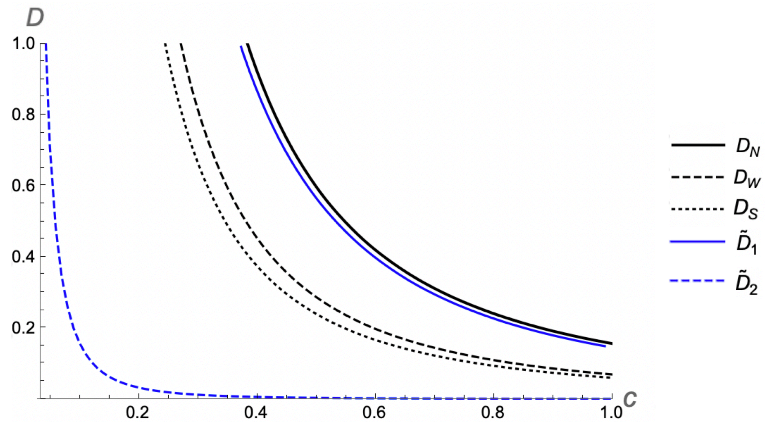


Figure A.1: Equilibrium Data Quantity $D$ under No, Weak and Strong Data Neutrality Regulations vs. the Equilibrium Data Quantity $\tilde{D}_j$ under Alternative Remedy: $\alpha = 1/2$, $\bar{q} = 2$, and $\bar{\theta} = 3/2$.

# B Omitted Details for the Equilibrium Price and Quantity Derivations in Section 3.1

From the utility specification, $u_{ij} = \theta_i \bar{q}_j (1 + d_j^\alpha) - p_j$, two marginal valuations ($\theta^H$ making a consumer indifferent between buying from high- and low-quality sellers, and $\theta^L$ making a consumer indifferent between buying from low-quality seller and not buying) are given by:

$$\theta^H = \frac{p_1 - p_2}{\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}; \quad \theta^L = \frac{p_2}{1 + d_2^\alpha}. \tag{B.1}$$

Then, the market share for each seller is given as follows.

$$x_1(\mathbf{p}, \mathbf{d}) = \frac{1}{\bar{\theta}} \left(\bar{\theta} - \theta^H\right); \quad x_2(\mathbf{p}, \mathbf{d}) = \frac{1}{\bar{\theta}} \left(\theta^H - \theta^L\right). \tag{B.2}$$

Given the market share, each seller's first order condition with respect to price, denoted as $\zeta_{p_j}$, is given as follows.

$$\zeta_{p_1} = 1 - \frac{2p_1 - p_2}{\bar{\theta}[\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)]} = 0; \quad \zeta_{p_2} = \frac{1}{\bar{\theta}} \left[\frac{p_1 - 2p_2}{\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)} - \frac{2p_2}{1 + d_2^\alpha}\right] = 0. \tag{B.3}$$

Solving two first order conditions in (B.3) yields the following prices:

$$p_1(\mathbf{d}) = \frac{2\bar{q}\bar{\theta}(1 + d_1^\alpha)[\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)]}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}; \quad p_2(\mathbf{d}) = \frac{\bar{\theta}(1 + d_2^\alpha)[\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)]}{4\bar{q}(1 + d_1^\alpha) - (1 + d_2^\alpha)}. \tag{B.4}$$

Consequently, the market share is given as follows.

$$x_1(\mathbf{d}) = \frac{2\bar{q}(1+d_1^\alpha)}{4\bar{q}(1+d_1^\alpha)-(1+d_2^\alpha)}; \quad x_2(\mathbf{d}) = \frac{\bar{q}(1+d_1^\alpha)}{4\bar{q}(1+d_1^\alpha)-(1+d_2^\alpha)}. \tag{B.5}$$

# C Robustness with Respect to Match Concavity and Consumer Dispersion

In this appendix, we examine the robustness of our main results with respect to two important model primitives: the concavity of the match function, captured by $\alpha$ (which determines the marginal impact of data on targeting quality), and the dispersion of consumer valuations, captured by $\bar{\theta}$ (which determines the intensity of downstream price competition).

For comparative statics, we numerically solve for the equilibrium and visualize some of the results below, assuming that $c = 1/2$ and $\bar{q} = 2$ and focusing on the case without data neutrality. Note that the comparative statics show that our central mechanism is not knife-edge: changes in match concavity and consumer dispersion can materially shift equilibrium data provision, market outcomes, and welfare, but they do so precisely through the same trade-off emphasized in the main text—between the competition-enhancing effect of broader data access and the platform's incentive to reduce the amount of data supplied under neutrality.

First, the larger the marginal returns to data (with a larger $\alpha$) are, the greater the amount of data that the platform wants to provide in the market under no regulation (i.e., $D_N$ increases). Additionally, the more dispersed consumers are, the greater the amount of data in the market because the downstream market becomes less competitive with more dispersed or differentiated consumer types, which allows the integrated firm to charge a higher price when providing better targeting with a larger amount of data. The upward pressure on $D_N$ arising from the larger marginal returns to data becomes greater as consumers are more dispersed: more dispersed consumer distribution makes the platform's willingness to provide data more sensitive to the increase in the marginal match benefit, as depicted in the left panel of Figure C.1.

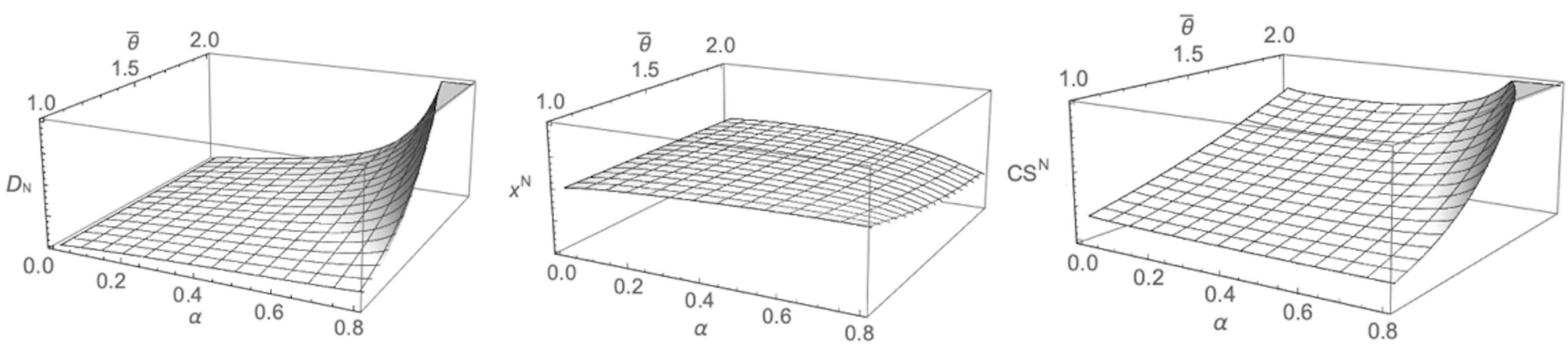


Figure C.1: Equilibrium $D$, $X$, and $CS$ under No Data Neutrality Regulation as $\alpha$ and $\bar{\theta}$ Change.

Second, the total size of the market, denoted as $X^N = x_1^N + x_2^N$, can expand or shrink as $\alpha$ and $\bar{\theta}$ change. The dispersion of consumers only indirectly affects the total market share through its effect on $D_N$, and we showed that a greater dispersion leads to a larger amount of data. Under no data neutrality, this greater amount of data is offered only to the affiliated seller. Thus, both more dispersed consumers and larger marginal returns to data increase the differentiation of the two sellers. On the one hand, larger $\bar{\theta}$ and $\alpha$ lead the unaffiliated seller to charge a lower price due to competitive pressure, which expands the market. On the other hand, more targeting differentiation leads to soft price competition, allowing the unaffiliated seller to raise its price. As depicted in the middle panel of Figure C.1, as consumers become more dispersed, the market size-reducing effect is sufficiently large to dominate the market size-increasing effect, such that the total market share decreases.

Finally, when consumers are more dispersed, it results in greater consumer welfare: even if more dispersion partly reduces consumer welfare due to soft price competition, its data amount-increasing effect is sufficiently large to make consumers better off. Additionally, as the marginal returns to data become larger, consumers become better off, and such positive effects of a larger $\alpha$ become greater as $\bar{\theta}$ increases, as depicted in the right panel of Figure C.1. That is, the combined effects of larger $\alpha$ and $\bar{\theta}$ on a greater amount of data are sufficiently large to offset any negative effects from the market size-reducing effect.

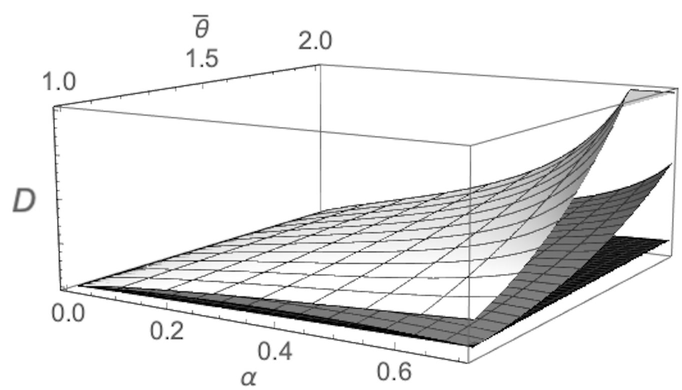


Figure C.2: Equilibrium $D$ under Different Regulatory Regimes as $\alpha$ and $\bar{\theta}$ Change: $\bar{q} = 2$ and $c = 1/2$
*Notes:* White: No regulation, Gray: Weak data neutrality, Black: Strong data neutrality.

For either weak or strong data neutrality, we find qualitatively similar results as above. For example, as shown in Figure C.2, which depicts $D_N$, $D_W$, and $D_S$, the effects of $\alpha$ and $\bar{\theta}$ on the equilibrium amount of data are qualitatively similar between different scenarios. Thus, the welfare and policy implications depend on many different aspects of the market, such as how large the marginal returns to data are and how dispersed consumers are. If consumers are less differentiated and the marginal match benefits are relatively negligible, there is no significant difference in the equilibrium amounts of data between the with and without data neutrality scenarios, such that the competition-enhancing effect is more likely to dominate the data-amount-reducing effect; data neutrality is more likely to be welfare-enhancing than no regulation as $\bar{\theta}$ and $\alpha$ become smaller. However, if the marginal returns to data are relatively large and consumers are sufficiently differentiated, the negative aspects of data neutrality, compared to no regulation, in terms of less available data become more salient, thereby making data neutrality less socially desirable.