Paying Consumers for Their Data: An Economic Analysis of Data Acquisition and Digital Privacy*

Zhijun Chen†

January 24, 2025

Abstract

Digital firms offer digital products to consumers and collect consumer data as a by-product of their usage. This data acquisition generates both data-monetization revenue and data-driven consumer benefits, while imposing privacy costs on consumers. The paper explores compensation schemes for consumer data, focusing on the interdependence between data collection and digital product provision, and examining the role of cross-subsidization in compensation mechanisms. We analyze the optimal compensation scheme for a monopolistic digital firm, examine the impact of data acquisition on competition, and investigate personalized pricing in the context of consumer privacy choices. Our findings offer valuable policy implications for digital privacy regulations and competition policies related to data collection.

Key words: Consumer Data, Digital Privacy

JEL codes: D47, L11, L40, K21

*I gratefully acknowledge financial support from the Australian Research Council (grant number DP210102015). I thank Jacques Cremer and Patrick Rey for their helpful comments, and audiences at numerous seminars for their insightful comments.

†Monash University. Email: zhijun.chen@monash.edu.

1 Introduction

1.1 Motivation

Recent developments in data analysis technology and computational infrastructure have facilitated enormous growth in the scale and precision of consumer data acquisition.[1] These advances have sparked a trend in digital transformation, enabling firms to explore new products and services and adopt innovative business models. Consequently, consumer data has become a fundamental resource for the modern digital economy.

Consumer data is traded as a commodity in a market with tremendous value. The market for trading consumer data, known as the consumer data broker market, is rapidly growing and plays a crucial role in providing businesses with valuable insights into consumer behavior, preferences, and interests. According to a 2023 report on the global data broker market, it was valued at US$319.030 billion in 2021 and is expected to grow at a rate of 7.96% over the forecast period, reaching US$545.431 billion by 2028.[2]

Despite the immense value of consumer data and the rapid growth of the data broker market, there is a notable absence of a direct market for consumer data acquisition. Consumer data is primarily generated through the use of digital products offered by platforms such as search engines, online marketplaces, social networks, app stores, video-sharing platforms, and smart devices. These platforms typically provide their products to consumers "for free," subsequently collecting and monetizing consumer data without offering direct monetary compensation to users.

However, this free-product-for-data business model has been recently criticized and challenged by economists and legal scholars, who argue that digital platforms should share a portion of their substantial annual profits with consumers. There is growing advocacy for offering monetary payment to consumers for their data, as highlighted in the pioneering article by Posner and Weyl (2018).[3] In the influential report, "*Market Design for Personal Data*" (2023), Bergemann

[1] Consumer data refers to the behavioral, demographic, and personal information trail that consumers leave behind as a result of their internet use. The terms "consumer data" and "personal data" are often used interchangeably. However, we prefer to use "consumer data" to emphasize the commercial purpose of such data.

[2] See https://www.knowledge-sourcing.com/report/global-data-broker-market

[3] Andrew Yang, an American businessman and politician, recently launched the Data Dividend Project to try to establish data as property rights (https://www.datadividendproject.com/). However, his project has been criticized as useless (https://www.vice.com/en/article/935358/andrew-yangs-data-dividend-isnt-radical-its-useless). A growing number of companies have also endeavored to upend the existing

et al. point out that “users generate a resource of tremendous value–personal information–and yet firms extract this resource without payment (other than the provision of digital goods and services in exchange). This exchange stands in sharp contrast to what we see in other markets, in which who control resources are paid for their extraction or use.”

The report advocates for establishing a market for data collection, a proposal that aligns with recent digital privacy regulations, such as the General Data Protection Regulation (GDPR) and the Digital Markets Act (DMA). These regulations require digital platforms to obtain consumer consent for data collection while continuing to offer their digital products even if consent is denied.

At the core of this advocacy and debate are two fundamental questions: What is the economic nature of consumer data, and how should consumers be compensated for their data?

In this paper, we explore these questions. Consumer data is a unique commodity, defined by two fundamental characteristics. First, it is generated as a by-product when consumers use digital products, emphasizing the interdependence between digital products and data acquisition. This interdependence makes it technically challenging to assign control rights over data to consumers. Second, the acquisition of consumer data imposes privacy costs on individuals, creating a negative externality that necessitates a well-designed compensation mechanism to address these costs.

We develop a theoretical model to analyze data acquisition and consumer compensation, incorporating the key features of data collection. In this model, a digital firm offers a digital product as the primary good, providing value to consumers while generating consumer data as a by-product of its use. The cost of providing the digital product increases with its quality (or value) and follows a convex curve, reflecting economies of scale. Data acquisition generates two types of benefits: (1) data-monetization revenue, earned by the digital firm through targeted advertising, personalized pricing, and the direct sale of consumer data; and (2) data-driven benefits, which consumers enjoy through the enhanced and personalized functionality of the digital product. However, data acquisition also imposes a privacy cost on consumers, which varies across individuals.

---

relationship between consumers and tech companies by offering monetary compensation for their data (https://www.washingtonpost.com/technology/2023/02/06/consumers-paid-money-data/).

1.2 Results

Digital firms often bundle their digital products with consumer data collection as the default option. However, regulatory compliance requires them to also offer their digital products on a stand-alone basis for consumers who decline data sharing. Consumers who choose the stand-alone product derive its base value, whereas those who opt for the bundle benefit from enhanced functionality by sharing their data.

First, we characterize the optimal scheme for a monopoly digital firm adopting mixed bundling, where consumers with relatively high privacy costs choose the stand-alone product, while those with lower privacy costs purchase the bundle at a discounted price. Through this mixed bundling scheme, the monopoly captures the total benefits of data acquisition while compensating each consumer for their data with a payment equal to the privacy cost of the marginal consumer who is indifferent between the two options. If the revenue from data monetization is sufficiently large, the firm optimally provides a monetary subsidy to consumers who choose the bundle, in addition to offering the "free" digital product. This creates cross-subsidization between the two digital activities. Compared to a benchmark scenario without data collection, the monopoly earns higher profits, offers a more valuable product, and enables consumers to derive greater utility by sharing their data. In cases where digital firms face a non-negative price constraint, offering the bundle for free becomes the constrained optimal solution.

The analysis of the monopoly firm's optimal scheme yields several policy implications.

Impact of digital privacy regulations. In November 2023, Meta introduced a binary "pay or consent" model to comply with GDPR requirements, charging consumers a fee for using Facebook and Instagram if they do not consent to data collection while continuing to offer the bundle for free. However, in July 2024, the European Commission (EC) determined that Meta's "pay or consent" model violated the DMA and mandated that Meta provide free access to its digital products without monetizing consumer data. Essentially, the DMA model allows consumers who opt out of data collection to free-ride on digital products, which negatively impacts the provision of such products. Compared to Meta's model, which aligns with the optimal mixed bundling scheme under a non-negative-price constraint, our analysis indicates that the DMA model results in lower product quality and reduced participation in data sharing. This leads to decreased revenue for digital firms and diminished incentives for investment in digital products.[4] Conversely, the DMA model may increase consumer surplus by raising the

[4] These findings align with evidence from recent empirical studies, such as Aridor et al. (2020) and Jia et al. (2021).

baseline utility for all consumers, particularly when the data-monetization revenue is sufficiently large.

Data property rights. Bergemann et al. (2023) advocate for granting consumers data property rights and propose a competitive market of data intermediaries that collect and monetize data on consumers' behalf. A data intermediary, while monetizing consumer data, does not provide digital products and therefore cannot generate data-driven consumer benefits. It exerts competitive pressure on the digital firm only if it is efficient in data monetization. In such cases, competition for data collection enhances both product quality and consumer surplus when the digital firm continues to collect data. However, if the intermediary is highly efficient in data monetization, it dominates the competition for data collection, leaving the digital firm to focus exclusively on the digital product. In this equilibrium, consumers benefit from data-monetization revenues provided by the intermediary but lose the data-driven benefits previously offered by digital firms. This mechanism results in higher consumer surplus but lower product quality. The breakdown of the twofold benefits of data acquisition ultimately leads to a welfare loss.

Second, we analyze the impact of data acquisition on competition, focusing on two applications.

Data driven mergers. In the recent wave of data-driven mergers, such as the controversial Google-Fitbit merger in 2021 and the Amazon-iRobot merger in 2022, competition authorities have expressed concerns that dominant digital platforms may leverage data acquisition to foreclose competitors. This paper explores two opposing effects of data acquisition on competition. First, data acquisition generates substantial benefits, which can be leveraged by the merged firm as a competitive advantage, thereby intensifying competition. Second, the heterogeneity of consumer privacy costs in data acquisition facilitates product differentiation, which can soften competition. The interplay of these effects leads to varied impacts on competition and consumer welfare. When the merged firm's advantage over its rival is relatively small, its commitment to pure bundling softens competition, raising the rival's price and profit while reducing consumer welfare.[5] In such scenarios, banning pure bundling can enhance consumer welfare. Conversely, when the merged firm has a significant advantage over its rival, the merger intensifies competition, reducing digital product prices, increasing consumer surplus, and ultimately enhancing overall welfare.

Data privacy protection. Since the enactment of GDPR, digital platforms have increasingly

[5] A review of the pricing history for Fitbit and Apple smartwatches suggests that the merger did not intensify competition between these products.

standardized their data collection practices and enhanced consumer data protection. A notable example is Apple, which introduced privacy labels in December 2020 and launched a new operating system in 2021 emphasizing data privacy. We analyze a competition game where two digital firms choose their business models—whether to engage in data acquisition—before entering a pricing game. Our analysis shows that differentiation in business models arises as the unique subgame perfect Nash equilibrium: one firm opts for data acquisition (engaging in pure bundling), while the other abstains. Apple's strategic positioning as a leader in data privacy protection exemplifies this equilibrium, despite its capacity for data acquisition. By distinguishing its business model from competitors, Apple effectively softens competition. As a potential remedy, prohibiting pure bundling could enhance competition in the digital product market.

Third, we extend our analysis to scenarios where consumers exhibit heterogeneous preferences for data-driven benefits, and a dominant digital firm leverages data collection to implement personalized pricing. This approach enables the firm to extract consumer surplus by tailoring prices to match each consumer's utility from outside options. While personalized pricing can intensify competition and reduce the rival's profitability, the dynamics shift when consumers retain the choice to share their data and their privacy sensitivity remains private information.

In competition with a rival firm using uniform pricing, a neutrality result emerges: the rival's price and profit remain unchanged regardless of whether the dominant firm employs personalized or uniform pricing.[6] Compared to uniform pricing, personalized pricing benefits consumers with lower data-driven preferences while disadvantaging those with higher preferences. This pricing strategy reduces overall consumer surplus and total social welfare, despite increasing the dominant firm's profitability.

The welfare effects change fundamentally when the dominant firm personalizes offers based on heterogeneous data-monetization revenues from each consumer. In this case, personalized pricing encourages participation from consumers with higher data-monetization value, allowing the dominant firm to generate greater revenue while also benefiting these consumers. The neutrality result persists, but personalized pricing increases the dominant firm's profits, improves consumer surplus, and enhances total social welfare.

1.3 Related Literature

This paper contributes to the extensive literature on data collection and intermediation. The first stream of this literature views consumer data as an informative signal used to predict consumer

[6] We demonstrate this result under the assumption of a uniform distribution of privacy costs.

preferences, enabling digital firms to offer personalized pricing and product recommendations. Acemoglu et al. (2022) show that digital platforms tend to over-collect data due to information externalities and identify conditions under which shutting down data markets could enhance welfare. Bergemann et al. (2022) propose a model of data intermediation to examine the incentives for sharing individual data in the presence of information externalities. Ichihashi (2021) investigates how complements or substitutes to consumer signals influence the equilibrium price of individual data under information externalities.

In settings without information externalities, Ichihashi (2020) explores personalized pricing and product recommendations, demonstrating that sellers benefit from committing not to use consumer information for pricing. Ichihashi (2023) develops a dynamic model of consumer privacy and platform data collection, highlighting a cross-period effect on privacy protection in data collection. Bergemann and Bonatti (2024) examine revenue-optimal mechanisms for selling consumer data to competing sellers through managed advertising campaigns that efficiently match products and preferences. Argenziano and Bonatti (2024) analyze data linkages, where a data-collecting firm interacts with a data-receiving firm, endogenizing linkage formation under digital privacy regulations.

This paper distinguishes itself from the above literature in two key aspects. First, it conceptualizes consumer data as a by-product of using digital products, emphasizing the interdependence between these two components. This interdependence facilitates cross-subsidization in compensation schemes and creates complementarity between digital activities. Second, while the existing literature often focuses on microfoundations of the benefits and costs of data collection, this paper adopts a reduced-form approach to these benefits, incorporating heterogeneity in privacy costs. This simplification enables an analysis of compensation schemes that treats consumer data and digital products as a bundled offering, providing a distinct perspective on the impacts of data acquisition and deriving novel policy implications.

The second stream of literature focuses on digital privacy protection and data property rights, examining the social benefits and privacy costs of data collection. Choi et al. (2019) present a model that incorporates information externalities and investigates the impact of GDPR on digital privacy protection. Fainmesser et al. (2021) develop a framework in which digital platforms determine the scale of data collection and associated privacy protection strategies, noting that data breaches impose privacy costs on users. Jullien et al. (2020) study data leakages as reduced-form negative consequences of information diffusion. Markovich and Yehezkel (2023) analyze the equilibrium outcomes of data collection under different data control regimes, focusing on three distinct benefits and privacy costs while accounting for two forms of heterogeneity. Dosis and Sand-Zantman (2023) use a mechanism design approach to explore how assigning property rights over consumer data affects market outcomes. Chen (2022) treats data and analytics as distinct strategic variables and analyzes the optimal mechanisms for data acquisition to maximize social welfare. A closely related paper is Markovich and Yehezkel (2024), which explores a digital platform's strategic choice among a data-based model, a subscription-based model, or a hybrid of both, accounting for network effects.[7]

These studies primarily assume that the provision of digital products is exogenously given, focusing on the social benefits and privacy costs of data collection. In contrast, this paper treats digital products and data collection as a bundled offering and underscores the inherent cross-subsidization between two digital activities—providing digital products and acquiring data—in designing consumer compensation schemes.

Our paper also contributes to the growing literature on personalized pricing, initiated by the seminal work of Thisse and Vives (1988). Shaffer and Zhang (2002) explore personalized pricing when one firm enjoys a brand advantage over its competitor, while Chen and Iyer (2002) examine scenarios where firms must advertise to reach consumers before implementing personalized pricing. Montes, Sand-Zantman, and Valletti (2019) investigate whether data intermediaries should sell consumer data exclusively when firms use it for price personalization. Chen et al. (2020) analyze competitive personalized pricing in settings where consumers take proactive measures to protect their data. Chen et al. (2022) study the impact of data-driven mergers on personalized product recommendations and personalized pricing strategies. Rhodes and Zhou (2024a) address personalized pricing in a general oligopoly model, identifying conditions under which it harms firms while benefiting consumers.

These studies, however, largely overlook consumer privacy choices.[8] In contrast, our paper examines personalized pricing in a framework where consumers can choose whether to share their data, and where privacy costs are heterogeneous. We identify its neutrality impact on competitors and uncover varying implications for consumer surplus and social welfare, depending on the nature of heterogeneous preferences and benefits.

[7] A detailed discussion of the relationship between the two papers is provided in Sections 4 and 5.

[8] Rhodes and Zhou (2024b) is an exception. The detailed relationship with their paper is discussed in Section 5.

2 The Model

2.1 Setup

A digital firm supplies a digital product at a quality level $q \in [0, +\infty)$. The firm incurs a total cost $C(q)$, which is increasing, convex, and satisfies $C(0) = 0$, reflecting economies of scale in the supply of the digital product. The market consists of a continuum of consumers, with the total population normalized to a unit mass. Consumers have homogeneous preferences for the quality of the digital product, deriving a base utility q. Their reservation value for outside options is normalized to zero for simplicity.

Consumer data is generated through the use of the digital product. This data, processed by the digital firm at zero marginal cost, is treated as a by-product. For clarity, the digital product is referred to as the primary product (A), and consumer data as the secondary product (D). The quantity of data collected from each consumer is normalized to 1.

Data collection generates two types of benefits:

- **Data monetization revenue:** The digital firm can monetize consumer data through various channels, including the sale of targeted advertisements, the provision of data-driven services to other businesses, and direct sales of products or services to consumers. Let $r \geq 0$ denote the monetary value of a data record. In the baseline setting, we assume r to be identical across consumers, while its heterogeneity is examined in Section 5.

- **Data driven consumer benefit:** The digital firm can also use consumer data to enhance the value of its digital product by integrating it into its digital ecosystem and personalizing its functionality. The enhanced value of the digital product is expressed as bq, where $b \geq 1$ represents the multiplier reflecting the data-driven consumer benefit. We assume b is homogeneous across consumers in the baseline analysis but account for its heterogeneity in Section 5.

In this setting, the digital firm selects the quality level q and determines a corresponding price p, while treating both r and b as exogenously given.

Consumers, however, incur a privacy cost when their personal data is collected by the digital firm. This privacy cost, reflecting consumers' sensitivity to privacy, is represented by the parameter θ, which is distributed over the interval $\left[0, \bar{\theta}\right]$ according to the cumulative distribution function F, with a strictly positive density f. Assume $\bar{\theta}$ is sufficiently large such that the equilibrium results in partial participation—i.e., consumers with very high privacy sensitivity opt out of sharing their data.

The following assumption ensures the uniqueness of the equilibrium:

Assumption A: The hazard rate functions $h \equiv F/f$ and $k \equiv (1-F)/f$ are monotonic, where h is increasing and k is decreasing.

2.2 Discussion of modeling features

Data acquisition. The digital firm collects consumer data through various tracking technologies embedded in digital applications used by consumers.[9] The collected raw data, characterized by its unstructured nature and complex attributes, requires processing to extract valuable insights. We consider the digital firm as an entity that has developed comprehensive data analytics capabilities, encompassing infrastructure for data collection, storage, analysis, and interpretation.

The processed consumer data, comprising various attributes and types of information, is categorized and stored by the digital platform as a dataset (or data record) in accordance with industrial standards. For instance, Acxiom, a leading personal data broker, reports that an average consumer data record includes approximately 1,500 pieces of information or attributes. Data brokers typically sell these records in bulk, rather than by individual attributes, and the monetary value of personal data is significantly influenced by the size and completeness of the dataset.[10]

The digital firm can control the quantity of data collected from each consumer. However, the verifiability of the firm's commitment to a specific data scale depends on the regulatory environment. In the absence of regulations or standardized data acquisition practices, most digital firms collect consumer data without disclosing detailed information about the scale or nature of the data being gathered. Under such non-transparent practices, any commitment by a digital firm to a specific data scale becomes unverifiable. As a result, consumers rationally anticipate that digital firms will collect the maximum permissible amount of data, normalized to 1.

In an extension, we consider scenarios where the firm's commitment to a specific data scale

[9] One of the most common methods of data collection involves the use of cookies—small text files placed on a user's device by a web browser or other digital applications when visiting websites or using apps.

[10] For an in-depth discussion of the pricing factors and formulas for personal data, see Malgieri and Custers (2018).

is verifiable, either through oversight by an independent third party or a regulatory authority. In such cases, the firm can optimally determine and commit to a specific data scale.

Data monetization revenue. In addition to selling consumer data directly to brokers, digital firms monetize their data through sponsored content and targeted advertising. Advertising markets are typically two-sided, where advertisers aim to connect with users most likely to be interested in their products. Consumer data is collected using cookies, with each internet user associated with a unique cookie ID. Digital platforms monetize on this data by facilitating matching advertisements through auctions, predominantly first-price auctions.

In these auctions, advertisers participate indirectly via "Demand Side Platforms" (DSPs), which represent advertisers and manage the vast array of advertising opportunities online. Major platforms like Amazon, Facebook, and Google operate their own DSPs. When third-party DSPs are involved, they receive a user's cookie ID before bidding. This identifier enables DSPs to track the user's behavior, estimate preferences, and identify the advertiser best suited to the user. The DSP then submits a bid on behalf of the advertiser.

When a consumer visits the platform, its tracking technology identifies the user's cookie ID. The DSP generates a public signal for advertisers regarding the consumer's preferences, facilitating a match with the most relevant seller or advertiser. The revenue generated by this match is shared among the advertising intermediaries (DSPs) and the digital platform.

Recently, dominant platforms have introduced a new approach called managed campaigns for selling advertisements, including sponsored search, display advertising, and sponsored product listings. In a managed campaign, the platform charges a lump-sum fee for a bundle of realized consumer matches (a fixed advertising budget). Sellers specify high-level objectives for their campaigns, while the task of targeting individual consumers is handled by the platform's "auto-bidding" algorithms, such as Google Performance Max or Meta Advantage+. Bergemann and Bonatti (2024) analyze managed campaigns and show that the revenue-maximizing mechanism is a managed advertising campaign that efficiently matches products with consumer preferences.

The baseline analysis examines applications of data analytics that primarily rely on within-user data aggregation, aligning closely with business models for targeted advertising and personalized recommendations. In this case, the digital firm's data monetization revenue from one consumer is independent of revenues from other consumers. Our analysis can be extended to settings involving positive data externalities driven by cross-user data aggregation. A detailed exploration of these scenarios is provided in Online Appendix D.

Data driven benefits. We model data-driven consumer benefits as the enhanced value of the digital product, represented by the function bq, where the multiplier b captures the digital platform's comparative advantage in leveraging consumer data. This approach aligns with the development of AI-powered technology in digital products and reflects the advantage digital firms derive from their ecosystems. For example, following the Google/Fitbit merger, the Fitbit smartwatch integrates seamlessly with other Google devices and apps within its ecosystem. Users' health data can be shared with applications like Google Health, enabling personalized recommendations. Consequently, consumers derive substantial benefits from sharing their data.

Privacy costs. Privacy has been defined in various ways, as highlighted in a comprehensive review by Acquisti et al. (2016). Becker (1980) distinguishes between intrinsic and instrumental privacy preferences. Intrinsic preferences relate to the moral value of privacy, often tied to concepts of human dignity, autonomy, and independence (Westin, 1967; Schoeman, 1992). Instrumental preferences, by contrast, are endogenously determined by how private information is used in transactions. For example, consumers may be concerned that firms could exploit their data to implement personalized pricing based on their willingness to pay. Experimental evidence from Lin (2022) reveals significant heterogeneity across consumers in both intrinsic and instrumental privacy preferences.

The literature also diverges in its approaches to modeling the costs associated with privacy loss due to data collection. The first stream of literature views consumer data as private information related to consumer preferences, focusing on instrumental privacy preferences. In this framework, a loss of privacy results in the revelation of private information, enabling observers to form more precise posterior beliefs about an individual's preferences or behavior. Eilat, Eliaz, and Mu (2021) quantify privacy as the Kullback-Leibler divergence between prior and posterior belief distributions about an agent's type. Bonatti et al. (2023) develop a theory of privacy centered on the concavity or convexity of the indirect utility function with respect to market beliefs. In their model, a loss of privacy leads to a mean-preserving spread of posterior beliefs about an individual's type or behavior. Using Jensen's inequality, they demonstrate that the concavity (or convexity) of the indirect utility function determines whether privacy is valuable (or not valuable) to individuals.

The second stream of literature treats privacy concerns as intrinsic preferences, where a loss of privacy imposes a direct cost on consumers. Representative works include Markovich and Yehezkel (2023, 2024), Dosis and Sand-Zantman (2023), and Rhodes and Zhou (2024b). This paper aligns with the second stream by considering a consumer's privacy sensitivity as indicative of their intrinsic privacy preferences.

3 Monopoly digital firm

Suppose the digital firm operates as the monopoly supplier of the digital product A. Consider the benchmark case where the monopoly focuses solely on the primary good. It charges the monopoly price $\bar{p}_A = q$, earning a profit of $\Pi(q) = q - C(q)$. The monopoly firm chooses the optimal quality q^0, determined by $C'\left(q^0\right) = 1$, resulting in an equilibrium profit of $\Pi^0 = q^0 - C\left(q^0\right)$.

When the digital firm monetizes consumer data, it bundles data collection with its digital product. The firm can choose between two bundling strategies:

Pure Bundling: The firm offers only the bundle $\mathcal{B}$, ceasing to sell the primary good A on a stand-alone basis.

Mixed Bundling: The firm offers both the bundle $\mathcal{B}$ and the stand-alone product A separately, allowing consumers to choose between the two options.

3.1 The optimal scheme

The monopoly firm chooses mixed bundling as its optimal strategy. It offers the bundle scheme $S = \{bq, p\}$, where bq is the committed consumer utility and p denotes the price (or subsidy). If $p > 0$, the firm charges consumers a fee for the digital product. Conversely, if $p < 0$, the firm provides consumers with a monetary subsidy. In addition, the monopoly offers the stand-alone product scheme $S_A = \{q, p_A\}$. In equilibrium, the firm sets $p_A = q$ to extract the full consumer surplus for the stand-alone option.

Consumers derive a net utility $U(\theta) = bq - p - \theta$ from consuming the bundle, while receiving a net utility of zero if they opt for the stand-alone product A instead.[11] Therefore, consumers with their privacy sensitivity $\theta \leq \tau$ choose the bundle, where τ is the cut-off threshold for participation, defined as $\tau \equiv bq - p$.

Under this mixed bundling strategy, the monopoly's profit is expressed as:

$$\Pi(p, q) = F(\tau)(r + p) + (1 - F(\tau))\, q - C(q),$$

where $F(\tau)(r + p)$ represents the revenue generated from consumers who choose the bundle, $(1 - F(\tau))\, q$ captures the revenue from consumers choosing the stand-alone product, and $C(q)$ is the total cost for providing the digital product.

Using the pricing equation $p = bq - \tau$, which represents the transfer of value between the

[11] When indifferent, consumers prefer the stand-alone good A over the outside option.

firm and consumers, the firm can alternatively optimize its profit by choosing τ and q directly:

$$\Pi(\tau, q) = F(\tau)((b-1)q + r - \tau) + q - C(q). \tag{1}$$

In addition to the profit from supplying the digital product, $q - C(q)$, the monopoly extracts the added value of the bundle, $(b-1)q + r$, where $(b-1)q$ is the net data-driven benefit while r is the data-monetization revenue. In contrast, the firm compensates each consumer at τ, equal to equal to the privacy cost of the marginal consumer, τ, where $\theta = \tau$.

An increase in the quality q generates a marginal benefit of $(b-1)F(\tau) + 1$ from consumers, while the digital firm incurs a marginal cost of $C'(q)$. Equating these two terms determines the optimal quality $q_m^r(\tau)$ (where the subscript m stands for the optimum under mixed bundling):[12]

$$C'(q) = 1 + (b-1)F(\tau). \tag{2}$$

The optimal quality $q_m^r(\tau)$ is increasing in τ.

Meanwhile, the optimal participation threshold $\tau_m^r(q)$ is uniquely determined by the following first-order condition:

$$\phi(\tau) \equiv \tau + h(\tau) = (b-1)q + r, \tag{3}$$

where $\phi(\tau)$ is an increasing function.[13] Then, $\tau_m^r(q)$ is also increasing in q.

The equilibrium quality and participation threshold, denoted by q_m^* and τ_m^* respectively, are jointly determined by (2) and (3). We summarize the equilibrium outcomes in the following proposition:

Proposition 1 *Consider a monopoly digital firm that supplies a digital product and collects consumer data. There exists a unique equilibrium under mixed bundling, where consumers with $\theta < \tau_m^*$ choose the bundle, receiving a net utility of $U^*(\theta) = \tau_m^* - \theta$, while those with $\tau_m^* < \theta < \bar{\theta}$ purchase the stand-alone product, obtaining zero net utility. The equilibrium quality q_m^* and participation threshold τ_m^* are jointly determined by conditions* (2) *and* (3). *Furthermore:*

- *Data acquisition improves the product quality if and only if $b > 1$.*
- *Consumers opting for the bundle are compensated with a payment τ_m^*, which increases with b and r.*
- *The firm offers consumers a monetary subsidy if $r > h(\tau_m^*) + q_m^*$.*

[12] It is straightforward to verify that $\Pi(\tau, q)$ is concave in q.

[13] It is straightforward to verify that $\Pi(\tau, q)$ is quasi-concave in τ.

- *Data acquisition enhances the firm's profit, increases consumer surplus, and improves overall social welfare.*

Proof. See Online Appendix A. ■

Proposition 1 addresses the following key questions:

1. What is the impact of data acquisition on the quality of the digital product?

An increase in the data-driven multiplier b raises the marginal benefit of the digital product, thereby increasing its optimal quality. In contrast, an increase in per-consumer revenue r does not directly impact the quality, as shown in (2). However, r indirectly influences q by raising the optimal participation threshold τ. This indirect impact results in higher quality only if $b > 1$. In the absence of the data-driven benefit ($b = 1$), the equilibrium quality remains the same as without data collection, i.e., $q_m^* = q^0$. Therefore, under mixed bundling, data acquisition improves the quality of the digital product if and only if $b > 1$.

The monopoly firm earns additional profit, $F(\tau)((b-1)q + r - \tau)$, from data acquisition. Based on (3), this additional per-consumer profit is equal to $h(\tau)$, and the firm's equilibrium profit is given by:

$$\Pi_m^* = h(\tau_m^*) F(\tau_m^*) + q_m^* - C(q_m^*). \tag{4}$$

2. How should consumers be compensated for their data?

Consumers with $\theta < \tau_m^*$ share their data. The monopoly offers these consumers a compensation scheme $S = \{bq_m^*, p_m^*\}$, from which each consumer receives a payment $\tau_m^* = bq_m^* - p_m^*$ and derives a net utility:

$$U_m^*(\theta) = bq_m^* - p_m^* - \theta = \tau_m^* - \theta. \tag{5}$$

From (3), the equilibrium compensation τ_m^* corresponds to a share of the revenue generated from data acquisition, $(b-1)q + r$. For instance, if θ is uniformly distributed over $[0, \bar{\theta}]$, then $h(\tau) = \tau$, and consumers obtain half of the revenue:

$$\tau_m^* = \frac{(b-1)q_m^* + r}{2}.$$

The monopoly implements the compensation scheme through two components: the enhanced consumer value from the digital product, bq_m^*, and a price p_m^*, given by

$$p_m^* = bq_m^* - \tau_m^* = h(\tau_m^*) + q_m^* - r. \tag{6}$$

This pricing reflects the gap between the combined profit from the bundle, $h(\tau_m^*) + q^*$, and the revenue from data monetization, r. The monopoly offers a monetary payment for data when $r > h(\tau_m^*) + q^*$.

The optimal scheme facilitates cross-subsidization between the two digital activities: providing the digital product and acquiring consumer data. If data acquisition provides no consumer benefit ($b = 1$), the firm earns revenue solely from data monetization. In this case, τ_m^* depends only on r, and the monopoly retains part of the revenue ($r > \tau_m^*$). Since q_m^* is unaffected by r while τ_m^* increases in r, the equilibrium price, $p_m^* = q_m^* - \tau_m^*$, becomes negative when r is sufficiently large. The firm then offers monetary payments in addition to the free access to its digital product.

Conversely, when consumers derive substantial benefits from data provision ($b > 1$), the monopoly can leverage these benefits as part of the compensation scheme, potentially offering consumers more than the revenue from data monetization ($\tau_m^* > r$). In this case, the monopoly might charge a positive price. In the leading example provided below, the equilibrium price becomes negative if and only if $(b+1) < r\left(\bar{\theta} - b(b-1)\right)$, which occurs when r is sufficiently large and/or b is relatively small.

3. What is the impact of data acquisition on social welfare?

Consumers with $\tau_m^* < \theta < \bar{\theta}$ opt for the stand-alone product, receiving the same utility as they would without data collection. However, consumers with $\theta < \tau_m^*$ choose the bundle, gaining higher net utility by sharing their data. Data acquisition generates social benefits, which the monopoly shares with consumers due to the heterogeneity in privacy costs. As a result, data acquisition enhances total social welfare.

Leading example: Assume θ is uniformly distributed in $\left[0, \bar{\theta}\right]$ and $C(q) = q^2/2$. Assume $\bar{\theta} > \max\{r, b^2\}$.

In mixed bundling, the equilibrium quality and participation threshold are given respectively by

$$q_m^* = \frac{(b-1)\,r + 2}{2\bar{\theta} - (b-1)^2},\ \tau_m^* = \frac{(b-1) + r\bar{\theta}}{2\bar{\theta} - (b-1)^2}.$$

The equilibrium quality is increasing in b and increasing in r if $b > 1$. The equilibrium price, given as

$$p_m^* = bq_m^* - \tau_m^* = \frac{(b+1) - r\left(\bar{\theta} - b(b-1)\right)}{2\bar{\theta} - (b-1)^2},$$

is negative if

$$r > \frac{b+1}{\bar{\theta} - b(b-1)}.$$

We briefly discuss two constrained optima before proceeding to policy implications, with detailed analysis provided in Online Appendix A.

Pure bundling. In many cases, digital firms adopt pure bundling as a uniform scheme for data collection and compensation, often in compliance with regulations such as non-discrimination requirements. Under pure bundling, the monopoly firm earns profit exclusively from the bundle, incentivizing it to attract more consumers to choose the bundle compared to mixed bundling, which positively impacts quality. However, the marginal benefit of supplying the digital product is lower than under mixed bundling, as the firm excludes consumers from purchasing the stand-alone product, resulting in a negative effect on product quality. The interplay of these opposing effects can lead to either higher or lower quality than under mixed bundling, depending on the parameters b and r. In a leading example, compared to mixed bundling, pure bundling results in higher product quality and greater consumer surplus when the revenue from data monetization r is sufficiently high.

Non-negative price constraint (NNPC). When the revenue from data monetization r is sufficiently high, it is optimal for the digital firm to offer monetary subsidies to consumers in addition to providing free digital products. However, implementing such payment schemes can be challenging due to the non-negative price constraint. First, digital firms may face financial limitations that restrict their ability to provide cash payments. Second, when the scale of data collection is unverifiable, consumers may exploit the system by creating multiple accounts to receive payments, leading to moral hazard issues. In such scenarios, digital firms may instead offer the bundle of digital products and data provision at zero price. Consequently, the widely adopted "free-product-for-data" business model may represent the optimal scheme under the NNPC. This constraint results in lower product quality and a reduced participation threshold, ultimately diminishing consumer surplus, the digital firm's profit, and overall social welfare.

Remark 1 *Markovich and Yehezkel (2024) examine a monopolist digital firm's choice among a data-based model (equivalent to pure bundling in our framework), a subscription-based model (digital product only), or a hybrid business model (mixed bundling) in the presence of network effects. Their analysis focuses on the "free-product-for-data" approach in the data-based model, where the bundle price is restricted to* $p = 0$*, representing a suboptimal scheme. This assumption leads to equilibrium outcomes that differ from those in our paper. Specifically, in their framework, when data-monetization revenue is relatively small, the monopolist opts for the subscription-based model, without data collection. In contrast, our analysis shows that the mixed bundling model is always optimal, regardless of the value of* r*, as the monopolist can leverage the bundle price* p *to effectively extract consumer surplus.*

3.2 The impact of data privacy regulation

The European Union (EU) introduced the General Data Protection Regulation (GDPR) in 2016 to enhance digital privacy protection. The GDPR grants consumers the right to control their data and manage their privacy through two key principles:

- The notice-and-consent rule: digital platforms must obtain specific and unambiguous consumer consent for data collection.
- The opt-out rule: digital platforms must allow users to access their products even if they refuse consent for data collection.

These principles require digital firms to offer their digital products on a stand-alone basis in addition to any bundles involving data collection. They are further reinforced under the EU's Digital Markets Act (DMA).[14] While the GDPR and the DMA require digital firms to offer stand-alone digital products to opt-out consumers, they do not specify whether these products must remain free. In the absence of such specifications, the EU has taken pre-GDPR business models as the default benchmark. Before GDPR, many digital firms adopted a "free-product-for-data" model, bundling digital products with data collection at zero price for the consumer. Consequently, EU regulators require firms to continue offering free stand-alone digital products to opt-out consumers, while providing additional compensation to consumers who opt into data sharing.

This interpretation is reflected in the European Commission's (EC) recent decision against Meta. Before GDPR, Meta bundled its free digital products, such as Facebook and Instagram, with free data provision. In November 2023, Meta introduced a binary "pay or consent" model in response to regulatory changes, offering EU users two options: either a monthly subscription fee to access an ads-free version of Facebook and Instagram, or free access to Facebook and Instagram with personalized ads. Under this model, consumers who opt into data sharing are offered the same terms as before GDPR, while those opting out are charged a subscription fee.

In July 2024, the EC found Meta's "pay or consent" model in violation of the DMA. The Commission stated:

"Meta's model does not allow users to opt for a service that uses less of their personal data but is otherwise equivalent to the 'personalized ads' based service. To ensure compliance with

[14] Article 5(2) of the DMA requires digital platforms designated as gatekeepers to seek users' consent before combining their personal data with primary digital products. If a user declines consent, the platform must provide a less personalized but equivalent alternative.

the DMA, users who do not consent should still get access to an equivalent service that uses less of their personal data, in this case for the personalization of advertising."[15]

While Meta claims to comply with GDPR and the DMA by offering mixed bundling, the core disagreement lies in how consumer data should be compensated. The EC requires Meta to offer opt-out consumers (i.e., these choose the stand-alone digital product) free access to Facebook and Instagram while compensating opt-in consumers (i.e., these choose the bundle) for their data. In contrast, Meta compensates opt-in consumers with free digital products while charging opt-out consumers a fee. This disagreement fundamentally affects the benchmark utility for opt-out consumers, which, in turn, influences the utility for opt-in consumers.

To explore the impact of GDPR on data collection and digital service provision, we analyze a monopoly firm's optimal pricing under the DMA requirement.

To comply with the DMA model, the firm must offer $S_A = \{q, 0\}$ for the stand-alone product, in addition to its bundle scheme $S = \{bq, p\}$. Opt-out consumers receive a net utility q from the stand-alone product, while opt-in consumers receive $U(\theta) = bq - p - \theta$ from the bundle. The cut-off threshold between these two options becomes $\tau = (b-1)q - p$.

Essentially, opt-out consumers are free-riding on digital products, causing a negative impact on the provision of the digital product. In particular, when data-sharing does not improve the consumer value from the digital product ($b = 1$), opt-out consumers can enjoy the full benefit from the digital product without sharing their data. Consequently, the firm has no incentives to provide the digital product (where the subscript d stands for DMA): $q_d^* = 0$.[16]

When $b > 1$, the free-riding from opt-out consumers raises the firm's cost for the digital product by $F(\tau)q$. The platform's profit is now given by

$$\begin{aligned} \Pi_d(q,\tau) &= F(\tau)((b-1)q + r - \tau) - C(q) \\ &= F(\tau)(bq + r - \tau) - F(\tau)q - C(q). \end{aligned}$$

The equilibrium quality q_d^* and participation threshold τ_d^* are determined jointly by the following FOCs:

$$C'(q) = (b-1)F(\tau), \tag{7}$$

$$\phi(\tau) = \tau + h(\tau) = (b-1)q + r. \tag{8}$$

[15] See https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3582

[16] For continuity of the equilibrium, we assume that the digital firm can monetize consumer data without providing the digital product, paying consumers at $p_d^* = -\tau_d^*$ for their data when $q_d^* = 0$.

In contrast, Meta offers two schemes for the bundle and the stand-alone digital services: $S_A = \{q, p_A\}$ and $S = \{bq, 0\}$. These may represent the optimal mixed bundling strategy under the non-negative price constraint (NNPC). Without NNPC, Meta would adopt the optimal mixed bundling strategy, with the equilibrium outcomes q_m^* and threshold τ_m^* determined in Subsection 3.1.

Comparing the equilibrium outcomes under these two scenarios, the following proposition highlights the impact of the DMA model on product quality and consumer surplus:

Proposition 2 *Suppose the digital firm is required to follow the DMA model. The firm has no incentives to provide digital products when $b = 1$. Assume $b > 1$:*

- *compared to the optimal mixed bundling without NNPC, the DMA model reduces the quality and the participation threshold: $q_d^* < q_m^*$ and $\tau_d^* < \tau_m^*$, while it can increase consumer surplus by raising the benchmark utility;*
- *compared to the equilibrium under NNPC, the DMA model compels the firm to offer a monetary payment for opt-in consumers, and reduces the product quality in the leading example.*

Proposition 2 indicates economic impact of the regulation such as GDPR and DMA.

First, the regulation leads to lower quality of digital products and less participation for data sharing, reducing digital firm's revenue. Suppose the digital firm is not subject to NNPC. Compared to the equilibrium under mixed bundling, the response function for the optimal participation threshold remains the same $\tau_d^r(q) = \tau_m^r(q)$, while the response function for the optimal quality decreases: $q_d^r(\tau) < q_m^r(\tau)$. This implies $q_d^* < q_m^*$ and $\tau_d^* < \tau_m^*$. When the digital firm faces the binding NNPC, the same outcome holds in the leading example. The negative impact becomes serious when b is close to 1, in which case the digital firm has little incentive to invest in digital products.

The negative impact on digital platforms' data collection and their profitability have been widely observed. One of the most common ways consumer data is collected is through the use of specifically placed small text files, called "cookies". An empirical study by Aridor et al. (2023) indicates a 12.5% drop in total cookies, while Johnson et al. (2020) find that the significant reduction of cookies reduces a platform's revenue by 52% from opt-out consumers.

The decrease of revenue has also led to a significant reduction of investment in the digital sector. Jia et al. (2021) study GDPR's impact on venture investment in new and emerging technology firms. They find that shortly after GDPR's rollout, the venture investment by digital firms in the EU dropped by more than 30% relative to their counterparts in the U.S. and the rest of the world. They also find that the negative effect of GDPR on investment appears particularly pervasive for firms relying heavily on consumer data, including those in the healthcare and finance sectors. The significant reduction of investments in the digital sector might cause a long-run negative effect on social welfare, since the innovation in data science becomes the main driving force of economic growth.

Second, the regulation drives the digital firms to change their business model, causing potential financial issues. When the digital firm faces the binding NNPC, it is optimal to charge zero price for the bundle and a fee for the stand-alone product. However, the DMA model compels the digital firm to offer a monetary payment for opt-in consumers, violating the NNPC.

In spite of the negative impact, the DMA model can increase consumer surplus by raising the benchmark utility for all consumers. Under the DMA model, opt-in consumers receive $U_d^*(\theta) = \tau_d^* - \theta + q_d^*$, while opt-out consumers receive q_d^*. In contrast, under the optimal mixed bundling, opt-in consumers receive $U^*(\theta) = \tau_m^* - \theta$ while opt-out consumers get zero. Thus, $U_d^*(\theta) > U_m^*(\theta)$ if $\tau_d^* + q_d^* > \tau_m^*$, in which case the DMA model increases consumer surplus. Using the leading example for further comparison, we find that the DMA model improves consumer surplus if

$$r > r_d \equiv \frac{b-1}{2\bar{\theta}\left(2\bar{\theta} + 2b - b^2 - 2\right)}.$$

3.3 Data property rights

There is a growing advocacy for granting consumers property rights over their personal data. However, significant technological challenges arise in allocating these rights directly to consumers. Consumer data is a by-product of engaging with digital products, and its value is realized only after being processed by digital firms. This value creation depends on the data analytics capabilities of firms, which constitute a core part of their assets. Consequently, consumers lack the means to monetize their data independently without access to these analytics tools.

Bergemann et al. (2023) propose the establishment of a competitive market for data intermediaries as a potential solution. Data intermediaries could install specialized software within consumers' web browsers to collect and monetize data on their behalf. We examine the economic implications of such a proposal.

Suppose such a market is implemented and consumers can earn data-monetization revenue r_i

from the intermediary. Following the regulation, the digital firm is obliged to supply the digital product on a stand-alone basis, in addition to the bundle. Let $S = \{bq, p\}$ and $S_A = \{q, p_A\}$ denote the firm's schemes for the bundle $\mathcal{B}$ and the stand-alone product A respectively. Given the non-rival nature of consumer data, exclusive dealing contracts are necessary to prevent consumers from selling their data through multiple channels.

The data intermediary mechanism introduces an additional option for consumers, complementing the mixed bundling choices. By combining the firm's stand-alone product A with data monetization from the intermediary, consumers achieve a net utility of $U_i(\theta) = q - p_A + r_i - \theta$ (where the subscript i denotes intermediary). This mix-and-match option serves as a perfect substitute for the pure bundle, from which consumers derive a net utility of $U_b(\theta) = bq - p - \theta$. Under the obligation of mixed bundling, the monopoly digital firm will optimally set $p_A = q$ to extract full consumer surplus from choosing the stand-alone product, resulting in $U_i(\theta) = r_i - \theta$. Consequently, the Bertrand-type competition between the bundle $\mathcal{B}$ and the mix-and-match determines the consumer's net utility from data sharing: $U(\theta) = \max\{U_b(\theta), U_i(\theta)\}$.

Consumers with $\theta \leq \tau_i$ choose to share their data, while those with $\theta > \tau_i$ opt for the stand-alone product, where the cut-off threshold is given by $\tau_i = \max\{bq - p, r_i\}$. Let q_i^* and τ_i^* denote the equilibrium quality and participation threshold, respectively. The outcome depends on the efficiency of data monetization by the intermediary (r_i), which leads to three distinct cases, as summarized in the following proposition.

Proposition 3 *Consider the data intermediary mechanism where consumers are entitled to earn data-monetization revenue r_i from a data intermediary:*

- *If $r_i \leq \tau_m^*$, the equilibrium outcome remains the same as under the optimal mixed bundling: $q_i^* = q_m^*$ and $\tau_i^* = \tau_m^*$.*
- *If $\tau_m^* < r_i \leq r + (b-1)\, q_i^*$, the digital firm collects consumer data. The data intermediary mechanism increases product quality and consumer surplus: $q_i^* > q_m^*$ and $\tau_i^* = r_i > \tau_m^*$.*
- *If $r_i > r + (b-1)\, q_i^*$, the data intermediary collects consumer data, and the digital firm focuses solely on the digital product. This mechanism results in a lower product quality but higher consumer surplus: $q_i^* = q^0 < q_m^*$ while $\tau_i^* = r_i > \tau_m^*$.*

Case (1): Inefficient Data Intermediary. If the data intermediary is inefficient in data monetization, it does not pose a competitive threat to the digital firm. The firm's optimal mixed bundling strategy sets a higher participation threshold, such that $\tau_m^* \geq r_i$. Consequently, the equilibrium outcome remains unchanged from the scenario without the data intermediary.

Case (2): Efficient Data Intermediary. When the data intermediary is efficient ($r_i > \tau_m^*$), competition from the intermediary imposes a binding constraint on the participation threshold, resulting in $\tau_i = r_i$. If the firm's competitive advantage from data acquisition $(r + (b-1)\,q)$ exceeds r_i, consumers with $\theta \leq \tau_i = r_i$ opt for the bundle. In this case, the firm's profit is given by:

$$\begin{aligned} \Pi_i &= F(\tau_i)\,(r + (b-1)\,q - \tau_i) + q - C(q) \\ &= F(r_i)\,(r + (b-1)\,q - r_i) + q - C(q)\,. \end{aligned}$$

The equilibrium quality q_i^* is determined by

$$C'(q) = 1 + (b-1)\,F(\tau_i) = 1 + (b-1)\,F(r_i)\,. \tag{9}$$

Compared to the first-order condition for quality under mixed bundling (2), it follows that $q_i^* > q_m^*$ as $\tau_i^* = r_i > \tau_m^*$. Competition from the data intermediary leads to greater participation in data acquisition compared to the optimal mixed bundling strategy. This increased participation facilitates an improvement in the quality of the digital product. As a result, the data intermediary mechanism enhances both product quality and consumer surplus.

Case (3): Highly Efficient Data Intermediary. If the data intermediary is highly efficient ($r_i > r + (b-1)\,q_i^*$), it becomes the primary entity collecting consumer data, leaving the digital firm to focus solely on the digital product. In this scenario, the firm chooses the optimal quality q^0 and earns a profit $\Pi^0 = q^0 - C\left(q^0\right)$. While the data intermediary mechanism increases consumer surplus, it reduces the quality of the digital product, resulting in a welfare loss. This welfare loss arises from the breakdown of the twofold benefits of data acquisition. By selling their data to the intermediary, consumers gain data-monetization revenues but forgo the data-driven benefits previously provided by digital firms, which collected and processed the data for enhanced product quality.

4 Data acquisition and competition

In recent years, there has been a surge in data-driven mergers, where dominant digital platforms acquire competitive producers of smart devices. Two particularly controversial examples are the Google-Fitbit merger in 2021 and the Amazon-iRobot deal in 2022. In both cases, the dominant platform already possessed advanced data analytics capabilities, enabling it to leverage data acquisition from smart devices. These mergers generate substantial data-driven consumer benefits by integrating smart devices into the platform's digital ecosystem, allowing consumers to personalize and enhance product functionality.

Competition authorities have raised concerns that dominant digital platforms may use their access to consumer data to hinder competition. Additionally, mergers involving such platforms often spark digital privacy concerns due to the extensive scale of consumer data collection. In this section, we investigate the impact of data acquisition on competition in digital products.

4.1 Analysis of data-driven mergers

An incumbent digital firm, having already established advanced data acquisition capabilities, merges with a producer of digital products. The incumbent (called it firm i) competes with firm e that supplies a substitute digital product of superior quality, where $q + \delta > q$ and $\delta > 0$ represents firm e's competitive edge in product quality. For simplicity, we assume that the quality levels of both firms' digital products are predetermined prior to competition, with their total costs treated as sunk.

In the absence of the merger, firm e would dominate the competition in providing the digital product by setting a price of $p_e = \delta$. Under this scenario, firm e earns a profit of $\Pi_e = \delta$, while consumers derive a net utility q.

Suppose firm i adopts a pure bundling strategy, offering its bundle at a price p_i, while firm e simultaneously offers its digital product at a price p_e^A. In this asymmetric competition, firm i derives a competitive advantage from data collection, denoted by $\beta \equiv (b-1)\,q + r$, whereas firm e benefits from its product quality advantage δ. Our analysis focuses on scenarios where firm i's data-driven advantage β exceeds firm e's product quality advantage δ, allowing firm i to achieve profitability.

Data acquisition has two contrasting effects on competition. On one hand, it provides significant benefits that firm i can exploit as a competitive advantage, thereby intensifying competition. On the other hand, the heterogeneity in consumer privacy costs fosters product differentiation, which can soften competition. The interaction between these opposing forces results in diverse impacts on both competition and consumer welfare.

Consumers derive a net utility of $U_i(\theta) = bq - p_i - \theta$ from firm i's bundle and $U_e = q + \delta - p_e^A$ from firm e's stand-alone product. Firm i's bundle and firm e's product are differentiated due to consumer heterogeneity in privacy sensitivity, θ, which leads to market segmentation. Consumers with $\theta \leq \tau_c$ choose the bundle, while those with $\theta > \tau_c$ opt for firm e's stand-alone product. The cut-off threshold τ_c (where the subscript c denotes competition) is given by:

$$\tau_c \equiv (b-1)\,q - \delta + p_e^A - p_i. \tag{10}$$

Accordingly, firm i earns a profit from the bundle, given by $\Pi_i = F(\tau_c)(r + p_i)$, while firm e earns a profit from its stand-alone product, given as $\Pi_e = (1 - F(\tau_c))\,p_e^A$. The firms' optimal pricing responses are jointly determined by the following FOCs:

$$p_i = h(\tau_c) - r, \tag{11}$$

$$p_e^A = k(\tau_c). \tag{12}$$

From (10), the equilibrium threshold τ_c^* is uniquely determined as the solution to:

$$\varphi(\tau_c) \equiv \tau_c + h(\tau_c) - k(\tau_c) = \beta - \delta, \tag{13}$$

where $\varphi(\tau_c)$ is an increasing function.[17] The equilibrium prices p_i^* and p_e^{A*} are determined by (11) and (12), respectively.

To ensure the existence of an interior optimum, we impose the following boundedness conditions on δ and β:

Assumption B: $\delta < \beta < \bar{\beta} \equiv \bar{\theta} + \frac{1}{f(\bar{\theta})} + \delta$. In addition $\delta < \frac{1}{f(0)}$.

The interplay of two opposing effects results in three distinct equilibrium outcomes, as summarized in the following proposition:

Proposition 4 *Consider the asymmetric competition after the data-driven merger. When Assumption B is satisfied, there exists a unique equilibrium in which firm i adopts pure bundling. Consumers with $\theta \leq \tau_c^*$ choose the bundle, while those with $\theta > \tau_c^*$ opt for firm e's digital product, where τ_c^* is given by* (13)*. Compared to the benchmark case without the merger, there exist two cut-off thresholds $\tilde{\beta}$ and $\hat{\beta}$, satisfying $\delta < \tilde{\beta} < \hat{\beta} < \bar{\beta}$, such that:*

- *when $\beta > \hat{\beta}$, asymmetric competition results in a lower price: $p_e^{A*} < \delta$, benefiting consumers while harming firm e;*
- *when $\beta < \hat{\beta}$, asymmetric competition leads to a higher price: $p_e^{A*} > \delta$, which harms consumers opting for firm e's digital product;*

[17] Assumption B ensures $\tau_c^* < \bar{\theta}$.

- *when $\beta < \tilde{\beta}$, asymmetric competition increases firm e's price and profit, with $p_e^{A*} > \delta$ and $\Pi_e^* > \delta$.*

Proof. See Online Appendix B. ■

The participation threshold τ_c^* increases with β, while firm e's price p_e^{A*} decreases with τ_c^*. When β is sufficiently large, firm e charges a price below its efficiency gain: $p_e^{A*} < \delta$. Consequently, there exists a unique value $\hat{\beta} \in (\delta, \bar{\beta})$ such that $p_e^{A*} > \delta$ if and only if $\beta < \hat{\beta}$. Furthermore, there exists a cutoff $\tilde{\beta} < \hat{\beta}$ such that $\Pi_e^* = (1 - F(\tau_c^*))\, p_e^{A*} > \delta$ if and only if $\beta < \tilde{\beta}$, in which case firm e earns a higher profit than before the merger.

Proposition 4 establishes the following implications:

(i). Firm i benefits from pure bundling. When its competitive advantage is relatively small (i.e., $\beta < \hat{\beta}$), committing to pure bundling softens competition and raises the rival's price. If firm i offered mixed bundling instead, Bertrand competition in the stand-alone products would drive firm e's price to $p_e^A = \delta$, making both firms worse off. However, when $\beta > \hat{\beta}$, offering mixed bundling does not affect the equilibrium outcome. Firm e's offer of its stand-alone product at a zero price imposes no competitive constraint on the entrant, since the latter charges $p_e^{A*} < \delta$.

(ii). Large advantage intensifies competition. When firm i's advantage is relatively large (i.e., $\beta > \hat{\beta}$), data acquisition intensifies competition, reducing the equilibrium price for the stand-alone product and increasing consumer surplus. Consumers who choose firm e's product receive $U_e^* = q + \delta - p_e^{A*} > q$, while those opting for the bundle obtain $U_i^*(\theta) = \tau_c^* - \theta + U_e^* > \tau_c^* - \theta + q$. In this scenario, the merger improves welfare.

(iii). Small advantage softens competition. When $\beta < \hat{\beta}$, data acquisition under pure bundling softens competition, harming consumers who choose the stand-alone product. For $\beta < \tilde{\beta} < \hat{\beta}$, this reduction in competition even leads to higher profits for firm e than before the merger. Although the merger benefits consumers choosing the bundle, the welfare loss for those opting for the stand-alone product can outweigh these gains. Consequently, consumer surplus may decrease, as occurs when $\beta < \underline{\beta}$ in the leading example. In such cases, banning pure bundling could improve consumer welfare.

Existing studies on data-driven mergers focus on the interconnection between data collection and data application markets, where the merged firm leverages consumer data for targeted advertising and personalized offers. This often places rivals without comparable data at a competitive disadvantage. However, as highlighted in the literature, such uses of consumer data frequently intensify competition, ultimately benefiting consumers.[18]

Chen et al. (2022) examine data-driven mergers that link data collection and application markets through consumption synergies, enabling the merged firm to employ consumer data for personalized offers and pricing. They find that while prices in the data collection market typically decrease, prices in the data application market rise as efficiency gains are extracted through personalized pricing. When consumption synergies are sufficiently large, such mergers can result in the foreclosure of competitors in both markets.

de Cornière and Taylor (2024) investigate mergers between firms operating in two data-connected markets, considering the possibility of pre-merger data trade. Their findings suggest that when data enhances product quality, the merger benefits consumers in both markets if pre-merger data trade is not possible; however, it harms consumers otherwise. Conversely, when data is used to extract consumer surplus in the product market, the merger reduces consumer welfare.

In contrast, this paper examines two opposing effects of data acquisition: intensified competition from the merged firm's data advantage and softened competition due to privacy-cost-driven product differentiation. Their interplay leads to varied impacts on competition and consumer welfare, as outlined above.

Meanwhile, evidence suggests that mergers do not necessarily intensify competition. For example, following the Google-Fitbit merger in 2020, the prices of Fitbit smartwatches have remained stable. The Fitbit Versa was priced between £139.00 and £169.00 in 2019,[19] and its updated version in 2024, the Fitbit Versa 4, is similarly priced, ranging from £149.99 to £179.00.[20] In comparison, Apple has maintained a consistent pricing strategy for its Apple Watch, introducing new versions annually at a steady price of $399 since 2020.

Remark 2 *A key concern in data-driven mergers is the potential for the merged firm to engage in below-cost pricing to foreclose competitors. Our analysis indicates that the incumbent adopts below-cost pricing when data monetization revenue is sufficiently large: $p_i^* < 0$ if $r > h(\tau_c^*)$.*[21] *This pricing strategy, however, does not occur when data monetization is prohibited. While banning data monetization eliminates below-cost pricing, it also reduces the equilibrium participation*

[18] See discussions in the literature on personalized pricing.

[19] See https://www.pricehistory.co.uk/currys/10178149/fitbitversa-black-aluminium-black

[20] See https://www.pricehistory.co.uk/product/1153743/fitbit-versa-4-smart-watch-blackgraphite?utm_source=chatgpt.com

[21] In the example with uniform distribution, $p_i^* < 0$ if and only if $r > \left((b-1)q + \bar{\theta} - \delta\right)/2$.

threshold τ_c^*, negatively affecting consumers who prefer the bundled offering. Furthermore, this prohibition raises firm e's price, as p_e^{A*} decreases with β. Consequently, banning data monetization benefits the competitor but harms all consumers.

4.2 Privacy protection and differentiation in business models

Since the enactment of GDPR, a growing number of digital platforms have sought to standardize their data collection practices and enhance consumer data protection. A prominent example is Apple's introduction of privacy labels in December 2020. Apple requires all app developers on its platform to disclose their data collection practices through privacy "nutrition" labels, providing consumers with a clear and transparent overview of how their data is handled.

In June 2021, Apple released new versions of its operating systems, signaling a heightened emphasis on digital privacy—a defining feature that sets Apple apart from Android and Windows competitors. Notable privacy-focused features introduced in iOS 15 and macOS Monterey include "No Tracking Pixels," "Private Relay," "Hide My Email," and the "App Privacy Report."[22]

While Apple possesses the capacity for data acquisition, its strategic positioning as a leader in data privacy protection effectively distinguishes its business model from competitors, reflecting an equilibrium outcome in the competition between digital firms.[23]

To further analyze this dynamic, consider the competition between two firms, firm i and firm e, both capable of data acquisition. Firm i, having already established expertise in data analytics, becomes the leader in data acquisition and bundles its digital product with data collection. Firm i possesses a comparative advantage in data monetization, generating higher revenue than its rival: $r_i = r > r_e = 0$. In this scenario, firm i derives aggregate benefits of $bq + r$ from data acquisition, while firm e generates only bq. Conversely, firm e offers a superior stand-alone digital product compared to firm i, with quality defined as $q + \delta > q$, where $\delta > 0$ represents firm e's competitive edge in product quality. To ensure firm e benefits from data acquisition in the absence of competition, assume $(b-1)\,q > \delta$.

This modelling approach captures key aspects of competition between digital platforms. For

[22] https://www.cnbc.com/2021/06/07/apple-is-turning-privacy-into-a-business-advantage.html

[23] Americans increasingly report that privacy influences their purchasing decisions. According to a 2020 Pew study, 52% of Americans chose not to use a product or service due to concerns about data protection. See https://www.pewresearch.org/short-reads/2020/04/14/half-of-americans-have-decided-not-to-use-a-product-or-service-because-of-privacy-concerns/

example, Google and Microsoft benefit from superior access to consumer data and earn higher revenues from data monetization, whereas Apple distinguishes itself by offering higher-quality digital products and prioritizing privacy protection.

The timing of the game is structured as follows:

Stage 1: Firm i announces its bundling strategy for offering both goods.

Stage 2: Firm e selects its business model and, if applicable, its bundling strategy.

Stage 3: Both firms simultaneously offer their pricing schemes.

Stage 4: Consumers observe the offered schemes and make their participation decisions.

The equilibrium outcome is presented below:

Proposition 5 *Consider competition between two digital firms, both capable of data acquisition. When Assumption B is satisfied, a unique subgame perfect Nash equilibrium (SPNE) exists in which firm i adopts pure bundling while firm e focuses exclusively on the digital product. The equilibrium outcome aligns with the characterization in Proposition 4.*

Proof. See Online Appendix B. ■

The above proposition shows that differentiation in business models emerges as the unique equilibrium outcome, even when both firms have the capability for data acquisition. This uniqueness is established through three steps. First, if firm i adopts pure bundling, firm e's best response is to focus solely on the digital product. Second, if firm i instead adopts mixed bundling, firm e's optimal strategy is to supply the stand-alone digital product. Third, by comparing the equilibrium outcomes in these two subgames, we conclude that firm i's optimal strategy is pure bundling.

Data acquisition thus facilitates differentiation in business models, which softens competition between the firms. This result remains robust even when both firms choose their business models simultaneously. To illustrate, consider a revised game in which both firms choose their business models and bundling strategies simultaneously, followed by a simultaneous pricing game. If firm i focuses solely on the digital product, firm e's best response is to engage in data acquisition and adopt pure bundling. Conversely, if firm i engages in data acquisition and offers a bundle, firm e's best response is to concentrate on the stand-alone product, irrespective of whether firm i adopts pure or mixed bundling. Consequently, there are three types of equilibria in this framework. In each equilibrium, the firms differentiate their business models, with one firm specializing exclusively in the digital product.

Remark 3 *Markovich and Yehezkel (2024) examine competition between an incumbent digital* *firm and an entrant, each deciding among a data-based model (i.e., pure bundling in our framework), a subscription-based model (i.e., digital product only), or a hybrid business model (i.e., mixed bundling). Their analysis emphasizes network effects, where a consumer joining platform j derives a benefit proportional to the network size, βn_j, with n_j representing the population of other consumers on the platform. The equilibrium choice of business models is shaped by the interplay of two forces: a competition-softening effect arising from consumer heterogeneity in privacy costs and a competition-intensifying effect driven by network effects. When network effects are strong, the latter dominates, leading the incumbent to adopt mixed bundling and exclude the entrant. Conversely, when network effects are weaker, the former prevails, with the incumbent choosing a subscription-based model and the entrant opting for a data-based model.*

5 Heterogeneous preferences and personalized pricing

5.1 Heterogeneous data-driven benefit

Consumers may exhibit heterogeneous preferences regarding data-driven benefits. For example, a consumer who has already integrated Amazon.com's Alexa and other smart devices into their home may derive significant additional benefits from purchasing a new smart device, such as a robotic vacuum cleaner. In contrast, a new customer to Amazon.com's smart home ecosystem might experience lower benefits from using its devices. In such cases, charging a uniform price for all consumers cannot achieve the efficient outcome for data collection.

Conversely, when a digital firm collects consumer data, it can leverage this information to infer individual preferences and offer personalized prices. Personalized pricing has the potential to achieve efficient outcomes but can also be used to extract consumer surplus. These two conflicting functions make personalized pricing a controversial practice.

A growing literature explores personalized pricing, with most studies assuming that firms have complete information about consumers' willingness to pay. Under such conditions, firms set personalized prices to fully capture each consumer's additional benefit from the product, leaving each consumer indifferent between accepting or rejecting the personalized offer. However, when consumers can decide whether to share their data and their privacy sensitivity remains private information, the digital firm must leave consumers a positive surplus as the information rents. In such scenarios, the implications of personalized pricing for profitability and consumer welfare require further examination.

We extend the baseline model by introducing heterogeneity in data-driven consumer benefits.

The data-driven benefit is expressed as $bq = (1+x)q$, where x represents a consumer's preference and is distributed over $[0, 1]$ according to a distribution function $G(\cdot)$ with a strictly positive density $g(\cdot)$; let μ_x denote its mean.

Consumers learn their preferences x and privacy sensitivity θ prior to making purchase decisions. We further assume that a consumer's preference for data-driven benefits is uncorrelated with their privacy sensitivity, and that the firm cannot infer privacy sensitivity from collected data. For analytical tractability, we focus on a uniform distribution of θ, characterized by $F(\theta) = \theta/\bar{\theta}$.

Consider the competition between firms i and e, where both offer perfect substitutes with identical quality q, but firm i benefits from an advantage in data acquisition, enhancing consumer value by xq. For simplicity, we further assume that the costs incurred by both firms in providing digital services are already sunk. Suppose firm i adopts a pure bundling strategy,[24] and announces its pricing scheme (personalized pricing or uniform pricing) before consumers make their choices.

Under personalized pricing, firm i learns each consumer's preference x through data acquisition and offers personalized prices for the bundle, $p_i(x)$, tailored to consumers with preference x. A consumer of type (θ, x) derives utility $U_i(\theta, x) = (1+x)q - p_i(x) - \theta$ when choosing firm i's bundle, while receiving utility $U_e = q - p_e^A$ from firm e's stand-alone digital product. Consumers with $\theta \le \tau(x)$ choose the bundle, whereas those with $\theta > \tau(x)$ opt for the stand-alone product. The cut-off threshold is defined as:

$$\tau(x) \equiv (1+x)q - p_i(x) - U_e.$$

Firm i's profit under personalized pricing is expressed as

$$\Pi_i = \frac{1}{\bar{\theta}} \int_0^1 (p_i(x) + r)\,\tau(x)\, dG(x) = \frac{1}{\bar{\theta}} \int_0^1 ((1+x)q + r - \tau(x) - U_e)\,\tau(x)\, dG(x).$$

For each consumer with preference x, firm i captures the data acquisition benefit, $(1+x)q + r$, while compensating consumers with $\tau(x) + U_e$. Consequently, when consumers have the ability to exercise privacy choices and their privacy sensitivity remains private information, firm i must leave positive surplus for each consumer.

Now suppose firm i charges a uniform price p_i for the bundle. A consumer of type (θ, x) derives utility $\hat{U}_i(\theta, x) = (1+x)q - p_i - \theta$. The cut-off threshold is then given by

$$\hat{\tau}(x) \equiv (1+x)q - p_i - U_e,$$

[24] As discussed in the previous section, firm i does not benefit from mixed bundling.

and firm i's profit is expressed as

$$\hat{\Pi}_i = \frac{1}{\bar{\theta}} \int_0^1 (p_i + r)\, \hat{\tau}(x)\, dx = \frac{1}{\bar{\theta}} \int_0^1 ((1+x)q + r - \hat{\tau}(x) - U_e)\, \hat{\tau}(x)\, dx.$$

Similarly, firm i captures the data acquisition benefit $(1+x)q + r$ while compensating each consumer with $\hat{\tau}(x) + U_e$.

We assume that $\bar{\theta}$ is sufficiently large to ensure partial participation for data acquisition from consumers $x \in [0,1]$ ($0 < \tau(x) < \bar{\theta}$ and $0 < \hat{\tau}(x) < \bar{\theta}$). Formally:

Assumption C: $\bar{\theta} > \bar{\theta}_l^x \equiv \max\{2q, 4\mu_x q\} + \frac{r}{2}$.

The comparison of equilibrium outcomes under personalized pricing and uniform pricing is summarized in the following proposition:

Proposition 6 *Consider a competitive setting between two firms, where consumers exhibit heterogeneous preferences for data-driven benefits, and firm i can implement personalized pricing. Assume x is distributed over $[0,1]$ while θ is uniform distributed over $\left[0, \bar{\theta}\right]$, with Assumption C satisfied. Compared to uniform pricing, personalized pricing:*

- *results in the same equilibrium price and profit for firm e, while increasing firm i's profit;*
- *provides higher utility from the bundle for consumers with $x < \mu_x$ but lower utility for those with $x > \mu_x$;*
- *reduces both consumer surplus and overall social welfare.*

Proof. See Online Appendix C. ■

Under the uniform pricing, firm i chooses p_i to maximize its expected profit, with the optimal price determined by

$$p_i + r = \int_0^1 \hat{\tau}(x)\, dG(x). \tag{14}$$

In contrast, under personalized pricing, firm i chooses $p_i(x)$ to maximize its profit from each consumer with preference x, resulting in the optimal price:

$$p_i(x) + r = \tau(x). \tag{15}$$

Given U_e, solving for (14) leads to the optimal uniform price:

$$\hat{p}_i^* = \frac{(1+\mu_x)\, q}{2} - \frac{U_e + r}{2}.$$

Meanwhile, solving for (15) gives the optimal personalized price:

$$p_i^*(x) = \frac{(1+x)q}{2} - \frac{U_e + r}{2}.$$

It follows that $\hat{p}_i^*$ is the expected value of $p_i^*(x)$:

$$\hat{p}_i^* = \int_0^1 p_i^*(x)\, dG(x).$$

The expected participation thresholds under the two pricing schemes are identical:

$$\int_0^1 \hat{\tau}^*(x)\, dG(x) = \int_0^1 ((1+x)q - \hat{p}_i^* - U_e)\, dG(x) = \int_0^1 ((1+x)q - p_i^*(x) - U_e)\, dG(x) = \int_0^1 \tau^*(x)\, dG(x).$$

Since firm e employs only uniform pricing, its optimal price is determined by the expected participation population. The property above implies that firm e's optimal price and equilibrium profit remain the same under both pricing schemes.

The neutrality result above also implies that consumers opting for firm e's digital product achieve the same net utility, U_e^*, under both pricing schemes. In contrast, consumers selecting the bundle receive $U_i^*(\theta, x) = U_e^* + \tau^*(x) - \theta$ under personalized pricing while $\hat{U}_i^*(\theta, x) = U_e^* + \hat{\tau}^*(x) - \theta$ under uniform pricing, where

$$\begin{aligned} \tau^*(x) &= \frac{(1+x)q}{2} - \frac{U_e - r}{2}, \\ \hat{\tau}^*(x) &= \frac{(1+x)q}{2} + \frac{(x-\mu_x)q}{2} - \frac{U_e - r}{2}. \end{aligned}$$

This implies that consumers with $x < \mu_x$ are better off under personalized pricing, while those with $x > \mu_x$ are worse off.

Since the participation thresholds $\tau(x)$ are upward-sloping, the participation population of consumers with $x > \mu_x$ exceeds that of those with $x < \mu_x$. As a result, consumer surplus is lower under personalized pricing. While firm i benefits from personalized pricing, the reduction in consumer surplus outweighs the firm's profit gain, leading to a decrease in total social welfare under personalized pricing.

Proposition 6 highlights key implications of personalized pricing in the presence of consumer privacy choices.

First, the literature commonly finds that personalized pricing by an incumbent intensifies competition with a rival firm using uniform pricing, thereby reducing the rival's profit.[25] However, in this setting, the rival firm earns the same profit under both pricing schemes, as firm i maintains identical expected participation thresholds in equilibrium:

$$\int_0^1 \hat{\tau}^*(x)\, dx = \int_0^1 \tau^*(x)\, dx.$$

[25] See the discussion of the related literature on personalized pricing.

This allows firm i to extract more surplus from consumers without intensifying competition, resulting in higher profits.

Second, personalized pricing is often associated with improved allocative efficiency and increased total social welfare compared to uniform pricing, as it functions similarly to first-degree price discrimination. However, this conclusion does not hold when consumers exhibit heterogeneous privacy sensitivity. While personalized pricing preserves the same expected participation threshold as uniform pricing, the welfare loss from consumers with relatively high valuations for data-driven benefits outweighs the gains from those with lower valuations and firm i's profit increase. Consequently, personalized pricing reduces total social welfare.

Remark 4 *The price neutrality result is derived under the assumption of the uniform distribution for θ. For more general distributions, the comparison becomes significantly more complex, making it difficult to establish a definitive sign. A brief discussion of the general case is provided in Online Appendix C.*

5.2 Heterogeneous data-monetization revenue

In the baseline analysis, we assumed identical data-monetization revenue r across consumers. Now, consider the case where r is heterogeneous. Let r be distributed over the interval $[0, 1]$ according to the distribution function $G(\cdot)$, with a strictly positive density function $g(\cdot)$, and let μ_r denote its mean. Additionally, assume that the distributions of r and θ are uncorrelated, while consumers derive a homogeneous data-driven benefit bq.

When consumers are uncertain about their individual value r from data monetization before selecting the bundle, the heterogeneity of r does not affect the cut-off threshold between the bundle and the stand-alone product. Consumers derive a net utility $U(\theta) = bq - p - \theta$ from consuming the bundle and U_e from the stand-alone product A. Consequently, the cut-off threshold remains unchanged: $\tau = bq - p - U_e$.

Consider a monopoly firm's optimal scheme under mixed bundling. The firm offers $S = \{bq, p\}$ for the bundle while charging $p_A = q$ for the stand-alone product. The firm's profit is given by:

$$\Pi = \int_0^1 F(\tau)(r + p)\, dG(r) + (1 - F(\tau))\, q - C(q).$$

Simplifying this expression using μ_r, the mean of r, the profit becomes

$$\Pi = F(\tau)(\mu_r + p) + (1 - F(\tau))\, q - C(q).$$

Thus, by substituting μ_r for r, the baseline analysis extends to this scenario, and the equilibrium results remain unchanged.

Consider a competitive setting between two firms, i and e, where firm i can implement personalized prices $p_i(r)$ based on a consumer's data-monetization value r and assume the firm's costs are sunk. For analytical tractability, let θ be uniformly distributed over $\left[0, \bar{\theta}\right]$. Using the same framework as for heterogeneous data-driven benefits, we can characterize the equilibrium under both personalized and uniform pricing. The following assumption ensures partial participation for data acquisition from consumers with $0 < \tau(r) < \bar{\theta}$:

Assumption D: $\bar{\theta} > \bar{\theta}_l^r \equiv 1 + \frac{bq}{2}$.

The neutrality result regarding the rival's pricing and profit holds in this scenario: personalized pricing does not influence the rival's pricing strategy. However, firm i achieves higher profits under personalized pricing compared to uniform pricing.

The welfare effects of personalized pricing differ fundamentally between heterogeneous data-driven benefits and heterogeneous data-monetization values. In the former case, firm i employs personalized pricing to extract consumer benefits but cannot fully exploit consumers due to the heterogeneity in privacy costs. In contrast, with heterogeneous data-monetization values, personalized pricing serves to incentivize participation from consumers with higher r, enabling firm i to generate greater revenue. This distinction results in divergent welfare outcomes.

Under uniform pricing, the consumer utility from the bundle is independent of r, and the equilibrium participation threshold remains constant:

$$\hat{\tau} = bq - p_i - U_e = \frac{bq - U_e + \mu_r}{2}.$$

In contrast, with personalized pricing, firm i charges lower prices to consumers with higher r, thereby raising the participation threshold as r increases:

$$\tau(r) = \frac{bq - U_e + r}{2}.$$

As a result, consumers with $r > \mu_r$ benefit from personalized pricing, while those with $r < \mu_r$ are worse off. Although the expected values of participation thresholds are identical across the two schemes—$\hat{\tau}$ equals the mean of $\tau(r)$—personalized pricing attracts a larger share of consumers with $r > \mu_r$ to the bundle. This increases consumer surplus overall.

In conclusion, personalized pricing enhances firm i's profits, consumer surplus, and total social welfare by effectively leveraging the heterogeneity in data-monetization values:

Proposition 7 *Consider competition between two firms, where the data-monetization revenues are heterogeneous across consumers, firm i can implement personalized pricing. Assume r is distributed over* $[0, z]$ *while* θ *is uniform distributed over* $\left[0, \bar{\theta}\right]$*, with Assumption D satisfied. Compared to uniform pricing, personalized pricing:*

- *results in the same equilibrium price and profit for firm e, while increasing firm i's profit;*
- *provides higher utility from the bundle for consumers with* $r > \mu_r$ *but lower utility for those with* $r < \mu_r$*;*
- *increases both consumer surplus and overall social welfare.*

Proof. See Online Appendix C. ■

Few papers in the literature on personalized pricing explore the interaction between data-driven personalization and consumer privacy, with the notable exception of Rhodes and Zhou (2024b). Our paper differs from theirs in the following key aspects:

Timing of Consumer Preference Revelation: They assume that consumers learn their preferences after making privacy choices (i.e., whether to share their data). A consumer who agrees to share data incurs a sunk privacy cost and subsequently receives a personalized price. In this scenario, a monopoly firm can use personalized pricing to fully extract consumer surplus, leaving consumers with a payoff equivalent to their outside option (e.g., a price cap under uniform pricing or the competitor's offer). By contrast, our model considers a setting where consumers learn their preferences before making privacy choices. This prevents the monopoly firm from fully extracting consumer surplus through personalized pricing.

Core Economic Forces: The main economic force in Rhodes and Zhou's model is a privacy-choice externality among consumers. When some consumers share their data, it not only affects the offers they receive but also influences the offers made to other consumers. Let σ denote the proportion of consumers who share their data. A consumer's net utility from sharing data is defined as the difference between the payoffs under the two choices: $\Delta(\sigma) = V_s(\sigma) - V_a(\sigma)$, where $V_s(\sigma)$ is the utility with data sharing, and $V_a(\sigma)$ is the utility without data sharing. Consumers with privacy costs $\theta \leq \Delta(\sigma)$ are willing to share their data, and the cut-off threshold σ^* is a fixed point satisfying $\sigma^* = F(\Delta(\sigma))$. Under monopoly, $\Delta(\sigma) = 0$, meaning the consumption benefit from data sharing is zero, as the monopoly can fully extract consumer surplus through personalized pricing. When privacy costs θ is distributed over $\left[0, \bar{\theta}\right]$, no consumers share their data in equilibrium ($\sigma^* = 0$).

In contrast, our model emphasizes compensation for privacy costs when a consumer's privacy sensitivity remains private information. While our analysis does not incorporate network externalities, the key insight remains valid in such settings: the monopoly cannot fully extract consumer surplus, as personalized prices are determined before consumers make their privacy choices. Consequently, the monopoly must share a portion of the benefits from data acquisition and personalized pricing with consumers. In equilibrium, consumers with $\theta \leq \tau^*(x)$ will share their data, where $\tau^*(x) > 0$.

6 Variant and extension

6.1 Cross-User Data Aggregation

A significant portion of big data applications focus on marketing and product or service recommendations. In these cases, digital platforms leverage predictive data analytics to forecast consumer behavior, estimate the likelihood of accepting a recommended product or service, and assess willingness-to-pay. These analytics heavily depend on a consumer's historical data (behavioral data) and other available activity profiles, utilizing behavior-based machine learning to deliver personalized product recommendations and pricing—commonly referred to as within-user data aggregation. For instance, Fitbit's premium service offers personalized health, sleep, and fitness recommendations based on data collected through Fitbit devices, as well as linked Google accounts following Fitbit's acquisition by Google.

In other cases, data analytics incorporates machine learning techniques that rely on cross-user data aggregation, drawing from demographic datasets across different users. Examples include Grammarly for spelling, grammar, and tone optimization; Cruise for autopilot technologies; and Deep Sentinel for home security solutions.[26]

The baseline analysis focuses on applications of data analytics primarily utilizing within-user data aggregation, aligning well with business models for personalized advertising, product recommendations, pricing, and related domains. A recent empirical study by Smith et al. (2023) evaluates the profitability of personalized pricing policies in settings with consumer-level panel data. The study explores pricing policies derived from various empirical models, including Bayesian hierarchical choice models, regularized regressions, neural networks, and nonparametric classifiers, using different data inputs. Across all models, information on consumers' purchase histories significantly improves profits, while demographic data has only a minor effect.

[26] For a discussion on within-user versus cross-user learning, see Hagiu and Wright (2021).

Nevertheless, our analysis can extend to environments featuring positive data externalities arising from cross-user data aggregation. The main insights and results remain applicable in these broader contexts, while detailed analysis is provided in Online Appendix D.

6.2 Verifiable data scale

In this extension, we explore scenarios where the firm's commitment to a specific data scale is verifiable, either through oversight by an independent third party or a regulatory authority.[27] Let $s \in [0, \bar{s}]$ denote the quantitative measure of the scale of a consumer dataset. The scale s is a real-valued function mapping a personal dataset to the set of positive real numbers ($[0, \bar{s}] \subset \Re_+$). The benefits from data acquisition increase with s. The digital firm's revenue from data monetization is denoted by $r(s)$, an increasing and concave function with $r(0) = 0$. Similarly, the data-driven consumer benefit can be expressed as $b(s)\,q$, where $b(s)$ is increasing and concave, with $b(0) = 1$. At the same time, consumer privacy costs, expressed as $s\theta$, also increase with s, where the privacy sensitivity θ is distributed over $\left[0, \bar{\theta}\right]$.

Consider a monopoly digital firm's compensation scheme under mixed bundling, offering the scheme $S = \{s, b(s)\,q, p\}$ for the bundle while charging $p_A = q$ for the stand-alone product. Consumers receive a net utility $U(\theta) = b(s)\,q - p - s\theta$ from the bundle and a net utility of 0 from the stand-alone product. The participation threshold is now defined as $\tau = (b(s)\,q - p)/s$. Using the price $p = b(s)\,q - s\tau$ as a transfer of value, the monopoly can capture the aggregate benefits $b(s)\,q+ r(s)$ while compensate each participant with a value equal to the privacy cost of the marginal consumer with $\theta = \tau$, $s\tau$, and its profit is given by

$$\Pi = F(\tau)\left((b(s) - 1)\,q + r(s) - s\tau\right) + q - C(q)\,.$$

A consumer derives a gross utility $b(s)\,q - p$ from the bundle. The monopoly can influence this value through three policy variables: the quality of the product q, the scale of data s, and the price (or subsidy) p. Optimization under the verifiability of the data scale introduces two new implications:

- Balancing benefits and privacy costs: While the benefits from data acquisition increase with s, so does the privacy cost. The monopoly's optimal choice of s balances the marginal benefit from data acquisition, $b_s(s)\,q + r_s(s)$, with the marginal privacy cost, τ.

[27] See Bergemann et al. (2023) for discussions on such commitment mechanisms.

- Complementarity between q and s. The optimal product quality, determined by $C'(q) = 1 + (b(s) - 1) F(\tau)$, increases with s, reflecting the complementarity between q and s.

The characterization of the optimal scheme under mixed bundling follows a similar approach to the baseline analysis, with detailed results provided in Online Appendix D.

Remark: Mechanism design.

Data acquisition under a verifiable data scale creates opportunities for discriminatory compensation schemes, where the digital firm can offer a menu of options tailored to consumers' privacy sensitivities. Chen (2022) employs a mechanism design framework to characterize the optimal incentive-compatible compensation scheme. However, this paper does not address type-dependent mechanisms.

In practice, implementing such optimal mechanisms poses significant policy challenges. While offering a menu of options (contracts) to consumers is a well-established practice in industries such as telecommunications and electricity, the application of similarly sophisticated schemes in the digital sector remains a formidable task. Existing regulations, such as the GDPR, establish foundational principles for consumer data collection, including lawfulness, fairness, transparency, purpose limitation, and data minimization. Yet, these regulations do not provide detailed guidelines for standardizing or categorizing data collection practices.

7 Conclusions

This paper addresses two key questions: What is the economic nature of consumer data, and how should consumers be compensated for their data? Digital firms offer digital products to consumers and collect consumer data as a by-product of their usage. This data acquisition generates both data-monetization revenue and data-driven consumer benefits, while imposing privacy costs on consumers. The paper explores compensation schemes for consumer data, focusing on the interdependence between data collection and digital product provision, and examining the role of cross-subsidization in compensation mechanisms. We analyze the optimal compensation scheme for a monopolistic digital firm, examine the impact of data acquisition on competition, and investigate personalized pricing in the context of consumer privacy choices. Our findings offer valuable policy implications for digital privacy regulations and competition policies related to data collection.

We conclude by acknowledging the main limitations of this study.

First, for simplicity, we treat data-monetization revenue as exogenously given and do not address the strategic sharing of data among digital firms. In contexts where data-monetization occurs through personalization and targeted advertising, strategic data-sharing practices could significantly influence the revenue generated from data acquisition.

Second, this paper does not account for the network externalities of data acquisition among consumers. While this simplification facilitates explicit equilibrium characterizations and welfare comparisons, it does not capture how data externalities might influence outcomes. We believe that the core insights regarding a monopolistic firm's behavior remain robust even in the presence of data externalities. However, such externalities could notably impact equilibrium outcomes in competitive scenarios involving asymmetric firms, especially when amplified by the invasive application of AI technologies. We aim to address these issues in future research.

References

Acemoglu, D., Makhdoumi, A., Malekian, A. and A. Ozdaglar (2022). "Too Much Data: Prices and Inefficiencies in Data Markets. *American Economic Journal: Microeconomics*, vol 14 (4): 218-256.

Acquisti, A., Taylor, C. and L. Wagman (2016). "The Economics of Privacy". *Journal of Economic Literature*, vol 54(2): 442-492.

Argenziano, R and A. Bonatti (2024). "Data Linkages and Privacy Regulation". *Working paper*, https://www.mit.edu/%7Ebonatti/linkages.pdf.

Aridor, G., Che, Y.-K. and T. Salz (2023). "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR". *Rand Journal of Economics*, vol 54 (4): 695-730.

Becker, G. (1980). "Privacy and Malfeasance: A Comment". *Journal of Legal Studies*, vol 9: 823-826.

Bergemann, D, and A. Bonatti (2024). "Data, Competition, and Digital Platforms". *American Economic Review*, vol 114 (8): 2553-2595.

Bergemann, D., Bonatti, A. and T. Gan (2022). "The Economics of Social Data". *Rand Journal of Economics,* vol 53: 263-296.

Bergemann, D, Crémer, J., Dinielli, D., Groh, C., Heidhues, P., Schäfer, M., Schnitzer, M., Scott Morton, F., Seim, K., and M. Sullivan (2023). "Market Design For Personal Data". *Yale Journal on Regulation*, vol 40 (3): 1056-1120.

Bonatti, A, Y. Huang and J.M. Villas-Boas (2024). "A Theory of the Effects of Privacy". *Working paper*, https://www.mit.edu/%7Ebonatti/privacytheory.pdf

Chen, Y., and G. Iyer (2002) "Consumer Addressability and Customized Pricing". *Marketing Science*, vol 21(2), 197-208.

Chen, Z. (2022), "Privacy Costs and Consumer Data Acquisition: An Economic Analysis of Data Privacy Regulation". *Working paper*, https://ideas.repec.org/p/mos/moswps/2022-07.html.

Chen, Z., C. Choe, and N. Matsushima (2020). "Competitive Personalized Pricing". *Management Science*, vol 66(9), 4003-4023.

Chen, Z., Choe, C., Cong, J. and N. Matsushima (2022). "Data-Driven Mergers and Personalization". *RAND Journal of Economics*, vol 53 (1): 3-31.

Choi, J. P., Jeon, D.-S. and B.-C. Kim (2019). "Privacy and Personal Data Collection with Information Externalities". *Journal of Public Economics*, vol 173: 113-124.

de Cornière, A, and G. Taylor (2024) "Data-Driven Mergers", *Management Science*, vol 70 (9): 6473-6482.

Dosis, A and W. Sand-Zantman (2023). "The Ownership of Data". *Journal of Law, Economics, and Organization*, vol 39 (3): 615-641.

Eilat, R., K. Eliaz, and X. Mu (2021). "Bayesian Privacy". *Theoretical Economics*, vol 16 (4):1557–1603.

Fainmesser, I. P., Galeotti, A. and R. Momot (2023). "Digital Privacy". *Management Science*, vol 69(6): 3157-3173.

Hagiu, A. and J. Wright (2021). "Data-Enabled Learning, Network Effects and Competitive Advantage". *Rand Journal of Economics,* vol 54 (4): 638-667.

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

Ichihashi, S. (2021). "The Economics of Data Externalities". *Journal of Economic Theory*, vol 196: 1-26.

Ichihashi, S. (2023). "Dynamic Privacy Choices". *American Economic Journal: Microeconomics*, vol 14 (2): 1-40.

Jia, J., Jin, G. Z. and L. Wagman (2021). "The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment". *Marketing Science*, vol 40: 661-684.

Johnson, G. A., Shriver, S. K. and S. Du (2020). "Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?" *Marketing Science*, vol 39: 33-51.

Jullien, B, Y. Lefouili, and M. Riordan (2020). "Privacy Protection, Security, and Consumer Retention". *Working paper*, https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2018/wp_tse_

Lin, T. (2022). "Valuing Intrinsic and Instrumental Preferences for Privacy". *Marketing Science*, vol 41 (4): 663-869.

Markovich, S, and Y. Yehezkel (2023). "For the Public Benefit: Data Policy in Platform Markets". *Working paper*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3945108.

Markovich, S, and Y. Yehezkel (2024). "Competing for Cookies: Platforms' Business Models in Data Markets with Network Effects". *Working paper*, https://www.tau.ac.il/~yehezkel/Business_Models.pd

Malgieri, G. and B. Custers (2018). "Pricing Privacy– The Right to Know the Value of Your Personal Data". *Computer Law and Security Review,* vol 34: 289-303.

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

Rhodes, A and J. Zhou (2024 a). "Personalized Pricing and Competition". *American Eco-*

nomic Review, vol 114 (7): 2141-2170.

Rhodes, A and J. Zhou (2024 b). "Personalization and Privacy Choice". *Working paper*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4795118.

Schoeman, F. (1992). *Privacy and Social Freedom.* Cambridge University Press, Cambridge and New York.

Shaffer, G., and Z. J. Zhang (2002). "Competitive One-to-One Promotions," *Management Science*, vol 48(9): 1143-1160.

Smith, A., S. Seiler, and I. Aggarwal (2023). "Optimal Price Targeting". *Marketing Science*, vol 42 (3): 429-636.

Thisse, J.-F., and X. Vives (1988). "On The Strategic Choice of Spatial Price Policy". *American Economic Review*, vol 78(1): 122-137.

Westin, A. (1967). *Privacy and Freedom.* New York: Atheneum Publishers.

Online Appendix

A The monopoly firm

A.1 Mixed bundling

Proof of Proposition 1

Suppose the monopoly digital firm adopts mixed bundling, offering two schemes: $S = \{bq, p\}$ and $S_A = \{q, p_A\}$ for the bundle $\mathcal{B}$ and the stand-alone product A. Consumers derive net utility $U(\theta) = bq - p - \theta$ from consuming the bundle and $U_A = q - p_A$ from the stand-alone A. Let $\tau \equiv bq - p - U_A$ denote the cut-off threshold. Consumers with $\theta \leq \tau$ choose the bundle, while those with $\theta > \tau$ opt for the stand-alone product A.

The monopoly firm's profit is expressed as:

$$\Pi = F(\tau)(r + p) + (1 - F(\tau))\, p_A - C(q).$$

Substituting $p = bq - \tau - U_A$ and $p_A = q - U_A$, the profit function can be rewritten as:

$$\begin{aligned} \Pi(q, \tau, U_A) &= F(\tau)(r + bq - \tau - U_A) + (1 - F(\tau))(q - U_A) - C(q) \\ &= F(\tau)(r + (b-1)q - \tau) + q - U_A - C(q). \end{aligned}$$

Since $\Pi(q, \tau, U_A)$ decreases in U_A, it is optimal to set $p_A = q$, implying $U_A = 0$.

Given $p_A = q$, the firm chooses q and τ to maximize

$$\Pi(q, \tau) = F(\tau)(r + (b-1)q - \tau) + q - C(q).$$

It is straightforward to verify that $\Pi(q, \tau)$ is concave in q. The optimal response $q_m^r(\tau)$ is uniquely determined by the first-order condition:

$$C'(q) = bF(\tau) + (1 - F(\tau)).$$

Clearly, $q_m^r(\tau)$ increases in τ.

Next, differentiating $\Pi(q, \tau)$ with respect to τ yields:

$$\begin{aligned} \frac{\partial \Pi(q, \tau)}{\partial \tau} &= f(\tau)(r + (b-1)q - \tau) - F(\tau) \\ &= f(\tau)[r + (b-1)q - \tau - h(\tau)]. \end{aligned}$$

Since the term in the brackets, $r + (b-1)q - \tau - h(\tau)$, is a decreasing function of τ and $f(\tau) > 0$, $\Pi(q, \tau)$ is quasi-concave in τ. Assume $\bar{\theta}$ is sufficiently large, the optimal response $\tau_m^r(q)$, as an interior solution, is uniquely determined by the FOC:

$$\phi(\tau) = \tau + h(\tau) = (b-1)\,q + r,$$

where $\phi(\tau)$ is an increasing function. Moreover, $\tau_m^r(q)$ increases in q.

Solving the two FOCs yields the equilibrium quality and participation threshold, q_m^* and τ_m^*, respectively. The equilibrium price for the bundle is then:

$$p_m^* = bq_m^* - \tau_m^* = h(\tau_m^*) + q_m^* - r.$$

The remaining details are presented in the main text.

Leading example. Assume θ is uniformly distributed over $\left[0, \bar{\theta}\right]$ and $C(q) = q^2/2$. Suppose $\bar{\theta} > \max\{r, b^2\}$.

Under mixed bundling, the FOCs are given by:

$$\begin{aligned} \bar{\theta} q &= (b-1)\,\tau + 1, \\ 2\tau &= (b-1)\,q + r. \end{aligned}$$

Solving these equations yields:

$$q_m^* = \frac{(b-1)\,r + 2}{2\bar{\theta} - (b-1)^2},\ \tau_m^* = \frac{(b-1) + r\bar{\theta}}{2\bar{\theta} - (b-1)^2}.$$

The equilibrium price is

$$p_m^* = bq_m^* - \tau_m^* = \frac{(b+1) - r\left(\bar{\theta} - b\,(b-1)\right)}{2\bar{\theta} - (b-1)^2},$$

The equilibrium price p_m^* is negative if:

$$r > \frac{b+1}{\bar{\theta} - b\,(b-1)}.$$

Finally, it is straightforward to verify that $\bar{\theta} > \max\{r, b^2\}$ ensures the interior optimum where $\tau_m^* < \bar{\theta}$.

A.2 Pure bundling

Suppose the monopoly firm is committed to pure bundling, offering a uniform bundle scheme $S = \{bq, p\}$ to all consumers. Consumers who choose the bundle receive a net utility $U(\theta) = bq - p - \theta$, while their reservation utility is zero. This results in the same cut-off threshold as under mixed bundling. The monopoly firm earns profit solely from the bundle, which is expressed as:

$$\Pi(\tau, q) = F(\tau)\,(bq + r - \tau) - C(q)\,. \tag{16}$$

The constraint of a uniform scheme has two opposing effects on consumer welfare:

Reduction in marginal benefit. It reduces the marginal benefit of the digital product by $(1 - F(\tau))$, as the monopoly is unable to extract surplus from consumers with $\theta > \tau$. This leads to a reduction in the optimal quality, $q_p^r(\tau)$, determined by (here the subscript p stands for the optimum under pure bundling):

$$C'(q) = bF(\tau). \tag{17}$$

Increase in participation threshold. The constraint raises the optimal participation threshold, $\tau_p^r(q)$, which is given by:

$$\phi(\tau) = \tau + h(\tau) = bq + r. \tag{18}$$

Compared to the optimal responses under mixed bundling, it follows that $q_p^r(\tau) < q_m^r(\tau)$ while $\tau_p^r(q) > \tau_m^r(q)$. Consequently, the equilibrium quality and participation threshold under pure bundling, q_p^* and τ_p^*, which are jointly determined by (2) and (18), can be higher or lower depending on the parameters b and r.

The following proposition summarizes the equilibrium outcomes:

Proposition 8 *Suppose the monopoly digital firm engages in pure bundling. In equilibrium, consumers with $\theta < \tau_p^*$ opt for the bundle, receiving a net utility of $U^*(\theta) = \tau_p^* - \theta$, while these with $\tau_p^* < \theta < \bar{\theta}$ do not use the digital product. The equilibrium quality q_p^* and participation threshold τ_p^* are jointly determined by conditions* (17) *and* (18)*. Data monetization under pure bundling results in a welfare loss and reduces product quality when b is close to* 1*. In the leading example, compared to mixed bundling, pure bundling results in higher product quality when $r > r_q$ and higher consumer surplus when $r > r_\tau$.*

Proposition 8 highlights the distinct effects of data acquisition under pure bundling compared to mixed bundling.

First, data acquisition can reduce the equilibrium quality of the digital product under pure bundling, as consumers with $\theta > \tau_p^*$ do not purchase the digital product. When $b = 1$, $q_p^* < q^0$. Consequently, there exists a cut-off threshold for b, below which $q_p^* < q^0$.

Second, while data acquisition benefits consumers with $\theta < \tau_p^*$, it leads to a welfare loss from consumers with $\theta > \tau_p^*$, who do not use the digital product under pure bundling.

The monopoly firm's equilibrium profit is given by

$$\Pi_p^* = h\left(\tau_p^*\right) F\left(\tau_p^*\right) - C\left(q_p^*\right). \tag{19}$$

It is noteworthy that, under pure bundling, data acquisition may not always be profitable when b and r are sufficiently small. The compensation provided to consumers, τ, is proportional to the exogenously determined data scale $\bar{s}$. From (16), the firm's per-consumer profit, $r-\tau$, can become negative when r is small. Since the digital firm cannot choose the optimal data scale, it is forced to lower the participation threshold τ to mitigate losses. This adjustment reduces both the quality of the digital product and the firm's profitability. As a result, the firm may earn less profit than it would without data monetization.

Leading example

Consider the leading example for illustration. The FOCs under pure bundling are given by

$$\begin{aligned} b\tau &= \bar{\theta}q, \\ 2\tau &= bq + r. \end{aligned}$$

Solving for the equilibrium yields

$$q_p^* = \frac{br}{2\bar{\theta} - b^2},\ \tau_p^* = \frac{\bar{\theta}r}{2\bar{\theta} - b^2},$$

while the equilibrium profit is

$$\Pi_p^* = \frac{r^2}{2\left(2\bar{\theta} - b^2\right)}.$$

In comparison, the equilibrium profit without data monetization is $\Pi^0 = 1/2$. Therefore, data monetization is profitable if and only if:

$$r > r_\pi \equiv \sqrt{2\bar{\theta} - b^2}.$$

Furthermore, compared to the equilibrium under mixed bundling:

- the quality of the digital product is higher (i.e., $q_p^* > q_m^*$) if

$$r > r_q \equiv \frac{2\left(2\bar{\theta} - b^2\right)}{b^2 - b + 2\bar{\theta}},$$

- the participation threshold is larger (i.e., $\tau_p^* > \tau_m^*$) if

$$r > r_\tau \equiv \frac{(b-1)\left(2\bar{\theta} - b^2\right)}{\bar{\theta}\,(2b-1)}.$$

  This latter condition always holds when $b = 1$.

Consumers who choose the bundle receive a net utility: $U_p^*(\theta) = \tau_p^* - \theta$. Thus, these consumers are better off under pure bundling if and only if $\tau_p^* > \tau_m^*$. Since other consumers receive zero surplus under both bundling strategies, it follows that consumer surplus is higher under pure bundling if $r > r_\tau$.

Under pure bundling, the monopoly firm is incentivized to offer consumers a monetary payment to attract more participants. The equilibrium price for the bundle is given by:

$$p_p^* = h\left(\tau_p^*\right) - r = -\frac{\left(\bar{\theta} - b^2\right) r}{2\bar{\theta} - b^2},$$

which becomes negative under the assumption of $\bar{\theta} > b^2$.

A.3 Non-negative-price constraint

Consider the firm's schemes $S = \{bq, p\}$ for the bundle and $S_A = \{q, p_A\}$ for the stand-alone product. Consumers derive a net utility $U(\theta) = bq - p - \theta$ from consuming the bundle, and $U_A = q - p_A$ from the stand-alone A. Let $\hat{\tau} \equiv (b-1)\,q - p + p_A$ denote the cut-off threshold. The monopoly chooses p, q, and p_A to maximize

$$\begin{aligned} \Pi &= F(\hat{\tau})(r+p) + (1 - F(\hat{\tau}))\,p_A - C(q) \\ &= F(\hat{\tau})(r + p - p_A) + p_A - C(q), \end{aligned}$$

subject to the constraint $p \geq 0$ and $q - p_A \geq 0$

The Lagrangian for this optimization program is given by

$$\mathcal{L} = F(\hat{\tau})(r+p) + (1 - F(\hat{\tau}))\,p_A - C(q) + \lambda p + \mu (q - p_A),$$

where $\lambda \geq 0$ and $\mu \geq 0$ are the multipliers.

The first-order conditions for the optimal q, p, and p_A are given by:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q} &= f(\hat{\tau})(b-1)(r + p - p_A) - C'(q) - \mu = 0, \\ \frac{\partial \mathcal{L}}{\partial p} &= (F(\hat{\tau}) + \lambda) - f(\hat{\tau})(r + p - p_A) = 0, \\ \frac{\partial \mathcal{L}}{\partial p_A} &= (1 - F(\hat{\tau}) - \mu) + f(\hat{\tau})(r + p - p_A) = 0. \end{aligned}$$

The FOC for the optimal quality,

$$f(\hat{\tau})(b-1)(r + p - p_A) = C'(q) + \mu,$$

implies $r + p - p_A > 0$.

Meanwhile, the FOC for p_A is given by

$$f(\hat{\tau})(r+p-p_A)=\mu-(1-F(\hat{\tau})),$$

which, given $r+p-p_A>0$, implies $\mu>0$. Therefore, the constraint $q-p_A\geq 0$ must be binding in the optimization, leading to $p_A=q$.

Substituting $p_A=q$ into the objective function, we obtain

$$\mathcal{L}=F(\hat{\tau})(r+p)+(1-F(\hat{\tau}))q-C(q)-\lambda p,$$

where $\hat{\tau}=bq-p$. The FOCs for the optimal q and p are given by

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial q} &= f(\hat{\tau})b(r+p-q)+(1-F(\hat{\tau}))-C'(q)=0,\\ \frac{\partial\mathcal{L}}{\partial p} &= (F(\hat{\tau})-\lambda)-f(\hat{\tau})(r+p-q)=0.\end{aligned}$$

Combining the two FOCs yields

$$(1-F(\hat{\tau}))+(F(\hat{\tau})-\lambda)b=C'(q).$$

When the constraint $p\geq 0$ is binding, $\lambda>0$. Compared to the optimal q_m^* without the constraint ($\lambda=0$), as given by

$$(1-F(\hat{\tau}))+F(\hat{\tau})b=C'(q),$$

it follows that $\hat{q}^*<q_m^*$. This also implies

$$\hat{\tau}^*=b\hat{q}^*<bq_m^*<bq_m^*-p_m^*=\tau_m^*,$$

where the last equality comes from $p_m^*<0$.

The monopoly firm's optimal price for the bundle becomes negative when $r>h(\tau^*)+q^*$. When the non-negative-price constraint binds, the firm sets $p=0$ for the bundle. Under this constraint, the cut-off threshold is given by $\hat{\tau}=bq$. The monopoly firm's profit is

$$\begin{aligned}\Pi &= F(\tau)r+(1-F(\tau))q-C(q)\\ &= q-C(q)+F(\tau)(r-q).\end{aligned}$$

The optimal quality $\hat{q}^*$ is determined by the first-order condition:

$$C'(q)=(1-F(\tau))+f(\tau)(r-q)b, \tag{20}$$

while the equilibrium threshold is given by

$$\hat{\tau}^*=b\hat{q}^*. \tag{21}$$

The non-negative-price constraint results in a lower quality level, $\hat{q}^* < q_m^*$, and a lower participation threshold, $\hat{\tau}^* < \tau_m^*$. Consequently, consumers opting for the bundle experience reduced consumer surplus:

$$\hat{U}^*(\theta) = \hat{\tau}^* - \theta < U_m^*(\theta) = \tau_m^* - \theta.$$

In contrast, consumers choosing the stand-alone digital product receive zero surplus, as in the case without the constraint. Therefore, the non-negative-price constraint leads to reductions in total consumer surplus, the digital firm's profit, and overall social welfare.

This analysis is summarized in the following proposition:

Corollary 1 *Suppose the monopoly digital firm faces a binding non-negative-price constraint. The constraint reduces the quality of services, the participation threshold, and the net utility for consumers choosing the bundle, without affecting other consumers.*

Leading example

In the leading example, the FOC for the optimal quality is given by

$$\left(\bar{\theta} - \tau\right) + (r - q)\, b = \bar{\theta} q.$$

Solving for the optimal quality yields

$$\hat{q}^* = \frac{\bar{\theta} + rb}{\bar{\theta} + 2b},$$

and the equilibrium threshold is

$$\hat{\tau}^* = b\hat{q}^* = b\left(\frac{\bar{\theta} + rb}{\bar{\theta} + 2b}\right).$$

A.4 Impact of digital privacy regulation

The DMA model

The equilibrium of the DMA model is characterized in the main text. In the leading example, the first-order conditions are:

$$\begin{aligned}(b-1)\,\tau &= \bar{\theta} q \\ (b-1)\, q + r &= 2\tau.\end{aligned}$$

Solving for the equilibrium leads to

$$\begin{aligned}\tau_d^* &= \frac{r\bar{\theta}}{2\bar{\theta} - (b-1)^2}, \\ q_d^* &= \frac{r(b-1)}{2\bar{\theta} - (b-1)^2}.\end{aligned}$$

Opt-in consumers receive

$$U_d^*(\theta) = \tau_d^* + q_d^* - \theta = \frac{\left(\bar{\theta} + b - 1\right) r}{2\bar{\theta} - (b-1)^2} - \theta,$$

while opt-out consumers obtain q_d^*. The consumer surplus under the DMA model is:

$$CS_d = \frac{1}{\bar{\theta}} \int_0^{\tau_d^*} \left(\tau_d^* - \theta\right) d\theta + q_d^* = \frac{\left(\tau_d^*\right)^2}{2\bar{\theta}} + q_d^* = \frac{1}{2\bar{\theta}} \left(\frac{r\bar{\theta}}{2\bar{\theta} - (b-1)^2}\right)^2 + \frac{r(b-1)}{2\bar{\theta} - (b-1)^2}.$$

Comparison

Recall that the equilibrium outcome under optimal mixed bundling is given by

$$q_m^* = \frac{(b-1) r + 2}{2\bar{\theta} - (b-1)^2}, \ \tau_m^* = \frac{(b-1) + r\bar{\theta}}{2\bar{\theta} - (b-1)^2}.$$

Consumer surplus under optimal mixed bundling is:

$$CS_m = \frac{1}{\bar{\theta}} \int_0^{\tau_m^*} \left(\tau_m^* - \theta\right) d\theta = \frac{\left(\tau_m^*\right)^2}{2\bar{\theta}} = \frac{1}{2\bar{\theta}} \left(\frac{(b-1) + r\bar{\theta}}{2\bar{\theta} - (b-1)^2}\right)^2.$$

Comparing the consumer surplus under two models, then $CS_d > CS_m$ if

$$2\bar{\theta}\left(2\bar{\theta} + 2b - b^2 - 2\right) r > (b-1),$$

which simplifies to

$$r > r_d = \frac{b-1}{2\bar{\theta}\left(2\bar{\theta} + 2b - b^2 - 2\right)}.$$

The Meta model

Without NNPC, Meta would adopt the optimal mixed bundling strategy, with the equilibrium quality q_m^* and the threshold τ_m^*. Suppose the NNPC constraint is binding, which occurs when $r > h\left(\tau_m^*\right) + q_m^*$. In this case, Meta offers two schemes for the bundle and the stand-alone digital services: $S = \{bq, 0\}$ and $S_A = \{q, p_A\}$, with $p_A = q$. The equilibrium quality $\hat{q}^*$ and participation threshold $\hat{\tau}^*$ are characterized in the above subsection.

In the leading example,

$$\hat{q}^* = \frac{\bar{\theta} + rb}{\bar{\theta} + 2b}.$$

Comparing the quality under two models, $\hat{q}^* > q_d^*$ if and only if

$$\frac{\bar{\theta} + rb}{\bar{\theta} + 2b} > \frac{r(b-1)}{2\bar{\theta} - (b-1)^2}.$$

which simplifies to

$$\bar{\theta}\left(2\bar{\theta} - (b-1)^2\right) + r(b+1)\left(\bar{\theta} - b^2 + b\right) > 0.$$

The above condition holds true under the assumption $\bar{\theta} > b^2$.

A.5 Data property rights

Suppose the data monetization revenue from the intermediary is r_i. Consumers who choose to share their data can receive

$$\max\{U(\theta) = bq - p - \theta, U_i(\theta) = r_i - \theta\},$$

while those opting for the stand-alone product obtain 0. They will share data if $\theta \leq \max\{\tau_i, r_i\}$, where $\tau_i = bq - p$.

Suppose the digital firm wins the competition, in which $\tau_i = bq - p \geq r_i$. The firm chooses q and τ_i to maximize its profit

$$\begin{aligned} \Pi_i &= F(\tau_i)(r + p) + (1 - F(\tau_i))\, q - C(q) \\ &= F(\tau_i)(r + (b-1)\, q - \tau_i) + q - C(q), \end{aligned}$$

subject to $\tau_i \geq r_i$.

Case (a). If $r_i \leq \tau_m^*$, the constraint is not binding. The optimization leads to the same equilibrium outcome as under optimal mixed bundle: $\tau_i^* = \tau_m^*$ and $q_i^* = q_m^*$.

Case (b). If $r_i > \tau_m^*$, the constraint is binding. In this case, the optimal cut-off threshold is determined by $\tau_i^* = r_i > \tau_m^*$. The digital firm's profit is given by

$$\Pi_i = F(r_i)(r + (b-1)\, q - r_i) + q - C(q).$$

The optimal quality q_i^* is determined by

$$C'(q) = 1 + (b-1)\, F(\tau_i) = 1 + (b-1)\, F(r_i).$$

Compared to the FOC for q under the mixed bundling (2), $q_i^* > q_m^*$ since $\tau_i^* = r_i > \tau_m^*$. Consumers choosing the bundle receive a net utility $U_i^*(\theta) = \tau_i^* - \theta = r_i - \theta$, while they obtain zero otherwise. Since $\tau_i^* > \tau_m^*$, the data intermediary mechanism increases the product quality and consumer surplus.

In the competition for data acquisition, the digital firm can offer at most $r + (b-1)\, q_i^*$. Therefore, the digital firm loses the competition when $r_i > r + (b-1)\, q_i^*$, in which case it will focus solely on the digital product. The firm chooses the optimal quality q^0 and earns a profit $\Pi^0 = q^0 - C\left(q^0\right)$. In this scenario, while the data intermediary mechanism increases consumer surplus, it reduces the quality of the digital product.

B Data acquisition and competition

B.1 Analysis of data-driven mergers

The equilibrium prices p_i^* and p_e^{A*} are characterized in the main text. The equilibrium threshold τ_c^* is determined by

$$\varphi(\tau_c) \equiv \tau_c + h(\tau_c) - k(\tau_c) = \beta - \delta,$$

where $\beta = r + (b-1)q$. For a given δ, the threshold τ_c increases in β. Then, the assumption $\beta < \bar{\beta} = \bar{\theta} + \frac{1}{f(\bar{\theta})} + \delta$ ensures $\tau_c^* < \bar{\theta}$.

Since p_e^{A*} decreases in τ_c, it also decreases in β. When $\beta \to \delta$, $\tau_c \to 0$, and $p_e^{A*} \to 1/f(0) > \delta$. Conversely, as $\beta \to \bar{\beta}$, $\tau_c \to \bar{\theta}$, and $p_e^{A*} \to 0$. Therefore, there exists a unique cut-off threshold $\hat{\beta} \in (\delta, \bar{\beta})$ such that $p_e^{A*} > \delta$ if and only if $\beta < \hat{\beta}$.

The equilibrium profits are given respectively by

$$\Pi_i^* = F(\tau_c^*) h(\tau_c^*), \ \Pi_e^* = p_e^{A*} (1 - F(\tau_c^*)).$$

Note that Π_e^* decreases in β. As $\beta \to \delta$, $\tau_c \to 0$, and $p_e^{A*} \to 1/f(0) > \delta$, implying $\Pi_e^* > \delta$. In contrast, when $\beta < \hat{\beta}$, $p_e^{A*} < \delta$, implying $\Pi_e^* < \delta$. Thus, there exists a cut-off threshold $\tilde{\beta} \in (\delta, \hat{\beta})$ such that $\Pi_e^* > \delta$ if and only if $\beta < \tilde{\beta}$.

Illustrative Example. Assume θ is uniformly distributed over $\left[0, \bar{\theta}\right]$. Then,

$$\tau_c^* = \frac{\bar{\theta} + \beta - \delta}{3}.$$

The equilibrium prices are

$$\begin{aligned} p_i^* &= \tau - r = \frac{\bar{\theta} + \beta - \delta}{3} - r, \\ p_e^{A*} &= \bar{\theta} - \tau = \frac{2\bar{\theta} - \beta + \delta}{3}. \end{aligned}$$

Then, $p_i^* < 0$ if

$$r > \frac{(b-1)q + \bar{\theta} - \delta}{2}.$$

Note that $p_e^{A*} > \delta$ if and only if $\beta < \hat{\beta} = 2\bar{\theta} - 2\delta$.

Firm e's profit is given by

$$\Pi_e^* = p_e^{A*} (1 - F(\tau_c^*)) = \frac{\left(p_e^{A*}\right)^2}{\bar{\theta}}.$$

Then, $\Pi_e^* > \delta$ if

$$p_e^{A*} = \left(\frac{2\bar{\theta} - \beta + \delta}{3}\right)^2 > \bar{\theta}\delta,$$

which simplifies to

$$\beta < \tilde{\beta} \equiv 2\bar{\theta} + \delta - 3\sqrt{\bar{\theta}\delta}.$$

Note that $\delta < 1/f(0) < \bar{\theta}$ ensures $\tilde{\beta} < \hat{\beta}$.

Remark: Note that, $p_e^{A*} \leq q + \delta$ requires

$$\bar{\theta} \leq \frac{(b+1)\, q + 2\delta + r}{2}.$$

If this condition does not hold, then $p_e^{A*} = q + \delta$, implying that $U_e^* = 0$. In this case, firm i operates as a monopoly under pure bundling, and the cut-off threshold is determined by $\tau_c = bq - p_i$.

Comparison of consumer surplus.

Before the merger, each consumer receives a net utility q, and the consumer surplus is equal to q. After the merger, consumers who choose the stand-alone product receive $U_e^* = q + \delta - p_e^{A*}$, while those opting for the bundle receive $U_i^* = U_e^* + \tau_c^* - \theta$. Assume a uniform distribution, the consumer surplus is expressed as

$$\begin{aligned} S_c &= \frac{1}{\bar{\theta}} U_e^* \left(\bar{\theta} - F(\tau_c^*)\right) + \frac{1}{\bar{\theta}} \int_0^{\tau_c^*} \left(U_e^* + \tau_c^* - \theta\right) d\theta = U_e^* + \frac{1}{\bar{\theta}} \frac{(\tau_c^*)^2}{2} \\ &= q + \delta - p_e^{A*} + \frac{1}{2\bar{\theta}} \left(\bar{\theta} - p_e^{A*}\right)^2 = q + \frac{1}{2\bar{\theta}} \left(\bar{\theta}^2 - 4\bar{\theta} p_e^{A*} + 2\delta\bar{\theta} + \left(p_e^{A*}\right)^2\right). \end{aligned}$$

Hence, the merger reduces consumer surplus if

$$\left(p_e^{A*}\right)^2 - 4\bar{\theta} p_e^{A*} + 2\delta\bar{\theta} + \bar{\theta}^2 < 0,$$

which simplifies to

$$p_e^{A*} > 2\bar{\theta} - \sqrt{3\bar{\theta}^2 - 2\delta\bar{\theta}}.$$

This holds when

$$\beta < \underline{\beta} \equiv 3\sqrt{3\bar{\theta}^2 - 2\delta\bar{\theta}} + \delta - 4\bar{\theta}.$$

B.2 Privacy protection and differentiation in business models

We have characterized the equilibrium in a subgame where firm i commits to pure bundling while firm e supplies only the digital product. In this equilibrium, each firm earns a positive profit. We now show that this is the unique subgame perfect Nash equilibrium. The proof proceeds in three steps, as outlined below:

Step 1. Suppose firm i engages in pure bundling. There exists an equilibrium in which firm e focuses solely on the digital product.

The equilibrium is characterized in the main text, where each firm earns a positive profit. We now show that there are no profitable deviations for firm e. Consider two cases:

Case (1). Suppose firm e engages in data acquisition and adopts the pure bundling strategy. Let p_i and p_e denote the prices for bundle offered by firm i and firm e, respectively. Consumers receive a net utility $U_i(\theta) = bq - p_i - \theta$ and $U_e(\theta) = bq - p_e - \theta$ from the bundle i and bundle e, respectively. Consumers prefer bundle i to bundle e if $p_i \leq p_e$. Since firm i earns a per-consumer revenue of $p_i + r$ while firm e earns p_e, Bertrand-type competition for the bundles leads to $p_i = p_e = 0$, in which firm i wins the competition, earning a profit of r, while firm e earns zero profit.

Case (2). Suppose firm e adopts mixed bundling by offering p_e for the bundle and and p_e^A for the stand-alone product. Consider the price competition in Stage 3. Let $p_b \equiv \min\{p_e, p_i\}$, the cut-off threshold is determined by

$$\tau_c = (b-1)\, q + p_e^A - p_b.$$

Betrand competition for the bundle leads to $p_b^* = p_e = p_i = 0$ in equilibrium, and firm e earns a profit only from the stand-alone product:

$$\Pi_e = p_e^A \left(1 - F\left(\tau_c\right)\right).$$

Let $\hat{\Pi}_e^*$ denote firm e's maximum profit under deviation, which increases with p_b by the envelope theorem. Then, since $p_b^* = 0$, it follows that $\hat{\Pi}_e^* < \Pi_e^* = p_e^{A*}\left(1 - F\left(\tau_c^*\right)\right)$.

Step 2. Suppose firm i engages in mixed bundling. There exists a unique equilibrium in which firm e focuses on the digital product only.

Suppose firm e supplies the digital product only at the price p_e^A. Then, firm i's offer of stand-alone product imposes a price cap for firm e, leading to $p_e^A \leq \delta$ in equilibrium. If firm e charges $p_e^A > \delta$, firm i can undercut the rival by setting $p_i^A = p_e^A - \delta - \varepsilon > 0$, earning a profit.

Consumers choosing the bundle receive $U_i(\theta) = bq - p_i - \theta$ while those opting for firm e's product obtains $U_e = q + \delta - p_e^A$. The cut-off threshold becomes

$$\tau_c = (b-1)\, q - \delta - p_i + p_e^A,$$

and firms profits are given by

$$\Pi_i = F\left(\tau_c\right)\left(r + p_i\right), \ \Pi_e = \left(1 - F\left(\tau_c\right)\right) p_e^A.$$

The equilibrium outcome can be characterized using the same approach as before. Firm e's equilibrium price is $p_e^{A*} < \delta$ when $\beta > \hat{\beta}$, while $p_e^{A*} = \delta$ when $\beta < \hat{\beta}$, which can be expressed by

$$p_e^{A*} = \min\{\delta, k(\tau_c)\}.$$

We now show that firm e cannot profitable deviate. Consider two cases.

Case (1). Suppose firm e engages in data acquisition and adopts the pure bundling strategy. Applying the same logic as in Step 1, firm e cannot benefit from pure bundling due to its competitive disadvantage in the bundle.

Case (2). Suppose firm e adopts mixed bundling in Stage 2. Consider the pricing game in Stage 3. Let p_i^A and p_e^A denote the prices for the stand-alone digital product A set by the firms i and e, respectively, along with the prices for the bundle p_i and p_e. In equilibrium, competition for the stand-alone product must lead to $p_e^A \leq \delta$. Similarly, competition for the bundle leads to $p_i \leq 0$. Consumers with $\theta \leq \tau_c = (b-1)q - \delta - p_i + p_e^A$ choose firm i's bundle, while those with $\theta > \tau$ opt for firm e's stand-alone product. The firms' profits are given by

$$\begin{aligned}\Pi_i &= F(\tau_c)(r + p_i), \\ \Pi_e &= (1 - F(\tau_c))\, p_e^A.\end{aligned}$$

Then, the optimal price p_e^A is given by

$$p_e^A = \min\{\delta, k(\tau_c)\}.$$

Since p_e^A and Π_e increase in p_i, competition for the bundle makes firm e worse off. Therefore, firm e does not benefit from mixed bundling.

Step 3. Comparing firm i's equilibrium profits under two bundling strategies, and noting that firm i's profit increases in p_e^{A*}, we can conclude that firm i is better off under pure bundling. Therefore, there exists a unique SPNE in which firm i adopts pure bundling while firm e focuses solely on the stand-alone product.

C Heterogeneous preferences and personalized pricing

C.1 Heterogeneous data-driven benefit

The data-driven benefit is expressed as $bq = (1+x)q$, where x represents a consumer's preference and is distributed over $[0,1]$ according to a distribution function $G(\cdot)$, with a strictly increasing density $g(\cdot)$. Let μ_x and σ_x denote the mean and variance of x, respectively, defined as

$$\mu_x \equiv \int_0^1 x dG(x), \ \sigma_x \equiv \int_0^1 x^2 dG(x).$$

Consumers are assumed to learn their preferences x and privacy sensitivity θ prior to making purchase decisions. In addition, the distributions of θ and x are uncorrelated.

The following assumption ensures $0 < \tau(x) < \bar{\theta}$ and $0 < \hat{\tau}(x) < \bar{\theta}$ for all $x \in [0,1]$.

Assumption C:

$$\bar{\theta} > \bar{\theta}_l^x \equiv \max\{2q, 4\mu_x q\} + \frac{r}{2}.$$

Personalized pricing

Suppose firm i engages in pure bundling, offering personalized prices $p_i(x)$ for its bundle. The cut-off threshold is given by

$$\tau(x) = (1+x)q - p_i(x) - U_e = xq - p_i(x) + p_e^A.$$

Focusing on the configuration with $0 < \tau(0) < \tau(1) < \bar{\theta}$, firm i's profit is given by

$$\Pi_i = \frac{1}{\bar{\theta}} \int_0^1 (p_i(x) + r)\,\tau(x)\, dG(x).$$

It is straightforward to verify that Π_i is concave in $p_i(x)$. Therefore, the optimal prices are uniquely determined by the first-order conditions.

Under personalized pricing, firm i can choose $p_i(x)$ to maximize its profit from each consumer, given by $(p_i(x)+r)\,\tau(x)\, g(x)$ for all x, leading to the following first-order condition:

$$p_i(x) + r = \tau(x).$$

Solving the first-order condition yields

$$p_i(x) = \frac{xq - r + p_e^A}{2}.$$

Meanwhile, firm e chooses p_e^A to maximize its profit

$$\Pi_e = \frac{p_e^A}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \tau(x)\right) dG(x),$$

subject to $p_e^A \leq q$.

Consider two cases:

Case (1). The optimal price is an interior optimum: $p_e^{A*} \leq q$.

In this scenario, firm e's optimal price is determined by:

$$p_e^A = \bar{\theta} - \int_0^1 \tau(x)\, dG(x) = \bar{\theta} - \int_0^1 \left(\frac{xq + r + p_e^A}{2}\right) dG(x).$$

Solving for the optimal price leads to

$$p_e^{A*} = \frac{2\bar{\theta} - r - \mu_x q}{3}.$$

For $p_e^{A*} < q$, the condition is

$$\bar{\theta} < \bar{\theta}_u^x \equiv \frac{(3 + \mu_x) q + r}{2}.$$

The equilibrium personalized prices are given by:

$$p_i^* (x) = \frac{xq - r + p_e^A}{2} = \frac{xq}{2} + \frac{\bar{\theta}}{3} - \frac{2r}{3} - \frac{\mu_x q}{6},$$

and the equilibrium participation threshold is

$$\tau^* (x) = p_i (x) + r = \frac{xq}{2} + \frac{\bar{\theta}}{3} + \frac{r}{3} - \frac{\mu_x q}{6}$$

The equilibrium arises when $\tau (0) > 0$, which requires

$$\bar{\theta} > \frac{\mu_x q}{2} - r.$$

Additionally, $\tau (1) < \bar{\theta}$ holds when

$$\bar{\theta} > \frac{(3 - \mu_x) q}{4} + \frac{r}{2}.$$

Then, the equilibrium exists under Assumption C.

Case (2). Firm e's optimal price is bounded by q: $p_e^{A*} = q$.

This occurs when $\bar{\theta} > \bar{\theta}_u^x$. In this case firm e charges the monopoly price, extracting the full consumer surplus from the stand-alone product (i.e., $U_e = 0$). In contrast, firm i operates as a monopoly supplier of the bundle, in which the equilibrium personalized price and participation threshold are given by

$$\begin{aligned} p_i^* (x) &= \frac{(1 + x) q - r}{2}, \\ \tau^* (x) &= \frac{(1 + x) q + r}{2}. \end{aligned}$$

In this case, $0 < \tau^* (0) < \tau^* (1) < \bar{\theta}$ holds under Assumption C.

The equilibrium profits of both firms are given by

$$\begin{aligned} \Pi_i^* &= \frac{1}{\bar{\theta}} \int_0^1 (\tau^* (x))^2 \, dG (x), \\ \Pi_e^* &= \frac{p_e}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \tau (x)\right) dG (x) = \frac{\left(p_e^{A*}\right)^2}{\bar{\theta}}. \end{aligned}$$

Consumers choosing firm e's product receive a net utility $U_e^* = q - p_e^{A*} \geq 0$. In contrast, consumers choosing the bundle receive $U_i^*(\theta, x) = U_e^* + \tau^*(x) - \theta$. Finally, the consumer surplus is given by

$$\begin{aligned} S^* &= \frac{U_e^*}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \tau^*(x)\right) dx + \frac{1}{\bar{\theta}} \int_0^1 \int_0^{\tau^*(x)} \left(U_e^* + \tau^*(x) - \theta\right) d\theta dx \\ &= U_e^* + \frac{1}{\bar{\theta}} \int_0^1 \int_0^{\tau^*(x)} \left(\tau^*(x) - \theta\right) d\theta dx = U_e^* + \frac{1}{\bar{\theta}} \int_0^1 \frac{\left(\tau^*(x)\right)^2}{2} dx. \end{aligned}$$

Uniform pricing

Suppose now firm i charges the uniform price p_i for the bundle, in which a consumer with type (θ, x) receives $\hat{U}_i(\theta, x) = (1 + x)q - p_i - \theta$. The cut-off threshold now becomes

$$\hat{\tau}(x) \equiv (1 + x)q - p_i - U_e = xq - p_i + p_e^A.$$

Firm i's profit is given by

$$\hat{\Pi}_i = \frac{(p_i + r)}{\bar{\theta}} \int_0^1 \hat{\tau}(x)\, dG(x).$$

The optimal uniform price is determined by

$$p_i + r = \int_0^1 \hat{\tau}(x)\, dG(x) = \int_0^1 \left(xq - p_i + p_e^A\right) dG(x).$$

Solving the FOC leads to

$$p_i = \frac{\mu_x q + p_e^A - r}{2}.$$

The optimal participation threshold is

$$\hat{\tau}(x) = xq - p_i + p_e^A = \frac{(2x - \mu_x)\, q + p_e^A + r}{2}.$$

Meanwhile, firm e choose p_e^A to maximize the following profit

$$\hat{\Pi}_e = \frac{p_e^A}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \hat{\tau}(x)\right) dG(x),$$

subject to $p_e^A \leq q$.

Consider two cases:

Case (1). Interior optimum with $p_e^{A*} < q$.

This occurs when $\bar{\theta} > \bar{\theta}_u^x$. In this scenario, firm e's optimal uniform price is determined by

$$p_e^A = \bar{\theta} - \int_0^1 \hat{\tau}(x)\, dG(x) = \bar{\theta} - p_i - r.$$

Solving for the equilibrium prices yields

$$
\begin{aligned}
\hat{p}_i^* &= \frac{1}{3}\bar{\theta} - \frac{2}{3}r + \frac{1}{3}\mu_x q, \\
\hat{p}_e^{A*} &= \frac{2\bar{\theta} - r - \mu_x q}{3}.
\end{aligned}
$$

The equilibrium price $\hat{p}_e^{A*} < q$ if $\bar{\theta} < \bar{\theta}_u^x$. The equilibrium participation threshold is

$$\hat{\tau}^* (x) = \frac{1}{3}\bar{\theta} + \frac{1}{3}r + xq - \frac{2}{3}\mu_x q.$$

This equilibrium arises when $\hat{\tau}^* (0) > 0$ and $\hat{\tau}^* (1) < \bar{\theta}$. Notably, $\hat{\tau}^* (0) > 0$ requires

$$\bar{\theta} > 2\mu_x q - r,$$

while $\hat{\tau}^* (1) < \bar{\theta}$ holds if

$$\bar{\theta} > \frac{(3 - 2\mu_x) q}{2} + \frac{r}{2}.$$

Thus, the equilibrium exists under Assumption C.

Case (2). The boundary optimum with $p_e^{A*} < q$.

This happens when $\bar{\theta} > \bar{\theta}_u^x$. In this case, firm i's optimal uniform price is given by

$$\hat{p}_i^* = \frac{(1 + \mu_x) q - r}{2},$$

and the participation threshold is

$$\hat{\tau}^* (x) = \frac{(1 + 2x - \mu_x) q + r}{2}.$$

Note that $0 < \hat{\tau}^* (0) < \hat{\tau}^* (1) < \bar{\theta}$ holds under Assumption C.

The equilibrium profits for both firms are given respectively by

$$
\begin{aligned}
\hat{\Pi}_i^* &= \frac{1}{\bar{\theta}} \left(\int_0^1 \hat{\tau}^* (x) \, dG (x) \right)^2 = \frac{1}{\bar{\theta}} (\hat{p}_i^* + r)^2, \\
\hat{\Pi}_e^* &= \frac{\hat{p}_e^{A*}}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \hat{\tau}^* (x) \right) dG (x) = \frac{\left(\hat{p}_e^{A*} \right)^2}{\bar{\theta}}.
\end{aligned}
$$

Consumers choosing the bundle receive $\hat{U}_i^* (\theta, x) = \hat{U}_e^* + \hat{\tau}^* (x) - \theta$, while those opting for firm e's product receive $\hat{U}_e^*$. The consumer surplus is

$$\hat{S}^* = \hat{U}_e^* + \frac{1}{\bar{\theta}} \int_0^1 \frac{(\hat{\tau}^* (x))^2}{2} dG (x).$$

Comparison

Firm e charges the same price under both schemes: $p_e^{A*} = \hat{p}_e^{A*}$, earning the same profit: $\Pi_e^* = \hat{\Pi}_e^*$. This implies the same consumer utility from firm e's product: $U_e^* = \hat{U}_e^*$. For simplicity, we will use $U_e = U_e^* \geq 0$ as a parameter in the comparison.

Firm i's prices under two schemes are given by

$$\begin{aligned} p_i^*(x) &= \frac{(1+x)q}{2} - \frac{U_e + r}{2} = \frac{xq}{2} + \frac{q - U_e - r}{2}, \\ \hat{p}_i^* &= \frac{\mu_x q}{2} + \frac{q - U_e - r}{2}. \end{aligned}$$

Thus, $\hat{p}_i^*$ is the mean of $p_i^*(x)$:

$$\hat{p}_i^* = \int_0^1 p_i^*(x)\, dG(x).$$

The participation thresholds under two schemes are given by

$$\begin{aligned} \tau^*(x) &= \frac{xq}{2} + \frac{q - U_e + r}{2}, \\ \hat{\tau}^*(x) &= xq - \frac{\mu_x q}{2} + \frac{q - U_e + r}{2}. \end{aligned}$$

Consumers with $x < \mu_x$ are better off under personalized pricing: $\tau^*(x) > \hat{\tau}^*(x)$ if and only if $x < \mu_x$.

Denoting by $\omega \equiv \frac{q - U_e + r}{2}$, firm i's profits under two schemes are expressed as:

$$\begin{aligned} \Pi_i^* &= \frac{1}{\bar{\theta}} \int_0^1 (\tau^*(x))^2\, dG(x) = \frac{1}{\bar{\theta}} \int_0^1 \left(\frac{xq}{2} + \omega\right)^2 dG(x) = \frac{1}{\bar{\theta}} \left(\frac{\sigma_x q^2}{4} + \mu_x q\omega + \omega^2\right), \\ \hat{\Pi}_i^* &= \frac{1}{\bar{\theta}} \left(\int_0^1 \hat{\tau}^*(x)\, dG(x)\right)^2 = \frac{1}{\bar{\theta}} \left(\frac{\mu_x q}{2} + \omega\right)^2 = \frac{1}{\bar{\theta}} \left(\frac{\mu_x^2 q^2}{4} + \mu_x q\omega + \omega^2\right). \end{aligned}$$

Since $\sigma_x > \mu_x^2$, it follows that $\Pi_i^* > \hat{\Pi}_i^*$.

Consumer surplus under personalized pricing and uniform pricing is expressed as

$$\begin{aligned} S^* &= U_e + \frac{1}{2\bar{\theta}} \int_0^1 (\tau^*(x))^2\, dG(x) = U_e + \frac{1}{2\bar{\theta}} \left(\frac{\sigma_x q^2}{4} + \mu_x q\omega + \omega^2\right), \\ \hat{S}^* &= U_e + \frac{1}{2\bar{\theta}} \int_0^1 (\hat{\tau}^*(x))^2\, dG(x) = U_e + \frac{1}{2\bar{\theta}} \left(q^2 \left(\sigma_x - \frac{3\mu_x^2}{4}\right) + \mu_x q\omega + \omega^2\right). \end{aligned}$$

Therefore, consumer surplus is lower under personalized pricing: $S^* < \hat{S}^*$.

Finally, the total social welfare under personalized pricing is expressed as:

$$\begin{aligned} W^* &= \Pi_i^* + \Pi_e^* + S^* = \Pi_e^* + U_e^* + \frac{3}{2\bar{\theta}} \int_0^1 (\tau^*(x))^2\, dG(x) \\ &= \Pi_e^* + U_e^* + \frac{1}{2\bar{\theta}} \left(\frac{3\sigma_x q^2}{4} + 3\mu_x q\omega + 3\omega^2\right), \end{aligned}$$

where the total social welfare under uniform pricing is given by:

$$\begin{aligned}\hat{W}^* &= \hat{\Pi}_i^* + \Pi_e^* + \hat{S}^* = \Pi_e^* + U_e^* + \frac{1}{2\bar{\theta}}\int_0^1 (\hat{\tau}^*(x))^2 dx + \frac{1}{\bar{\theta}}\left(\int_0^1 \hat{\tau}^*(x)\, dx\right)^2 \\ &= \Pi_e^* + U_e^* + \frac{1}{2\bar{\theta}}\left(q^2\left(\sigma_x - \frac{\mu_x^2}{4}\right) + 3\mu_x q\omega + 3\omega^2\right).\end{aligned}$$

Since

$$\hat{W}^* - W^* = \frac{q^2}{2\theta}\left(\frac{\sigma_x - \mu_x^2}{4}\right) > 0,$$

the total social welfare is lower under personalized pricing than under uniform pricing: $W^* < \hat{W}^*$.

C.2 Heterogeneous data-monetization revenue

Suppose r is distributed over $[0,1]$ according to the distribution function $G(\cdot)$, with a strictly positive density $g(\cdot)$. Let μ_r and σ_r denote the mean and variance of r, respectively, defined as

$$\mu_r = \int_0^1 r dr,\ \sigma_r = \int_0^1 r^2 dr.$$

Consider the competition between two firms i and e, where firm i can offer personalized prices $p_i(r)$ based on a consumer's data-monetization value r. Consumers are unaware of their value from data monetization before choosing the bundle.

Assume the firms' costs are sunk, and for analytical tractability, θ is uniformly distributed over $\left[0,\bar{\theta}\right]$.

Personalized pricing

Firm i engages in pure bundling and offers personalized prices $p_i(r)$ for the bundle. The following assumption ensure the existence of equilibrium with partial participation for all consumers: $0 < \tau(x) < \bar{\theta}$ for all $x \in [0,x]$.

Assumption D:

$$\bar{\theta} > \bar{\theta}_l^r \equiv 1 + \frac{bq}{2}.$$

A consumer of type (θ, r) receive $U_i(\theta, r) = bq - p_i(r) - \theta$ from firm i's bundle while $U_e = q - p_e^A$ from firm e's stand-alone product. The consumer chooses the bundle if

$$\theta \leq \tau(r) = bq - p_i(r) - U_e.$$

Firm i's profit is given by

$$\Pi_i = \frac{1}{\bar{\theta}}\int_0^1 (r + p_i(r))\, \tau(r)\, dG(r).$$

The optimal personalized price is determined by

$$p_i(r) = \tau(r) - r.$$

Solving for the optimal $p_i(r)$ yields

$$p_i(r) = \frac{bq - U_e - r}{2} = \frac{(b-1)q + p_e^A - r}{2}.$$

The optimal participation threshold is

$$\tau(r) = \frac{bq - U_e + r}{2} = \frac{(b-1)q + p_e^A + r}{2}.$$

Meanwhile, firm e chooses $p_e^A \leq q$ to maximize its profit

$$\Pi_e = \frac{p_e^A}{\bar{\theta}} \int_0^1 \left(\bar{\theta} - \tau(r)\right) dG(r).$$

Consider two scenarios.

Scenario (1). Interior optimum with $p_e^{A*} < q$.

This occurs when

$$\bar{\theta} < \bar{\theta}_u^r \equiv \frac{(b+2)q}{2} + \frac{\mu_r}{2}.$$

In this case, firm e's optimal price is given by:

$$p_e^A = \bar{\theta} - \int_0^z \tau(r)\, dG(r) = \bar{\theta} - \int_0^z \left((b-1)q - p_i(r) + p_e^A\right) G(r).$$

Solving for the optimal p_e^A gives

$$p_e^{A*} = \frac{2\bar{\theta} - (b-1)q - \mu_r}{3}.$$

Then, $p_e^{A*} < q$ when $\bar{\theta} < \bar{\theta}_u^r$.

The equilibrium personalized price and threshold are given by

$$\begin{aligned} p_i^*(r) &= \frac{(b-1)q + p_e^{A*} - r}{2} = \frac{\bar{\theta} + (b-1)q}{3} - \frac{\mu_r}{6} - \frac{r}{2}, \\ \tau^*(r) &= \frac{(b-1)q + p_e^{A*} + r}{2} = \frac{\bar{\theta} + (b-1)q}{3} - \frac{\mu_r}{6} + \frac{r}{2}. \end{aligned}$$

Finally, $0 < \tau^*(0) < \tau^*(1) < \bar{\theta}$ holds under Assumption D.

Scenario (2): Boundary optimum with $p_e^{A*} = q$.

This arises when $\bar{\theta} > \bar{\theta}_u^r$. In this scenario, firm i's equilibrium price and the equilibrium participation threshold are given by

$$\begin{aligned} p_i^*(r) &= \frac{bq - r}{2}, \\ \tau^*(r) &= \frac{bq + r}{2}. \end{aligned}$$

Then, $0 < \tau^*(0) < \tau^*(1) < \bar{\theta}$ is satisfied under Assumption D.

Firms' equilibrium profit under personalized pricing are given by

$$\begin{aligned} \Pi_i^* &= \frac{1}{\bar{\theta}} \int_0^1 (\tau^*(r))^2 \, dG(r), \\ \Pi_e^* &= \frac{p_e^A}{\bar{\theta}} \int_0^z \left(\bar{\theta} - \tau(r)\right) dG(r) = \frac{\left(p_e^{A*}\right)^2}{\bar{\theta}}. \end{aligned}$$

Consumers choosing firm e's product receive a net utility U_e^*. In contrast, consumers choosing the bundle receive $U_i^*(\theta, r) = U_e^* + \tau^*(r) - \theta$. The consumer surplus is:

$$S^* = U_e^* + \frac{1}{\bar{\theta}} \int_0^1 \int_0^{\tau^*(r)} (\tau^*(r) - \theta) \, d\theta dG(r) = U_e^* + \frac{1}{\bar{\theta}} \int_0^1 \frac{(\tau^*(r))^2}{2} dG(r).$$

Uniform pricing

Firm i offers uniform price p_i for its bundling, with the cut-off threshold $\hat{\tau} = bq - p_i - U_e$. Firm i's profit is given by

$$\hat{\Pi}_i = \frac{\hat{\tau}}{\bar{\theta}} \int_0^z (r + p_i) \, dG(r) = \frac{\hat{\tau}(p_i + \mu_r)}{\bar{\theta}}.$$

The optimal price is:

$$p_i = \frac{bq - U_e - \mu_r}{2} = \frac{(b-1)q + p_e^A - \mu_r}{2}.$$

The equilibrium threshold is:

$$\hat{\tau} = bq - p_i - U_e = \frac{bq - U_e + \mu_r}{2} = \frac{(b-1)q + p_e^A + \mu_r}{2}.$$

Thus, $\hat{\tau}$ is equal to the mean of $\tau(r)$:

$$\hat{\tau} = \int_0^1 \tau(r) \, dG(r).$$

Similarly, firm e chooses $p_e^A \leq q$ to maximize the following profit:

$$\hat{\Pi}_e = \frac{1}{\bar{\theta}} p_e^A \left(\bar{\theta} - \hat{\tau}\right).$$

Scenario (1): Interior optimum with $\hat{p}_e^{A*} < q$.

This occurs when $\bar{\theta} < \bar{\theta}_u^r$. The optimal p_e^A is determined by

$$p_e^A = \bar{\theta} - \hat{\tau}.$$

Solving for the equilibrium prices gives

$$\begin{aligned}
\hat{p}_e^{A*} &= \frac{2\bar{\theta} - (b-1)\,q - \mu_r}{3}, \\
\hat{p}_i^{*} &= \frac{\bar{\theta} + (b-1)\,q - 2\mu_r}{3}
\end{aligned}$$

The equilibrium threshold is

$$\hat{\tau}^* = \frac{\bar{\theta} + (b-1)\,q + \mu_r}{3}.$$

The equilibrium exists when $\hat{\tau}^* < \bar{\theta}$, which holds under Assumption D.

Scenario (2): Boundary optimum with $\hat{p}_e^{A*} = q$.

This happens when $\bar{\theta} > \bar{\theta}_u^r$. In this scenario, firm i's equilibrium price and the participation threshold are given by

$$\begin{aligned}
\hat{p}_i^{*} &= \frac{bq - \mu_r}{2} \\
\hat{\tau}^* &= \frac{bq + \mu_r}{2}.
\end{aligned}$$

The equilibrium arises when $\hat{\tau}^* < \bar{\theta}$, which is satisfied under Assumption D.

The equilibrium profits for both firms are given respectively by

$$\hat{\Pi}_i^* = \frac{(\hat{\tau}^*)^2}{\bar{\theta}},\ \hat{\Pi}_e^* = \frac{\left(\hat{p}_e^{A*}\right)^2}{\bar{\theta}} = \frac{\left(\bar{\theta} - \hat{\tau}^*\right)^2}{\bar{\theta}}.$$

Consumers choosing the bundle receive $\hat{U}_i^*(\theta, r) = \hat{U}_e^* + \hat{\tau}^* - \theta$, while those opting for the entrant's product receive $\hat{U}_e^*$. The consumer surplus is

$$\hat{S}^* = \hat{U}_e^* + \frac{(\hat{\tau}^*)^2}{2\bar{\theta}}.$$

Comparison

Comparing the prices in two schemes, we have $\hat{p}_e^{A*} = p_e^{A*}$. Therefore, firm e's equilibrium price and profit remain the same across both schemes. Consequently, consumers' utility from the stand-alone product, U_e, does not change. We use U_e as the parameter in the comparison.

Firm i's prices under two schemes are given by:

$$\begin{aligned} p_i^*(r) &= \frac{bq - U_e - r}{2}, \\ \hat{p}_i^* &= \frac{bq - U_e - \mu_r}{2}. \end{aligned}$$

Therefore, $\hat{p}_i^*$ is equal to the mean of $p_i^*(r)$:

$$\hat{p}_i^* = \int_0^1 p_i^*(r)\, dG(r).$$

The participation thresholds under two schemes are:

$$\begin{aligned} \tau^*(r) &= \frac{bq - U_e + r}{2}, \\ \hat{\tau}^* &= \frac{bq - U_e + \mu_r}{2}. \end{aligned}$$

Therefore

$$\hat{\tau}^* = \int_0^1 \tau^*(r)\, dG(r).$$

Comparing the two thresholds, consumers with $r > \mu_r$ are better off under personalized pricing, while those with $r < \mu_r$ are worse off: $\tau^*(r) > \hat{\tau}^*$ if and only if $r > \mu_r$.

Let $\rho \equiv \frac{bq-U_e}{2}$. Firm i's profits under two schemes are given by

$$\begin{aligned} \Pi_i^* &= \frac{1}{\bar{\theta}} \int_0^z (\tau^*(r))^2\, dG(r) = \frac{1}{\bar{\theta}} \left(\rho^2 + \frac{\sigma_r}{4} + \rho\mu_r \right), \\ \hat{\Pi}_i^* &= \frac{(\hat{\tau}^*)^2}{\bar{\theta}} = \frac{1}{\bar{\theta}} \left(\rho^2 + \frac{\mu_r^2}{4} + \rho\mu_r \right). \end{aligned}$$

Since $\sigma_r > \mu_r^2$, it follows that $\Pi_i^* > \hat{\Pi}_i^*$.

Consumer surplus under personalized pricing and uniform pricing is expressed as

$$\begin{aligned} S^* &= U_e + \frac{1}{\bar{\theta}} \int_0^1 \frac{(\tau^*(r))^2}{2} dr = U_e + \frac{1}{2\bar{\theta}} \left(\rho^2 + \frac{\sigma_r}{4} + \rho\mu_r \right), \\ \hat{S}^* &= U_e + \frac{1}{\bar{\theta}} \frac{(\hat{\tau}^*)^2}{2} = U_e + \frac{1}{2\bar{\theta}} \left(\rho^2 + \frac{\mu_r^2}{4} + \rho\mu_r \right). \end{aligned}$$

Therefore, consumer surplus is higher under personalized pricing: $S^* > \hat{S}^*$. This also implies that the total social welfare is higher under personalized pricing than under uniform pricing: $W^* > \hat{W}^*$.

D Variant and Extension

D.1 Cross-User Data Aggregation

Consider a variant of the baseline model in which the digital firm's revenue depends on the aggregate amount of data collected, denoted by S. Assuming the normalized quantity of data from each consumer, the platform's revenue function can be expressed as $R(F(\tau))$, where $S = F(\tau)$. The function $R(\cdot)$ is increasing and concave, with $R(0) = 0$. Moreover, cross-user data aggregation generates higher total revenue than in the baseline model, i.e., $R(F(\tau)) > rF(\tau)$.

Now, consider a monopoly digital firm adopting mixed bundling, offering $\{bq, p\}$ for the bundle while charging $p_A = q$ for stand-alone product. Consumers receive $U(\theta) = bq - p - \theta$ from the bundle while zero from the stand-alone product.

The participation threshold $\tau = bq - p$ remains the same as in the baseline model. Using $p = bq - \tau$ to represent the transfer of value between the platform and consumers, the platform's optimization can be reformulated as choosing τ and q to maximize:[28]

$$\Pi_e(\tau, q) = R(F(\tau)) + F(\tau)((b-1)q - \tau) + q - C(q).$$

Under cross-user data aggregation, the platform's data-monetization revenue becomes $R(F(\tau))$, compared to $rF(\tau)$ in the baseline setting. While the change in the revenue function does not directly affect the optimization of quality, the optimal quality q_e^r is determined by the same first-order condition (FOC) as in the baseline model:

$$C'(q) = 1 + (b-1)F(\tau). \tag{22}$$

Thus, data acquisition enhances the quality of the digital product if $b > 1$.

In contrast, increasing participation τ affects the marginal benefits from both data monetization and digital product provision. The optimal participation threshold $\tau_e^r(q)$ is given by

$$\psi(\tau) \equiv \tau + h(\tau) - R'(F(\tau)) = (b-1)q, \tag{23}$$

where $\psi'(\tau) = 1 + h'(\tau) - R''(F(\tau))f(\tau) > 0$ by assumption.

Comparing the best response $\tau_e^r(q)$ with $\tau_m^r(q)$, defined by $\tau + h(\tau) = (b-1)q + r$, the relative magnitude of $\tau_e^r(q) > \tau_m^r(q)$ depends on $R'(F(\tau))$. Since $R(F(\tau)) > rF(\tau)$ for $\tau \in (0, \bar{\theta})$ while $R(F(\tau)) = rF(\tau)$ at $\tau = 0$, there exists a threshold $\tilde{\tau} > 0$, defined by $R'(F(\tilde{\tau})) = r$, such that $R'(F(\tau)) > r$ for $\tau < \tilde{\tau}$. If $\tau_e^r(q) < \tilde{\tau}$, then $\tau_e^r(q) > \tau_m^r(q)$, indicating that cross-user data aggregation increases the quality of service when $b > 1$.

The equilibrium quality and participation threshold, denoted as q_e^* and τ_e^* respectively, are jointly determined by (22) and (23). Summarizing the analysis yields the following corollary:

Corollary 2 *Consider a variant of the baseline setting where the digital firm's revenue from data monetization depends on cross-user data aggregation. Suppose a monopoly digital firm*

[28] The subscript e denotes the scenario with data externalities under cross-user data aggregation.

engages in mixed bundling. The equilibrium quality q_e^* and participation threshold τ_e^* are jointly determined by (22) and (23). Data acquisition enhances product quality if and only if $b > 1$. Compared to the baseline setting, cross-user data aggregation increases both the participation threshold and product quality if $\tau_e^* < \tilde{\tau}$.

D.2 Verifiable data scale

Let $s \in [0, \bar{s}]$ denote the quantitative measure of the scale of a consumer dataset. Let $r(s)$ denote the revenue from data monetization and $b(s)\, q$ represent the data-driven consumer benefit. The privacy cost is expressed as $s\theta$, where the privacy sensitivity θ is distributed over $\left[0, \bar{\theta}\right]$. The following regularity assumption ensures the uniqueness of the equilibrium:

Assumption E: $r(s)$ is an increasing and concave function with $r(0) = 0$, while $b(s)$ is an increasing and concave function with $b(0) = 1$.

Consider a monopoly firm's compensation scheme under mixed bundling, offering the scheme $S = \{s, b(s)\, q, p\}$ for the bundle while charging $p_A = q$ for the stand-alone product. Consumers receive a net utility from the bundle $U(\theta) = b(s)\, q - p - s\theta$, and a net utility 0 from the stand-alone product. The participation threshold is now defined as:

$$\tau = \frac{b(s)\, q - p}{s}.$$

The monopoly's profit is expressed as:

$$\Pi = F(\tau)\,(r(s) + p) + (1 - F(\tau))\, q - C(q)\,.$$

A consumer obtains a gross value $b(s)\, q - p$ from the bundle. The firm can adjust this value through three policy variables: the quality of the digital product q, the scale of data, s, and the price (or subsidy) p. For any given data scale s, q and p are complements to consumers. However, they play different roles in the platform's compensation scheme due to different marginal costs. Since the marginal benefit and cost of the monetary payment is always equal to 1, the digital firm can use the price p as a transfer of value between the platform and consumers. Substituting $p = b(s)\, q - s\tau$, the monopoly can instead choose s, τ, and q to maximize its profit:

$$\Pi(s, \tau, q) = F(\tau)\,((b(s) - 1)\, q + r(s) - s\tau) + q - C(q)\,.$$

The monopoly is engaged in two digital activities: providing digital product and collecting consumer data. These two digital activities are interdependent, as the marginal consumer value from digital product $b(s)$ (weakly) increases with s. Using the price p as a transfer of value, the monopoly can capture the aggregate benefits $b(s)\,q+\ r(s)$ while compensate each participant with a value equal to the privacy cost of the marginal consumer with $\theta = \tau,\ s\tau$.

The marginal revenue from providing the digital product is $1 + (b(s) - 1)\,F(\tau)$, while the marginal cost is $C'(q)$. It is easy to verify that Π is concave in q. Then, the optimal product quality as a response function of s and τ, $q^r(s,\tau)$, is uniquely determined by the following first-order condition (FOC):

$$C'(q) = 1 + (b(s) - 1)\,F(\tau)\,. \tag{24}$$

Since $b_s(s) > 0$, consumer data and the product quality are complements to generate consumer value. Therefore, the optimal quality increases with the scale of consumer data: $q^r(s,\tau)$ increases in s.

By contrast, data collection and monetization yield a marginal revenue $b_s(s)\,q + r_s(s)$ at a marginal cost τ. Since the profit function Π is concave in s, the optimal data scale $s^r(\tau, q)$ is the unique solution to the following FOC:

$$b_s(s)\,q + r_s(s) = \tau. \tag{25}$$

The response function $s^r(\tau, q)$ decreases in τ, and increases in q as $b_s(s) > 0$.

The marginal benefit from increasing participation threshold τ is $f(\tau)\,((b(s) - 1)\,q + r(s) - s\tau)$ while the marginal cost is $sF(\tau)$, reflecting the increasing payment to all participants. The optimal participation threshold $\tau^r(s,q)$ is the unique solution to the following FOC:

$$\phi(\tau) = \tau + h(\tau) = \frac{(b(s) - 1)\,q + r(s)}{s}. \tag{26}$$

The equilibrium data scale, s^*, product quality q^*, and participation threshold τ^* are jointly determined by (24), (25), and (26). Using (26), the equilibrium price is given by

$$p^* = b(s^*)\,q^* - s^*\tau^* = s^* h(\tau^*) - r(s^*). \tag{27}$$

Then, the monopoly firm pays consumers for their data if the revenue from data monetization $r(s^*)$ exceeds the aggregate revenue from both activities, $s^* h(\tau^*)$.

From (26), in equilibrium, the firm's revenue from each participant, $b(s)\,q + r(s) - s\tau$, is equal to $sh(\tau)$. Thus, its equilibrium profit is given by

$$\Pi^* = F(\tau^*)\,s^* h(\tau^*) - C(q^*)\,. \tag{28}$$

Consumers with $\theta < \tau^*$ choosing the bundle and receive the net utility

$$U^*(\theta) = s^*(\tau^* - \theta)\,,$$

while those with $\tau^* < \theta < \bar{\theta}$ receive zero.

We summarize equilibrium outcomes in the following proposition:

Proposition 9 *Suppose a monopoly digital firm's commitment to a specific data scale is verifiable, and Assumption E is satisfied. There exists a unique equilibrium with mixed bundling, where consumers with* $\theta < \tau^*$ *choose the bundle, receiving* $U^*(\theta) = s^*(\tau^* - \theta)$*, while those with* $\theta > \tau^*$ *opt for the stand-alone product and obtain zero. The equilibrium data scale* s^**, quality* q^**, and participation threshold are jointly determined by* (24)*,* (25)*, and* (26)*, while the equilibrium price* p^* *is given by* (27)*. The monopoly firm's equilibrium profit is given by* (28)*.*