The Ownership of Data*

Anastasios Dosis † Wilfried Sand-Zantman‡

September 2020

Abstract

We study the effects of property rights over the use of data on market outcomes. To do so, we consider a model in which a monopolistic firm offers a service to a set of heterogeneous users. The use of the service generates valuable data, but data extraction entails a privacy cost for users. A trade-off emerges between under-processing and over-monetization of data. We show that both the firm and users prefer the users (the firm) to own the rights for low (high) values of data. We further discuss the robustness of our results when allowing more possible contracts for the data owner and show that the main trade-off is robust to these extensions.

KEYWORDS: Ownership; Data; Imperfect Competition; Privacy

JEL CLASSIFICATION: D82; D83; D86; L12; L19; L49

*We are grateful to Bernard Caillaud, Giacomo Calzolari, Gorkem Celik, Alexandre de Cornière, Bruno Jullien, Thomas Mariotti, Bertin Martens, Maryam Saeedi, and Marta Troya Martinez for many useful remarks. We also thank participants in the ESSEC Economics Internal Seminar; the French Association of Law and Economics Congress; the TSE Digital Economy Conference; the University of Montpellier; the 11th Conference on Digital Economics at Telecom ParisTech; the International Industrial Organization Conference (IIOC) 2019; the Bergen Center for Competition Law and Economics (BECCLE) Conference 2019; the 2019 JRC-TSE Conference on the Economics of AI and Data; the Midwest Economic Theory Conference 2019; the SAET 2019 Conference; and the European Association Research in Industrial Economics (EARIE) 2019 conference for their insightful comments and discussions. Wilfried Sand-Zantman acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). All remaining errors are our own.

†ESSEC Business School and THEMA, 3 Av. Bernard Hirsch, B.P. – 50105, Cergy, 95021, France; Email: dosis@essec.edu

‡Toulouse School of Economics and ESSEC Business School (THEMA), 3 Av. Bernard Hirsch, B.P. – 50105, Cergy, 95021, France; Email: wsandz@tse-fr.eu

1 INTRODUCTION

The rise of the internet has been accompanied by a heated debate on how data should be used and controlled. Currently, large firms specialize in collecting, processing and re-selling personal data to firms that operate in various markets. By default, companies control all generated data, and concerns over the optimality of this arrangement have arisen in public discourse over time.[1] Importantly, such concerns are relevant not only to consumers but also equally to firms that benefit from monetizing personal data.[2] This paper addresses the economic impact of the allocation of property rights to either side of the market.

The concerns related to consumer privacy and the importance of establishing clear property rights on data are exemplified by the recent open letter by Business Roundtable CEOs to congressional leaders.[3] In this letter, CEOs of leading companies, such as Amazon's Jeff Bezos, IBM's Ginni Rometty, and Best Buy's Corie Barry, urged policymakers to pass "*...a comprehensive federal consumer data privacy law to strengthen consumer trust and establish a stable policy environment in which new services and technologies can flourish within a well-understood legal and regulatory framework. Innovation thrives under clearly defined and consistently applied rules*". Elsewhere in the letter, the CEOs state that "*We are also united in our belief that consumers should have meaningful rights over their personal information and that*

[1]See, among others, "Should Consumers Be Able to Sell Their Own Personal Data?", Wall Street Journal, 13 Oct 2019; "We need to own our data as a human right and be compensated for it", the Economist, Open Future, Jan 21, 2019; "Draining Data Moats: What Happens When Consumers Take Control Of Their Own Data?", Forbes, 21 Aug 2019.

[2]As early as the 1990s, Laudon (1996) and Hagel III and Rayport (1997) laid the groundwork for the debate on the ownership of data. For instance, Laudon (1996) advocates that consumers should own their data and that market mechanisms should be established to prevent privacy invasion. Furthermore, Hagel III and Rayport (1997) argue that as the information collected by firms becomes more valuable, consumers will demand a share of the generated value, and firms should be ready to enter into costly and complex negotiations with them. More recently, Berinato (2014) proposes a "New Deal" for data, where he argues that allowing customers to own their data "...gives customers a stake in the new data economy; that will bring first greater stability and then eventually greater profitability as people become more comfortable sharing data".

[3]See, for instance, "Business Roundtable CEOs Call on Congress to Pass Comprehensive, Nationwide Consumer Data Privacy Law", Business Round Table, September 10, 2019

companies that access this information should be held consistently accountable under a comprehensive federal consumer data privacy law".

On the regulatory side, the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the US provide individuals with more control over their data. Nonetheless, the current state of the law does not directly address the question of ownership, and firms and regulators circumvent many of the obstacles by avoiding direct references to ownership; instead, they focus on the potential controls exercised by a data subject or the limitations of a data controller (see Ritter and Mayer, 2017).[4] In short, the legal and other difficulties in defining data ownership notwithstanding, investigating the impact of the different legal statuses of data on firms' profits and consumers' welfare is a relevant debate everywhere.

Model and Results Preview. To study this question, we develop a simple two-period model in which a monopolistic firm (a website in many cases) interacts with a set of heterogeneous consumers (users). The firm offers a service and sets subscription fees that consumers need to pay to access the provided service. The fees consumers pay depend on their usage, and a consumer's type determines the combination of fee and usage that the consumer prefers. Key to our analysis is that usage in the first period generates data that can be exploited in the second period. The value generated by the data depends on how important this data is but also on an effort level chosen by the firm. Although data generates value that can be monetized, the collection and use of data entail a privacy cost for users. We define property rights as the ability to control the amount of data collected

[4]The GDPR defines rights such as the need for informed consent, the right to access and extract data, the right to be forgotten, and some duties for data holders and controllers. In theory, the GDPR seems to considerably restrict the use of data; in practice, the GDPR does not strictly prevent firms from processing data. Indeed, Article 6 of the GDPR offers a particularly subjective assessment of the lawfulness of processing data (see Duch-Brown, Martens, and Mueller-Langer, 2017). Article 6 states, among other things, that the processing of data is lawful if it "... *is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.*".

and to monetize it. In this setting, we highlight a central trade-off between two forms of contractual incompleteness: the first linked to the firm's decision to process the data and the second related to the limit to data monetization.

As a benchmark, we first characterize the firm's profit-maximizing offer to consumers when data extraction, processing, and monetization are contractible. We show that the firm offers a usage schedule that is increasing in the consumer's type, defined here as the willingness to pay; under this schedule, all data for relatively low-type consumers, but only part of the data for relatively high-type consumers, is extracted and monetized. This result stems from the fact that for low-type consumers, the cost of privacy is low, and hence the firm can afford to monetize all data without sufficiently distorting usage; for high-type consumers, the cost of privacy is high, which restricts the amount of data monetization.

We then turn to the characterization of the firm's optimal offers when none of the parties can commit regarding its future behavior about data processing and monetization. This can occur, for instance, if data processing and monetization are not contractible. Indeed, it might be easy for third parties, such as a court of law, to verify whether personal information is used but impossible to verify the precise content of the data extracted and whether or how it is processed. Similarly, it might be difficult for third parties to verify the exact value of data as this is affected by many variables that are impossible to describe and include in a contract.

If the firm owns the rights, it has full control over the collection and use of the data of all consumers.[5] Users rationally anticipate that any information generated in the first period will be monetized in the second period. As a result, the firm's optimal offer entails lower usage for high-type consumers relative to the benchmark; for low-type consumers,

[5] As we argued above—and in light of recent regulations—in reality, there are limits to the amount of data firms can extract. Nonetheless, consumers' limited understanding of how data is collected and used blurs these limits. For instance, Chakravorti (2020) provides compelling arguments of what has come to be known as a "privacy paradox": users' strong privacy concerns are not reflected in their online behavior.

usage and data extraction are the same as in the benchmark. If consumers own the rights, they have full control over the extraction and monetization of their data. Nonetheless, in this case, the firm has no incentive to invest in data processing, and hence, the value consumers can reap in the market by monetizing their data is lower than when the firm owns the rights. The firm's profit-maximizing usage offer and the consumers' optimal data monetization policy resemble those when monetization is contractible. Nevertheless, relative to the case in which data monetization is contractible, the firm's offer often entails lower usage for low-type consumers. Regarding the consumers' data monetization policy, it remains everywhere lower than in the benchmark due to the consumers' limited ability to exploit the data's full potential in the data market.

We compare the welfare in the two ownership regimes and show that the ranking depends on the impact of contractual incompleteness leading to under-processing or over-utilization of data. We demonstrate that, for relatively low (high) values of data, both the firm and users prefer the latter (former) to own the rights. Intuitively, for relatively low values of data, the cost of under-processing is offset by the cost of privacy, and consumer's rights dominate. By contrast, for relatively high values of data, the firm is willing to compensate the users for opting for high usage; the resulting decrease in subscription fees, alongside the high value of data, offsets the cost of privacy.

We extend the baseline model in two directions. First, we study the possibility that, when the firm owns the rights, it offers to let consumers pay for their privacy. This is equivalent to allowing the firm to sell two types of good as is commonly observed in practice: (i) a good that provides access only to the service while leaving all rights from data generation to the firm; and (ii) a good that provides access to the service and prevents further usage of the data generated in the first place. Indeed, we show that the firm's profit-maximizing offer entails low usage with full data extraction for low-type consumers and high usage with no data extraction for high-type consumers.

We demonstrate that the option for privacy increases both the firm's profit and consumers' welfare relative to the regime in which no such option is available. We then compare this regime with the regime in which the consumers own the rights. We show that when the consumers are unable to earn any money from monetizing their data, the option for privacy increases both the firm's profit and consumers' welfare relative to the regime in which consumers own the rights. When consumers can monetize their data (albeit imperfectly), the option for privacy leads to a higher profit and consumer surplus as long as the value of data is sufficiently high.

Second, we consider the case in which consumers can only decide whether to sell their entire dataset at a uniform price instead of fine-tuning the amount and type of data they sell. This is conceptually equivalent to a situation according to which there is a market for privacy and consumers can sell to third parties the right to use their data. In this case, we show that the firm's profit is lower relative to the case in which consumers can select how much data to sell. Nonetheless, we also show that consumers earn exactly the surplus they could earn if they could fine tune the amount and type of data they sell.

Literature Review. Our work builds on the literature that studies the effect of property rights on economic allocations. Following Coase (1960) (or, more formally, Dasgupta, Hammond, and Maskin, 1979), it is well known that when the firm's contracting possibility set is large enough, property rights are irrelevant for economic allocations. This is indeed the case in our model when data extraction is contractible (Section 3). This leads us to consider some contractual incompleteness — in particular, the inability to commit ex ante to the intensity of data processing and monetization — to make a meaningful comparison between different property rights regimes. In our model, as in Grossman and Hart (1986), property rights are defined as residual control rights.

Our contribution is also related, albeit imperfectly, to the literature on privacy. Hermalin and Katz (2006) links privacy to the right to (i) compel information disclosure or (ii) conceal information. They show that, with complete contracts, the outcome is the same in both cases. The authors also show that more or less privacy may benefit or harm consumers. Unlike our paper, in which the information revealed in one market is potentially used in other markets, in Hermalin and Katz (2006), there is a unique market. In this sense, our paper is related to the literature on dynamic pricing on the internet (Acquisti, 2006) or the impact of information on market outcomes (Acquisti, Taylor, and Wagman, 2016).

Jones and Tonetti (2018) is the article most closely related to the research question pursued here. They develop a dynamic general equilibrium growth model in which the data is a byproduct of consumption. Each unit of data can be used not only by the firm that generated it but also by other firms through a market in which the firm can sell its data. When firms control the data, they tend to limit data sharing for fear of increased competition; when consumers own their data, they restrict data sharing due to the privacy cost. The authors emphasize the non-rivalrous aspect of data and the gains for growth when consumers control their data. Our approach focuses on the role of ownership over the amount of data created through its impact on the contract between firms and consumers. In other words, we are more interested in the consequences of the allocation of ownership on the firm's and consumer's strategies than in the general ex post consequences for economic efficiency.

Finally, our setting relates to the literature on standard two-sided markets (Rochet and Tirole, 2003; Armstrong, 2006). Indeed, the firm in our model is reminiscent of a platform, although we mostly analyze the interaction of the platform with one side of the market (the firm with consumers). Nevertheless, one can see how the pricing strategy on one side (e.g., the number of users, price of the service) changes when the value of data on the other side increases.

Organization of the Paper. In Section 2, we present the model. In Section 3, we study the firm's profit-maximizing usage and data extraction offer when all dimensions of data use are contractible. In Section 4, we assume that data extraction, processing, and monetization are not contractible and compare welfare when the firm owns the rights with that when consumers own the rights and derive our central trade-off. In Section 5, we discuss various extensions of our model. Section 6 concludes the paper. All proofs are relegated to an appendix at the end of the paper.

2 The Model

A. Agents

We consider a bilateral relationship between a firm and a unit mass of consumers. The firm provides a service, and each consumer decides her consumption level $q \geq 0$. One can regard q as the intensity of usage or the time a consumer spends on the website; hence, for future convenience, we refer to q as *usage* and assume throughout that this variable is contractible. Crucial in what follows is that consumers differ in their valuations of usage. In particular, the gross utility of a consumer from usage q is given by $u(\theta, q) = \theta q - q^2/2$, where θ denotes the type of consumer. We assume that θ is uniformly distributed in $[0, 1]$.[6] We further assume that θ is a consumer's private information and that the outside option of every consumer is zero.[7]

B. Data Extraction: Benefits and Costs

Benefits.—A specific feature of our model is that usage in one period generates valuable data that can be exploited in subsequent periods. In dynamic frameworks, the way agents interact in one period affects their behavior in subsequent periods (Fudenberg and Tirole,

[6]Note that all the results of Sections 3 and 4 extend to any general distribution function. It is necessary to assume a uniform distribution to derive results in the extensions. More details are provided in the appendix.

[7]Because the firm is a monopolist, we consider a context in which there is uncertainty regarding some of the characteristics of the consumers. This guarantees that part of the gains from trade are captured by consumers. An alternative but somewhat equivalent approach would be to have a unique offer for all potential consumers.

2000). Such potential dynamics have been at the center of recent discussions on the economics of the internet (Acquisti and Varian, 2005; Fudenberg and Villas-Boas, 2012). In this paper, we assume that the interaction between the firm and consumers generates potential informational value.

In particular, we assume that each unit of usage generates one unit of data. Under this assumption, if the amount of data extracted is denoted by e, there is a natural technological constraint $e \leq q$. The monetary value of a unit of data is denoted $\alpha > 0$. This monetary value of data could depend on the intensity of competition in this market—that is, the extent to which the information generated by the service substitutes or complements information owned by other firms—or the breadth of the market for data—that is, the number of other firms interested in acquiring the data. Importantly, we assume that the full monetary value of data cannot be generated for free. Instead, only a share $\delta \in [0, 1]$ of the data can be valuable without any further processing by the firm. This condition could, for instance, correspond to the market value of raw data. The remaining share $(1 - \delta)$ can be properly monetized only if the firm processes the data (i.e., exerts some effort); however, such processing entails a per unit cost equal to $k \geq 0$. The idea of costly data processing is also consistent with the recent data regulations mentioned in the introduction. For instance, one of the fundamental rights defined by the GDPR is that of *data portability*, according to which a user can "port" any generated data from one firm to another. Nonetheless, this only applies raw data, that is, data that is not processed by the firm that generated it. In sum, the value of data $B_v(\alpha, e)$, where $v \in \{\delta, 1\}$, is linear in the amount of data monetized, with $B_\delta(\alpha, e) = \delta\alpha e$ in the first case and $B_1(\alpha, e) = (\alpha - k)e$ in the second case.

Costs.—Another key element of our model is that, although data extraction creates value, it is costly for consumers. There are at least two relevant categories of costs related to data extraction. The first category includes every cost related to the loss of privacy that accompanies the use of personal data. These psychological costs are a consequence of a pure preference for privacy and are therefore independent of economic outcomes (Westin, 1967). The second category is related to the economic impact that data extraction has on the secondary market, particularly the possibility of finest price discrimination using personal data. We summarize these two costs in a single function and assume that when e is extracted and monetized, a consumer suffers disutility equal to $C(e) = \gamma e^2/2$, where $\gamma > 0$.

Discussion.—Our specifications of the benefit and cost of data extraction imply that the net value of data is strictly concave in e. We must admit that it remains an open question whether the net value of data exhibits increasing or decreasing returns to scale. Regarding the value, if one considers classical Bayesian updating, changes in decisions are more likely for the first than the last signals received. If we interpret e as pieces of information, it seems natural to assume that the benefits are concave (Agrawal, Gans, and Goldfarb, 2018). Although the linearity assumption simplifies the analysis, most of our results extend to the case of strictly concave benefits. It is unclear whether privacy costs are convex or concave in the information revealed. Regarding price discrimination, one might expect that the marginal impact of information is decreasing and, thus, that concavity prevails. By contrast, the psychological cost can be either concave, if an agent suffers more during the first breaches of privacy, or convex, if an agent suffers only when very precise and intimate information is discovered. Nevertheless, most papers (as in Jones and Tonetti, 2018) assume that this cost is convex, and we follow their lead by imposing the same assumption.[8]

[8]Note that when the net value is convex, the data extraction policy is different, but the main trade-off between the firm owning the rights and consumer owning the rights remains unaltered. What will be crucial is the fact that there can be over-monetization of data in the former case and under-monetization in the latter case.

3 Contractible Benchmark

As a benchmark, in this section, we consider the firm's profit-maximizing offer by assuming that data extraction, processing and monetization are all contractible. In this case, as the firm can commit to every dimension of its contract, any issue related to data ownership is irrelevant for the firm's maximum profit and consumer surplus.

Consider first the firm's decision to invest in data processing. Even if this decision is taken ex post (i.e., after the data is generated), this can be specified in the contract signed ex ante. This implies that the firm needs to simply choose between a net value $\delta\alpha$ per unit of data, when the data is not processed, or a net value $\alpha - k$ when the data is processed. Because of the linearity assumption of the benefit and processing cost, this decision does not depend on the quantity of monetized data induced by the contract. Therefore, the following lemma directly follows.

Lemma 1. *Suppose that data processing and monetization are contractible. Then, the firm processes any data if and only if $\alpha \geq \bar{\alpha}$, where $\bar{\alpha} = \frac{k}{1-\delta}$.*

We therefore need to consider two regimes depending on whether the firm prefers to process the data or not.

We now turn our attention to the firm's ex ante optimal contract. We focus without loss of generality on direct revelation mechanisms, in which consumers truthfully report their type and a pre-defined price-usage-data extraction triplet is implemented as a function of the report (see Laffont and Martimort, 2002 for details). Let a mechanism be denoted by $(t(\cdot), q(\cdot), e(\cdot))$. The indirect utility of type θ from truthfully announcing her type is given by

$$U(\theta) = \theta q(\theta) - q^2(\theta)/2 - t(\theta) - C(q(\theta)) \tag{1}$$

Using standard techniques in information economics, we can show that the mechanism is incentive compatible only if usage is non-decreasing in θ and $\dot{U}(\theta) = q(\theta)$. By integrating the latter condition, we obtain

$$U(\theta) = U(0) + \int_0^\theta q(\tau) d\tau \tag{2}$$

This expression represents type θ's information rent, which is the share of the surplus that must be left to type θ to truthfully reveal her type. The firm therefore has incentives to minimize this amount. As the payoff is increasing in θ, the firm will bind the participation constraint of the lowest type—that is, $U(0) = 0$.

The profit of the firm consists of the sum of the transfers it collects from consumers and the benefit from data monetization.[9] Therefore, the firm's profit is equal to

$$\int_0^1 \Big(t(\theta) + B_v(\alpha, e(\theta)) \Big) d\theta$$

with $v \in \{\delta, 1\}$. Using (1), this can be rewritten as

$$\int_0^1 \Big[\theta q(\theta) - q^2(\theta)/2 + B_v(\alpha, e(\theta)) - C(e(\theta)) - U(\theta) \Big] d\theta$$

If we substitute (2) and employ the fact that $U(0) = 0$, the profit is given by

$$\int_0^1 \Big[\theta q(\theta) - q^2(\theta)/2 + B_v(\alpha, e(\theta)) - C(e(\theta)) - \int_0^\theta q(\tau) d\tau \Big] d\theta$$

which, provided that $\int_0^1 \Big(\int_0^\theta q(\tau) d\tau \Big) d\theta = (1-\theta) q(\theta)$, can be rewritten as

$$\int_0^1 \Big[(2\theta - 1) q(\theta) - q^2(\theta)/2 + B_v(\alpha, e(\theta)) - C(e(\theta)) \Big] d\theta.$$

It is convenient to let

$$\phi_v(q) = \max_{e \leq q} B_v(\alpha, e) - C(e) \tag{3}$$

represent the maximum net value that is generated in the secondary market when usage in the first period is equal to q. For notational simplicity, we further let $\bar{e}_v$ denote the

[9] Although irrelevant for the results of this section, we assume here that the firm owns the data. If consumers owned the data, the firm could appropriate the generated surplus by offering higher fees.

amount of data that maximizes $B_v(\alpha, \cdot) - C(\cdot)$ when $e < q$. Function $\phi_v(q)$ is strictly increasing in q for $q < \bar{e}_v$ and constant for $q \geq \bar{e}_v$. By using (3), we can write the profit as

$$\int_0^1 \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2 + \phi_v(q(\theta))\Big] d\theta \tag{4}$$

The firm's objective then is to select a usage schedule $q(\cdot)$ to maximize (4) subject to monotonicity (usage is non-decreasing) and feasibility (usage is positive for every θ) constraints. Therefore, provided that type θ purchases (i.e., $q(\theta) \geq 0$), the optimal usage schedule is implicitly defined by

$$q(\theta) = (2\theta - 1) + \phi_v'(q(\theta)). \tag{5}$$

The impact of α on the firm's ex ante optimal contract is dual. First, when α exceeds $\bar{\alpha}$, the firm chooses to process the data (see Lemma 1) in order to generate the full value of data on the secondary market. This means that v increases from δ to 1. Second, an increase in α increases the firm's incentive to monetize a higher share of data. For the low-usage consumers, the marginal benefit of monetization always exceeds the marginal privacy cost, and all the data is monetized. For the high-usage consumers, the marginal benefit from the data monetization does not always exceed the marginal privacy cost, and only part of the data is monetized.

We summarize the firm's profit-maximizing offer in the following proposition.

Proposition 1. *Suppose that data processing and monetization are contractible. Then, there exists $\hat{\theta}_v \in [0, 1]$, where $v = \delta$ if $\alpha < \bar{\alpha}$ and $v = 1$ if $\alpha \geq \bar{\alpha}$, such that type-θ's usage in the firm's optimal mechanism is given by*

$$q_v^*(\theta) = \begin{cases} \max\{q_v(\theta), 0\}, & \text{if } \theta < \hat{\theta}_v \\ q_0(\theta), & \text{if } \theta \geq \hat{\theta}_v \end{cases},$$

where $q_1(\theta) = (2\theta - 1 + \alpha - k)/(1 + \gamma)$, $q_\delta(\theta) = (2\theta - 1 + \delta\alpha)/(1 + \gamma)$ and $q_0(\theta) = 2\theta - 1$. Moreover, type-$\theta$'s data extraction is given by $\max\{q_v(\theta), 0\}$ if $\theta < \hat{\theta}_v$ and $\bar{e}_v$ if $\theta \geq \hat{\theta}_v$.

For any unit value of data, the firm decides on its optimal mechanism by trading off this value with the privacy cost consumers incur. For high-type consumers, higher usage is preferred to data extraction, and the firm chooses to restrict the share of data monetized. For low-type consumers, the opposite holds because the privacy cost associated with the marginal unit of data monetized is small.

Another interesting element in this benchmark case concerns the cut-off type, who is indifferent between buying the service or not, denoted by $\underline{\theta}_v$. Provided that $q_v(\cdot)$ is increasing in α for every θ, for any data processing decision, $\underline{\theta}_v$ is decreasing in α. Intuitively, when α increases, the firm (i) trades with more consumers and (ii) offers more usage to all of the consumers with whom it trades. The first result is reminiscent of two-sided markets (Rochet and Tirole, 2003; Caillaud and Jullien, 2003; Armstrong, 2006). Indeed, in a standard two-sided market, an increase in the potential value on one side of the market is accompanied by an increase in the size of the other side. In our case, the higher the value in the secondary market, the stronger the firm's incentive to accommodate a larger share of consumers in the primary market.

A last interesting remark is that whenever α is sufficiently high—i.e., larger than either $1/\delta$ for $v = \delta$ or $1 - k$ for $v = 1$—all consumers purchase the product (i.e., $\underline{\theta}_v = 0$). In this case, the optimal mechanism entails subsidization of consumers to join the firm, consume, and generate data. Intuitively, when the data is very valuable, the firm has an incentive to attract as many consumers as possible. Because low types have low (or at the limit no) willingness to pay for the service, the only way to attract customers is by subsidizing them.[10]

4 Firm's Right vs. Consumer's Right

In this section, we assume that data extraction, processing and monetization are not contractible and investigate the economic consequences of allocating the rights to one of the two sides of the market. We say that we are in a firm's right regime when the firm owns the exclusive rights over any generated data. As we explained in the introduction, this

[10] This result is related to Amelio and Jullien (2012), who identify problems and propose solutions regarding the subsidization of consumers in two-sided markets.

is the regime that *de facto* prevails in most developed countries. Indeed, although recent regulations (e.g., GDPR, CCPA) restrict data extraction and monetization, it is still natural to consider firm's right as the status quo. Nonetheless, we assume that the firm is unable to commit to any action regarding the data generated by consumers. In what follows, we define as consumer's right regime the ability given to consumers to directly monetize any generated data. As is the case in the firm's right regime, we assume that consumers are unable to commit to a data extraction policy but can manage their data directly in the data market.

A. Firm's Right

As in the contractible data-extraction benchmark, the firm can offer different contracts that condition the price paid on usage.[11] Unlike the contractible data-extraction benchmark, the firm cannot commit to a data extraction policy. Given that data extraction is costless for the firm, the firm extracts all the generated data from all of its customers. Consequently, in what follows, we take $e = q$. Rationally expecting this, type θ knows that for usage $q \geq 0$, her gross utility is given by $\theta q - q^2/2 - C(q)$. Regarding the data processing, that is, the decision to invest in order to increase the monetary value of data, the same decision rule as in the benchmark prevails. Therefore, the two benefit functions for the firm are $B_\delta(\alpha, q) = \delta\alpha q$ for $\alpha < \bar{\alpha}$ and $B_1(\alpha, q) = (\alpha - k)q$ for $\alpha \geq \bar{\alpha}$.

We now turn to the analysis of the contract $(t(\cdot), q(\cdot))$ offered to consumers. The indirect utility of type θ from truthfully announcing her type is as in (1). The expected profit of the firm is given by

$$\int_0^1 \left[t(\theta) + B_v(\alpha, q(\theta))\right] d\theta,$$

[11] If the firm can only offer a fixed contract, the main features that appear below are preserved, albeit in different form.

which, by analogy to the benchmark, is equal to

$$\int_0^1 \Big[(2\theta-1)q(\theta) - q^2(\theta)/2 + B_v(\alpha, q(\theta)) - C(q(\theta))\Big] d\theta \tag{6}$$

Therefore, the objective of the firm is to select a usage schedule $q(\cdot)$ to maximize (6) subject to monotonicity (usage is non-decreasing) and feasibility (usage is positive for every θ) constraints.

Lemma 2. *Suppose that data processing and monetization are not contractible and the firm owns the rights. Then, type-θ's usage in the firm's optimal mechanism is given by* $\max\{0, q_v(\theta)\}$*, where* $v = \delta$ *if* $\alpha < \bar{\alpha}$ *and* $v = 1$ *if* $\alpha \geq \bar{\alpha}$*, with* $q_v(\theta)$ *defined in Proposition 1.*[12] *Moreover, the firm extracts and monetizes all the data from all the consumers who purchase.*

The firm's maximum profit in this regime is then given by

$$\Pi^F = \int_{\underline{\theta}_v}^1 \Big[(2\theta-1)q_v(\theta) - q_v^2(\theta)/2 + B_v(\alpha, q_v(\theta)) - C(q_v(\theta))\Big] d\theta$$

where $\underline{\theta}_v$ is the cut-off type, who is indifferent between purchasing the product or not as was defined in the contractible data-extraction benchmark. Moreover, provided that the consumer surplus consists of the sum of the informational rents of consumers, we have

$$CS^F = \int_{\underline{\theta}_v}^1 q_v(\theta)(1-\theta) d\theta$$

Inefficiencies.—When the firm owns the rights, the main potential inefficiency relative to the contractible data-extraction benchmark is the excessive data monetization for the high-type consumers. Such excessive monetization of data results in usage that is strictly lower when data extraction is not contractible and the firm owns the rights relative to the contractible data-extraction benchmark. This is due to the fact that high-type consumers anticipate ex ante that all their data will be monetized and are hence reluctant to opt for

[12] Note that this usage schedule is increasing in θ; therefore, this schedule satisfies the (second-order) incentive compatibility condition.

high usage. Therefore, the firm offers a mechanism that entails lower usage for high-type consumers relative to the contractible data-extraction benchmark. Such distortion is exacerbated the more important privacy is for consumers (the higher the γ). Regarding the low-type consumers, no such inefficiency arises given that in any case, all their data is monetized.

As seen in Proposition 1, the highest type such that full data extraction and monetization are optimal depends on the value of data α. In particular, if α is large enough, it is optimal to monetize all the data of all the consumers (regardless of the decision to invest in data processing). Thus, firm's right does not entail any inefficiency in this case. For future reference, it is useful to be more precise on this point. Suppose first that data processing does not bring much additional value (δ large) or, equivalently, that the processing cost is high (k large). This means that the firm invests in data processing only when the value of the data is large. Since, in this case, it is optimal to monetize all the data for all consumers, there will be no inefficiency as soon as data processing occurs. Suppose instead that data processing generates considerable value (δ small) or, equivalently, that the processing cost is low (k small). This means that the firm may be willing to invest in data processing even when the value of the data is low. Thus, there can be over-monetization both with and without data processing. This discussion, which is derived from Proposition 1 and Lemma 2, is formalized in the following corollary.

Corollary 1. *When the firm owns the rights, data monetization is excessive if and only if $\alpha \leq \hat{\alpha}$, where (i) $\hat{\alpha} > \bar{\alpha}$ if $\delta < \frac{\gamma}{\gamma+k}$ and (ii) $\hat{\alpha} \leq \bar{\alpha}$ if $\delta \geq \frac{\gamma}{\gamma+k}$.*

B. Consumer's Right

Presentation.—We now consider the case in which the consumers directly manage their data in the data market. Until recently, the idea of consumers directly accessing the data market seemed unrealistic. The unlikelihood of this scenario was mainly due to the fact that most consumers lack the knowledge and expertise to directly negotiate with other parties that have interest in their data (e.g., data brokers, firms that sell complementary or competing products). Nonetheless, two recent developments provide compelling arguments that this is not currently true.[13] First, as mentioned above, recent regulations, such as the GDPR, give consumers the right of "data portability", which allows them to transfer the data generated as a byproduct with their interaction with one firm to any other firm. Second, recently established platforms such as `www.people.io`, `datacoup.com` or `wibson.org` offer tools that help users obtain part of the financial returns generated by their data in exchange for a fee. It is also reasonable to assume that if consumers were indeed allocated the rights over their data, this industry would further flourish because data brokers would have an incentive to directly negotiate with consumers, who would be much better equipped for such negotiations.

In this section, we assume that consumers own the rights and can obtain a monetary value of their data. Nonetheless, provided that the data is a byproduct of a consumer's interaction with the firm, consumers still depend on the firm that generates the data. Moreover, as we argued above, consumers can have access to and control the raw data, that is, the share δ of the potential value. The remaining share $1-\delta$ crucially depends on the firm's decision to process this data. Nevertheless, the impossibility to write explicit contracts on the data processing destroys the firm's ex post incentive to process the data, as shown in the following lemma.

Lemma 3. *Suppose consumers own the rights and that neither processing nor monetization are contractible. Then, there is no ex ante contract that induces the firm to invest in data processing.*

As discussed in the contractible data-extraction benchmark, $\phi_\delta(q)$ represents the maximum net value that the consumers can reap in the market for data when the data is not processed. In order to maximize this value, consumers monetize all their data (i.e., $e=q$) if $q \leq \bar{e}_\delta$ but only part of their data (i.e., $\bar{e}_\delta < q$) if $q > \bar{e}_\delta$.

[13]See, among others, "As of Today, European Consumers Can Profit From Selling Their Own Personal Data", *Forbes*, 11 October 2018.

The Firm's Optimal Offer.—If the firm offers a contract $(t(\cdot), q(\cdot))$, a consumer of type θ selects an optimal data extraction e (lower than q) to maximize $\phi_\delta(\cdot)$. The *ex ante* utility of type θ, conditional on this anticipated decision, is

$$U(\theta) = \theta q(\theta) - q^2(\theta)/2 + \phi_\delta(q) - t(\theta) \tag{7}$$

As in the firm's right regime, to satisfy the incentive constraint, the contract proposed by the firm must be such that usage is non-decreasing in θ and the informational rent profile given by $\dot{U}(\theta) = q(\theta)$.

Taking into consideration the incentive and participation constraints, the expected profit of the firm is the sum of the payment of all its customers, that is, $\int_0^1 t(\theta)d\theta$. Using (7), it can be rewritten as

$$\int_0^1 \Big[\theta q(\theta) - q^2(\theta)/2 + \phi_\delta(q(\theta)) - U(\theta)\Big] d\theta$$

which, by analogy to the contractible benchmark, is equal to

$$= \int_0^1 \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2 + \phi_\delta(q(\theta))\Big] d\theta. \tag{8}$$

This expression reflects the fact that the firm can incorporate in the price set in the primary market any additional value consumers can earn in the secondary market.

Lemma 4. *Suppose consumers own the rights and that neither processing nor monetization are contractible. Then, type-θ's usage in the firm's optimal mechanism is given by*

$$q_*(\theta) = \begin{cases} \max\{0, q_\delta(\theta)\}, & \text{if } \theta < \hat{\theta}_\delta, \\ q_0(\theta), & \text{if } \theta \geq \hat{\theta}_\delta \end{cases}$$

where $q_\delta(\theta)$, $q_0(\theta)$ and $\hat{\theta}_\delta$ are defined in Proposition 1. Moreover, consumers in $[0, \hat{\theta}_\delta]$ monetize all their data, whereas consumers in $(\hat{\theta}_\delta, 1]$ monetize only an amount of data equal to $\bar{e}_\delta$.

The advantage of having consumers directly manage their data in the data market is the flexibility to monetize only part instead of all of their data. Recall that in the regime in which the firm owns the rights, its inability to commit to a data extraction policy precludes such partial monetization. Instead, this flexibility creates an added value both for the firm and the consumers. Note that the cut-off type $\hat{\theta}_\delta$ is chosen by the firm in its design of the offer $(t(\cdot), q(\cdot))$. Since the value generated on the market for data for this consumer is the same regardless of the usage schedule chosen, the first-period profit for this consumer is also the same under both usage schedules. We can then write the firm's maximum profit as

$$\Pi^C = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \Big[(2\theta - 1)q_\delta(\theta) - q_\delta^2(\theta)/2 + B_\delta(\alpha, q_\delta(\theta)) - C(q_\delta(\theta))\Big] d\theta$$
$$+ \int_{\hat{\theta}_\delta}^{1} \Big[(2\theta - 1))q_0(\theta) - q_0^2(\theta)/2 + \phi_\delta(\bar{e}_\delta)\Big] d\theta$$

and the consumer surplus as

$$CS^C = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} q_\delta(\theta)(1-\theta)d\theta + \int_{\hat{\theta}_\delta}^{1} q_0(\theta)(1-\theta)d\theta$$

Inefficiencies.—When the consumers own the rights, the main potential inefficiency relative to the contractible data-extraction benchmark stems from the underlying impossibility that the parties can directly contract with the firm processing the data. In some cases, this prevents the parties from exploiting the full potential value of the data. Because of this reduced value of data, consumers decrease the amount of data they monetize, and the firm optimally decreases the low-type consumers' usage. Notably, any inefficiency only prevails in so far as data processing is optimal. Therefore, we can state the following corollary.

Corollary 2. *When consumers own the rights, the usage of low-type consumers is inefficient if and only if* $\alpha > \bar{\alpha}$.

C. Optimal Allocation of Property Rights

In this subsection, we study the optimal allocation of property rights over the extraction and monetization of data; that is, we compare the profit and the consumer surplus under the two regimes. We explained above that allocating the rights to the firm may result in excessive data monetization and low usage for high-type consumers, whereas allocating the rights to consumers discourages the firm from processing the data and, hence, may result in low usage for low-type consumers. Therefore, the optimal allocation of property rights is obtained by exploiting the trade-off between excessive data monetization (and low usage for high-type consumers) when the firm owns the rights and no data processing (and low usage for low-type consumers) when the consumers own the rights.

First, we know that for $\alpha < \bar{\alpha}$, the firm never processes the data, and hence, the only difference between the two regimes stems from the excessive data monetization when the firm owns the rights. Therefore, it is optimal for the consumers to own the rights for low values of data. Conversely, when the value of data is large enough, that is, for $\alpha \geq \max\{\bar{\alpha}, k + \gamma\}$, the firm always processes the data when it has the rights—as is the case in the contractible data extraction benchmark—but never when the consumers own the rights; hence, the only difference between the two regimes stems from the inability to force the firm to process the data when consumers own the rights. Therefore, in this case, the firm's right is optimal.

The most interesting case arises when $\bar{\alpha} < \alpha < k + \gamma$. This can only occur when $\delta < \frac{\gamma}{k+\gamma}$, that is, when data processing generates enough additional value. We now focus on this case.

Consider first the firm's profit. We show in the appendix (see proof of Proposition 2) that (i) the firm's profit is strictly increasing in α under any regime, and (ii) the profit under the firm's right regime increases faster in α than the profit under the consumer's right regime. Indeed, under either regime, an increase in α has a dual positive impact on the firm's profit. First, for a given offer, this condition allows the firm to generate more value, either directly—when it owns the right—or indirectly by increasing the price it charges the consumers—when the consumers own the rights. Second, it allows the firm to offer higher usage to consumers. Nonetheless, in the consumer's right regime, the positive impact of an increase in α is mitigated by (i) the fact that there is less data monetization (i.e., consumers with types greater than $\hat{\theta}_\delta$ do not monetize all their data) and (ii) the lower value generated for any unit of data monetized. This implies that an increase in the value of data is more profitable for the firm when the firm owns the rights than when the consumers own the rights.

As the ability of consumers to monetize their data alone decreases (that is, as δ decreases), the return the firm obtains indirectly in the data market when consumers own the rights also decreases. In the extreme case in which $\delta = 0$, the consumers are unable to generate alone any monetary returns in the data market when they own the rights. This condition, however, does not necessarily mean that the firm's profit is always higher when the firm owns the rights than when the consumers own the rights. Indeed, even if the firm can gain more by monetizing the data, it suffers from its inability to commit to limit data exploitation. Therefore, the optimal ownership regime for the firm's profit depends on the value of the data. When this value is sufficiently high (i.e., exceeds some threshold), the firm's profit is higher when it owns the rights, whereas when this value is low (i.e., is below the threshold mentioned above), the firm's profit is higher when the consumers own the rights.

A related result prevails for the consumer surplus. Note first that since the entire value generated in the market for data accrues to the firm through the fee charged to access the service (i.e., $t(\cdot)$), the consumer surplus solely depends on the level of usage. Indeed, it is the level of usage that determines the informational rents and, hence, consumer surplus. In particular, the consumer surplus is increasing is usage; hence, the comparison of the consumer surplus under the two regimes is only based on the comparison of the usage schedules under the two regimes.

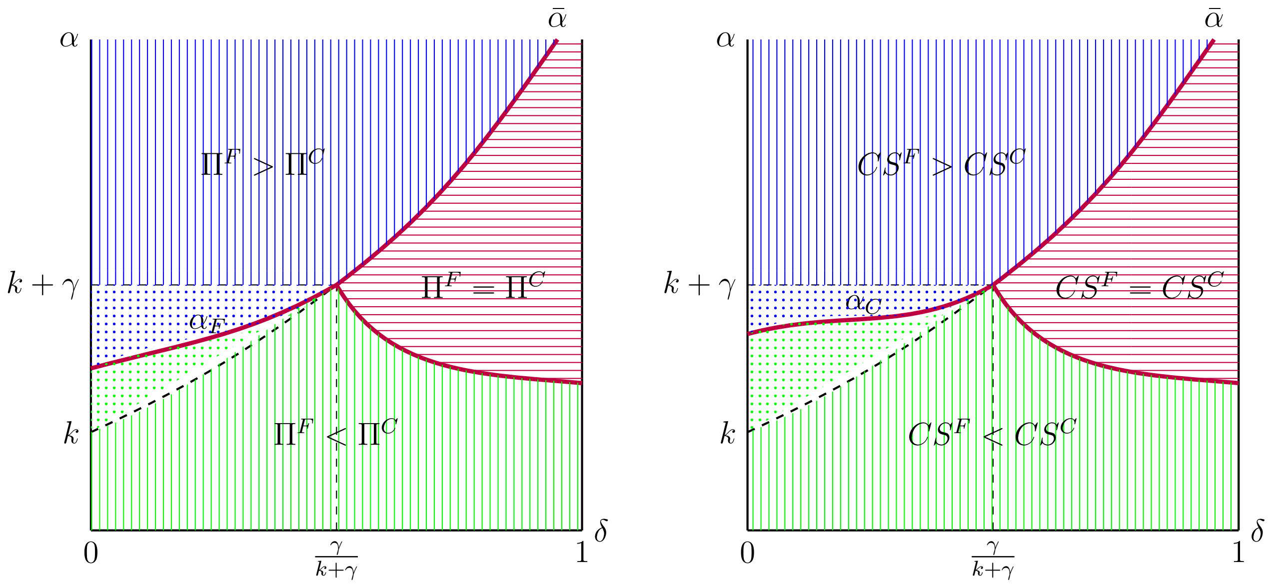
Similarly to the profit, we show in the appendix (see proof of Proposition 2) that (i) the consumer surplus is strictly increasing in α under any regime and (ii) the consumer surplus under the firm's right regime increases faster in α than the consumer surplus under the consumer's right regime. The intuition is in line with that described above for the profit. When the value of data increases, the firm is more likely to foster usage, which, as we argued above, is beneficial to consumers. However, this incentive crucially depends on the share of value of data that the firm can capture in the market for data. When the consumers own the rights, the value that the firm can capture in this market is lower; therefore, the firm has less incentive to foster usage relative to when it can fully exploit the data.

Proposition 2. *Suppose that data processing and monetization are not contractible. Then,*

(i) when the impact of processing on the value of data is high (i.e., for $\delta < \frac{\gamma}{k+\gamma}$), there exist $\alpha_F, \alpha_C \in (\bar{\alpha}, k+\gamma)$, both strictly increasing in δ, such that profit and consumer surplus respectively are higher under consumer's right than under firm's right if and only if $\alpha \leq \alpha_F$ and $\alpha \leq \alpha_C$, respectively;

(ii) when the impact of processing on the value of data is low (i.e., for $\delta \geq \frac{\gamma}{k+\gamma}$), both the firm's profit and consumer surplus are higher under consumer's right than under firm's right for low values of α, equal under both ownership regimes for intermediate values of α, and lower under consumer's right than under firm's right for high values of α (i.e., $\alpha \geq \bar{\alpha}$).

For both the firm and the consumers, the key element is the extent to which their interaction can generate some value. Because of the inability to commit ex ante to ex post actions (e.g., data processing and monetization), each ownership regime induces a particular form of inefficiency. Under the firm's right regime, the firm optimally decides on the data processing but monetizes an excessive amount of data. Such an excessive monetization is anticipated by the consumers and inevitably forces the firm to reduce the usage it offers to the high-type consumers. Under the consumer's right regime, conditional on the firm not processing the data, the amount of data monetized is ex post optimal, but the firm has no incentive to process the data. This results in lower levels of usage for low-type consumers relative to the contractible data-extraction benchmark. The trade-off between the two forms of inefficiency, i.e., under-processing vs. over-utilization, depends on the value of data: when α is low, consumer's right is optimal, whereas when α is high, firm's right is optimal. Figure 1 illustrates these results; the left (resp. right) panel compares the firm's profit (consumer surplus) under the two ownership regimes.



(a) Firm's Profit (b) Consumer Surplus

Figure 1: Graphical Illustration of Proposition 2

5 Extensions and Robustness

A. Firm's Right and Paying for Privacy

An alternative to transferring the rights to consumers is to allow the firm to offer an option such that the consumer pays for her privacy. Under such a regime, the firm owns the rights and offers two goods: (i) a good that provides access only to the service (and hence leaves the firm with all rights to any data generated) and (ii) a good that provides access to the service but guarantees that any data generated by the interaction between the firm and the consumer is not used in the future.[14] In practical terms, platforms such as YouTube or certain game apps usually offer versions of the same product: a *basic* version that uses data to promote products to users through advertising and a *premium* version that is advertising free and therefore limits data collection.

As above, we focus without loss of generality on contracts (or direct revelation mechanisms) in which consumers report their type and a pre-defined price-usage-extraction offer is implemented. Because the option for privacy is a binary decision, if consumers do not choose to opt for privacy, all the data they generate is indeed extracted and monetized. Therefore, the option for privacy amounts to choosing $e(\theta) \in \{0, q(\theta)\}$. When $e(\theta) = q(\theta)$, the firm retains the right to use the data of type θ, which, due to its inability to commit, implies that it monetizes the data *ex post*; when $e(\theta) = 0$, the firm gives up every right to extract and monetize the data of type θ.

Let $(t(\cdot), q(\cdot), e(\cdot))$ denote the offer made to consumers. The indirect utility of type θ from truthfully announcing her type is

$$U(\theta) = \theta q(\theta) - q^2(\theta)/2 - t(\theta) - C(e(\theta)) \tag{9}$$

As in the benchmark, we can show that the mechanism is incentive compatible only if usage is non-decreasing in θ and $\dot{U}(\theta) = q(\theta)$. The firm's expected profit is

$$\Pi = \int_0^1 \big(t(\theta) + B_v(\alpha, e(\theta))\big) d\theta,$$

which, by substituting (9) and re-arranging terms, is equal to

$$\Pi = \int_0^1 \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2 + B_v(\alpha, e(\theta)) - C(e(\theta))\Big].d\theta \tag{10}$$

Therefore, the objective of the firm is to find a usage schedule $q(\cdot)$ and a privacy offer $e(\cdot)$ that maximize its profit (10) subject to the monotonicity constraint ($q(\cdot)$ is non-decreasing) and feasibility constraints ($q(\theta) \geq 0$ and $e(\theta) \in \{0, q(\theta)\}$ for every θ). This profit-maximizing offer is described in the following lemma.

[14] Akcura and Srinivasan (2005) propose a related analysis in which a monopolist can resell some information and commit to doing so ex ante. Casadesus-Masanell and Hervas-Drane (2015) analyze a competitive version where the level of information disclosure differs among firms.

Lemma 5. *Suppose that the firm owns the rights and can propose that consumers pay for their privacy. Then, there exists* $\hat{\theta}_v^P \in [\hat{\theta}_v, 1]$*, where* $v = \delta$ *if* $\alpha < \bar{\alpha}$ *and* $v = 1$ *if* $\alpha \geq \bar{\alpha}$*, such that type-*θ*'s usage in the firm's optimal mechanism is given by*

$$q_*^P(\theta) = \begin{cases} \max\{0, q_v(\theta)\}, & \textit{if } \theta < \hat{\theta}_v^P, \\ q_0(\theta), & \textit{if } \theta \geq \hat{\theta}_v^P \end{cases}$$

where $q_v(\theta)$ *and* $q_0(\theta)$ *are defined in Proposition 1. Moreover, the firm monetizes all the data of consumers in* $[0, \hat{\theta}_v^P]$ *and no data for consumers in* $(\hat{\theta}_v^P, 1]$*.*

Despite the option for privacy, inefficiencies still prevail in the data monetization process. Indeed, to maximize the surplus, the firm should monetize all the data for low-type consumers but not all the data for high-type consumers. However, the firm's inability to precisely fine-tune the data monetization leads the firm to monetize more data than is optimal for some high-type consumers, namely, those whose type is between $\hat{\theta}_v$ and $\hat{\theta}_v^P$, and monetize no data for consumers whose type is greater than $\hat{\theta}_v^P$. The maximum firm's profit and consumer surplus are given by

$$\Pi_v^P = \int_{\underline{\theta}_v}^{\hat{\theta}_v^P} \Big((2\theta - 1)q_v(\theta) - q_v^2(\theta)/2 + B_v(\alpha, q_v(\theta)) - C(q_v(\theta))\Big)d\theta + \int_{\hat{\theta}^P}^{1} \Big((2\theta - 1)q_0(\theta) - q_0^2(\theta)/2\Big)d\theta$$

$$CS^P = \int_{\underline{\theta}_v}^{\hat{\theta}_v^P} q_v(\theta)(1-\theta)d\theta + \int_{\hat{\theta}_v^P}^{1} q_0(\theta)(1-\theta)d\theta$$

By construction, both the firm and consumers are better off under the option for privacy relative to the regime in which the firm owns the rights but no option for privacy is available.[15] The option for privacy equips the firm with an additional contracting tool to mitigate its commitment problem. This allows the firm to increase the price it charges.

[15]Notably, the possibility for the firm to offer an option for privacy nests as a subcase the case in which no such option is available. Indeed, the firm can always refrain from offering any option for privacy, which is equivalent to the case in which no option for privacy is available.

Regarding consumers, the option for privacy allows higher usage and therefore a larger share of the generated surplus.

We now turn our attention to the comparison between the regime in which the firm owns the rights and can offer an option for privacy with the consumer's right regime. As in Section 4, this comparison depends on parameters and, in particular, (i) how efficient consumers are when they directly monetize their data (i.e., δ) and (ii) the value of data in the secondary market (i.e., α).

One special case that is different from Section 4 is that in which the consumers cannot monetize any piece of data alone, that is, when $\delta = 0$. Recall that in Section 4, we argued that even in this case, consumer's right is preferred to firm's right whenever the value of data is low enough. Indeed, although the consumers were unable to monetize any data, for low values of α the inefficiency due to excessive data monetization on behalf of the firm was sufficient to offset the inefficiency due to the inability of consumers to earn any value from their data. This is no longer the case if the firm has the privacy option. Indeed, the firm can always offer all consumers privacy and hence monetize no data. This implies that, whenever $\delta = 0$, for $\alpha \leq \bar{\alpha}$, the allocation of rights is irrelevant for the firm's profit and consumer surplus.

Some other cases are similar to Section 4. For instance, when the value of data is high, the firm always processes the data when it has the rights—and optimally monetizes all the data from all consumers. Therefore, the option for privacy does not introduce any trade-off that is meaningfully different relative to the regime in which the firm has no ability to offer consumers the option to pay for their privacy.

Similarly to Section 4, the most interesting case arises when $\bar{\alpha} < \alpha < k + \gamma$ (and $\delta < \frac{\gamma}{k+\gamma}$). As we explained above, introducing the option for privacy benefits the firm because it partially mitigates the commitment problem regarding the data over-utilization. Analogously, the option for privacy benefits the consumers because high-type consumers can enjoy higher levels of usage, which increases consumer surplus. Therefore, one can expect that introducing the option for privacy makes the regime in which the firm owns the rights preferable to the consumer's right regime more often. This case is stated in the following proposition.

Proposition 3. *Suppose that data processing and monetization are not contractible, but when the firm owns the rights, it can offer an option for privacy. Then,*

(i) when consumers cannot monetize their data alone (i.e., for $\delta = 0$), firm's right is weakly better than consumer's right both for firms and consumers for any value of data;

(ii) when the impact of data processing on the value of data is high (i.e., for $0 < \delta < \frac{\gamma}{k+\gamma}$),there exist $\tilde{\alpha}_F \in (\bar{\alpha}, \alpha_F)$ and $\tilde{\alpha}_C \in (\bar{\alpha}, \alpha_C]$, both strictly increasing in δ, such that the firm's profit and consumer surplus are higher under consumer's right than under firm's right if and only if $\alpha \leq \tilde{\alpha}_F$ and $\alpha \leq \tilde{\alpha}_C$, respectively;

(iii) when the impact of data processing on the value of data is low (i.e., for $\delta \geq \frac{\gamma}{k+\gamma}$), introducing an option for privacy does not alter the ranking between the ownership regimes either for the firm or for the consumers.

There are two main discrepancies between Propositions 2 and 3. The first concerns the case in which consumers cannot monetize any piece of data alone (i.e., $\delta = 0$). Then, because the firm, through the option for privacy, can basically offer to monetize the data of no consumer, it can replicate the outcome attained under consumer's right. However, the firm can also generate more profit when it has the rights and adds value to the data through processing (that is when $\alpha > \bar{\alpha}$). The second difference concerns the case in which $\delta < \frac{\gamma}{k+\gamma}$. In this case, the option of privacy allows the firm to better accommodate the consumer's taste for privacy, leading to higher profit and higher consumer's surplus. As a consequence, firm's right is preferred to consumer's right more often both by firms and consumers with the option for privacy than without.

B. Market for Data with a Uniform Price

In this subsection, we consider the case in which consumers own the rights but cannot fine-tune the amount of data they sell; instead, they can sell either all their data at a uniform price or no data at all. This is equivalent to allowing consumers to sell the right to access their data rather than the precise data content. We suppose that the decision to monetize the data is taken after the data is generated. Equivalently, one can assume that consumers cannot commit to monetizing their data before purchasing the service.

As in the baseline model, we consider a competitive data market—that is, there are many data brokers interested in purchasing the database. This implies that firms purchasing the data (e.g., data brokers) earn zero profits. Let P denote the price at which consumers sell their data. For any price P, the consumers who are willing to sell their data are those with a usage $q(\theta)$ such that $P \geq C(q(\theta))$. Let $\hat{\theta}_M$ denote the cut-off type such that $P = C(q(\hat{\theta}_M))$. We assume that only those consumers who have indeed purchased the service can sell their data; consumers who do not purchase the service have no data to sell.

As explained in Section 4, provided that the data is not processed by the firm that generates it, the benefit that can be directly extracted is given by $B_\delta(\alpha, q(\theta))$. Let $\underline{\theta}_M$ denote the cut-off type below which no type purchases the service. The zero-profit condition in the market for data boils down to

$$P = \int_{\underline{\theta}_M}^{\hat{\theta}_M} \frac{B_\delta(\alpha, q(\theta))}{\hat{\theta}_M - \underline{\theta}_M} d\theta \tag{11}$$

To explain (11), note that $P \times (\hat{\theta}_M - \underline{\theta}_M)$ is the cost of purchasing the data at price P from the share of consumers $(\hat{\theta}_M - \underline{\theta}_M)$, whereas $\int_{\underline{\theta}_M}^{\hat{\theta}_M} B_\delta(\alpha, q(\theta))d\theta$ is the benefit from monetizing the data of those consumers.

We now turn to the interaction between the firm and the consumers in the first period. The utility of those consumers who anticipate selling their data on the market (i.e., $\theta \leq$

$\hat{\theta}_M$) is equal to

$$U(\theta) = \theta q(\theta) - q^2(\theta)/2 - t(\theta) + P - C(q(\theta)),$$

whereas the utility of those consumers who anticipate that they will not be active on the market for data (i.e., $\theta \leq \hat{\theta}_M$) is equal to

$$U(\theta) = \theta q(\theta) - q^2(\theta)/2 - t(\theta).$$

In both cases, to derive the optimal firm's offer, we focus on direct revelation mechanisms. The conditions for this offer to be incentive compatible are the same as above and given by the conditions $\dot{U}(\theta) = q(\theta)$ and $q(\theta)$ increasing in θ. The firm's profit can be written as

$$\Pi = \int_{\underline{\theta}_M}^{\hat{\theta}_M} \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2 + P - C(q(\theta))\Big] d\theta + \int_{\hat{\theta}_M}^{1} \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2\Big] d\theta$$

which, by employing (11), boils down to

$$\begin{aligned} \Pi &= \int_{\underline{\theta}_M}^{\hat{\theta}_M} \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2 + B_\delta(\alpha, q(\theta)) - C(q(\theta))\Big] d\theta \\ &\qquad + \int_{\hat{\theta}_M}^{1} \Big[(2\theta - 1)q(\theta) - q^2(\theta)/2\Big] d\theta \end{aligned} \tag{12}$$

Let us compare the profit in this regime with the profit in the consumer's right regime, in which consumers can choose how much data to sell and therefore sell this data at a personalized price. Since the market price is set at the average data value, the profit the firm can make on the consumers that sell their data is the same as the profit obtained in the standard regime with consumer's right. By contrast, when consumers do not sell their data on the market, no value is created, so the firm cannot retrieve any revenue from these consumers. This condition implies that the profit is reduced in this regime relative to the consumer's right regime in which consumers can choose how much data to sell.

As far as consumers are concerned, the analysis slightly differs. Indeed, what matters for consumer surplus is the extent to which the change from personalized price to a uniform price modifies the usage schedules offered by the firm. It turns out that, even if the price proposed by the firm changes, there is no difference in the usage schedules between the regimes with consumer's right and the regime with uniform price. Indeed, by inspection of (12), one can see that the firm still offers $q_\delta(\cdot)$ to low-type consumers who sell their data and $q_0(\cdot)$ to high-type consumers who do not sell their data. Moreover, in spite of the absence of data monetization for the high-type consumers, the cut-off type between the two usage schedules, $\hat{\theta}_M$, is the same as that obtained in the standard regime with consumer's right, $\hat{\theta}_\delta$. Intuitively, even if consumers directly decide on whether they should sell their data or not, the firm indirectly controls this decision through its offer $(t(\cdot), q(\cdot))$. Therefore, the firm also implicitly selects the cut-off type $\hat{\theta}_M$ to maximize its profit. Since the value generated on the data market for the cut-off type is zero, as $P = C(q(\hat{\theta}_M))$, the cut-off type is chosen such that the firm's first-period profit is the same under both usage schedules. This is precisely the profit-maximization condition in the consumer's right regime studied in Section 4. Provided that the usage schedules in the consumer's right regime and the regime with uniform price are the same, the consumer surpluses in the two regimes are also the same.

We summarize this discussion below.

Proposition 4. *Suppose that data processing and monetization are not contractible but, rather, consumers can only sell all their data at a uniform price or no data at all. Then, the firm's profit is lower relative to the case in which consumers can select the amount of data they can sell. By contrast, consumer surplus is the same in both cases.*

6 Conclusion

Data is currently ubiquitous in the economy. Therefore, the optimal allocation of rights over its exploitation seems paramount. From our point of view, the debate regarding the ownership of data stems from the firms' inability to costlessly contract on the way data is processed and monetized. This is due to the fact that both data processing and monetization are technical issues that make commitment and verifiability difficult to realize.

However, these issues have a distinct impact on the outcomes when the firms own the property rights over the data as opposed to when users own the property rights. This implies that different allocations of rights lead to different offers made by firms and, consequently, differences in firms' profits and consumer surplus. The optimal allocation of rights crucially depends on the value of the data or equivalently on the relative weight between the market in which the data is generated and the market in which it is used. When the former is more important, consumers should own the rights to their data; when the latter is more important, it may be welfare improving for firms to own the rights.

Our approach was developed under a set of assumptions, some of which should be relaxed in future research. First, we assumed that the firm in the primary market that trades with consumers and therefore generates—or co-generates—data is a monopolist. It would be interesting to determine whether our results hinge on this assumption, that is, how competition in the primary market affects the optimal allocation of rights. On the one hand, competition would certainly restrict firms' ability to extract consumer surplus. On the other hand, it is at this stage rather unclear whether this would encourage the allocation of rights to firms or consumers.

We also assumed that usage is contractible, with the firm offering a tariff conditional on usage. Nonetheless, in many real-world applications, firms set a subscription fee for the service, and consumers each freely decide on their usage. This assumption on contractibility simplifies the analysis but is not crucial for our results. Nevertheless, a model in which firms could only set a subscription fee would be better to explore questions related to the strategies of firms on the internet, particularly in a competitive setting, and their consequences for consumers.

Finally, we assumed that the data generation process involved only two parties—one firm and one consumer—per transaction. Nevertheless, it is common for data generation to involve more than two parties. For example, in the automobile industry, car manufacturers and software producers compete over the rights to use data on drivers' behavior (see Agrawal, Gans, and Goldfarb (2018)). In social media, information concerns groups rather than individuals. Therefore, the simple dichotomy between firm rights and consumer rights is not sufficient to study optimal ownership. In a multi-firm or multi-agent environment, new challenges arise, and thinking about the ownership of data is even more important for helping firms to design their strategy.

A APPENDIX

A. Proof of Proposition 1

From Lemma 1, we know that $v = \delta$ if $\alpha < \bar{\alpha}$ and $v = 1$ if $\alpha \geq \bar{\alpha}$. We also know that for $q < \bar{e}_v$, $\phi'_v(q) > 0$, whereas for $q \geq \bar{e}_v$, $\phi'_v(q) = 0$. Therefore, the question is to examine when, indeed, $q(\theta) < \bar{e}_v$. Assuming that $q(\theta) < \bar{e}_v$, Eq. (5) yields an optimal usage schedule equal to $q_v(\theta)$, whereas assuming that $q(\theta) \geq \bar{e}_v$, Eq. (5) yields an optimal usage schedule equal to $q_0(\theta)$, where $q_1(\theta) = (2\theta - 1 + \alpha - k)/(1+\gamma)$, $q_\delta(\theta) = (2\theta - 1 + \delta\alpha)/(1+\gamma)$ and $q_0(\theta) = 2\theta - 1$. Note, however, that $q_v(\theta) < \bar{e}_v$ if and only if $\theta < \hat{\theta}_v$, where $\hat{\theta}_\delta = \min\{\frac{1}{2}\Big(\frac{\delta\alpha}{\gamma} + 1\Big), 1\}$ and that $\hat{\theta}_1 = \min\{\frac{1}{2}\Big(\frac{\alpha - k}{\gamma} + 1\Big), 1\}$. Therefore, the optimal mechanism is such that $q^*_v(\theta) = e^*_v(\theta) = \max\{q_v(\theta), 0\}$ for $\theta < \hat{\theta}_v$ and $q^*_v(\theta) = q_0(\theta) \geq e^*_v(\theta) = \bar{e}_v$ for $\theta \geq \hat{\theta}_v$, for $v \in \{\delta, 1\}$.

Q.E.D.

B. Proof of Lemma 2

We argued in the text that when the firm owns the rights, it extracts and monetizes the data of all the consumers. We showed in Lemma 1 that the firm processes the data if and only if $\alpha \geq \bar{\alpha}$. Therefore, type-θ's usage in the firm's optimal mechanism is given by $\max\{0, q_v(\theta)\}$, where $q_v(\theta)$ is as in Proposition 1.

Q.E.D.

C. Proof of Corollary 1

It suffices to identify conditions such that $e_v^*(\theta) < \max\{0, q_v(\theta)\} < q_0(\theta)$. This is equivalent to showing that (i) $\hat{\theta}_1 < 1$ with data processing and (ii) $\hat{\theta}_\delta < 1$ without data processing. Using the expressions in Proposition 1, it is directly observable that (i) $\hat{\theta}_1 < 1$ if and only if $\alpha < \gamma + k$ and that (ii) $\hat{\theta}_\delta < 1$ if and only if $\alpha < \gamma/\delta$. Moreover $\bar{\alpha} < \gamma + k$ if and only if $\delta < \frac{\gamma}{\gamma+k}$ and $\bar{\alpha} > \frac{\gamma}{\delta}$ if and only if $\delta > \frac{\gamma}{\gamma+k}$. Therefore, if we denote $\hat{\alpha} = \gamma + k$ for $\delta < \frac{\gamma}{\gamma+k}$ and $\hat{\alpha} = \frac{\gamma}{\delta}$ for $\delta > \frac{\gamma}{\gamma+k}$, the result directly follows.

Q.E.D.

D. Proof of Lemma 3

Suppose that there is a contract (t, q) such that the anticipated value per unit of data is α and that the consumers sell an anticipated amount of data $\hat{e} \leq q$. This means that a consumer with type θ obtains utility

$$\theta q - \theta q^2/2 + \alpha\hat{e} - C(\hat{e}) - t$$

For the contract to be consistent with the belief that the firm will process the data, it must maximize the firm's net profit $t - k\hat{e}$ subject to the standard incentive compatibility and participation conditions. This means that the contract proposed by the firm should be similar, when the variables are contractible, to that derived in the benchmark case. The quantity is then given by $q_1(\theta)$, and the price t equals the surplus minus the informational rents (R). This price can be written as

$$t = \theta q_1 - \theta q_1^2/2 + \phi_1(q_1) - R$$

However, ex post, the firm has no incentive to invest in data processing as it has been paid ex ante and only incurs some costs by investing. This means that the agent's utility would be given by

$$U = R - \phi_1(q_1) + \phi_\delta(q_1) < R.$$

Because of this hold-up problem, consumers would never accept to pay price t that incorporates the value of processed data. Therefore, if the consumers own the rights, the firm never processes the data.

Q.E.D.

E. Proof of Lemma 4

The proof is similar to that of Proposition 1. We know that for $q < \bar{e}_\delta$, $\phi'_\delta(q) > 0$, whereas for $q \geq \bar{e}_\delta$, $\phi'_\delta(q) = 0$. The optimal usage schedule is found by point-wise maximization of (8). Assuming that $q(\theta) < \bar{e}_\delta$, the first-order condition yields $q_\delta(\theta)$, whereas assuming that $q(\theta) \geq \bar{e}_\delta$, the first-order condition yields $q_0(\theta)$. Note, however, that $q_\delta(\theta) < \bar{e}_\delta$ if and only if $\theta < \hat{\theta}_\delta$. Therefore the optimal usage schedule of type θ is $q_\delta(\theta)$ for $\theta < \hat{\theta}_\delta$ and $q_0(\theta)$ for $\theta \geq \hat{\theta}_\delta$.

Q.E.D.

F. Proof of Proposition 2

PRELIMINARIES. The proof of this proposition is written for any distribution over $[0, 1]$, with density $f(\cdot)$ and cdf $F(\cdot)$. Let $h(\cdot) = (1 - F(\cdot))/f(\cdot)$ denote the *inverse hazard ratio* and assume that $h(\cdot)$ is non-increasing in θ. This *monotone hazard rate property* is rather weak and satisfied for most distributions. Using a more general proof illustrates that our results extend beyond the simple case of a uniform distribution over $[0, 1]$. It also allows a better understanding of some important equations.

For any general distribution function, the expressions specified in Proposition 1 are modified as follows:

$$q_0(\theta) = \theta - h(\theta), \quad q_\delta(\theta) = \frac{1}{1+\gamma}(\theta - h(\theta) + \delta\alpha), \quad \text{and} \quad q_1(\theta) = \frac{1}{1+\gamma}(\theta - h(\theta) + \alpha - k).$$

Moreover, $\underline{\theta}_\delta$ and $\underline{\theta}_1$ are such that $q_\delta(\underline{\theta}_\delta) = 0$ and $q_1(\underline{\theta}_1) = 0$, whereas $\hat{\theta}_\delta$ and $\hat{\theta}_1$ are given by $\hat{\theta}_\delta - h(\hat{\theta}_\delta) = \frac{\delta\alpha}{\gamma}$ and $\hat{\theta}_1 - h(\hat{\theta}_1) = \frac{\alpha-k}{\gamma}$.

PROOF OF (i). As we argued in the text (and in light of Corollaries 1 and 2), for any $\alpha > k + \gamma$, the usage and data extraction in the firm's optimal mechanism when the firm owns the rights are equal to those in the contractible data-extraction benchmark, which implies that both the firm's profit and consumer surplus are higher when the firm owns the rights than when the consumers own the rights. Moreover, when $\alpha < \bar{\alpha}$, the usage in the firm's optimal mechanism and the data extraction chosen by consumers when the consumers own the rights are equal to those in the contractible data-extraction benchmark, which implies that both the firm's profit and consumer surplus are higher when the consumers own the rights than when the firm owns the rights. Therefore, below, we study the case in which $\alpha \in [\bar{\alpha}, k + \gamma]$.

Firm's Profit.—The corresponding firm's profits under the two different regimes are given by

$$\Pi^F = \int_{\underline{\theta}_1}^{1} \left[(\theta - h(\theta) + \alpha - k)q_1(\theta) - (1+\gamma)\frac{q_1^2(\theta)}{2}\right] f(\theta)d\theta$$

and

$$\Pi^C = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \left[(\theta - h(\theta) + \delta\alpha)q_\delta(\theta) - (1+\gamma)\frac{q_\delta^2(\theta)}{2}\right] f(\theta)d\theta + \int_{\hat{\theta}_\delta}^{1} \left[(\theta - h(\theta))q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_\delta(\bar{e}_\delta)\right] f(\theta)d\theta$$

We proceed as follows. First, we compare the profits under the two regimes for the extreme values $\alpha = \bar{\alpha}$ and $\alpha = k + \gamma$. Then, we compare the first derivatives of the profit functions with respect to α.

Consider first the difference of the firm's profits under the two regimes ($\Pi^F - \Pi^C$) for $\alpha = \bar{\alpha}$. In this case, we have $\underline{\theta}_\delta = \underline{\theta}_1$ and $q_\delta(\theta) = q_1(\theta)$. Therefore, using the above expressions for the profits,

$$\Pi^F - \Pi^C = \int_{\hat{\theta}_\delta}^1 \Bigg(\Big[(\theta - h(\theta) + \alpha - k) q_1(\theta) - (1+\gamma) \frac{q_1^2(\theta)}{2} \Big] - \Big[(\theta - h(\theta)) q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_1(\bar{e}_1) \Big] \Bigg) f(\theta) d\theta$$

$$= \int_{\hat{\theta}_\delta}^1 \Bigg(\Big[(\theta - h(\theta)) q_1(\theta) - \frac{q_1^2(\theta)}{2} + (\alpha - k) q_1(\theta) - \gamma \frac{q_1^2(\theta)}{2} \Big] - \Big[(\theta - h(\theta)) q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_1(\bar{e}_1) \Big] \Bigg) f(\theta) d\theta$$

We know, however, that $\phi_1(\bar{e}_1) = \max_e (\alpha - k)e - \gamma \frac{e^2}{2}$ and that $q_0(\theta) = \arg\max_q (\theta - h(\theta))q - \frac{q^2}{2}$. Therefore, for $\alpha = \bar{\alpha}$, $\Pi^F - \Pi^C < 0$.

Consider now the difference in the firm's profits ($\Pi^F - \Pi^C$) when $\alpha = k + \gamma$. It is easy to see that in this case, $\hat{\theta}_1 = 1$, which means that when the firm owns the rights, it extracts and monetizes all the data from all its customers. Moreover, since $\delta < \frac{\gamma}{\gamma + k}$, $\hat{\theta}_\delta < 1$, which means that when the consumers own the rights, only those consumers with type $[\underline{\theta}_\delta, \hat{\theta}_\delta]$ monetize all their data. Because the firm's profit when the firm owns the rights is equal to its profit in the contractible data extraction benchmark—which is the maximum profit possible—whereas the firm's profit when the consumers own the rights differs from that in the contractible data-extraction benchmark, for $\alpha = k + \gamma$, $\Pi^F - \Pi^C > 0$.

We now find the derivatives of the profit functions with respect to α. The derivative of Π^F with respect to α is given by

$$\frac{d\Pi^F}{d\alpha} = -\frac{d\underline{\theta}_1}{d\alpha} \underbrace{\Big[(\underline{\theta}_1 - h(\underline{\theta}_1) + \alpha - k) q_1(\underline{\theta}_1) - (1+\gamma) \frac{q_1^2(\underline{\theta}_1)}{2} \Big]}_{=0} f(\underline{\theta}_1) + \int_{\underline{\theta}_1}^1 q_1(\theta) f(\theta) d\theta$$

$$= \int_{\underline{\theta}_1}^1 q_1(\theta) f(\theta) d\theta \geq 0$$

Moreover, the derivative of Π^C with respect to α is given by

$$\begin{aligned}\frac{d\Pi^C}{d\alpha} &= \frac{d\hat{\theta}_\delta}{d\alpha}\Big[(\hat{\theta}_\delta - h(\hat{\theta}_\delta) + \delta\alpha)q_\delta(\hat{\theta}_\delta) - (1+\gamma)\frac{q_\delta^2(\hat{\theta}_\delta)}{2}\Big]f(\hat{\theta}_\delta) \\ &- \frac{d\underline{\theta}_\delta}{d\alpha}\underbrace{\Big[(\underline{\theta}_\delta - h(\underline{\theta}_\delta) + \delta\alpha)q_\delta(\underline{\theta}_\delta) - (1+\gamma)\frac{q_\delta^2(\underline{\theta}_\delta)}{2}\Big]}_{=0}f(\underline{\theta}_\delta) \\ &+ \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta}\delta q_\delta(\theta)f(\theta)d\theta + \int_{\hat{\theta}_\delta}^{1}\frac{d\phi_\delta(q_\delta(\hat{\theta}_\delta))}{d\alpha}f(\theta)d\theta \\ &- \frac{d\hat{\theta}_\delta}{d\alpha}\Big[(\hat{\theta}_\delta - h(\hat{\theta}_\delta))q_0(\hat{\theta}_\delta) - \frac{q_0^2(\hat{\theta}_\delta)}{2} + \phi_\delta(q_\delta(\hat{\theta}_\delta))\Big]f(\hat{\theta}_\delta)\end{aligned}$$

Because $q_0(\hat{\theta}_\delta) = q_\delta(\hat{\theta}_\delta) = \frac{\delta\alpha}{\gamma}$ and $\phi_\delta(q_\delta(\hat{\theta}_\delta)) = \frac{q_\delta^2(\hat{\theta}_\delta)}{2}\gamma$, the following equality holds

$$(1+\gamma)\frac{q_\delta^2(\hat{\theta}_\delta)}{2} = \frac{q_0^2(\hat{\theta}_\delta)}{2} + \phi_\delta(q_\delta(\hat{\theta}_\delta))$$

which implies that

$$\frac{d\Pi^C}{d\alpha} = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta}\delta q_\delta(\theta)f(\theta)d\theta + \int_{\hat{\theta}_\delta}^{1}\frac{d\phi_\delta(q_\delta(\hat{\theta}_\delta))}{d\alpha}f(\theta)d\theta$$

Moreover, because $\frac{d\phi_\delta(q_\delta(\hat{\theta}_\delta))}{d\alpha} = \delta q_\delta(\hat{\theta}_\delta)$, we obtain

$$\frac{d\Pi^C}{d\alpha} = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta}\delta q_\delta(\theta)f(\theta)d\theta + \delta q_\delta(\hat{\theta}_\delta)(1 - F(\hat{\theta}_\delta)) \tag{13}$$

Provided that $\underline{\theta}_\delta \geq \underline{\theta}_1$, $q_\delta(\theta) \leq q_1(\theta)$ and $\frac{d\hat{\theta}_\delta}{d\alpha} \geq 0$, it is directly observable that $\frac{d\Pi^C}{d\alpha} > 0$ and $\frac{d\Pi^C}{d\alpha} \leq \frac{d\Pi^F}{d\alpha}$.

Therefore, given that $\Pi^F < \Pi^C$ for $\alpha = \bar{\alpha}$, $\Pi^F > \Pi^C$ for $\alpha = k + \gamma$, Π^F and Π^C are continuous in α, and $\frac{d\Pi^C}{d\alpha} \leq \frac{d\Pi^F}{d\alpha}$ for every $\alpha \in [\bar{\alpha}, k+\gamma]$, there exists $\alpha_F \in (\bar{\alpha}, k+\gamma)$ such that the firm's profit is higher under the firm's right than under the consumer's right if and only if $\alpha \geq \alpha_F$.

Note, finally, that the first derivative of Π^C with respect to δ is given by

$$\frac{d\Pi^C}{d\delta} = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta}\alpha q_\delta(\theta)f(\theta)d\theta + \alpha q_\delta(\hat{\theta}_\delta)(1 - F(\theta)) \geq 0.$$

Therefore, Π^C is increasing in δ, which means that $\Pi^F - \Pi^C$ decreases with δ. This implies that α_F increases with δ.

Consumer Surplus.—The consumer surplus under the two regimes is given by

$$CS^F = \int_{\underline{\theta}_1}^{1} q_1(\theta)(1 - F(\theta))d\theta$$

and

$$CS^C = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} q_\delta(\theta)(1 - F(\theta))d\theta + \int_{\hat{\theta}_\delta}^{1} q_0(\theta)(1 - F(\theta))d\theta$$

To compare the consumer surplus under the two regimes, we follow a method similar to that we followed to compare the profit under the two regimes.

In particular, we first evaluate the consumer surplus for the extreme values $\alpha = \bar{\alpha}$ and $\alpha = k + \gamma$. For $\alpha = \bar{\alpha}$, we have $\underline{\theta}_\delta = \underline{\theta}_1$, $q_\delta(\theta) = q_1(\theta)$ and $q_0(\theta) > q_\delta(\theta)$. Therefore, for $\alpha = \bar{\alpha}$, it is clear that $CS^F < CS^C$. For $\alpha = k + \gamma$, $q_1(\theta) > q_\delta(\theta)$ and $q_1(\theta) > q_0(\theta)$ for all $\theta < 1$ and $q_1(\theta) = q_0(\theta)$ for $\theta = 1$. This implies that, for $\alpha = k + \gamma$, $CS^F - CS^C > 0$.

The first derivative of CS^F with respect to α is given by

$$\frac{dCS_F}{d\alpha} = \int_{\underline{\theta}_1}^{1} \frac{dq_1(\theta)}{d\alpha}(1 - F(\theta))d\theta$$

which, given that $\frac{dq_1(\theta)}{d\alpha} = \frac{1}{1+\gamma} > 0$, is positive. Moreover, using the fact that $q_\delta(\hat{\theta}_\delta) = q_0(\hat{\theta}_\delta)$, the first derivative of CS^C with respect to α is given by

$$\frac{dCS_C}{d\alpha} = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{dq_\delta(\theta)}{d\alpha}(1 - F(\theta))d\theta$$

which, given that $\frac{dq_\delta(\theta)}{d\alpha} = \frac{\delta}{1+\gamma} > 0$, is also positive. We argued above that $\underline{\theta}_1 \leq \underline{\theta}_\delta \leq \hat{\theta}_\delta \leq 1$. Therefore, $\frac{dCS_F}{d\alpha} \geq \frac{dCS_C}{d\alpha}$.

In sum, given that $CS^F < CS^C$ for $\alpha = \bar{\alpha}$, $CS^F > CS^C$ for $\alpha = k + \gamma$, CS^F and CS^C are continuous in α, and $\frac{dCS_C}{d\alpha} \leq \frac{dCS_F}{d\alpha}$ for every $\alpha \in [\bar{\alpha}, k + \gamma]$, there exists $\alpha_C \in (\bar{\alpha}, k + \gamma)$ such that the firm's profit is higher under firm's right than under consumer's right if and only if $\alpha \geq \alpha_C$.

Finally, note that CS^F does not depend on δ, whereas, again with $q_\delta(\hat{\theta}_\delta) = q_0(\hat{\theta}_\delta)$,

$$\frac{dCS_C}{d\delta} = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{dq_\delta(\theta)}{d\delta}(1 - F(\theta))d\theta \tag{14}$$

, which is positive because $\frac{dq_\delta(\theta)}{d\delta} = \frac{\alpha}{1+c} > 0$. We then conclude that α_C increases with δ.

PROOF OF (ii). From Corollary 1, we know that data monetization is inefficient for any α less that $\hat{\alpha} \leq \bar{\alpha}$. As there is no data processing in this case, and because of the excessive data monetization under firm's right, the usage is higher when consumers have the right. As this usage corresponds to the contractible benchmark characterized in Proposition 1, the profit is also higher under consumer's right.

When $\alpha \geq \bar{\alpha}$, it implies that $\alpha \geq \hat{\alpha}$ (still from Corollary 1). Therefore, under firm's right, data processing and monetization are optimal. The firm's profit is therefore equal to its profit in the contractible data-extraction benchmark. As far as consumer surplus is concerned, the usages are given by q_1 with firm's right and $\max\{q_0(\theta), q_\delta(\theta)\} = q_\delta(\theta) < q_1(\theta)$. Therefore, the consumer surplus is higher under firm's right.

Finally, for $\alpha \in [\hat{\alpha}, \bar{\alpha}]$, there is full monetization with both ownership regimes but no data processing. Therefore, the usage schedules, the profit, and the consumer surplus are equal under both ownership regimes.

Q.E.D.

G. Proof of Lemma 5

When the firm owns the rights, its objective can be split into two groups depending on whether consumers opt for privacy or not. If type θ does not opt for privacy, the firm's profit from this consumer is the same as when the firm owns the rights and no option for privacy is available. Therefore, the optimal usage is equal to either $q_1(\theta)$, if $\alpha \geq \bar{\alpha}$, or $q_\delta(\theta)$, if $\alpha < \bar{\alpha}$. If type θ opts for privacy, the optimal usage is equal to $q_0(\theta)$. Let

$$\Phi_1(\theta) = (2\theta - 1)q_1(\theta) - \frac{q_1^2(\theta)}{2} + (\alpha - k)q_1(\theta) - \gamma\frac{q_1^2(\theta)}{2}$$

and

$$\Phi_\delta(\theta) = (2\theta - 1)q_\delta(\theta) - \frac{q_\delta^2(\theta)}{2} + \delta\alpha q_\delta(\theta) - \gamma\frac{q_\delta^2(\theta)}{2}$$

denote the firm's profit from type θ when this type does not opt for privacy and $\alpha \geq \bar{\alpha}$ or $\alpha < \bar{\alpha}$, respectively. Moreover, let

$$\Phi_0(\theta) = (2\theta - 1)q_0(\theta) - \frac{q_0^2(\theta)}{2}$$

denote the firm's profit from type θ when this type opts for privacy. For simplicity, we also let

$$G_v(\theta) = \Phi_v(\theta) - \Phi_0(\theta)$$

denote the difference in the firm's profit from type θ when this type opts for privacy or not.

Note first that, by the envelope theorem, the derivative of $G_v(\theta)$ with respect to θ is

$$G_v'(\theta) = 2(q_v(\theta) - q_0(\theta)),$$

, which is strictly positive if $\theta < \hat{\theta}_v$.

Moreover, because $q_0(\underline{\theta}_0) = 0$ and $\underline{\theta}_v \leq \underline{\theta}_0$, it is true that $q_0(\underline{\theta}_v) \leq 0$, which implies that $G(\underline{\theta}_v) = \Phi_v(\underline{\theta}_v) - \Phi_0(\underline{\theta}_v) = 0 - \Phi_0(\underline{\theta}_v) \geq 0$.

Given that $G(\theta)$ is positive for $\theta = \underline{\theta}_v$, strictly increasing for $\theta < \hat{\theta}_v$ and strictly decreasing for $\theta > \hat{\theta}_v$, two distinct cases exist. When $G_v(1) < 0$, there exists $\hat{\theta}_v^P \in (\hat{\theta}_v, 1)$ such that $G(\theta) \geq 0$ iff $\theta \leq \hat{\theta}_v^P$, whereas when $G_v(1) \geq 0$, it is true that $G(\theta) \geq 0$ for every θ. Therefore, in any case, there exists $\hat{\theta}_v^P \in (\hat{\theta}_v, 1]$ such that the firm sets (i) $e(\theta) = q(\theta)$ and $q(\theta) = q_1(\theta)$ for every $\theta < \hat{\theta}_v^P$ and (ii) $e(\theta) = 0$ and $q(\theta) = q_0(\theta)$ for every $\theta \geq \hat{\theta}_v^P$.

We can characterize $\hat{\theta}_v^P$ more precisely. Indeed, $\Phi_v(\theta) = (1+\gamma)\frac{q_v^2(\theta)}{2}$ and $\Phi_0(\theta) = \frac{q_0^2(\theta)}{2}$. Since $\hat{\theta}_v^P$ is such that $V_1(\hat{\theta}_v^P) = V_0(\hat{\theta}_v^P)$, we obtain

$$(1+\gamma)\frac{q_v^2(\hat{\theta}_v^P)}{2} = \frac{q_0^2(\hat{\theta}_v^P)}{2}$$

Since $q_1(\theta) = \frac{q_0(\theta)+\alpha-k}{1+\gamma}$ and $q_\delta(\theta) = \frac{q_0(\theta)+\delta\alpha}{1+\gamma}$ for all θ, this leads to

$$q_0(\hat{\theta}_1^P) = \frac{\alpha - k}{\gamma}[1 + \sqrt{1+\gamma}] \text{ and } q_0(\hat{\theta}_\delta^P) = \frac{\delta\alpha}{\gamma}[1 + \sqrt{1+\gamma}]$$

Using the definition of $q_0(\theta)$, we obtain $\hat{\theta}_1^P = \frac{1}{2}\Big(1 + \frac{\alpha-k}{\gamma}(1 + \sqrt{1+\gamma})\Big)$ and $\hat{\theta}_\delta^P = \frac{1}{2}\Big(1 + \frac{\delta\alpha}{\gamma}(1 + \sqrt{1+\gamma})\Big)$.

Q.E.D.

H. Proof of Proposition 3

PROOF OF (i). For $\delta = 0$, the data has no value if the firm does not process it. For $\alpha \leq \bar{\alpha}$, the firm does not process the data. Nonetheless, because the firm can basically offer to all consumers that it will not use any data, its profit is at least equal to that it can earn under consumer's right. Moreover, because the usage schedule is the same under both regimes (i.e., $q_0(\cdot)$), consumer surplus is the same under both regimes. For $\alpha > \bar{\alpha}$, the firm processes the data when it has the right, which creates an extra value compared to the consumer's right regime. Moreover, because the usage schedule for some consumers is $q_1(\cdot)$, which is greater than $q_0(\cdot)$, the consumer surplus is greater when the firm has the rights than when the consumers have the rights. We conclude that when $\delta = 0$, the profit and consumer surplus are weakly greater under firm's right than under consumer's right.

PROOF OF (ii). As we argued in the text (and in light of Corollaries 1 and 2), when $\delta > 0$ and for any $\alpha > k + \gamma$, the usage and data extraction in the firm's optimal mechanism when the firm owns the rights are equal to those in the contractible data-extraction benchmark, which implies that both the firm's profit and consumer surplus are higher when the firm owns the rights than when the consumers own the rights. Moreover, when $\alpha < \bar{\alpha}$, the usage in the firm's optimal mechanism and the data extraction chosen by consumers when the consumers own the rights are equal to those in the contractible data-extraction benchmark, which implies that both the firm's profit and consumer surplus are higher when the consumers own the rights than when the firm owns the rights. Therefore, as in Proposition 2, below, we study the case in which $\alpha \in [\bar{\alpha}, k+\gamma]$.

Firm's Profit.— To compare the firm's profit under the two regimes, we evaluate the profit for the extreme values of $\alpha = \bar{\alpha}$ and $\alpha = k+\gamma$. We then find the first derivatives of Π^P and Π^C with respect to α.

Consider first the difference in the firm's profit under the two regimes ($\Delta\Pi = \Pi_1^P - \Pi^C$). This is given by

$$\begin{aligned}
\Delta\Pi &= \int_{\underline{\theta}_1}^{\hat{\theta}_1^P} \Big[(2\theta - 1 + \alpha - k)q_1(\theta) - (1+\gamma)\frac{q_1^2(\theta)}{2}\Big]d\theta + \int_{\hat{\theta}_1^P}^{1} \Big[(2\theta-1)q_0(\theta) - \frac{q_0^2(\theta)}{2}\Big]d\theta \\
&- \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \Big[(2\theta-1)q_\delta(\theta) - \frac{q_\delta^2(\theta)}{2} + \delta\alpha q_\delta(\theta) - \gamma\frac{q_\delta^2(\theta)}{2}\Big]d\theta \\
&\qquad - \int_{\hat{\theta}_\delta}^{1} \Big[(2\theta-1)q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_\delta(\bar{e}_\delta)\Big]d\theta
\end{aligned}$$

Note that for $\alpha = \bar{\alpha}$, we have $\underline{\theta}_\delta = \underline{\theta}_1$ and $q_\delta(\theta) = q_1(\theta)$. Moreover, we showed in Lemma 5 that $\hat{\theta}_\delta \leq \hat{\theta}_1^P$. Therefore,

$$\begin{aligned}
\Delta\Pi = &\int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \Big[(2\theta - 1 + \bar{\alpha} - k)q_1(\theta) - (1+\gamma)\frac{q_1^2(\theta)}{2}\Big]d\theta + \int_{\hat{\theta}_1^P}^{1} \Big[(2\theta-1)q_0(\theta) - \frac{q_0^2(\theta)}{2}\Big]d\theta \\
- &\int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \Big[(2\theta-1)q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_\delta(\bar{e}_\delta)\Big]d\theta - \int_{\hat{\theta}_1^P}^{1} \Big[(2\theta-1)q_0(\theta) - \frac{q_0^2(\theta)}{2} + \phi_\delta(\bar{e}_\delta)\Big]d\theta
\end{aligned}$$

We know, however, that $\phi_\delta(\bar{e}_\delta) = \max_e (\bar{\alpha} - k)e - \gamma\frac{e^2}{2}$ and $q_0(\theta) = \arg\max_q (\theta - h(\theta))q - \frac{q^2}{2}$. Therefore, for $\alpha = \bar{\alpha}$, $\Delta\Pi < 0$.

It is not difficult to provide the analytical form of the $\Delta\Pi$. Following straightforward algebra, this is given by

$$\Delta\Pi = \frac{1}{12}\Big\{\frac{(\alpha-k)^3}{(\sqrt{1+\gamma}-1)^2} - \frac{3\gamma\delta^2\alpha^2 + (\gamma-1)\delta^3\alpha^3}{\gamma^2}\Big\} \tag{15}$$

;

therefore, the first derivative of $\Delta\Pi$ with respect to α is given by

$$\frac{d\Delta\Pi}{d\alpha} = \frac{1}{12}\Big\{\frac{3(\alpha-k)^2}{(\sqrt{1+\gamma}-1)^2} - \frac{6\gamma\delta^2\alpha + 3(\gamma-1)\delta^3\alpha^2}{\gamma^2}\Big\} \tag{16}$$

For future reference, we would like to compute the values of $\Delta\Pi$ and $\frac{d\Delta\Pi}{d\alpha}$ for $\alpha = k$. Although $\alpha = k$ is outside of the "relevant range" (recall that this is $\alpha \geq \frac{k}{1-\delta}$), examining the value of these expressions at $\alpha = k$ helps us study the behavior of $\Delta\Pi$ in the "relevant range" (i.e., $\alpha \geq \frac{k}{1-\delta}$). Consider, then, $\Delta\Pi$ and $\frac{d\Delta\Pi}{d\alpha}$ when $\alpha = k$. In either case, the first term is zero; therefore, the sign depends on the second term, that is, the signs of $3\gamma + (\gamma-1)\delta\alpha$ (for $\Delta\Pi$) and $2\gamma + (\gamma-1)\delta\alpha$ (for $\frac{d\Delta\Pi}{d\alpha}$). It follows that for $\gamma \geq 1$, both terms are positive. For $\gamma < 1$, note that because $\alpha < k + \gamma$ and $\delta < \frac{\gamma}{k+\gamma}$, it is true that $\alpha < \frac{\gamma}{\delta}$; therefore, $\alpha < \frac{2}{1-\gamma} \cdot \frac{\gamma}{\delta}$. This implies that for any γ, both $3\gamma + (\gamma-1)\delta\alpha$ and $2\gamma + (\gamma-1)\delta\alpha$ are positive; therefore, both $\Delta\Pi$ and $\frac{d\Delta\Pi}{d\alpha}$ are strictly negative for $\alpha = k$.

From Equation (16), it is clear that the second derivative of $\Delta\Pi$ is linear in α. Since, $\Delta\Pi$ is continuous, negative for $\alpha = k$ and positive for α large, there is an odd number of α values such that $\Delta\Pi = 0$. The linearity of the second derivative of $\Delta\Pi$ implies that we know that $\Delta\Pi$ will be either first convex and then concave or the reverse. However, this convexity changes only once. The slope is negative for $\alpha = k$, so if there were 3 solutions to $\Delta\Pi = 0$, the slope should be increasing (to reach a positive value for $\Delta\Pi$), then decreasing (to obtain negative value again) and finally increasing ($\Delta\Pi$ is positive for α large). This trend would require more than one change in the convexity/concavity of $\Delta\Pi$. As there is only one change, this is not possible. As the same argument can be made for any odd number of root above 3, this implies that there can be only one solution such that $\Delta\Pi = 0$. We know that the profit with the privacy option is always higher than without, so this solution $\tilde{\alpha}_F$ is such that $\tilde{\alpha}_F \in (\bar{\alpha}, \alpha_F)$.

Regarding the behavior of $\tilde{\alpha}_F$ with respect to δ, from the above reasoning, we know that the derivative of $\Delta\Pi$ w.r.t. α at $\alpha = \tilde{\alpha}_F$ is positive. Moreover, looking at Equation (15), it is easy to see that the derivative of $\Delta\Pi$ w.r.t. δ has the same sign as $-2\gamma - (\gamma-1)\delta\alpha$.

From the above, we know that it is negative. Therefore, it directly follows that $\tilde{\alpha}_F$ is increasing in δ.

Consumer Surplus.— Note first that the privacy option allows the firm to propose a higher usage schedule. This implies that consumers are better off when the option for privacy is allowed in the firm's right case. In what follows, we characterize more precisely the new cut-off value of α above which firm's right with the privacy option is better for consumers than consumer's right.

The consumer surplus under the two regimes is given respectively by

$$CS^P = \int_{\underline{\theta}_1}^{\hat{\theta}_1^P} q_1(\theta)(1 - F(\theta))d\theta + \int_{\hat{\theta}_1^P}^{1} q_0(\theta)(1 - F(\theta))d\theta \tag{17}$$

and

$$CS^C = \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} q_\delta(\theta)(1 - F(\theta))d\theta + \int_{\hat{\theta}_\delta}^{1} q_0(\theta)(1 - F(\theta))d\theta \tag{18}$$

To compare the consumer surplus, note first that, for $\alpha = \bar{\alpha}$, $\underline{\theta}_1 = \underline{\theta}_\delta$ and $q_1(\theta) = q_\delta(\theta)$. We also know that $\hat{\theta}_1^P > \hat{\theta}_\delta$ and that $q_0(\theta) > q_1(\theta)$ for any $\theta > \hat{\theta}_\delta$. Therefore, for $\alpha = \bar{\alpha}$, it is true that $CS^P - CS^C < 0$. Furthermore, we argue that the option of paying for privacy allows the firm to offer higher schedules compared to the case in which no such option is possible. This implies that $CS^P \geq CS^F$ for every α. We also argue in Proposition 2 that for $\alpha = k + \gamma$, $CS^F > CS^C$. Therefore, it is true that $CS^P - CS^C > 0$.

We now want to compute $\frac{d(CS^P - CS^C)}{d\alpha}$. Note first that $\underline{\theta}_1 \leq \underline{\theta}_\delta < \hat{\theta}_\delta < \hat{\theta}_1^P \leq 1$. Then, we have

$$CS^P - CS^C = \int_{\underline{\theta}_1}^{\underline{\theta}_\delta} q_1(\theta)(1-\theta)d\theta + \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \Big(q_1(\theta) - q_\delta(\theta)\Big)(1-\theta)d\theta + \int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \Big(q_1(\theta) - q_0(\theta)\Big)(1-\theta)d\theta.$$

Therefore,

$$\frac{d(CS^P - CS^C)}{d\alpha} = \frac{d\underline{\theta}_\delta}{d\alpha} q_1(\underline{\theta}_\delta)(1-\underline{\theta}_\delta) - \frac{d\underline{\theta}_1}{d\alpha} q_1(\underline{\theta}_1)(1-\underline{\theta}_1) + \int_{\underline{\theta}_1}^{\underline{\theta}_\delta} \frac{dq_\delta(\theta)}{d\alpha}(1-\theta)d\theta$$
$$+ \frac{d\hat{\theta}_\delta}{d\alpha}[q_1(\hat{\theta}_\delta) - q_\delta(\hat{\theta}_\delta)](1-\hat{\theta}_\delta) - \frac{d\underline{\theta}_\delta}{d\alpha}[q_1(\underline{\theta}_\delta) - q_\delta(\underline{\theta}_\delta)](1-\underline{\theta}_\delta)$$
$$+ \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{d(q_1(\theta) - q_\delta(\theta))}{d\alpha}(1-\theta)d\theta$$
$$+ \frac{d\hat{\theta}_1^P}{d\alpha}\Big[q_1(\hat{\theta}_1^P) - q_0(\hat{\theta}_1^P)\Big](1-\hat{\theta}_1^P) - \frac{d\hat{\theta}_\delta}{d\alpha}[q_1(\hat{\theta}_\delta) - q_0(\hat{\theta}_\delta)](1-\hat{\theta}_\delta)$$
$$+ \int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \frac{d(q_1(\theta) - q_0(\theta))}{d\alpha}(1-\theta)d\theta.$$

Using the facts that $q_1(\underline{\theta}_1) = q_\delta(\underline{\theta}_\delta) = 0$ and $q_\delta(\hat{\theta}_\delta) = q_0(\hat{\theta}_\delta)$, we obtain

$$\frac{d(CS^P - CS^C)}{d\alpha} = \int_{\underline{\theta}_1}^{\hat{\theta}_\delta} \frac{dq_1(\theta)}{d\alpha}(1-\theta)d\theta + \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{d(q_1(\theta) - q_\delta(\theta))}{d\alpha}(1-\theta)d\theta$$
$$+ \frac{d\hat{\theta}_1^P}{d\alpha}\Big[q_1(\hat{\theta}_1^P) - q_0(\hat{\theta}_1^P)\Big](1-\hat{\theta}_1^P) + \int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \frac{d(q_1(\theta) - q_0(\theta))}{d\alpha}(1-\theta)d\theta.$$

Moreover, using the expression of the usage schedule, we have $\frac{dq_1}{d\alpha}(\theta) = \frac{1}{1+\gamma}$, $\frac{d(q_1(\theta)-q_\delta(\theta))}{d\alpha} = \frac{1-\delta}{1+\gamma}$, $\frac{dq_0(\theta)}{d\alpha} = 0$, $\frac{d\hat{\theta}_1^P}{d\alpha} = \frac{1}{2\gamma}[1+\sqrt{1+\gamma}]$, and $q_1(\hat{\theta}_1^P) - q_0(\hat{\theta}_1^P) = -(\alpha-k)\frac{\sqrt{1+\gamma}}{1+\gamma}$. Therefore,

$$\frac{d(CS^P - CS^C)}{d\alpha} = \int_{\underline{\theta}_1}^{\hat{\theta}_\delta} \frac{(1-\theta)}{1+\gamma}d\theta + \int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{(1-\delta)}{1+\gamma}(1-\theta)d\theta$$
$$-(\alpha-k)\frac{\sqrt{1+\gamma}}{1+\gamma}\frac{[1+\sqrt{1+\gamma}]}{2\gamma}(1-\hat{\theta}_1^P) + \int_{\hat{\theta}_\delta}^{\hat{\theta}_1^P} \frac{1}{1+\gamma}(1-\theta)d\theta.$$

This leads to

$$\frac{d(CS^P - CS^C)}{d\alpha} = \frac{\alpha-k}{4\gamma}\Big[(1+\sqrt{1+\gamma})^2\frac{\alpha-k}{2\gamma} - \delta^2\Big(1 + \frac{\alpha-k}{2\gamma}\delta(\gamma-1)\Big)\Big]$$
$$= \frac{\alpha-k}{4\gamma}\Big[\frac{\alpha-k}{2\gamma}\Big((1+\sqrt{1+\gamma})^2 - \delta^3(\gamma-1)\Big) - \delta^2\Big].$$

Let $A = (1+\sqrt{1+\gamma})^2 - \delta^3(\gamma-1)$. Since $\delta \leq 1$, then $A \geq 0$ and $\frac{d(CS^P-CS^C)}{d\alpha} \geq 0$ iff $\alpha \geq k + \frac{2\delta^2}{A}\gamma$.

We know that $CS^P - CS^C < 0$ for $\alpha = \bar{\alpha}$, $CS^P - CS^C$ is continuous in α and $CS^P - CS^C > 0$ for $\alpha = k + \gamma$. We also know that $CS^P > CS^F$ for $\alpha < k + \sqrt{1+\gamma} - 1$ (in which case, $\hat{\theta}_1^P < 1$) and $CS^P = CS^F$ for $\alpha \geq k + \sqrt{1+\gamma} - 1$ (in which case, $\hat{\theta}_1^P = 1$). Therefore, two distinct cases exist: (i) if $\alpha_C \geq k + \sqrt{1+\gamma} - 1$, $CS^P \geq CS^C$ iff $\alpha \geq \alpha_C$; whereas (ii) if $\alpha_C < k + \sqrt{1+\gamma} - 1$, $CS^P \geq CS^C$ if $\alpha \geq \tilde{\alpha}$, where $\tilde{\alpha} < \alpha_C$. Therefore, in any case, there exists $\tilde{\alpha}_C \in (\alpha, \alpha_C]$ such that $CS^P \geq CS^C$ iff $\alpha \geq \tilde{\alpha}_C$.

Last, given that CS^P does not depend on δ,

$$\frac{d(CS^P - CS^C)}{d\delta} = -\frac{dCS^C}{d\delta} = -\int_{\underline{\theta}_\delta}^{\hat{\theta}_\delta} \frac{dq_\delta(\theta)}{d\delta}(1-\theta)d\theta \leq 0.$$

Therefore, $\tilde{\alpha}_C$ is increasing in δ.

PROOF OF (iii). When the value of data is low, introducing the option for privacy does not have any impact on the ranking of regimes. Indeed, if $\alpha < \bar{\alpha}$, then the consumers surplus is optimal under consumer's right. Instead, if $\alpha \geq \bar{\alpha} > k + \gamma$, the firm's right regime is optimal.

Q.E.D.

I. Proof of Proposition 4

Note first that the usage schedules, $q_\delta(\theta)$ and $q_0(\theta)$, are the solutions of the point-wise maximization of (12). Then, to derive the threshold $\hat{\theta}_M$, one should note the profit function taking P as given. Using, then, the fact that $q(\theta) = q_\delta(\theta)$ for $\theta \leq \hat{\theta}_M$ and $q(\theta) = q_0(\theta)$ for $\theta > \hat{\theta}_M$, the first-order condition to maximize the profit with respect to $\hat{\theta}_M$ leads to

$$\Big((2\hat{\theta}_M - 1)q_\delta(\hat{\theta}_M) - q_\delta^2(\hat{\theta}_M)/2 + P - C(q_\delta(\hat{\theta}_M))\Big) - \Big((2\hat{\theta}_M - 1)q_0(\hat{\theta}_M) - q_0^2(\hat{\theta}_M)/2\Big) = 0$$

Since $P = C(q_\delta(\hat{\theta}_M))$, we obtain at equilibrium

$$\Big((2\hat{\theta}_M - 1)q_\delta(\hat{\theta}_M) - q_\delta^2(\hat{\theta}_M)/2\Big) - \Big((2\hat{\theta}_M - 1)q_0(\hat{\theta}_M) - q_0^2(\hat{\theta}_M)/2\Big) = 0 \Rightarrow q_\delta(\hat{\theta}_M) = q_0^2(\hat{\theta}_M)$$

This condition is similar to the one characterizing $\hat{\theta}_\delta$, which here means that $\hat{\theta}_M = \hat{\theta}_\delta$. Since the usage schedules are the same in the regime with consumer's right and with a market with uniform price, the consumer surplus is the same in both regimes.

Q.E.D.

References

Acquisti, Alessandro, "Ubiquitous Computing, Customer Tracking, and Price Discrimination," in *Ubiquitous and Pervasive Commerce* (Springer, 2006).

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, "The Economics of Privacy," *Journal of Economic Literature*, 54 (2016), 442–92.

Acquisti, Alessandro, and Hal R Varian, "Conditioning prices on purchase history," *Marketing Science*, 24 (2005), 367–381.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press, 2018).

Akcura, Tolga, and Kannan Srinivasan, "Customer Intimacy and Cross-Selling Strategy," *Management Science*, 51 (2005), 1007–1012.

Amelio, Andrea, and Bruno Jullien, "Tying and Feebies in Two-Sided Markets," *International Journal of Industrial Organization*, 30 (2012), 436–446.

Armstrong, Mark, "Competition in Two-Sided Markets," *The RAND Journal of Economics*, 37 (2006), 668–691.

Berinato, Scott, "With big data comes big responsibility," *Harvard Business Review*, 92 (2014), 20.

Caillaud, Bernard, and Bruno Jullien, "Chicken & egg: Competition among intermediation service providers," *RAND journal of Economics*, 34 (2003), 309–328.

Casadesus-Masanell, Ramon, and Andres Hervas-Drane, "Competing with Privacy," *Management Science*, 61 (2015), 229–256.

Chakravorti, Bhaskar, "Why It's So Hard for Users to Control Their Data." *Harvard Business Review Digital Articles*, (2020), 2 – 6, available at: http://search.ebscohost.com.ezp.essec.fr/login.aspx?direct=true&db=bth&AN=141655838&site=bsi-live.

Coase, Ronald H, "The Problem of Social Cost," *Journal of Law and Economics*, 3 (1960), 1–44.

Dasgupta, Partha, Peter Hammond, and Eric Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *The Review of Economic Studies*, 46 (1979), 185–216.

Duch-Brown, Nestor, Bertin Martens, and Frank Mueller-Langer, "The Economics of Ownership, Access and Trade in Digital Data," *JRC Digital Economy Working Paper 2017-01*, (2017), available at: https://ssrn.com/abstract=2914144.

Fudenberg, Drew, and Jean Tirole, "Customer Poaching and Brand Switching," *RAND Journal of Economics*, 31 (2000), 634–657.

Fudenberg, Drew, and J Miguel Villas-Boas, "Price Discrimination in the Digital Economy," *The Oxford handbook of the digital economy*, (2012), 254.

Grossman, Sanford J, and Oliver D Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94 (1986), 691–719.

Hagel III, John, and Jeffrey F Rayport, "The coming battle for customer information," *The McKinsey Quarterly*, (1997), 64.

Hermalin, Benjamin E, and Michael L Katz, "Privacy, Property Rights and Efficiency: The Economics of Privacy as Secrecy," *Quantitative Marketing and Economics*, 4 (2006), 209–239.

Jones, Charles, and Christopher Tonetti, “Nonrivalry and the Economics of Data,” *Papers, Society for Economic Dynamics*, (2018).

Laffont, Jean-Jacques, and David Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton University Press, 2002).

Laudon, Kenneth C, “Markets and privacy,” *Communications of the ACM*, 39 (1996), 92–104.

Ritter, Jeffrey, and Anna Mayer, “Regulating Data as Property,” *Duke Law & Technology Review*, 16 (2017), 220–277.

Rochet, Jean-Charles, and Jean Tirole, “Platform Competition in Two-Sided Markets,” *Journal of the European Economic Association*, 1 (2003), 990–1029.

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