Competition and the Two Margins of Privacy

David Bounie, Antoine Dubus, Patrick Waelbroeck

Competition and the Two Margins of Privacy*

David Bounie,†Antoine Dubus‡ and Patrick Waelbroeck§

January 30, 2024

Abstract

We analyze the relationship between privacy protection and market competition. We consider a model where firms collect data to price discriminate consumers in a competitive product market, and we distinguish two margins of privacy. Firms strategically choose the number of consumers on whom they collect data – the extensive margin of privacy – as well as the precision of information – the intensive margin of privacy. We show that policymakers can efficiently protect both margins of privacy and consumer surplus by safeguarding the intensive margin. Indeed, restricting the amount of information that firms have on each consumer (the intensive margin) also induces firms to collect data on fewer consumers, thereby protecting the extensive margin of privacy. This softens the intensity of competition, but also reduces rent extraction by firms, and total consumer surplus increases.

*We thank for useful remarks and comments members of the French Data Protection Agency (CNIL), and participants to the EARIE conference. Patrick Waelbroeck thanks for insightful discussions members of the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris. This research has also been conducted within the Chair "Digital Finance" under the aegis of the Risk Foundation, a joint initiative by Groupement des Cartes Bancaires CB, Caisse des Dépôts, INSEE, Télécom Paris and University of Paris 2 Panthéon – Assas.

†CREST, Télécom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; david.bounie@telecom-paris.fr.

‡Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 – 8092, Switzerland; antoine1dubus@gmail.com.

§CREST, Télécom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; patrick.waelbroeck@telecom-paris.fr.

1 Introduction

Privacy protection and competition are two central considerations in today's digital economy. Data-driven services are developing rapidly, they have become integral to various digital markets and are essential for making informed decisions, enhancing user experiences, and driving innovation. However, they have also raised concerns about privacy and the need for effective data protection regulations.

Privacy regulations play a crucial role in ensuring that individual's personal information is safeguarded and that their privacy rights are respected. However, regulating privacy can also impact competition in many contrasting ways. It may first encourage competition by leveling the playing field between firms that offer different levels of privacy protection when some of them use the data collected to compete aggressively (Niebel, 2021), and competition authorities have often imposed mandated data sharing between competitors.[1] But it can also have unintended consequences such as decreasing competition in related technology markets by adding technical and administrative barriers to entry (Peukert et al., 2022). In addition, data can allow firms to price discriminate consumers, thereby reducing their surplus.

Privacy protection and competition are therefore intertwined in the digital economy, and striking a balance is crucial to ensure competition and the protection of individuals' privacy rights. Yet, up to now, both regulations have been applied separately. In the Facebook/Whatsapp merger, the European Commission expressly stated that "any privacy-related concerns flowing from the increased concentration of data (...) do not fall within the scope of the EU competition law rules but within the scope of the EU data protection rules". This stance was confirmed in early 2016 by Commissioner Vestager who declared in a speech that "privacy and competition concerns should be considered separately" (de Moncuit, 2018).

Recently, however, the need for co-operation and coordination across data protection agencies and competition authorities has been flagged by several reports. Experts argue that a strict separation between these two areas of law may result in

[1] See for instance the recently enacted Data Act in the European Union.

suboptimal outcomes for consumer welfare.[2] For instance, many academics have argued that the acquisition of WhatsApp by Facebook and of Fitbit by Google have granted Facebook and Google better information on their users (reducing their privacy), thereby consolidating their dominant positions in their respective markets (Valletti and Zenger, 2019; Bourreau et al., 2020).

The objective of this article is to provide a framework to analyze the relationship between privacy and competition, and to study how protecting consumer privacy can harm or benefit consumer surplus. We consider a market where firms collect data to increase their competitive power. Firms are located at the extremities of a Hotelling line, and they strategically choose the number of consumers on whom they collect information and the number of data points that they collect.

This framework captures two dimensions of consumer privacy. On the one hand, privacy depends on the amount of information a firm has on individual consumers – namely the intensive margin of privacy. But privacy can also depend on whether information allows firms to identify more consumers – the extensive margin of privacy. In this model, firms use data to better identify the willingness to pay of consumers for their products, and a firm that collects more information about consumers will be able to charge them personalized prices, reducing their surplus. Yet, using consumer data to charge targeted prices will also increase the competitive pressure on the market, which benefits consumers (Thisse and Vives, 1988).

Using this framework, we show that when fewer data are collected on individuals – a protection of the intensive margin – firms charge lower prices and extract less surplus from consumers. In reaction firms collect data on consumers located closer to their location to maximize rent extraction. Hence, protecting privacy at the intensive margin also preserves the extensive margin of privacy, and we show that total consumer surplus also increases in this case.

On the contrary, regulators willing to protect the extensive margin of privacy

[2]As stated by Lina Khan, the Chair of the Federal Trade Commission, there is a "growing recognition that persistent commercial data collection implicates competition as well as privacy"(Statement of Chair Lina M. Khan Regarding the Report to Congress on Privacy and Security, October 1, 2021).

have to balance increased privacy protection with a loss of consumer surplus. Indeed, when fewer consumers are identified by firms, the intensity of competition decreases and consumers pay higher prices. Hence, depending on the margin of privacy considered, the objective of a data protection agency may be aligned with or opposite to the mission of a competition authority.

This article contributes to the economic literature and policy debates on two main points. First, we introduce a methodological distinction between the two margins of privacy, which is new in the literature. To this aim, we propose a flexible model in which two competing firms endogenously choose the number of consumers on whom they collect data and the amount of information that they collect on each consumer. This model allows us to analyze the relation between each margin of privacy and consumer surplus. Consumer privacy has extensively been analyzed in previous research, but most articles focus on one margin of privacy, providing incomplete analysis of data protection regulations and their interactions with competition law. For instance, in Lee et al. (2011), Taylor and Wagman (2014), Belleflamme and Vergote (2016), Montes et al. (2019), and Braulin (2023), firms charge personalized prices to consumers. Firms compete more fiercely when they have access to consumer information, and protecting consumer privacy softens market competition and reduces consumer surplus. Liu and Serfes (2004) analyze a framework in which firms can acquire information with exogenous precision. They show that consumer surplus can increase or decrease depending on the precision of information. Building on their model, we endogenize the collection of data by firms and we explicitly distinguish between both margins of privacy to identify their impact on competition and consumer surplus. Secondly, we call for greater collaboration between data protection agencies and competition authorities, as regulating the two margins of privacy may have unintended consequences on consumer surplus and lead to inefficient regulations.

The remainder of the article is organized as follows. We describe the model in Section 2. We characterize the equilibrium in Section 3. Section 4 provides an analysis of the relation between consumer data collection, price discrimination and consumer surplus. Section 5 concludes.

2 Model

We consider a model of competition à la Hotelling in a product market.[3] Consumers are uniformly distributed on a unit line $[0, 1]$. They purchase one product from two competing firms that are located at the two extremities of the line.

Firms can collect information that partitions consumer demand into segments. They use this information to charge personalized prices to consumers on whom they have information, and set a homogeneous price for the remaining consumers.

2.1 Consumers

Consumers buy one product at a price p_1 from Firm 1 located at 0, or at a price p_2 from Firm 2 located at 1. Consumers located at $x \in [0, 1]$ receive a utility V from purchasing the product, but incur a cost $t > 0$ of consuming a product that does not perfectly fit their taste x. Therefore, buying from Firm 1 (resp. from Firm 2) incurs a cost tx (resp. $t(1-x)$). Consumers choose the product that gives the highest level of utility:[4]

$$u(x) = \begin{cases} V - p_1 - tx & \text{if buying from Firm 1,} \\ V - p_2 - t(1-x) & \text{if buying from Firm 2.} \end{cases}$$

2.2 Firms

Firms strategically collect data that segments a share of the consumer demand into k segments. This information allows them to price discriminate these consumers. We describe the data collection strategy of each firm and the structure of information, then we characterize the resulting pricing strategy of each firm.[5]

[3] This simple model of horizontal differentiation is commonly used in the literature to analyze the impact of information on consumer surplus and on the profits of the firms (Thisse and Vives, 1988; Elliott et al., 2022; Ali et al., 2023).

[4] We assume that the market is covered, so that all consumers buy at least one product from the firms. In terms of model primitives, this assumption requires that $V \geq 2t$.

[5] Note that in this model, consumers are passive and firms choose the amounts of data they collect according to a cost-benefit analysis. Hence, in the terms of Dosis and Sand-Zantman (2023), firms have ownership over the data generated by consumers. Other models consider optimal privacy regulations when consumers have control over their information and choose whether to disclose it (Zogheib, 2023).

Extensive margin of privacy.

Firm 1 and Firm 2 respectively choose the shares of consumers $[0, x_1]$ and $[1 - x_2, 1]$ on whom they collect data. The shares x_1 and x_2 of identified consumers correspond to the extensive margin of privacy. Firm θ ($\theta = 1, 2$) then charges personalized prices on consumers on whom it has collected information, and a homogeneous price on the share $1 - x_\theta$ of unidentified consumers.

Firm θ chooses x_θ according to two opposite effects of information on its profits. On the one hand, a larger share of identified consumers allows a firm to extract more surplus, which increases its profits. On the other hand, identifying consumers also increases competition because a firm has information on consumers that are closer to its competitor, and thus can lower prices for these consumers (Thisse and Vives, 1988). This competition effect lowers the profits of the firms. Hence, when choosing the share of identified consumers x_θ, a firm will balance the rent extraction and the competition effects of information.

Intensive margin of privacy.

Firm 1 and Firm 2 collect endogenous amounts of information on consumers on $[0, x_1]$ and $[1 - x_2, 1]$ respectively.[6] Data allow Firm θ to partition the demand of identified consumers into k_θ segments of size $\frac{x_\theta}{k_\theta}$.[7] This ratio corresponds to the precision of information collected by Firm θ which increases with k_θ for a given x_θ. Hence, the numbers of consumer segments k_1 and k_2 collected by each firm characterize the intensive margin of privacy.

A firm can then third-degree price discriminate consumers on whom it has information by charging different prices on different segments. For instance, when $k_1 = 2$, information is coarse, and Firm 1 can only distinguish whether consumers belong to $[0, \frac{x_1}{2}]$ or to $[\frac{x_1}{2}, x_1]$. At the other extreme, when k_1 converges to infinity, Firm 1 knows the exact location of each consumer on $[0, x_1]$, which corresponds to first-degree price discrimination.

[6] In practice, firms can collect such consumer information through various means, such as online cookies and pixels allowing them to observe the behavior of prospective buyers.

[7] Modeling information through a partition of the consumer demand was first introduced by Liu and Serfes (2004).

Collecting more data allows firms to locate consumers more precisely. Considering firms that set prices on different segments of the consumer demand contrasts with the existing literature that assumes that firms can perfectly identify consumers. This allows us to analyze the effect of more precise information (finer segments) on profits and consumer surplus.

The k_θ segments of size $\frac{x_\theta}{k_\theta}$ form a partition illustrated in Figure 1. Both firms compete on the same unit line, but for clarity we represent the partitions of Firm 1 and Firm 2 on two separate lines.

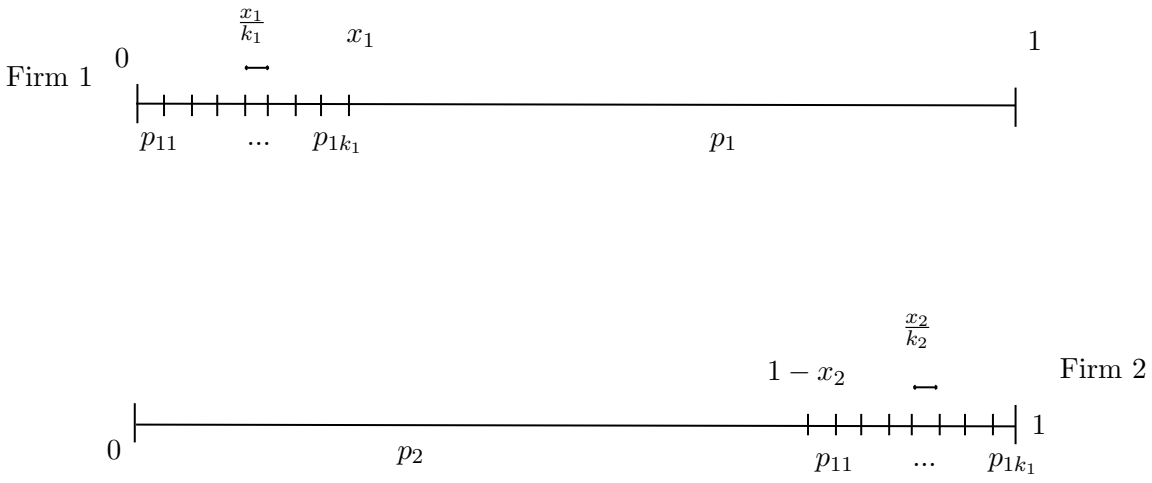

Figure 1: Data collection, Firms 1 and 2

Cost to collect consumer data.

Firms incur a cost to collect consumer data which increases with the number of identified consumers and with the number of segments collected. The cost of collecting information encompasses various dimensions of the activity of the firms, such as installing trackers and online cookies allowing it to observe the behavior of web users, or storing and handling data to eventually build detailed consumer scores (see Varian (2018) for a detailed discussion on the structure of the costs to collect data).

We denote this cost by $c(k_\theta, x_\theta)$, which we assume to be multiplicative and quadratic in k_θ and x_θ and equal to $c \cdot k_\theta^2 \cdot x_\theta^2, \quad c > 0$. A firm will thus balance the benefits from collecting more segments on a large share of consumers and the

cost to collect such data. Focusing on this functional form allows us to obtain closed-form solutions, but all our results hold with more general cost structures (see Appendix A.1).

Profits.

Firms choose the shares x_1 and x_2 of consumers on which they will collect data and charge personalized prices, as well as the number of segments k_1 and k_2 collected on these consumers. We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment that it price discriminates. $d_{\theta i}$ depends on the size of the segment which is defined by the number of data collected k_θ: for Firm 1, $d_{1i} = \frac{x_1}{k_1}$, and for Firm 2 $d_{2i} = \frac{x_2}{k_2}$. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta n}$:[8]

$$\pi_\theta(p_{\theta i}, p_\theta) = \sum_{i=1}^{k_\theta} d_{\theta i} p_{\theta i} + p_\theta d_\theta, \tag{1}$$

We assume that firms set prices in two stages. First, Firm 1 and Firm 2 simultaneously set homogeneous prices p_1 and p_2 on the whole unit line. Secondly, Firm θ sets a personalized price on each consumer segment on $[0, x_\theta]$, with $p_{\theta i}$ being the price on the ith segment from the origin. Then consumers observe prices. When setting the competitive price p_θ, Firm θ already knows which consumers it price discriminates, and thus charges p_θ accordingly.

Sequential pricing decision avoids the non-existence of Nash equilibrium in pure strategy, and allows an informed firm to charge consumers a higher price. This practice is common in the literature and is supported by managerial evidence. For instance, Acquisti and Varian (2005) use sequential pricing to analyze intertemporal price discrimination with incomplete information on consumer demand. Jentzsch et al. (2013), Lam et al. (2020) and Dubus (2021) also focus on sequential pricing where a higher personalized price is charged to identified consumers after a firm sets a uniform price.[9]

[8] In our simple setting, firms use consumer data to charge targeted prices. Hence, we abstract from other uses of data such as improved matching or product personalization.

[9] Sequential pricing is also common in business practices (see also Fudenberg and Villas-Boas (2006)). Recently, Amazon has been accused of showing higher prices for Amazon Prime subscribers, who pay an annual fee for unlimited shipping services, than for non-subscribers (Lawsuit alleges Amazon charges Prime members for "free" shipping, Consumer affairs, August

2.3 Timing.

The data collection strategies of the firms take place in two separate stages. Firms first choose the share of consumers on whom they collect information, and then collect an endogenous number of segments on these consumers. This two-stage approach allows us to distinguish the two strategic choices of firms that correspond to the two margins of privacy.

The timing of the game is the following:

- Stage 1: Firm θ chooses the share x_θ of consumers on whom they collect information.
- Stage 2: Firm θ collects k_θ segments of consumers on the share x_θ of identified consumers.
- Stage 3: Firm θ sets prices p_θ on the segment of unidentified consumers.
- Stage 4: Each firm charges personalized prices to identified consumers by setting $p_{\theta i}$ $(i \in [1, x_\theta(k_\theta)])$.

We analyze in Section 3 how a change in the intensity of competition impacts the data strategies of the firms, impacting in turn the two margins of privacy. We then analyze in Section 4 how a change in x_θ and k_θ – the two margins of privacy – impacts market competition and consumer surplus.

3 Equilibrium Analysis

We solve the game backward. We first characterize the equilibrium prices in stages 3 and 4 when firms price discriminate x_1 and x_2 consumers respectively. We then analyze the data strategies of each firm.

Prices and demand

At the price setting stage, Firm $\theta = 1, 2$ has collected data allowing to partition a share x_θ of close-by consumes in k_θ segments. Firm θ sets a price $p_{\theta i}$ for each

29 2017). Thus Amazon first sets a uniform price, and then increases prices for high-value consumers who are better identified when they join the Prime program.

segment of size $\frac{x_\theta}{k_\theta}$, and a unique price p_θ on the rest of the unit line. Firm θ sells to the x_θ consumers on whom it has collected data, and has demand on each segment equal to $d_{\theta i} = \frac{x_\theta}{k_\theta}$. The corresponding prices are computed using the indifferent consumer located on the extremity of the segment, $\frac{ix_\theta}{k_\theta}$. Hence, personalized prices satisfy the following relations in equilibrium:

$$p_{\theta i} = p_{-\theta} + t - 2t\frac{ix_\theta}{k_\theta}.$$

Let d_1 denote the demand for Firm 1 (resp. d_2 the demand for Firm 2) on the large segment $[x_1, 1]$ where firms compete (resp. on $[0, 1 - x_2]$). d_1 is determined by the location of the indifferent consumer: $\hat{x} = \frac{p_2 - p_1 + t}{2t}$ and $d_1 = \hat{x} - x_1 = \frac{p_2 - p_1 + t}{2t} - x_1$ (resp. $d_2 = \frac{p_1 - p_2 + t}{2t} - x_2$).

Profits of the firms.

Using these expressions for prices and demands, we can write the profits of the firms as:

$$\pi_\theta = \sum_{i=1}^{k_\theta} d_{\theta i} p_{\theta i} + d_\theta p_\theta = \sum_{i=1}^{k_\theta} \frac{x_\theta}{k_\theta}\left(p_{-\theta} + t - 2t\frac{ix_\theta}{k_\theta}\right) + \left(\frac{p_{-\theta} - p_\theta + t}{2t} - x\right) p_\theta - c\, k_\theta^2\, x_\theta^2.$$

These profits are equal to the sum of profits on each segment of size $\frac{x_\theta}{k_\theta}$ where Firm θ has collected information, and the competitive profits on the remaining segment where Firm θ charges a homogeneous price.

Prices and demands in equilibrium.

We characterize the equilibrium pricing decision of the firms by applying first-order conditions on π_θ with respect to p_θ. By doing so, we can write prices in equilibrium:

$$p_\theta = t[1 - \frac{2}{3}x_{-\theta} - \frac{4}{3}x_\theta], \quad p_{\theta i} = 2t - \frac{4}{3}x_{-\theta}t - \frac{2}{3}x_\theta t - 2\frac{itx_\theta}{k_\theta},$$

and we obtain the following demands in the competitive segment in equilibrium:

$$d_\theta = \frac{1}{2} - \frac{2}{3}x_\theta - \frac{1}{3}x_{-\theta}.$$

Plugging these expressions in the profits of the firms we obtain the profits of Firm θ as the difference of its revenues with data $r(x_\theta, k_\theta)$ and the cost to collect $c(x_\theta, k_\theta)$:

$$\begin{aligned}\pi^*_\theta(x_\theta, k_\theta) &= \sum_{i=1}^{k_\theta} \frac{x_\theta}{k_\theta}\left(p_{-\theta} + t - 2t\frac{i x_\theta}{k_\theta}\right) + \left(\frac{p_{-\theta} - p_\theta + t}{2t} - x_\theta\right) p_\theta - c\, k_\theta^2\, x_\theta^2, \\ &= \underbrace{\frac{t}{2} - \frac{7}{9}\, x_\theta^2\, t + \frac{2t}{9}\, x_{-\theta}^2 - \frac{4}{9}\, x_\theta\, x_{-\theta}\, t + \frac{2}{3}\, x_\theta\, t - \frac{2}{3}\, x_{-\theta}\, t - \frac{x_\theta^2 t}{k_\theta}}_{\text{Revenues with data } r(x_\theta, k_\theta)} - \underbrace{c\, k_\theta^2\, x_\theta^2}_{\substack{\text{Cost to collect} \\ \text{data } c(x_\theta, k_\theta)}}\end{aligned}$$

The profits of the firms depend on the amounts of data that they collect k_1 and k_2, and on the number of consumers on whom data is collected x_1 and x_2.

Number of segments collected

The number of segments collected by Firm θ is found by maximizing its profits with respect to k_θ. For Firm θ we have:

$$\frac{\partial \pi_\theta(x_\theta, k_\theta)}{\partial k_\theta} = \frac{x_\theta^2 t}{k_\theta^2} - 2\, c\, k_\theta\, x_\theta^2 \implies k^*_\theta = \left(\frac{t}{2c}\right)^{1/3}$$

The number of segments collected in equilibrium increases with the transport cost t. As t increases, the potential for rent extraction of the firms increases, and they benefit from collecting more segments. Moreover, as expected, k^*_θ decreases with the data collection cost. It is important to stress that k^*_θ does not depend on the value of x_θ when more consumers are identified (x_θ increases). The benefits from increased rent extraction perfectly match the loss resulting from an increase in the total data collection cost. This is a property of our cost function, and we show that our results are robust to a general specification of the data collection cost in Appendix A.1.

Number of consumers identified

Turning to the number of consumers identified in equilibrium, we solve for the optimal value of x^*_θ by maximizing profits with respect to x_θ.[10]

Simple comparative statics allow us to show that x^*_θ increases with k_θ. When a firm collects more segments, the marginal gain from collecting information on a

[10] Denoting $\alpha = 21k_{-\theta} + 27$, the numbers of consumers identified as functions of k_1 an k_2 are equal to: $x_\theta = \frac{(5k_{-\theta}+9)k_\theta t^2 + 9ck^3_{-\theta}k_\theta t}{((15k_{-\theta}+21)k_\theta+\alpha)t^2 + (c\alpha k^3_\theta + 21ck^3_{-\theta}k_\theta + 27ck^3_{-\theta})t + 27c^2k^3_{-\theta}k^3_\theta}$.

larger share of consumers increases and x_θ^* increases as well. In turn, this reduces the incentives of Firm $-\theta$ to identify consumers, and $x_{-\theta}^*$ decreases with k_θ.

Substituting the equilibrium number of segments collected in x_θ^* we obtain the following expression:

$$x_\theta^* = \frac{1}{3} - \frac{27c}{5 \cdot 2^{4/3} \cdot c^{1/3} \cdot t^{2/3} + 21 \cdot 2^{5/3} \cdot c^{2/3} \cdot t^{1/3} + 81 \cdot c}.$$

Hence, the equilibrium number of consumers identified by each firm decreases with the data collection cost c.

4 Privacy Protection and Consumer Surplus

We now analyze how the protection of each margin of privacy by a regulator impacts competition and consumer surplus $CS(x_1, x_2, k_1, k_2)$. A regulation that protects the intensive margin of privacy sets a limit $\overline{k}$ on the amount of data that firms can collect. Similarly, protecting the extensive margin of privacy sets a limit over the number of consumers $\overline{x}$ on whom a firm can collect data. To analyze the impacts of such regulations on consumer surplus, we perform comparative statics on consumer surplus with respect to x_θ and to k_θ.

4.1 Protecting the Intensive Margin of Privacy

A data protection agency can protect privacy at the intensive margin by setting a limit $\overline{k}$ over the number of segments that a firm can collect. This may also impact the incentives of firms to collect data on more consumers.

The impact of a limit on k_θ on the choice of x_θ depends on whether both variables are strategic complements or substitutes. Taking the cross derivative of the profit function we can write:

$$\frac{\partial^2 \pi_\theta(x_\theta, k_\theta)}{\partial k_\theta \partial x_\theta} = 2t\frac{x_\theta}{k_\theta^2} - 4ck_\theta x_\theta.$$

The cross derivative is positive for $\overline{k} < k_\theta^* = \left(\frac{t}{2c}\right)^{1/3}$. Thus, for any value of $\overline{k}$ that limits the number of segments that firms can collect, reducing the number of segments collected to $\overline{k}$ also lowers the value of x_θ.

Protecting the intensive margin of privacy has then two opposite effects on consumer surplus. On the one hand, firms collect coarser segments which limits their ability to extract rent from consumers. On the other hand, fewer consumers are identified resulting in lower intensity of competition between firms. We show that the first effect dominates the second, and protecting the intensive margin of privacy increases consumer surplus.

Proposition 1

When a regulator protects the intensive margin of privacy by reducing the number of consumer segments k_θ collected by Firm θ:

- *Firms have incentives to collect data on fewer consumers, protecting their privacy at the extensive margin.*
- *Total consumer surplus increases.*

Proof: see Appendix A.2.

This proposition highlights the efficiency of regulations aiming at limiting the number of segments collected by firms, as doing so also protects their privacy at the extensive margin, as well as aggregate consumer surplus.[11]

4.2 Protecting the Extensive Margin of Privacy

A data protection agency may also protect privacy at the extensive margin by limiting the ability of firms to identify more consumers. We study in this section how a reduction in the number of consumers identified by firms may also impact the number of segments they collect depending on the complementarity between x_θ and k_θ. When both variables are strategic substitutes, protecting the extensive margin of privacy exerts a negative impact on the intensive margin by increasing the incentives of firms to collect finer segments.

On the contrary, when x_θ and k_θ are strategic complements a reduction of the value of x_θ also reduces the incentives of the firms to collect data, protecting

[11] Note that this result stems from the complementarity of variables x_θ and k_θ, an effect that is robust to changing to cost function to an additive form such as: $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$ where α and β are positive.

privacy at the intensive margin. In our setting, the number of segments collected in equilibrium does not vary with x_θ, which is therefore the main relevant variable to analyze welfare.

Suppose initially that Firm 1 has collected information on x_1 consumers, and Firm 2 has information on x_2 consumers. If the share of consumers identified by Firm 1 is reduced by $\epsilon > 0$ so that $[x_1 - \epsilon, x_1]$ are unidentified, consumer surplus varies according to two effects:

1. A rent extraction effect: Firm 1 charges a uniform price to consumers on $[x_1 - \epsilon, x_1]$, which increases their surplus.

2. A competitive effect: Firm 1 increases its price on $[x_1 - \epsilon, 1]$, which reduces the competitive pressure on Firm 2. In turn, Firm 2 increases its price, which has a negative effect on the surplus of consumers over the whole unit line.

Overall, the second effect always dominates the first, and aggregate consumer surplus decreases when fewer consumers are identified. Indeed, the rent extraction effect only increases surplus on the ϵ consumers who are not identified anymore, while the competitive effect affects the whole Hotelling line. Therefore, the extensive margin of privacy is deeply related to the competitive impact of information, which implies a necessary trade-off between the protection of privacy and consumer surplus.

Moreover, when the share of identified consumers varies, we have shown that the number of segments collected remains equal to k^*_θ. Hence, segments are smaller when x_θ decreases, allowing Firm θ to better extract their surplus, which further reduces total consumer surplus.

Proposition 2

A regulation protecting the extensive margin of privacy by reducing the number of consumers x_θ on whom data is collected by Firm θ:

- *Reduces total consumer surplus.*

- *Increases the surplus of consumers on $[x_\theta - \epsilon, 1]$.*

Proof: See Appendix A.3.

Proposition 2 shows a general negative relation between privacy protection at the extensive margin and consumer surplus. When high-valuation consumers are identified, their surplus decreases. The intensity of competition increases on the rest of the line, which benefits low-valuation consumers. Hence, Proposition 2 highlights the difficulties to protect at the same time the extensive margin of privacy and consumer surplus.

5 Conclusion

We have analyzed a model where firms collect consumer data along two dimensions: they choose on how many consumers they collect information - the *extensive margin of privacy* - and the amount of information they collect on these consumers - the *intensive margin of privacy*. Using this framework, we have explored the interaction between privacy protection and market competition. The main finding of this article is to show that a regulation that restricts the number of consumer segments collected by firms leads them to collect data on fewer consumers, and results in an increase in total consumer surplus; protecting the intensive margin preserve also the extensive margin of privacy. On the contrary, a regulation that exclusively safeguards the extensive margin of privacy by reducing the number of consumers from whom data is collected is detrimental to total consumer surplus.

These findings are particularly worth stressing in light of recent regulations that target the way firms collect and use consumer data, and how they impact competition in product markets. Competition authorities have indeed recently taken measures to intensify competition by increasing the amount of personal data accessible to firms.[12] For instance, the Digital Market Act has enacted a data access right under which large digital platforms must grant their competitors access to their data. Our results support the idea that such a policy has a positive impact on the intensity of product-market competition, and increases the surplus of consumers. However, because more firms have access to consumer data, such

[12]The Digital Market Act is the EU law to make digital markets fairer and more contestable. The DMA is supported by a theoretical literature arguing that mandatory data sharing can foster competition between firms (Prüfer and Schottmüller, 2021; Krämer and Schnurr, 2022).

a regulation reduces consumer privacy at the extensive margin, and the benefits derived from consumer surplus should be cautiously balanced with the loss of privacy.

Finally, in recent years, data protection agencies have also adopted measures to limit the amount of consumer data collected by firms, thereby protecting privacy at the intensive margin. Such a data minimization principle is enacted for instance in the California Consumer Privacy Act and the Health Insurance Portability and Accountability Act in the US, or in the European GDPR.[13] Our results suggest that limiting the number of consumer data that a firm can collect may give firms incentives to collect data on fewer consumers. Hence, protecting the intensive margin may also protect privacy at the extensive margin. Our analysis has primarily focused on data for price discrimination within a model of horizontal differentiation. However, data serves various other purposes, including online search and targeted advertising, as analyzed by De Corniere and De Nijs (2016) and Bird and Neeman (2023). The impact of privacy protection may differ depending on the specific use of data. Further investigation is required to examine how the methodological distinction we have introduced between the intensive and extensive margins of privacy applies to these alternative contexts.

References

Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science 24*(3), 367–381.

Ali, S. N., G. Lewis, and S. Vasserman (2023). Voluntary disclosure and personalized pricing. *The Review of Economic Studies 90*(2), 538–571.

Belleflamme, P. and W. Vergote (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters 149*, 141–144.

Bird, D. and Z. Neeman (2023). The effect of privacy on market structure and prices. *The Journal of Law, Economics, and Organization*, ewad031.

Bourreau, M., C. Caffarra, Z. Chen, C. Choe, G. S. Crawford, T. Duso, C. Genakos, P. Heidhues, M. Peitz, T. Rønde, et al. (2020). *Google/Fitbit will monetise health data and harm consumers.* Centre for Economic Policy Research.

[13]The data minimization principle is expressed in Article 5(1)(c) of the GDPR and Article 4(1)(c) of Regulation (EU) 2018/1725, which provide that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".

Braulin, F. C. (2023). The effects of personal information on competition: Consumer privacy and partial price discrimination. *International Journal of Industrial Organization 87*, 102923.

De Corniere, A. and R. De Nijs (2016). Online advertising and privacy. *The RAND Journal of Economics 47*(1), 48–72.

de Moncuit, A. (2018). In which ways should privacy concerns serve as an element of the competition assessment. Technical report, Speech at the European Commission.

Dosis, A. and W. Sand-Zantman (2023). The ownership of data. *The Journal of Law, Economics, and Organization 39*(3), 615–641.

Dubus, A. (2021). Behavior-based price discrimination with strategic customer targeting.

Elliott, M., A. Galeotti, A. Koh, and W. Li (2022). Market segmentation through information.

Fudenberg, D. and J. M. Villas-Boas (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems 1*, 377–436.

Jentzsch, N., G. Sapi, and I. Suleymanova (2013). Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization 31*(2), 131–144.

Krämer, J. and D. Schnurr (2022). Big data and digital markets contestability: Theory of harm and data access remedies. *Journal of Competition Law & Economics 18*(2), 255–322.

Lam, W., P. Belleflamme, and W. Vergote (2020). Competitive imperfect price discrimination and market power. *Marketing Science*.

Lee, D.-J., J.-H. Ahn, and Y. Bang (2011). Managing consumer privacy concerns in personalization: a strategic analysis of privacy protection. *Mis Quarterly*, 423–444.

Liu, Q. and K. Serfes (2004). Quality of information and oligopolistic price discrimination. *Journal of Economics & Management Strategy 13*(4), 671–702.

Montes, R., W. Sand-Zantman, and T. Valletti (2019). The value of personal information in online markets with endogenous privacy. *Management Science 65*(3), 1342–1362.

Niebel, C. (2021). The impact of the general data protection regulation on innovation and the global political economy. *Computer Law & Security Review 40*, 105523.

Peukert, C., S. Bechtold, M. Batikas, and K. Tobias (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science 2022*.

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

Taylor, C. and L. Wagman (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization 34*, 80–84.

Thisse, J.-F. and X. Vives (1988). On the strategic choice of spatial price policy. *The American Economic Review*, 122–137.

Valletti, T. M. and H. Zenger (2019). Increasing market power and merger control. *Competition Law & Policy Debate 5*(1), 26–35.

Varian, H. (2018). Artificial intelligence, economics, and industrial organization. Technical report, National Bureau of Economic Research.

Zogheib, J.-M. (2023). Competition, privacy, and multi-homing.

A Appendix

A.1 General Data Collection Costs

We discuss how the structure of the data collection cost impacts the relation between the number of segments collected by a firm and the share of consumers it identifies. We consider two general classes of cost functions that characterize general properties of data such as increasing and decreasing returns to scale, and multiplicative and additive costs with respect to k_θ and x_θ.

We show how different cost structures impact our results. We can write the profit function of a firm in the following way:

$$\pi_\theta(k_\theta, x_\theta) = f(x_\theta) - \frac{x_\theta^2 t}{k_\theta} - c(k_\theta, x_\theta).$$

Hence the first degree derivative of the profits with respect to k_θ can be written as:[14]

$$\frac{\partial \pi_\theta(k_\theta, x_\theta)}{\partial k_\theta} = \frac{x_\theta^2}{k_\theta^2} - \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

FOCs imply that

$$x_\theta^2 = k_\theta^2 \cdot \frac{\partial c(k_\theta, x_\theta)}{\partial k_\theta}.$$

Multiplicative cost functions.

We first consider the class of multiplicative cost functions $c(k_\theta, x_\theta) = c \cdot k_\theta^\alpha \cdot x_\theta^\beta$.[15] Coefficient c corresponds to a productivity factor resulting for instance from

[14] It is clear that FOCs can be applied as long as the cost function is convex w.r.t k_θ or concave and such that $\frac{\partial^2 c(k_\theta, x_\theta)}{\partial k_\theta^2} > -2\frac{x_\theta^2}{k_\theta^3}$ with x_θ constant.

[15] In this case, the second order condition can be written $c \cdot k_\theta^{\alpha+1} \cdot \alpha(\alpha - 1) > -2x_\theta^{2-\beta}$. This condition is clearly satisfied for $\alpha \geq 1$.

different technologies used by an industry. We consider $\alpha \geq 1$ so that second order conditions hold. Such a specification characterizes both increasing and constant returns to scale w.r.t. k_θ, and increasing and decreasing returns to scale w.r.t. x_θ, depending on the values of $\alpha > 1$ and $\beta > 0$. Considering the FOCs, we have:

$$x_\theta^{2-\beta} = c \cdot \alpha \cdot k_\theta^{\alpha+1}$$
$$k_\theta^* = (c \cdot \alpha)^{\frac{-1}{\alpha+1}} \cdot x_\theta^{\frac{2-\beta}{\alpha+1}}$$

Hence, k_θ^* increases with x_θ when $\beta < 2$ and k_θ^* decreases with x_θ when $\beta > 2$. In the baseline model where $c(.)$ is proportional to x_θ^2, $2 - \beta$ simplifies and k_θ^* does not depend on x_θ. Then, $\beta = 2$ is the threshold value above which k_θ and x_θ are strategic substitutes and below which they are strategic complements.

Additive cost functions.

We consider now the class of costs functions where the costs associated with x_θ and k_θ are additive: $c(k_\theta, x_\theta) = c \cdot (k_\theta^\alpha + x_\theta^\beta)$.[16] In this case the FOC w.r.t. k_θ implies:

$$x_\theta^2 = c \cdot \alpha \cdot k_\theta^{\alpha+1}$$
$$k_\theta^* = c^{\frac{-1}{\alpha+1}} \cdot x_\theta^{\frac{2}{\alpha+1}}$$

And k_θ always increases with x_θ.

Market outcomes with general cost functions.

The main results of the analysis hold as consumer surplus increases with higher level of privacy at the intensive margin and decreases with a higher level of privacy at the extensive margin.

Moreover, we identify another effect when k_θ and x_θ are strategic complements. When far away consumers are identified by firms, the profitability of data decreases, which lowers rent extraction and increases consumer surplus. Overall the two effects go in opposite directions, and their magnitude depends on the structure of the data collection cost.

A.2 Proof of Proposition 1

We analyze the variation of consumer surplus with the number of segments collected by a firm.

[16] In this case, the second order condition can be written $c \cdot k_\theta^{\alpha+1} \cdot \alpha(\alpha - 1) > -2x_\theta^2$. This condition is clearly satisfied for $\alpha \geq 1$.

Consumer surplus

Consumer surplus when Firm 1 identifies x_1 consumers with k_1 segments and Firm 2 identifies x_2 consumers with k_2 segments is defined as follows.

$$\begin{aligned} CS(x_1,x_2,k_1,k_2) = &\sum_{i=1}^{k_1}[\int_0^{\frac{x_1}{k_1}} V - 2t[1-\frac{1}{3}x_1-\frac{2}{3}x_2-\frac{ix_1}{k_1}]-tx\mathrm{d}x] \\ &+\int_{x_1}^{\frac{1}{2}+\frac{x_1}{3}-\frac{x_2}{3}} V - t[1-\frac{4}{3}x_1-\frac{2}{3}x_2]-tx\mathrm{d}x \\ &+\int_{\frac{1}{2}+\frac{x_1}{3}-\frac{x_2}{3}}^{1-x_2} V - t[1-\frac{2}{3}x_1-\frac{4}{3}x_2]-tx\mathrm{d}x \\ &+\sum_{i=1}^{k_2}[\int_0^{\frac{x_2}{k_2}} V - 2t[1-\frac{1}{3}x_2-\frac{2}{3}x_1-\frac{ix_2}{k_2}]-tx\mathrm{d}x] \\ &= V + t[-\frac{5}{4}+\frac{17}{18}x_1^2+\frac{17}{18}x_2^2+x_1x_2]+\frac{1}{2}\frac{x_1^2t}{k_1^2}+\frac{1}{2}\frac{x_2^2t}{k_2^2} \end{aligned} \tag{2}$$

The number of consumers identified as a function of the number of segments collected is equal to:

$$x_\theta = \frac{(5k_{-\theta}+9)k_\theta t^2 + 9ck_{-\theta}^3k_\theta t}{((15k_{-\theta}+21)k_\theta+\alpha)t^2+(c\alpha k_\theta^3+21ck_{-\theta}^3k_\theta+27ck_{-\theta}^3)t+27c^2k_{-\theta}^3k_\theta^3}$$

Substituting these values in $CS(x_1,x_2,k_1,k_2)$, we can show that surplus decreases with the number of segments collected by a firm.

A.3 Proof of Proposition 2

We have seen that the number of segments collected in equilibrium does not depend on the number of consumers identified, and the derivative of $CS(x_1,x_2,k_1,k_2)$ with respect to x_1 is equal to

$$\frac{\partial CS}{\partial x_1} = \frac{17tx_1}{9} + tx_2 + \frac{x_1t}{k_1^2},$$

which is always greater than zero.

A direct comparison of prices shows that consumers in segments $\frac{ix_\theta}{k_\theta} \leq \frac{1}{2} - \frac{2x_{-\theta}}{3} - \frac{x_\theta}{3}$ are charged a higher price when they are identified than without information.