Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Economics

Ciclo 32

Settore Concorsuale: 13/A1 - Economia Politica
Settore Scientifico Disciplinare: SECS-P/01 - Economia Politica

Data, Competition, and Consumer Privacy in Digital Markets

Presentata da: Francesco Clavorà Braulin

Coordinatore Dottorato:
Prof.ssa Maria Bigoni

Supervisore:
Prof. Emilio Calvano

Esame finale anno 2020

Abstract

In digital markets personal information is pervasively collected by firms. In my thesis I examine the role of consumer data in these markets, the strategies adopted by firms when their actions are conditioned on this data and the implications for consumer privacy of such practices. In the first chapter I study data ownership and product customization in the light of the competition policy issue of exclusive access to non rival but excludable data about consumer preferences. I show that an incumbent firm does not have an incentive to sell an exclusively held dataset with a rival firm, but instead it has an incentive to trade a customizing technology with the other firm. In the second chapter I investigate the effects of consumer information on the intensity of competition. In a two dimensional duopoly model of horizontal product differentiation, firms can use information on consumer preferences to practice price discrimination. I contrast a full privacy and a no privacy benchmark with an information regime in which firms are able to target consumers only partially. When data is partially informative, firms are always better-off with price discrimination and an exclusive access to user data is not necessarily a competition policy concern. From a consumer protection perspective, the policy recommendation is that the regulator should promote either no privacy at all or full privacy. However, the effects of information on privacy at the individual level are ambiguous: some consumers are made better-off while other consumers are worse-off, and even under individual targeting, there are winners and losers. The ambiguous impact on consumer surplus suggests the design of a more nuanced approach to individual consumer privacy. In the third chapter I introduce an upstream data broker that perfectly observes either only one or both dimensions of consumer information and sells this data to downstream competing firms for price discrimination purposes. When the seller exogenously holds a partially informative dataset, an exclusive allocation arises. Instead, when the dataset held is fully informative, the data broker trades information non exclusively but each competitor acquires consumer data on a different dimension. Finally, I show that when data collection is made endogenous non exclusivity is robust provided that collection costs are not too high. Otherwise, only partial information is collected in equilibrium, which triggers an exclusive allocation of data among competing firms. Therefore, the competition policy suggestion is that exclusivity should not be banned *per se*, but it is rather data differentiation in equilibrium that rises market power in competitive markets. Upstream competition is sufficient to ensure that both firms get access to consumer information, even though in equilibrium they are informed about different attributes which is detrimental to consumers.

Contents

List of Figures

List of Tables

Introduction

Information has a crucial role in economics. The consequences on the society and the economy of the advent of digital markets, which provide a limitless offer of services and products to final users and customers, have prompted a hot policy debate about a particular type of information: data on consumer preferences or willingness to pay of customers or, more broadly, personal information. Firms can successfully profile final consumers and sort them in more or less precise categories. Moreover, individual profiling is increasingly feasible thanks to the pervasive collection of fine data on consumers. Exploitation of this huge amount of information allows firms to target advertising, to customize products and personalize recommendations, to strategically set dynamic prices and ideally even to price discriminate.

The pervasive availability of these enormous amounts of data, which is one (if not the pivotal one) of the pillars sustaining the extremely rapid growth of the digital economy, has clearly a crucial impact on the functioning of markets, especially the online ones. The strategic interactions between platforms, data collectors, firms and consumers pose new regulatory issues and competition policy concerns yet to be fully understood, provided that the applications based on data collection and exploitation are countless. The main concern is whether all these practices are pro-competitive or anti-competitive and what are the implications in terms of consumer protection for the privacy of online users, which supply their data, the fundamental input in digital markets. Clearly, data is not only directly supplied by consumers but it is also indirectly generated by the interaction between users and firms. The understanding of the impact of these practices on market competition, the potential market failures generated by firms that strategically collect and exclusively hold data and the aforementioned consumer protection concerns, of which privacy is just one of the many dimensions involved, is of great relevance. Ultimately, digitization has not only a direct impact on competition and the players active in digital markets, but it shapes more broadly many aspects of our modern societies.

The economic literature on digital markets and, more broadly, on digitization is growing fast, and this thesis makes a contribution to the stream of this literature which is more focused on the aspects of data and competition. In particular, my thesis is mainly concerned with data that can be related to a specific data subject, which is the person that can be identified by means of such data. However, in the first chapter I start with a wider approach. The initial chapter focuses on "de facto" data ownership: there are many instances in which firms treat data as proprietary, whereas consumers completely lose control over their personal information. A scenario in which a single firm holds enormous amounts of data about consumers clearly grants to the data holder a non trivial competitive advantage. These situations of "de facto" data ownership could create barriers to entry and, if the firm is in a dominant position, the market could eventually tip. Data is modeled both as a vertical shifter of the product or service value and as a personalized add-on which is customized for each consumer. With the help of a theoretical model I investigate the incentives of a data owner that has exclusive access to this information to share or trade the informational asset with other firms in the market. Data portability is one of the remedies that can limit the detrimental effects in a data-driven market with a dominant data owner.

The second chapter investigates the effects of information on competition. The primary research question is about the relationship between information and industry profits, in a scenario in which firms can use increasingly precise data about consumers' willingness to pay to practice price discrimination. Personalized prices will range between group price discrimination and individual price targeting. As a by-product of this analysis I will also investigate the effects of enforcing more or less privacy in the market on consumer surplus, in order to dig deeper into the consumer privacy consequences of price personalization. Price discrimination seems not to be so ubiquitous in digital markets, at least for the moment. This observation does not mean that it is not scientifically relevant to assess its consequences on market outcomes, provided that, despite the limited empirical evidence of such practice, price discrimination designed by autonomous agents is increasingly feasible and it is a theoretically interesting topic. From a consumer protection point of view but also with respect to competition policy it is crucial to understand whether personalization of prices can be beneficial for consumers and under which conditions, or if it is simply a profit maximizing strategy that uniquely benefit the firms. In the second chapter a hump-shaped relationship between information and profits is characterized: firms are collectively better-off with access to data that is only partially informative whereas each player would individually benefit from having exclusive access to maximally precise data about consumers' valuations. However, when all firms set tailored prices at the individual level, industry profit falls below the aggregate payoff that firms get with uniform pricing. Consumer surplus follows the inverse pattern and the model delivers the familiar intuition that, from a consumer protection point of view, no privacy should be promoted whereas *less* privacy - with respect to the baseline no information scenario - is bad for customers. In other words, if there is no privacy regulation in the market, then starting to regulate it is detrimental to consumers; however, if there is some form of imperfect privacy regulation and firms are able to get partial information on consumers, then a full privacy regulation should be welcomed. However, an important caveat also emerges: even though it is confirmed that, once we allow firms to collect and exploit some data, the best policy is to disclose information completely and symmetrically to firms, there are always winners and losers among consumers in all the types of games analyzed. As a result, no privacy could be the worst case for types that hold very polarized valuations for the products offered by the competing firms.

In the third chapter the richness of the setup proposed in the second part of this thesis is exploited to investigate the flows of information between firms. The data market is populated by many collectors, brokers and other players that gather multidimensional data on consumers: user profiles consisting of many potentially relevant attributes are created and sold to marketers and retailers. An initial discussion of data sharing between competing firms is developed, but the main concern is about the sale of information by an external seller. The goal is to shed light on the incentives of a monopolistic data broker to trade data. In other words, I investigate whether an exclusive or a non exclusive data allocation among competing firms will be induced in the subgame perfect Nash equilibrium of the game. From a competition policy point of view, this research question is of primary importance: exclusivity is often associated with a damage for the rival (uninformed) firms and the final consumers. It is therefore crucial to understand what type of data triggers exclusivity and whether this outcome is indeed bad for the market participants. I will show instead that, under some conditions, exclusivity not only is not detrimental to other players but it could be even more welcomed than a scenario in which all firms have *some* data about consumers which is only partially informative. Therefore, exclusivity is not *per se* bad and it should not be banned *a priori*. Rather, the market power of firms in competitive settings increases when there is data differentiation in equilibrium. The analysis is then extended to account for costly investment in data collection and, finally, also to competing data brokers, in order to check whether competition upstream could be the solution to some of the concerns that emerge in the case of a monopolistic data broker. Each chapter will include a specific introduction, a presentation of the close literature and a discussion of the main contributions, along with some policy illustrations.

Chapter 1

Data Ownership

The digitization of the economy has led to an accumulation of informational assets in the hands of a plethora of firms, intermediaries, organizations and institutions. However in some cases excludable assets carrying information on different data subjects and stakeholders are controlled and managed as proprietary, a quite well established practice that rises policy concerns. Clearly, when it comes to digital markets, it is evident that exploitation of data and algorithms by on-line firms generates an added value in the form of customized services and products. However, when exclusively held, more data and/or better algorithms are likely to grant a competitive advantage over rival firms and could also lead to an increase in market power. Rivals' foreclosure from the access to valuable information or leveraging data to create entry barriers are rising concerns in digital markets (Duch-Brown et al., 2017). Therefore it is not surprising that data ownership is one the topics around which the policy debate on data revolves (Drexl et al., 2016). In particular, the term "de facto" data ownership was devised, meaning that on-line firms, once they collect consumer information, consider data as a proprietary business asset. Facebook conducted "surveys of the usage of mobile apps by customers, and apparently without their knowledge. This knowledge helped them to decide which companies to acquire, and which to treat as a threat"[1]. Moreover, Facebook adopted an aggressive strategy towards some apps, "with the consequence that denying access to data led to failure of that business"[2]. Thus, who owns or should own user generated data?

I consider a firm that initially has a competitive advantage in digital markets and can trade an "informational asset" with competitors. Provided that this advan-

[1] `https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Note-by-Chair-and-selected-documents-ordered-from-Six4Three.pdf`

[2] *Ibid*, at 1.

tage is modelled as a non replicable asset, I contrast two alternative scenarios in which an informational asset can generate an added value: (i) exclusive access to consumer data or (ii) a proprietary algorithm that is not decodable or observable by competitors[3]. I investigate the incentives of the advantaged firm (for instance, an incumbent) to sell its asset to a horizontally differentiated competing firm (entrant). In particular, the first part of the chapter models data-driven customization as a uniform vertical shifter for all consumers whereas the second part investigates customization in the form of personalized product design. In the first case more data allow firms to deliver a better quality for *all* users, such as an improved user interface, while in the second case *each* user receives a personalized better recommendation by a sophisticated algorithm.

The goal of this work is to analyze the impact of exclusive data or algorithms on competition in digital markets. I derive industry profits, consumer surplus and overall welfare in the two scenarios, in order to make a welfare assessment of the characterized equilibria and shed light on the optimal strategy of the dominant digital firm.

Access to data is a hot topic in the policy debate on Big Data and digital platforms, and also the competition law literature has widely addressed this concern (Tucker and Wellford, 2014; Graef et al., 2015; Ezrachi and Stucke, 2016; Stucke and Grunes, 2016). The paper is related to the economic literature on firms' incentives to sell content or innovations to competing firms, competitive product customization in horizontally differentiated markets and data sales in digital environments. Relatedly to the way in which data is sold - the valuable input which resembles a premium content in other contexts - profound similarities can be found with works on competition in Pay-TV markets: Armstrong (1999) shows that, in a fixed-fee environment, exclusive provision of a premium content is preferred by an upstream seller, whereas by means of a per-subscriber fee, even though the seller would be better-off by supplying exclusively on of the buyers, the best he can achieve, relying on a credible procedure, is to supply both downstream distributors so that non-exclusive provision emerges; Weeds (2016) shows that, in a static setting in which a vertically integrated producer of a premium content directly competes in the downstream market, per-subscriber fees sustain non-exclusivity, but she also shows that the incumbent rights' holder prefers exclusivity in a dynamic model; Harbord and Ottaviani (2002), allowing for re-selling in the downstream market, find that (i) a

[3]More generally, it would be possible and fruitful to merge the two cases in a unique theoretical setup.

directly competing firm that has acquired the premium content does not resell it for a fixed payment but trades it for a per-subscriber resale charge, whereas (ii) the upstream rights' holder always has an incentive to sell the asset exclusively for a fixed payment no matter how subsequently the reselling takes place.

The literature on licensing of a cost-reducing innovation is close to this work as well, at least to the extent that we can find similarities with our quality-enhancing input. Useful references are Kamien and Tauman (1986), Katz and Shapiro (1985) and Sen and Tauman (2007), in particular for the cases in which the innovator is an insider of the industry: exclusivity can emerge in case of licensing by means of a fixed fee, whereas licensing to rivals is dominant in case of royalties.

Finally, interesting insights about product matching and consumer preferences can be found in works on matching information, one-to-one marketing and product design ("product fit", as in Wattal et al. (2009), where personalization turns out to be profitable for both firms when basic products are relatively similar ex-ante; instead, when products are enough differentiated, personalization is not necessarily profit enhancing). These works pertain to a management and marketing literature. "One-to-one" marketing is defined as the practice of using information technologies to treat customers on an individual basis by tailoring individually products, services (core) and other interactions (auxiliary) at the customer level. One of the several applications we can find in digital markets is the system of personalized recommendations provided by several platforms, such as music or video streaming services. Syam and Kumar (2006) define customization as the firms giving to consumers the possibility to influence the production process of a certain product, that consequently will be individually unique. They assume that even if products are customized, a uniform price is charged for them. Mendelson and Parlaktürk (2008) consider not only mass customization but also the possibility to practice differential pricing, and contrast this case to a setting in which only uniform prices are allowed. Here the idea is to consider as well this "product-uniqueness" granted by customization to each customer, but we are interested mainly in the role of the informational input that firms exploit to perform such customization: namely, in this model, firms use personal data exogenously collected to match their products to individuals' tastes. A well-known idea is that consumers must provide their personal information to the firm in the process of designing the customized product, or alternatively the firm itself must acquire this data somewhere in order to be able to offer tailored products.

The contribution of this chapter is to show that a data advantage, or more generally a form of digital competitive advantage, self-reinforces and leads to an increase in dominance, which is measured in this model by the dynamic of the market shares. The result is similar in spirit to the model proposed by Prufer and Schottmüller (2017), but the mechanism is different given that their consider "data-driven" indirect networks effects and in their paper there is not the possibility to trade data or foreclose access to data.

The chapter is organized as follows. In Section 1.2, I present a model with exclusivity in data whereas, in Section 1.3, I study a model with a proprietary customizing technology held by the incumbent firm.

1.1 The general model

Consider two providers $i = 1, 2$, located at the extremes of a unitary line, offering an horizontally differentiated service to users at a subscription price p_i. Marginal costs are normalized to zero. There is a unitary mass of heterogeneous consumers with preferences identified by $x \in [0, 1]$ and with a reservation value for the basic version of the service equal to v. On top of the standard version of the service, firms can offer a *data-driven* incremental utility, which comes from data exploitation or customization of the service based on individual preferences. In general, we consider a situation in which firms can offer an additional feature which improves users' experience and, in doing so, we model firm 1 as being able to always deliver a greater incremental utility to each consumer in the market. This asymmetry may reflect scenarios in which firm 1 is, for instance, an incumbent in the market and firm 2 is an entrant. The general form of the net utility of a consumer x accessing the service offered by firm i located in x_i at a price p_i therefore is equal to

$$u_i = v - t|x - x_i| - p_i + \Delta u_i \tag{1.1}$$

where $t > 0$ is the transportation cost and x_i is either zero or one, respectively for firm 1 and firm 2. Equation (1.1) represents the standard utility à la Hotelling plus an incremental term. Following standard procedures we derive the general expression for the indifferent consumer which is

$$x(p_1, p_2) = \frac{t + p_2 - p_1 + \Delta}{2t}$$

where $\Delta = \Delta u_1 - \Delta u_2$ measures the competitive advantage given by the digital asset exclusively held by the dominant firm[4] with $\Delta u_1 > 0$ and $\Delta u_2 \geq 0$. The incumbent may decide whether to sell or not this exclusive asset to the competitor. In other words, we investigate if a dominant firm has an incentive to leverage its exclusive access to consumer data and under which conditions. The timing of the game is sequential:

1. The seller makes a take-it-or-leave-it offer for the asset to the rival;
2. Firms set prices $p_i \in \mathbb{R}_+$;
3. Consumers decide which provider to get the service from.

In this chapter, we assume that price discrimination based on consumer preferences x is not feasible, so that firms are restricted to set uniform prices in all contingencies. In the following, we look for the subgame perfect Nash equilibrium of the game and the analysis of the symmetric and asymmetric subgames will be instrumental to derive the relevant payoffs of the players involved in the sequential game, in order to characterize the optimal choice of the dominant firm at the first stage. The next section focuses on exclusive data access while Section 1.3 investigates the case of a proprietary customizing technology.

1.2 Exclusive access to data

In the policy debate on competition in digital markets, a common theme is that more data deliver an added-value service, and that when there are situations of exclusivity then an equal level playing field may be threatened with detrimental consequences for the quality of services and products, for market efficiency and, in turn, for final consumers. We consider a scenario in which relevant data is initially held by only one firm whereas the rival seeks access to this data. Even tough often data is considered to be not rival and not excludable (Lambrecht and Tucker, 2017), in many real world situations access to some particular data may be practically excluded, as in the case of the Facebook platform and its third party applications. The observation that, when information is in the hands of a data controller, this firm treats collected consumer data as proprietary is at the basis of this part of the model. The surging debate about data property rights has crystallized the current practice using the term "de facto" data ownership. In the following, when a firm is in

[4] The advantage may be due to previously collected data in excess with respect to the rival firm.

possession and control of consumer data, the terms data holder and data controller will be used interchangeably.

Formally, when one firm exploits data then an added value is granted to consumers which access the service from that firm; else, the firm can offer just the basic version of the product. This add-on is modeled as a vertical shifter of the value of the service, and it is measured by a parameter $\delta \geq 0$ which is uniform for all consumers. In principle, any firm having access to suitable data would be able to guarantee this extra utility to its customers. Let us denote with $q_i = \{0, 1\}$ the quantity of data that firm i can access. Thus, q_i captures whether firm i has or not access to a dataset containing information about the unitary mass of consumers. The incremental utility in eq. (1.1) is specified linearly as $\Delta u_i = \delta q_i$. The "status quo" of the market is such that firm 1 has access to data by default ($q_1 = 1$) whereas firm 2 has no data ($q_2 = 0$). Therefore firm 1 can always offer an added value equal to $\Delta u_1 = \delta$ to its customers, whereas its competitor can only propose

$$\Delta u_2 = \begin{cases} 0 & \text{if } q_2 = 0 \\ \delta & \text{if } q_2 = 1 \end{cases}$$

where the value of q_2 is determined at the first stage. The parameter δ is assumed to lie in the interval $0 \leq \delta < 3t$ in order to avoid market tipping in favour of the dominant firm. The add-on delivered through data exploitation needs to be not "too" large, otherwise the uninformed rival would rather exit from the market and there would be no room for the trade of the digital asset. If this is the case, then exclusive data is effectively a strong barrier to entry. Notice that the market is fully covered in equilibrium if we assume that $v \geq \frac{3t-\delta}{2}$.

In line with the literature on data sales, information is sold on a fixed fee basis: the data holder has a database containing consumer data and posts a price for it. There is evidence that information on consumers is traded in the form of lists sold for a given price per thousand or per million consumers[5]. For consistency throughout the chapter, the same assumption will be maintained when focusing on the trade of a customizing technology between competing firms. We remark once again that the type of data considered in this part of the model is exclusively available to the dominant firm, and it cannot be easily acquired elsewhere in the market for data. In

[5]Data brokers such as Towerdata, Acxiom, DawexSystems, Nielsen, Intelius and many more are specialized in creating data packages containing individuals' information. Then, these datasets are sold to client firms. For instance, Acxiom sells comprehensive consumer lists (`https://www.acxiom.com/what-we-do/infobase/`).

the following, we will analyze the relevant subgames that follow the first stage and then characterize by backward induction the optimal strategy of the data controller.

1.2.1 Exclusive Case

Suppose that firm 1 maintains "de facto" data ownership: the data allocation is $\{q_1 = 1, q_2 = 0\}$. Net utilities of a consumer located at x are $u_1(x) = v + \delta - tx - p_1$ when going to outlet 1 and $u_2(x) = v - t(1-x) - p_2$ at outlet 2. The expression for the indifferent consumer writes

$$x^{EX} = \frac{1}{2} + \frac{p_2 - p_1}{2t} + \frac{\delta}{2t}$$

and firms' objective functions are $\pi_1 = p_1 x^{EX}$ and $\pi_2 = p_2(1 - x^{EX})$. Solving for the first order conditions yields the reaction functions

$$p_1(p_2) = \frac{p_2 + t + \delta}{2} \qquad p_2(p_1) = \frac{p_1 + t - \delta}{2}.$$

and the Nash equilibrium prices are

$$p_1^{EX} = t + \frac{\delta}{3} \qquad p_2^{EX} = t - \frac{\delta}{3}.$$

Given that the indifferent consumer simplifies to $x^{EX} = \frac{1}{2} + \frac{\delta}{6t}$, the expressions for profits are

$$\pi_1^{EX} = \frac{(3t+\delta)^2}{18t} \qquad \pi_2^{EX} = \frac{(3t-\delta)^2}{18t}$$

and an interior solution is ensured by the condition $\delta < 3t$ (i.e. the service providers are differentiated enough even in presence of an added-value component). Finally, consumer surplus can be written as

$$\begin{aligned} CS^{EX} &= CS_1^{EX} + CS_2^{EX} \\ &= \int_0^{x^{EX}} \left(v + \delta - tx - p_1^{EX}\right) dx + \int_{x^{EX}}^1 \left(v - t(1-x) - p_2^{EX}\right) dx \end{aligned}$$

where

$$CS_1^{EX} = \left(\frac{1}{2} + \frac{\delta}{6t}\right)\left(v + \frac{7}{12}\delta - \frac{5}{4}t\right)$$

$$CS_2^{EX} = \left(\frac{1}{2} - \frac{\delta}{6t}\right)\left(v + \frac{5}{12}\delta - \frac{5}{4}t\right).$$

1.2.2 Non Exclusive Case

Symmetric access to data implies $q_1 = q_2 = 1$. Net utilities change to $u_1(x) = v + \delta - tx - p_1$ and $u_2(x) = v + \delta - t(1-x) - p_2$ and the indifferent consumer moves to

$$x^{NE} = \frac{1}{2} + \frac{p_2 - p_1}{2t}.$$

Firms maximize with respect to prices the objective functions $\pi_1^{NE} = p_A x^{NE}$ and $\pi_2^{NE} = p_2(1 - x^{NE})$ which yield the symmetric price Nash equilibrium $p_1^{NE} = p_2^{NE} = t$. The market is equally split and consumer allocation among competing firms is efficient. Profits are $\pi_1^{NE} = \pi_2^{NE} = \frac{t}{2}$ while the expression for the consumer surplus is

$$\begin{aligned} CS^{NE} &= \int_0^{\frac{1}{2}} \left(v + \delta - tx - p_1^{NE}\right) dx + \int_{\frac{1}{2}}^{1} \left(v + \delta - t(1-x) - p_2^{NE}\right) dx \\ &= v + \delta - \frac{5}{4}t. \end{aligned}$$

The data-driven added-value is competed away and all consumers access a better service.

1.2.3 First stage analysis

We assume that firm 1 has all the bargaining power and it makes a take-it-or-leave-it offer to the competitor at the first stage of the sequential game. Notice that the result that we will present in Proposition 1 holds even if it is the data buyer to submit an offer to the data holder[6], so that the assumption on the bargaining power is immaterial. In addition, provided that when $\delta = 0$ the data holder is indifferent between selling and keeping data, the tie breaking rule is that exclusivity is preferred in that case.

If the data collector trades data with firm 2, he can get at most T from the

[6] Indeed, we can equivalently say that it is the data buyer which submits an offer to the original data holder, who can accept or refuse to sell. Consider a firm that has to decide whether to provide a basic service or a data-driven one to a consumer that may be interested in being served by that firm. The firm makes an access request to the company that holds data offering a price for information. This price can be at most equal to its maximal willingness to pay. In the event that this price is below the minimum offer the seller would accept, data trade does not take place.

transaction, where $T = \pi_2^{NE} - \pi_2^{EX}$. T represents firm 2's maximal willingness to pay for the data. Through this transfer, firm 1 can entirely extract the extra profit made by the rival with respect to a scenario without data. Indeed, in this context, this extra profit represents the value of information to firm 2. The optimal strategy of firm 1 therefore depends on the following inequality:

$$\pi_1^{EX} - \pi_1^{NE} \gtreqless T = \pi_2^{NE} - \pi_2^{EX} \tag{1.2}$$

which can be rewritten in a perhaps more intuitive way as

$$\pi_1^{EX} + \pi_2^{EX} \gtreqless 2\pi^{duopoly}$$

where $2\pi^{duopoly} = \pi_1^{NE} + \pi_2^{NE}$. The straightforward intuition is that if the additional gain made under exclusivity is larger than the price for information that 2 is willing to pay, then the data holder will never trade data with the rival. We can state our first result as follows.

Proposition 1. *The data holder always keeps data exclusively.*

Proof. Firstly, we show that under exclusivity firm 1 is always better-off. The right-hand side of (1.2) is

$$\pi_1^{EX} - \pi_1^{NE} = \frac{\delta(6t + \delta)}{18t}$$

which is positive for any $\delta > 0$. Secondly, we show that exclusivity is detrimental to the firm without access to data. In other words, we determine under which conditions firm 2 is willing to pay a positive price for data:

$$\pi_2^{NE} > \pi_2^{EX} \quad \iff \quad 0 < \delta < 6t$$

As long as firm 2 is in the market (i.e. $\delta < 3t$), it always has a positive willingness to pay for the data. In particular, the maximum price that firm 2 is ready to pay is the right-hand side of (1.2) which writes

$$T = \pi_2^{NE} - \pi_2^{EX} = \frac{\delta(6t - \delta)}{18t}.$$

Finally, we show that this transfer is not sufficient to compensate firm 1 for the loss in profits incurred when also the competitor can deliver an added value service:

$$\left(\pi_1^{EX} - \pi_1^{NE}\right) - T = \frac{\delta^2}{9t} > 0.$$

This expression is clearly always greater than zero[7]. ■

The rival firm finds it always profitable to acquire data but the loss of the competitive advantage suffered by the data holder overcomes the price obtained for the data: the benefit from receiving the transfer T is outweighed by the profit deriving from a larger market share and a higher price set under exclusivity. Notice that the price that would make firm 1 indifferent between exclusivity and non exclusivity should be at least as large as $P = \pi_1^{EX} - \pi_1^{NE} > T$. Since we have shown that P is always greater than T, there cannot be room for data trade in this context. Therefore, it is worth saying that we can equivalently state the result in terms of firm 1 charging a price P that is too high for firm 2 or 2 offering a price T for data that falls below firm 1's willingness to accept. In the exclusive equilibrium, total industry profits are equal to $\pi_{tot}^{EX} = t + \frac{\delta^2}{9t}$ whereas under non exclusivity $\pi_{tot}^{NE} = t$ holds. In the latter case, the value of data is totally appropriated by consumers. Data sales on the basis of a fixed fee do not allow the data controller to extract the value of data in a competitive environment[8]. The optimal strategy of the dominant firm is therefore to maintain exclusivity over the informational asset and compete in the market against a firm not able to offer a better service. In the next section we analyze the implications for welfare and consumers.

1.2.4 Welfare Analysis

The subgame perfect Nash equilibrium we have characterized so far involves no trade of data and a unique firm able to offer an added value service. In turn, not all the users can access a better service in the exclusive equilibrium and data are underutilized in comparison to a scenario in which both firms can use data as an input to improve their service. The non exclusive case is the first best, so that the total welfare is maximized: efficiency holds, overall transportation costs are minimized and all consumers access the better service. However, the first best never arises in

[7]The above expression simply implies that $\pi_1^{EX} > \pi_1^{NE} + T$ is always verified.

[8]The data holder could do even better than in this exclusive equilibrium in an alternative contracting environment. Assume data sales by means of a per-subscriber fee. Then non exclusivity arises as the equilibrium of the sequential game: firm 1 extracts all the value generated through data exploitation, firm 2's profit is constant and does not depend on δ, and the fee is entirely passed to users (see the extension proposed in Section 1.2.5).

equilibrium. We move to a welfare assessment of the equilibrium and we contrast it with the first best. Inefficiencies arise from two distinct sources: welfare lies below its maximum level both when some consumers acquire the product from the farthest outlet and when too many consumers get just a basic product.

Proposition 2. *Data exclusivity is inefficient, given that too few users consume an added value product.*

Proof. The first source of inefficiency is the asymmetry in market shares which trivially causes an increase in transportation costs. However, there is an additional inefficiency. As a first step, define $\tilde{x}$ as the location of the marginal consumer that maximizes social welfare under exclusivity, that is, already taking into account that some users will acquire the product at the farthest outlet. We can write the social welfare as

$$\begin{aligned} W &= \int_0^{\tilde{x}} (v + \delta - tx)dx + \int_{\tilde{x}}^1 (v - t(1 - x))dx \\ &= v - \frac{t}{2} + (t + \delta)\tilde{x} - t\tilde{x}^2 \end{aligned}$$

with $\partial W / \partial \tilde{x} = (t + \delta) - 2t$ and $\partial^2 W / \partial \tilde{x}^2 = -2t < 0$. Being the welfare function concave in $\tilde{x}$, we can derive the socially efficient allocation of consumers when only one firm offers and added value product. The second best is $\tilde{x} = \frac{t+\delta}{2t}$. Recalling that in equilibrium the marginal consumer is located at $x^{EX} = \frac{1}{2} + \frac{\delta}{6t}$, it is immediate to see that under exclusivity:

$$\tilde{x} = \frac{t + \delta}{2t} > x^{EX} = \frac{1}{2} + \frac{\delta}{6t}.$$

Too few consumers access a better service. ■

Exclusivity determines two inefficiencies that move in different directions. In particular, for $0 < \delta < t$ we have $\tilde{x} < 1$: the inefficiency coming from an increase in transportation weighs more than the inefficiency arising when someone does not get the better service, and the marginal consumer is interior. For $t \leq \delta < 3t$, as the value of the basic service overcomes the disutility from transportation costs, we find that the entire market should be covered by the data controller under exclusivity. Therefore, the exclusive equilibrium is inefficient in the sense that not all consumers get a service valued $v + \delta$; moreover, overall transportation costs borne by users increase. Indeed, it is possible to show that overall consumer surplus is lower under exclusivity. The relevant expressions are:

$$CS_{tot}^{EX} = v + \frac{\delta}{2} + \frac{\delta^2}{36t} - \frac{5}{4}t$$
$$CS_{tot}^{NE} = v + \delta - \frac{5}{4}t$$

and the comparison between them boils down to:

$$CS_{tot}^{NE} > CS_{tot}^{EX} \quad \Longleftrightarrow \quad 0 < \delta < 18t$$

which is trivially verified given our initial assumption on the parameters of the model. In aggregate, a non exclusive data allocation would be beneficial for users; moreover, it is possible to show that each user has a strictly positive utility gain. It is also worth to highlight that consumers are better-off for different reasons. Those who were already consuming an added value service continue to get it but at a lower price: p_1^{NE} if they remain at firm 1 and p_2^{NE} if they switch to the rival. On the other hand, users that were served by firm 2, even though they are charged a slightly higher price, have now the possibility to access the better service.

1.2.5 Extensions

Here we present several modifications of the baseline model, such as data portability, which provide an effective way to induce a non exclusive allocation. However, consumers are not necessarily the winners under non exclusivity, even though they consume a better service.

Third party data broker

We propose a modification of the "de facto" data ownership scenario. Suppose that firm 1 is no more the unique data provider, but that there is data broker active in the market for data, firm 3, that has previously collected information about users. We assume that the business model of firm 3 is uniquely that of selling information to firms, with the implication that the data broker will not be a competitor in the market for services, so that downstream competition still takes place only among firm 1 - with data - and firm 2 - without data. The novelty now is that firm 2 can get data not only from firm 1, but also elsewhere. We are interested in understanding how 1's incentives to sell data change when there is an alternative data collector and what will be the final price for data.

Again we have to deal with two scenarios: downstream competition can take place in an asymmetric fashion or in a symmetric one. The first scenario arises when firm 2 has not access to data, either because (i) it is not willing to acquire data (i.e. its willingness to pay for data is negative) or (ii) both data holders - 1 and 3 - are not willing to accept its offer, in case of positive willingness to pay. In the light of the results derived in the previous sections we can argue that: case (i) does not arise as we focus on a situation in which both firms are active in the market (i.e. $0 < \delta < 3t$), so that firm 2 must be willing to offer in any contingency a positive price for data; case (ii) hardly occurs because even though firm 1 would not trade data, it is rational for firm 3 to accept any price for data slightly above zero and make a positive profit, rather than exit the game with a zero payoff. The reason is that the outsider is merely interested in the gains made through data selling, whereas firm 1, which is also a competitor of firm 2, also weighs the effects on its profit from selling data to a rival firm.

The implication is that we should see a symmetric equilibrium with respect to data allocation. Therefore we focus directly on the second scenario, trying to understand which is the data seller to whom downstream firm 2 turns to. Both data holders, the dominant firm and the third party seller, know that 2's profit increases from π_2^{EX} to π_2^{NE} when it can access data. Therefore it is common knowledge that its maximal willingness to pay is $T = \pi_2^{NE} - \pi_2^{EX}$. As we have shown before, firm 1 would never be willing to sell at this price, but now a crucial role is played by the third party seller.

Suppose that the data broker charges exactly T for the data: firm 2 pays the price and gets the data, affecting also the profit of firm 1 that reduces from π_1^{EX} to π_1^{NE}. Firm 1 can clearly improve over this situation, given that anyway it is forced to make only the symmetric duopoly profits: it can undercut the price charged by the broker and make an offer equal to $T - \varepsilon$ to the rival 2, where ε is small enough. Both competitors are clearly better-off compared to the non exclusive case in which data is provided by the third party. In turn, firm 3 can make a new proposal to 2 equal to $T - 2\varepsilon$. The undercutting unravels until the price for data equals zero, and firm 2 is indifferent between which firm to get data from.

The main implication of this simple extension is that when at least two firms hold the same type of data and there is a single buyer, the price for information is driven to zero. This result is totally in contrast with our Proposition 1: whenever it is the unique data holder, firm 1 never has an incentive to sell its data; introducing a competing data seller, firm 1 is willing to sell data at any price. Of course, the result depends also on the fact that there are no costs of data collection, which could instead be a reasonable assumption: data is a valuable input and it is "produced" at some cost.

Per-subscriber fee

Suppose that the data holder can perfectly monitor the number of users accessing 2's service. In this case there is the possibility for firm 1 to charge a per-subscriber fee s for each user enjoying the rival's added-value service. In practice, the fee enters into the rival's maximization problem as the marginal cost of delivering the "premium" service . Upon acquiring data, firm 2 will therefore maximize

$$\max_{p_2} \ (p_2 - s)(1 - x^{NE})$$

whereas firm 1's optimization problem becomes

$$\max_{p_1} \ p_1(x^{NE}) + s(1 - x^{NE}).$$

Taking the first order conditions and solving for the equilibrium prices we get $p_1^{NE} = p_2^{NE} = t+s$, so that the equilibrium allocation of consumers among outlets is efficient, with associated profits

$$\pi_1^{NE} = s + \frac{t}{2} \qquad \pi_2^{NE} = \frac{t}{2}.$$

Equilibrium profits of the incumbent are monotonic in s. Moreover, the social value of data is exactly equal to δ, provided that the total mass of consumers is equal to one. Therefore, firm 1 will set the optimal per-subscriber fee so as to extract all the social value from the users of the service: it must be that $s^* = \delta$ at the optimum[9], with firm 1 getting all the surplus that is generated through the data-driven added value product. Suppose instead that the fee is set below δ: the incumbent can increase its payoff by slightly increasing the per-subscriber charge and nonetheless firm 2 will be better-off with access to data rather than without it. Likewise, a fee higher than δ cannot be an equilibrium, as in this case firm 2 is worse-off when acquiring data. Those who are worse-off in this scenario are clearly the users, given that the fee is entirely passed to them through the prices $p_i^{NE} = t + \delta$ where $s^* = \delta$. Indeed, total consumer surplus decreases by an amount exactly equal to δ, compared

[9]This argument is quite similar to the licensing of a cost reducing innovation: the licensor will optimally set the royalty rate exactly equal to the reduction in the marginal cost.

to a scenario in which data is sold on a fixed fee basis:

$$CS_1^{NE} = CS_2^{NE} = v - \frac{5}{4}t.$$

What we could expect, in the light of Armstrong (1999), is that the controller of the input embodying the competitive advantage would prefer non exclusivity under a per-subscriber fee scenario. As argued above, we set $s^* = \delta$ in equilibrium. Nothing changes for firm 2, provided that its profit does not depend on δ. On the other hand, we have to evaluate the difference between π_1^{EX} and π_1^{NE}, which boils down to

$$\pi_1^{EX} - \pi_1^{NE} = \frac{\delta}{18t}(\delta - 12t).$$

Exclusivity dominates only if $\delta > 12t$. However, if this is the case, then the result is trivial as firm 2 is never active in the market for these values of the parameters. Instead, focusing on the more interesting case in which both firms are active in the market (i.e. $\delta < 3t$), it is immediate to see that non exclusivity always arises as the unique equilibrium of the game. The data holder sells consumer data to the rival and charges an optimal per-subscriber fee equal to δ, which is entirely passed on to consumers.

Data portability regime

An alternative regime that we contrast with the market "status quo" of data ownership is a data portability regime. We assume that consumers have some control over their data in order to enforce a data sharing, if they wish so, between the original data controller and a new firm that may need this user data as well. Nonetheless, the original collector cannot be obliged to delete this information: users have an active role up to the possibility of "porting" their personal data to a new service provider. The implication is that the transition from a data allocation $\{q_1 = 1, q_2 = 0\}$ to a symmetric one with $\{q_1 = 1, q_2 = 1\}$ is not conditional on firm 1's willingness to trade data, but depends on consumers' choices. We model a scenario in which users have the right to port their data. It is common knowledge that this regulation is in place. Similarly to the baseline model, there is a data holder, firm 1, competing in the market for data-driven services with a rival 2 that has not access to data. However, we slightly change the timing of the game:

1. Firms make two types of offers to consumers, a price for the product with data and a price for the basic product, and commit to these offers;

2. Consumers choose which offer to accept[10].

The users' choices depend on the net surplus derived from consumption, provided that their options become:

- Consume from 1 a basic service valued v;
- Consume from 1 a better service $v + \delta$;
- Consume from 2 without porting data a service valued v;
- Consume from 2 and port data to obtain a better service $v + \delta$.

The decision to port data corresponds to an obligation for firm 1 to share data with the rival[11]. Depending on the type of service and the identity of the chosen provider, the net utility of a consumer located in x is

$$u(x) = \begin{cases} v - tx - p_1^N \\ v + \delta - tx - p_1^D \\ v - t(1-x) - p_2^N \\ v + \delta - t(1-x) - p_2^D \end{cases}$$

where the superscripts N and D identify a service provided without or with data, respectively. The novel intuition is that the initial asymmetry in data access is completely irrelevant, as consumers have the possibility to port data to firm 2. In other words, they always have the option to consume an added-value service, irrespective of the initial allocation of consumer data among competing firms. In a symmetric equilibrium with a service of basic quality only, the price equilibrium is $p_1^N = p_2^N = t$. It follows that the maximum price that a firm can charge for the access to a premium version, which delivers an added quality measured by δ, is $p_i^D = t + \delta$. However, it is possible to show that a candidate equilibrium such that $p_1^D = p_2^D = t + \delta$ is not sustainable, as firms have an incentive to undercut on

[10]Notice that it is not necessary for firm 2 to hold data at the first stage in order to be able to propose a better service: the provision of the added-value service is contextual to the user decision to consume this type of service at stage 2, as the user will request (or the firm, on his behalf) to enforce his data portability right.

[11]An interpretation could be that only if a consumer decide to get 2's added-value service, he enforces his data portability right, so that firm 2 has access only to a fraction of the total information. Instead, we assume that when data sharing takes place, the entire database about users is shared with the rival firm. This simplifies the analysis and makes it consistent with the previous regime.

the pricing of the added value component. The undercutting drives down the price so that $p_i^D = p_i^N$ in equilibrium. In a non exclusive equilibrium, which is *de facto* induced through the data portability regulation, δ is competed away.

Proposition 3. *Each firm offers the same price for both services,* $p_1^N = p_1^D = t$ *and* $p_2^N = p_2^D = t$*. All consumers access an added value service.*

Proof. See the argument given in the text. ∎

The equilibrium allocation of consumers among firms is efficient and the added value generated through data exploitation entirely benefits consumers.

1.3 Customizing technology

Firms compete for users by offering services that can be individually customized. The effectiveness of this data-driven customization comes from the fact that consumers who like more a particular service use it more intensely, disclosing more personal information to the firm. In turn, engagement grows and user-generated data increases as well. The scope for customization is therefore larger for more frequent users. However, customization depends not only on the amount of data available to the firm for each and every user but also on the technology available to that firm. Formally, suppose that there is a one to one mapping between horizontal preferences $(x, 1-x)$ and quantity of data that a consumer x generates to the two firms (q_1, q_2). For each type x, the mapping "preferences $\rightarrow$ data" for both firms is modeled as follows: distances $(x, 1-x)$ correspond to quantities of user-generated data $(q_1 = 1-x, q_2 = 1-(1-x) = x)$. The intuition is that close consumers generate more data to the firm, provided that they have a better match with that product or service, and perhaps they would use it more than the rival's service. For instance, the type located exactly at firm 1's address is identified by the triple $(x = 0, q_1 = 1, q_2 = 0)$. The information structure just described is exogenously given and allows us to abstract from the data collection stage.

Graphically, the representation of the mapping from horizontal preferences into the amount of user-generated data is given by the 45°line in Figure 1.1, where firm 1 is placed in the bottom-left corner and firm 2 in the top-right one. If $x < \frac{1}{2}$ it follows that $q_1 > q_2$, given that user x is closer to firm 1, whereas if $x > \frac{1}{2}$ then $q_1 < q_2$ holds. The extent - or level - of customization that firms can provide is captured by a parameter $\delta \geq 0$. When $\delta = 0$ user-generated data play no role in customization, and we are back in a standard city line model. This means that each

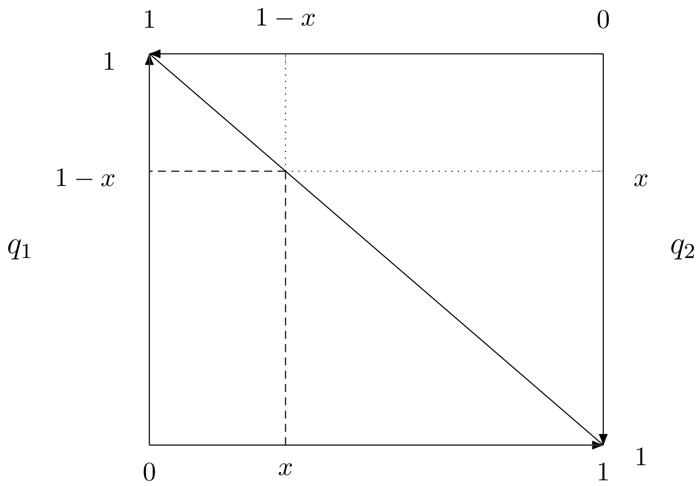


Figure 1.1: Mapping from preferences into data.

user gets an increase in his gross utility that depends on the amount of data held by the customizing firm and on the level of customization.

Similarly to the previous section, only firm 1 is able to exploit data about consumers in order to provide an added value service. Its competitive advantage is the proprietary customization technology. Instead, firm 2 does not have the ability to deliver a customized service, even though in possession of some data about consumers ($\Delta u_2 = 0, \forall x$ even though $q_2 > 0$). The intuition behind our formalization is that firm 1 holds a proprietary algorithm that makes customization effective. When $\delta = 0$ there is no scope in having this algorithm, but when customization is feasible (i.e. $\delta > 0$) then the incumbent has a competitive advantage. Firm 1 can sell an algorithm of precision $\alpha \in [0, 1]$ to the rival. It seems quite reasonable to assume that firm 1 cannot sell an algorithm that is even better than the one already used in house. Therefore, firm 1 is always able to deliver $\Delta u_1 = \delta q_1 = \delta(1 - x)$ to its customers whereas firm 2 can offer $\Delta u_2 = \alpha\delta q_2 = \alpha\delta x$. It follows that the net utilities of users upon going to firm 1 or 2 can be written as[12]

$$u_1(x) = v + \delta(1 - x) - tx - p_1$$
$$u_2(x) = v + \alpha\delta x - t(1 - x) - p_2.$$

In this section the analysis focuses on (i) the role of individually tailored cus-

[12]The novelty here is that the added-value is tailored upon consumer preferences. If we simply set $\Delta u_1 = \Delta u_2 = \delta$, uniform $\forall x$, we get the model with data exclusivity where the competitive advantage was modeled as a uniform vertical shifter.

tomization, (ii) the strategic and competitive effects at play when we switch from an exclusive scenario to a nonexclusive one in which also the initially "disadvantaged" player can exploit data by means of an algorithm of precision α and finally (iii) the relationship between the type of data firms are able to exploit and consumer demand. The main mechanism is the exploitation of different subsets of data, which relaxes price competition. In other words, customization based on different type of data leads to more differentiation in the market. The punchline is that consumer data collection and exploitation is a competitive device that, under some conditions, may be used by firms to increase or decrease product differentiation and, in turn, relax or intensify price competition.

1.3.1 Only one firm can customize

Suppose that firm 2 cannot customize its service. The customizing technology is exclusively in the hands of the incumbent. Consumer utilities can be rewritten as $u_1(x) = v + \delta(1-x) - tx - p_1$ and $u_2(x) = v - t(1-x) - p_2$. The indifferent user is located at

$$x^{EX}(p_1, p_2) = \frac{\delta + t + p_2 - p_1}{2t + \delta}.$$

Firms 1 and 2 maximize respectively the objective functions $\pi_1 = p_1 x^{EX}(p_1, p_2)$ and $\pi_2 = p_2(1 - x^{EX}(p_1, p_2))$. Taking the first order conditions, we get $p_1(p_2) = (\delta + t + p_2)/2$ and $p_2(p_1) = (\delta + p_1)/2$, and solving for prices we obtain

$$p_1^{EX} = t + \frac{2\delta}{3} \qquad p_2^{EX} = t + \frac{\delta}{3}.$$

The expression for the indifferent consumer simplifies to $x^{EX} = \frac{3t+2\delta}{3(2t+\delta)}$, and profits write

$$\pi_1^{EX} = \frac{(3t+2\delta)^2}{9(2t+\delta)} \qquad \pi_2^{EX} = \frac{(3t+\delta)^2}{9(2t+\delta)}.$$

1.3.2 Both firm customize

Suppose that firm 2 has acquired a customizing technology of unknown precision α at stage 1 from the incumbent firm. Following standard procedures, the indifferent consumer becomes

$$x^{NE}(p_1, p_2, \alpha) = \frac{\delta + t + p_2 - p_1}{2t + \delta(1+\alpha)}.$$

Firms' demands respectively are

$$D_1(p_1, p_2, \alpha) = \frac{\delta + t + p_2 - p_1}{2t + \delta(1+\alpha)} \qquad D_2(p_1, p_2, \alpha) = \frac{\alpha\delta + t + p_1 - p_2}{2t + \delta(1+\alpha)}$$

and the first order conditions associated to the maximization problems yield the best replies

$$p_1(p_2) = \frac{1}{2}(\delta + t + p_2) \qquad p_2(p_1) = \frac{1}{2}(\alpha\delta + t + p_1).$$

Solving for the non exclusive prices as functions of α and δ we get

$$p_1^{NE} = t + \frac{\delta(2+\alpha)}{3} \tag{1.3}$$

$$p_2^{NE} = t + \frac{\delta(1+2\alpha)}{3}. \tag{1.4}$$

Market shares as a function of α are determined accordingly to $x^* = \frac{3t+\delta(2+\alpha)}{3(2t+\delta(1+\alpha))}$. Notice that the market share of firm 1 is decreasing in α, ranging (from above) from $\frac{3t+2\delta}{3(2t+\delta)} > \frac{1}{2}$ to exactly one half as α increases from zero to one. However, both profits turn out to be positively related to α, even though the incumbent is loosing customers in favor of the rival when also firm 2 starts to customize its product.

$$\pi_1^{NE} = \int_0^{\frac{3t+\delta(2+\alpha)}{3(2t+\delta(1+\alpha))}} \left(t + \frac{\delta(2+\alpha)}{3}\right) dx = \frac{(3t+\delta(2+\alpha))^2}{9(2t+\delta(1+\alpha))}$$

$$\pi_2^{NE} = \int_{\frac{3t+\delta(2+\alpha)}{3(2t+\delta(1+\alpha))}}^{1} \left(t + \frac{\delta(1+2\alpha)}{3}\right) dx = \frac{(3t+\delta(1+2\alpha))^2}{9(2t+\delta(1+\alpha))}.$$

Industry profits are:

$$\pi_{tot} = \frac{18t^2 + 18(1+\alpha)\delta t + \delta^2(5\alpha^2 + 8\alpha + 5)}{9(2t+\delta(1+\alpha))}.$$

Finally, we report the expressions for consumer surplus:

$$\begin{aligned} CS_1 &= \int_0^{\frac{3t+\delta(2+\alpha)}{3(2t+\delta(1+\alpha))}} \left(v + \delta(1-x) - tx - \left(t + \frac{\delta(2+\alpha)}{3}\right)\right) dx \\ &= \frac{(3t+\delta(2+\alpha))(t(12v - \delta(7+11\alpha)) + \delta(6v(1+\alpha) - \alpha\delta(1+2\alpha)) - 15t^2}{18(2t+\delta(1+\alpha))^2} \end{aligned}$$

$$CS_2 = \int_{\frac{3t+\delta(2+\alpha)}{3(2t+\delta(1+\alpha))}}^{1} \left(v + \alpha\delta x - t(1-x) - \left(t + \frac{\delta(1+2\alpha)}{3} \right) \right) dx$$
$$= \frac{(3t+\delta(1+2\alpha))(t(12v-\delta(11+7\alpha))+\delta(6v(1+\alpha)-\delta(2+\alpha))-15t^2}{18(2t+\delta(1+\alpha))^2}.$$

Total consumer surplus therefore satisfies:

$$CS_{tot} = \frac{18(2t+\delta(1+\alpha))v - 27\delta t(1+\alpha) - \delta^2(2\alpha^2+5\alpha+2) - 45t^2}{18(2t+\delta(1+\alpha))}$$

1.3.3 Algorithm precision

The seller decides at stage 1 whether to trade or not its proprietary technology with the rival. Contextually, it optimally sets the precision of the customizing technology put on sale. Firm 2 is clearly interested in acquiring the algorithm only if customization is feasible (i.e. $\delta > 0$). Otherwise, if $\delta = 0$, not only the value of acquiring a customizing technology is null but also the incumbent firm has no competitive advantage at all. Therefore, we assume δ to be strictly positive from now on. Provided that the outcomes previously derived are functions of α, we can formalize the following result.

Proposition 4. *The proprietary firm has an incentive to offer an algorithm of maximal precision: $\alpha^* = 1$. Equilibrium prices are $p_1^* = p_2^* = t + \delta$. Industry profits are maximized and satisfy $\pi_1 + \pi_2 = t + \delta$.*

Proof. The first part of the proposition directly follows from the fact that firm 1's profit is monotonically increasing in $\alpha \in [0,1]$, with $\partial\pi_1/\partial\alpha > 0$ and $\partial^2\pi_1/\partial\alpha^2 > 0$.

$$\frac{\partial\pi_1}{\partial\alpha} = \frac{\delta(\alpha\delta+t)((\alpha+2)\delta+3t)}{9(\alpha\delta+\delta+2t)^2}$$

$$\frac{\partial^2\pi_1}{\partial\alpha^2} = \frac{2\delta^2(\delta+t)^2}{9(\alpha\delta+\delta+2t)^3}.$$

Next we show that for any value of α the incumbent has an incentive to sell the proprietary technology, as measured by the difference

$$\pi_1^{NE} - \pi_1^{EX} = \frac{\alpha\delta(3t^2+2(1+\alpha)t\delta+\alpha\delta^2)}{9(2t+\delta)(2t+\delta(1+\alpha))} > 0.$$

The quantity $\pi_1^{NE} - \pi_1^{EX}$ is obviously monotonically increasing in α as well. It is therefore immediate to show that firm 1 trades a customizing technology of precision $\alpha^* = 1$ with the rival. By substituting $\alpha^* = 1$ in (1.3) and (1.4) we get the equilibrium prices $p_i^* = t + \delta$ with associated profits $\pi_1 = \pi_2 = \frac{t+\delta}{2}$. ■

The allocation of consumers among outlets is efficient. Indeed, notice that as α increases from zero to one the incumbent's demand decreases. However, its price increases in the precision of the algorithm sold to the buyer. In equilibrium, the negative business stealing effect is more than offset by the strengthening of the rent extraction effect through prices. The novel result is that the incumbent firm benefits from bringing back competition to an equal level playing field.

The main mechanism works through the effects of individually tailored customization on product differentiation. As the scope for customizing the service offered to each consumer grows with δ, users clearly benefit from the added value given by a better match with the service. Contextually there is an increase in the prices charged by competing firms, so that a wider scope for customization strengthens the rent extraction effect. It is less straightforward to understand why symmetric customization technologies (i.e. $\alpha^* = 1$) reinforce the rent extraction effect, and not the other way around. It is useful to recall the version of the model in which the added value was modeled as a uniform vertical shift in gross utilities: upon moving to a symmetric scenario, in that setup prices were driven downwards and users got entirely the additional value generated by data exploitation. A uniform vertical add-on was entirely competed away. In this context, the standard intuition would be that an exclusive better algorithm should be a competitive advantage for the incumbent firm. Instead, non exclusivity benefits both competitors. This result hinges on the ability of firms, as α increases, to coordinate on a price increase. Symmetry in customizing technologies *relaxes* price competition. The optimal choice of the seller weighs the loss in terms of own market share when selling an increasingly precise algorithm to the rival against the benefit from the price increase. The latter more than offsets the decrease in demand. Notice that the elasticity of demand decreases with δ but it does not depend on α. Given the rival's price, for larger values of δ firm 1 can set an higher price. In addition, consider the strategic effect attached to α. As α increases, the best reply of firm 2 shifts outwards: given p_1, firm 2 charges an higher price. The incumbent firm internalizes this effect by selling a perfectly precise customizing technology.

1.3.4 Price of the algorithm

In addition to the optimal value of α, the seller must set a price for the algorithm at stage 1. The analysis is performed under the assumption that the incumbent firm has all the bargaining power in this transaction and it is able to set a fixed fee for the customizing device that entirely extracts the additional buyer's profit. We contrast firm 2's profit without the algorithm ($\alpha = 0$) with the profit made upon acquisition of the technology ($\alpha > 0$). The relevant profit difference is

$$\pi_2^{NE} - \pi_2^{EX} = \frac{(3t + \delta(1 + 2\alpha))^2}{9(2t + \delta(1 + \alpha))} - \frac{(3t + \delta)^2}{9(2t + \delta)} > 0 \tag{1.5}$$

which is positive and increasing in α.

Proposition 5. *The seller trades the proprietary algorithm at a price* $T = \frac{\delta(15t+7\delta)}{18(2t+\delta)}$.

Proof. We know from Proposition 4 that trade involves a device of precision $\alpha^* = 1$. The buyer always has a positive willingness to pay for the algorithm, and plugging α^* back into (1.5) we obtain

$$T = \pi_2^{NE} - \pi_2^{EX} = \frac{\delta(15t + 7\delta)}{18(2t + \delta)}.$$

■

The price charged by the seller exactly reflects the willingness to pay of the buyer for the maximally precise algorithm. Interestingly, firm 1 is never worse-off by selling the algorithm, even at a zero price, because anyway it enjoys an increase in its own profit.

1.3.5 Welfare analysis

We have characterized so far the unique equilibrium of the game. In this equilibrium the algorithm is always put on sale: firm 1 sells the maximally precise algorithm to the competitor and both firms customize their services in equilibrium. The characterized equilibrium is the socially efficient outcome. It is interesting to compare it, in terms of allocation of consumers among firms and type of product consumed, with the "status quo" scenario in which the technology remains proprietary.

Proposition 6. *When only one firm has access to the customizing technology, the outcome is inefficient. Too few users access a personalized product.*

Proof. Consider the indifferent consumer under any possible α and denote him as $\tilde{x}(\delta, \alpha)$. Social welfare is given by the following expression:

$$
\begin{aligned}
W &= \int_0^{\tilde{x}} (v + \delta - (\delta + t)x)\, dx + \int_{\tilde{x}}^1 (v - t + (\alpha\delta + t)x)\, dx \\
&= v - \frac{t + \alpha\delta}{2} + (t + \delta)\tilde{x} - \frac{1}{2}(2t + \delta(1 + \alpha))\tilde{x}^2.
\end{aligned}
$$

We have that $\partial W/\partial \tilde{x} = (t+\delta) - (2t+\delta(1+\alpha))\tilde{x}$ and $\partial^2 W/\partial \tilde{x}^2 = -(2t+\delta(1+\alpha)) < 0$. The above expression is concave in $\tilde{x}$ and it is maximized for $\tilde{x} = \frac{t+\delta}{2t+\delta(1+\alpha)}$. When $\alpha^* = 1$, $\tilde{x}$ is exactly one half. All the users consume a customized product and transportation costs are minimized. When we consider an asymmetric scenario it is possible to show that:

$$
x^{EX} < \tilde{x} \mid_{\alpha=0} \Leftrightarrow \frac{3t + 2\delta}{3(2t + \delta)} < \frac{t + \delta}{2t + \delta} \Leftrightarrow \delta + 2t > 0
$$

which is always true under our assumptions, implying that as long as only one firm customizes, "too few" users access a personalized service. This inefficiency reduces as the other firm acquires an algorithm of progressively increasing precision and vanishes for $\alpha^* = 1$.

■

The second best would require more consumers to switch to firm 1 when there is exclusivity. Therefore, when the algorithm cannot be traded, a second inefficiency emerges in addition to the increase in overall transportation costs borne by users: the actual location of the indifferent consumer under asymmetric competition lies to the left of the second best allocation of users among firms. This implies that when the algorithm is proprietary, firm 1 is not able to increase its demand of personalized services as it would be optimal. Finally, we briefly contrast consumer surplus in the equilibrium with $\alpha = 1$ against the initial scenario with $\alpha = 0$. We use the compact notation $CS(\alpha)$ to denote surplus in the two cases. For $t > 0.56$, $CS(1)$ is higher for any δ, whereas for relatively low transport costs the following holds: (i) for $0 < t \leq 0.56$ and $1.79t < \delta \leq 1$, we find that $CS(0) > CS(1)$; (ii) for $0 < t \leq 0.56$ and $0 < \delta < 1.79t$, it follows that $CS(1) > CS(0)$.

1.4 Main highlights

When the incumbent has exclusive access to a dataset with consumer data, the data holder has no incentives to trade data with the rival. A non exclusive data allocation would intensify price competition for the marginal user, driving down the prices. The added value coming from data exploitation is competed away and entirely passed in the hands of consumers. Instead, when personalized customization through algorithms is feasible and firms hold differentiated data, the firm holding a proprietary technology has an incentive to sell it to the competitor, provided that the transition to a symmetric scenario relaxes price competition. In this second scenario firms are able to extract part of the added value that customization delivers to consumers. In order to shed some light on the mechanisms behind our results, it is useful to focus on users' gross utilities, which reflect their valuations for the services conditional on preferences x, conveniently rewritten as

$$u_1(x) = v + \delta - (t + \delta)x$$

$$u_2(x) = v + \delta - (t + \alpha\delta)(1 - x)$$

where, in the symmetric equilibrium $\alpha^* = 1$, the utility from patronizing firm 2 simply becomes $v + \delta - (t + \delta)(1 - x)$. Customization has two effects: (i) a uniform vertical *shift* in each user's willingness to pay, independent from x and measured by δ, which reminds of the model with data exclusivity; and (ii) an increase in the heterogeneity of users' willingness to pay, measured by an increase in the slope of the utility equal to $t+\delta$. In other words, there is a *rotation* of consumers' valuations which leads to more *dispersed* valuations. Provided that t measures the extent of product differentiation in the market, individually tailored customization increases differentiation. Customized products are more differentiated than standard ones. The maximal level of differentiation is reached exactly when the maximally precise algorithm is sold[13]: firm 1 therefore has an incentive to sell the perfectly precise technology to the rival which is always willing to acquire it and commit to its use, enlarging the scope for surplus extraction from users. The intensity of price competition decreases in the level of customization.

To summarize, this chapter investigates the incentives of a firm active in a digital

[13] This result is totally different from literature on customized goods in horizontally differentiated markets, where the benefits from customization increase with the distance from the firm. Moreover, maximally customized products become completely undifferentiated, so that under full customization firms find themselves in pure Bertrand competition.

market to sell its competitive advantage, which comes in the form of exclusive data or a proprietary algorithm, to a rival firm. Results depend crucially on whether expected industry profits rise or decrease upon selling the advantage: in the former case, an exclusive equilibrium arises, whereas in the second case, a non exclusive equilibrium is found. My results can be reinterpreted in a duopoly setting with a third party data seller which is not a direct competitor in the downstream market. The exclusive and non exclusive equilibria do not change: the data broker would give data exclusively to only one downstream firm, but it would sell a maximally precise algorithm to both competitors, given that they customize differently to different sets of consumers.

Chapter 2

The Effects of Information on Competition and Privacy

Consumer data collection in on-line markets is pervasive and it poses several policy-relevant questions. One of the most intriguing is about the effects of consumer data exploitation on competition. Personal information is a crucial business asset for many on-line based firms and it is highly valuable when it allows firms to change their strategies in a profitable way[14]. For instance, it allows accurate consumer profiling which opens up the possibility of making personalized offers based on user characteristics (Stucke, 2018). Firms can therefore provide customized services and personalized recommendations, deliver more targeted advertising or even personalize prices shown to consumers. In particular, motivated by ubiquitous on-line data collection, personalized pricing is a topical area of research. Consumer privacy is a natural concern when firms can more or less accurately target final buyers through prices. In order to implement sophisticated pricing strategies, firms can collect consumer information by themselves or can acquire it from data brokers (Montes et al., 2018).

In this chapter I focus on the effects of information exploitation on profits and consumers surplus when competing firms use data to price discriminate among consumers but privacy can be partially enforced; in the third chapter I will introduce an upstream data seller and I will characterize the incentives to sell consumer data to competing firms. The novelty of the paper is in addressing price discrimination and data sales among firms in a two dimensional model of horizontal product differentiation in which only one or both dimensions of consumer private information

[14] `https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions`

can be observed. Initially I contrast a full privacy benchmark in which the downstream competing firms cannot price discriminate to another baseline scenario in which there is no privacy and individual price targeting is feasible. This second benchmark could be regarded more as a theoretical curiosity rather than a realistic scenario; yet, this type of price discrimination is increasingly feasible and it is crucial to understand all the possible impacts on market outcomes. Then, in the main part of the paper, I study the effects of a partial enforcement of consumer privacy. The intuition is that firms may know some consumers' characteristics but not everything about them: for instance, a regulation may prevent firms from exploiting complete consumer profiles by requesting some form of anonymization in order to preserve privacy or, simply, firms may not be able to infer perfectly each consumer's willingness to pay. Therefore, in contrast to the prevalent one dimensional literature on consumer privacy, I examine the effects of information on profits in a setting in which the information structure is slightly more complex and also more realistic.

To elaborate on the proposed model, consider as an example an online platform listing hotels which has access to data generated by consumers searching for an accommodation. Suppose that only information related to geo-localization can be collected, but this is not the entire set of information that concurs to determine the willingness to pay of each single customer: imagine that there exists another horizontal dimension that cannot be observed, such as the preference of each consumer for a relatively more quiet or lively neighbourhood[15]. Single hotels therefore seek access to consumer data in order to better tailor their offers, and it is likely that "far away" firms would be more aggressive when setting prices. As an alternative example, consider two online outlets selling technological products where the first retailer is specialized in MacBooks and the second one in personal computers equipped with an operating system provided by Microsoft. A data broker tracking the technical characteristics of each user's smartphone is likely to successfully infer the brand preferences of customers: a user browsing the web through an iPhone is more likely to buy another Apple product. This information is valuable to the competing websites. On the other side, consumers' preferences for a smaller sized but more portable laptop or for a large screen but heavier product may not be so easily observable.

When consumer data is two dimensional, symmetric access to full information (i.e. no privacy) strengthens the competition effect and firms are worse-off with price

[15] Alternatively, it would be interesting to consider a second vertical dimension of information, which would be a natural extension of this model to a setting à la Neven and Thisse (1987).

discrimination. Not surprisingly, the standard Bertrand competition argument in transportation costs holds also in two dimensions, even though some consumers are charged a higher personalized price than the full privacy Nash equilibrium price. Instead, I show that access to partial information about consumer preferences always increases profits. In particular, firms are better-off in *all* types of games: (i) when prices are conditioned on the same type of partial information, (ii) when prices are tailored on different dimensions of information, and (iii) when only one firm has exclusive access to partial information. In the latter case, the uninformed firm is not harmed by exclusivity but it is just indifferent with respect to the full privacy benchmark. The main mechanism that reinforces the rent extraction effect is the inability of competing firms to observe one of the two dimensions of consumer preferences: the standard Bertrand competition argument breaks down. When a firm observes only one dimension of information, it can rank consumers in terms of price elasticity of demand accordingly to that piece of information; at the same time, it is not considering that the mismatch in the unobserved dimension can be high or low when setting its price. As a result, the firm will price high to consumers close in the observed dimensions, and in equilibrium it will serve only those with a high willingness to pay in *both* dimensions, letting close consumers with a poor match in the unobserved dimension to inefficiently switch to the rival firm. Pricing above the Nash equilibrium uniform price more than offsets the loss of some customers at close locations in the observed dimension. These findings are in contrast with standard one dimensional information acquisition games, in which firms end up in a Prisoner's Dilemma situation: both firms price discriminate and make lower profits than absent price discrimination.

I also find that in equilibrium the impact of data exploitation on consumer surplus is ambiguous: some consumers are made worse-off but others are better-off. This ambiguous impact on consumers is there even under individual targeting. Moreover, under partial information an inefficient allocation of consumers among firms can arise and some customers end up buying a mismatched product. The ambiguity of the privacy consequences for final consumers do not allow to draw clear cut policy conclusions with respect to consumer protection but this result seems to suggest that a more nuanced approach to consumer privacy has to be considered.

Despite the fact that price discrimination could be considered just a theoretical exercise, there is evidence, although limited, of price discrimination in on-line markets (Hindermann, 2018). Consumer targeting revolves around user-based, technical (operating system) and location-based features. For instance, Hannak et al. (2014)

find limited evidence of price discrimination in the hotel sector, whereas Hupperich et al. (2018) bring some evidence also in the rental car sector. Dube and Misra (2017) investigate the empirical implications of price discrimination with high-dimensional data on customer features. They rely on experimental data and consider a large digital firm which employs machine learning techniques to target prices, showing that profits always increase with price discrimination whereas consumer surplus is almost unchanged. In a competitive setting in which firms have access to consumers' real-time and historical location data, Dube et al. (2017) find that profitability of price discrimination crucially depends on the competitor's response: a firm enjoys large profit gains when it targets the rival's location or when there is a price response in the same direction that softens price competition, whereas such gains are mitigated when prices move in different directions. In general, profits increase on average when price targeting is possible.

Economic literature on privacy and price discrimination formalizes a trade-off between a *rent extraction effect*, which is maximized when only one firm receives the data advantage, and a *competition effect*, which intensifies when information on consumer preferences is symmetrically allocated to firms and it drives down personalized price schedules below the level of the Nash equilibrium uniform price. Firms are worse-off when both have the ability to target consumers or, in other words, when there is no privacy in the sense of Taylor and Wagman (2014)[16]. The negative effect on profits of non exclusive information is one of the most robust results of the price discrimination literature (Thisse and Vives, 1988; Armstrong, 2006) under best-response asymmetry (Corts, 1998)[17]. The main takeaways of the economics of privacy literature, that are challenged in this chapter, are: (i) from a competition policy point of view, regulators should be concerned about exclusive allocations of consumer information in the data market, whereas (ii) from a consumer protection side, symmetric allocation of information hurts firms but benefits all consumers, implying that it would be optimal to have no privacy regulation at all.

This work is closely related to Baye et al. (2018) and Liu and Shuai (2013). The first paper, which builds on Jentzsch et al. (2013), proposes a two period model where consumers differ both in their geographical position and flexibility. Locations are perfectly observed by firms but individual transportation costs are not. However, first period purchases are imperfectly informative about consumer flexi-

[16]Belleflamme and Vergote (2016) show that consumers may be better-off with no privacy even under monopoly.

[17]More comprehensive literature reviews can be found in Acquisti et al. (2016) and Ganuza and Llobet (2018).

bility. They show that firms can be better-off by combining location information with behavioral data for price discrimination purposes when consumers are moderately heterogeneous in flexibility. Firms clearly rank consumers in the same way with respect to the degree of flexibility, and Baye et al. (2018) focus on the effects of combining additional data on flexibility with perfectly observable data on locations; in this paper, instead, the nature of data available to firms is different: transportation costs are homogeneous, but outlets rank consumers differently with respect to both dimensions of private information and I allow firms to hold both symmetric and asymmetric information sets. The second paper proposes a static two dimensional model of horizontal product differentiation, in which information allows firms to segment consumers in two groups along each dimension. Liu and Shuai (2013) find that, when both firms observe only one but the same dimension of private information, partial price discrimination rises industry profits; instead, when firms have partial information but on different dimensions, firms are again worse-off. My model builds on their setup, but in my paper consumer data is finer so that a continuum of consumers is identified along each dimension; in addition, I consider also the case of perfect price discrimination or, equivalently, no privacy. In contrast to their findings, I find that firms are always better-off under partial price discrimination, independently of the type of partial information held by the players. Finally, differently from Baye et al. (2018) and Liu and Shuai (2013), in the third chapter I will also investigate the incentives of an upstream data holder to sell data to competing firms.

In what follows I firstly set up the theoretical model and illustrate the two benchmark cases: either price discrimination is not feasible, so that there is full privacy, or firms perfectly identify each single consumer, implying that there is no privacy. Then I solve for the all relevant games with partial price discrimination that can emerge under any possible combination of the two dimensions, providing a welfare analysis and clarifying the relationship between the type of information structure held by the firms and industry profits. Finally, I propose several extensions that deliver additional results and intuitions.

2.1 The model

I consider an augmented version of the linear city model (Hotelling, 1929). There are two types of agents: consumers and two horizontally differentiated firms[18].

Consumers buy at most one unit of a product and get a gross utility $v > 0$ when buying the good whereas nothing otherwise. Their horizontal preferences (x, y) are two dimensional and orthogonal, uniformly distributed over the unit square $[0, 1] \times [0, 1]$, so that the total mass of consumers is one. Type (x, y) incurs in a linear transportation cost $t > 0$ when buying a product at a price p_i from firm i located in (x_i, y_i) and receives a net utility

$$u_i = v - p_i - t|x - x_i| - t|y - y_i|$$

where the two dimensions of product characteristics are qualified by the same degree of horizontal differentiation.

There are two competing firms $i = 1, 2$ exogenously located at $(0, 0)$ and $(1, 1)$, respectively. Types with a low (high) realization of both x and y are in firm 1's strong (weak) market, whereas types with a high (low) realization of x and y are in firm 2's strong (weak) market (Corts, 1988). Efficiency requires that consumers (x, y) with $x + y < 1$ buy from firm 1, whereas those with $x + y > 1$ acquire the product from firm 2. These conditions ensure that transportation costs are minimized. Throughout the main section of the paper, individual targeting based on both x and y is banned[19]. The timing of the sequential game is the following:

1. If firm i does not hold data, it sets a uniform price p_i; otherwise, it sets a personalized price conditional on the available information structure.

2. Consumers buy the product and payoffs are realized.

Firms observe the information allocation before the pricing game. Instead, consumers observe only the personalized price, if any, designed specifically for them. We present two privacy benchmarks: the full privacy ("*fp*") and the no privacy ("*np*") regimes. Then we study both the static games in which both firms access partial information, symmetrically or asymmetrically, and the game in which only one firm is partially informed. We solve for the static Nash equilibrium of each type of game.

[18] A data seller that can observe and sell information about consumer preferences will be introduce in the next chapter, along with the discussion of a competitive upstream market.

[19] The no privacy benchmark establishes what happens in this case.

The market is fully covered in equilibrium whenever $v \geq 2t$, and this assumption is maintained throughout the analysis.

2.1.1 Main results

This chapter is primarily concerned with the relationship between information and profits. The main findings can be graphically summarized as follows. In the panels below it is possible to appreciate the effects of increasing information in the market - from full privacy to no privacy - on profits. The novelty is the hump-shaped relationship between the amount of data available to the firms and the industry profits.

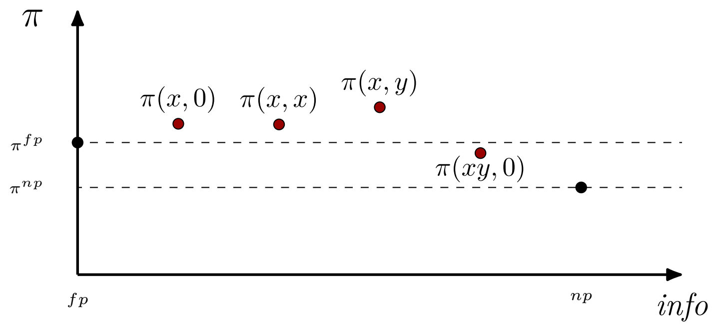


Figure 2.1: Industry profits.

Interestingly, notice that industry profits under partial information lie always above the full privacy profit level. Instead, whenever there is full exclusivity in the market and a single firm is able to individually target consumers, industry profits lie between the full privacy and the no privacy level. The analysis of the individual payoffs delivers additional insights.

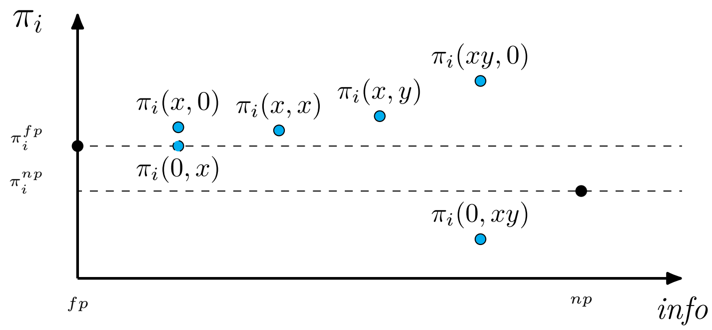


Figure 2.2: Individual profits

In any contingency in which there is partial information in the market, all firms, independently of which data they hold, are weakly better-off than under full privacy. In particular, an uniformed firm competing against a partially informed rival is able to secure the same payoff as under full privacy. The picture changes under full exclusivity: the informed firm gets the highest individual payoff but, considering also that industry profits decrease, this happens at the expenses of the uninformed rival. This type of negative externality does not arise instead under partial exclusivity.

2.1.2 Full privacy

The no information benchmark is useful for an evaluation of the effects on prices, profits and consumer surplus across the different types of games but also for a comparison with the outcomes of the one dimensional literature. Suppose that full privacy is enforced and firms compete in uniform prices. At given prices p_1 and p_2 there is an indifferent type y for each value of x, which implies that there exists a continuum of marginal consumers (x, y) defined as

$$v - p_1 - tx - ty = v - p_2 - t(1-x) - t(1-y) \;\Rightarrow\; \tilde{y}(x) = \frac{p_2 - p_1 + 2t(1-x)}{2t} \tag{2.1}$$

Both firms serve a positive fraction of consumers whenever $|p_i - p_j| < 2t$. When the price difference $p_1 - p_2$ is larger (lower) than $2t$ $(-2t)$ then firm 1 faces zero (unitary) demand. Firms' profits are $\pi_1 = p_1 D_1(p_1, p_2)$ and $\pi_2 = p_2 D_2(p_1, p_2)$.

Lemma 1. *When no firm has information, equilibrium prices are $p_1^* = p_2^* = t$ and each firm makes a profit equal to $\pi_i^{fp} = \frac{t}{2}$.*

Proof. See Appendix 2.4 ■

Marginal types are located along the bisector of the unit square $(y^*(x) = 1 - x)$ and each firm serves half of consumers. Since all consumers buy from the closest firm, efficiency is achieved. Industry profits are $\pi^{fp} = t$. Consumer surplus is

$$CS^{fp} = 2\int_0^{1-x}\int_0^1 \left(v - t - t(x+y)\right) f(x)f(y)dxdy = v - \frac{5t}{3}$$

and total welfare is equal to $v - \frac{2t}{3}$.

2.1.3 No privacy

Consider a scenario in which firms can observe both x and y, and can set a personalized price $p_i(x, y)$ for each consumer (x, y). This second benchmark is an extension of Taylor and Wagman (2014) to two dimensional private information. In their case, with consumers uniformly distributed along the unit line and with perfectly observable types x by both competing firms, discriminatory prices are driven downwards since firms compete for consumers at each location. The closest retailer charges just the saving in total transportation costs enjoyed by a consumer x when buying the product from that firm rather than the farthest retailer (Bhaskar and To, 2004). Price schedules are efficient and the personalized price charged to the consumer equidistant from both firms is just equal to the marginal cost. Given that each firm faces a pool of consumers which are relatively closer to its rival's location, retailers have a common incentive to tailor with lower discriminatory prices consumers located in their weak market. Since location based models are characterized by best response asymmetry in the horizontal dimension, this incentive is asymmetric across subsets of consumers, driving down all personalized prices (Thisse and Vives, 1988; Corts, 1998; Armstrong, 2006).

When there is no privacy, firms compete at the individual level. The indifference condition writes

$$p_1(x,y) + tx + ty = p_2(x,y) + t(1-x) + t(1-y)$$

and a generic consumer (x, y) buys from firm 1 if

$$p_1(x,y) \leq p_2(x,y) + t(1-2x) + t(1-2y).$$

Since both firms know exactly the location of each consumer, they can set very aggressive prices to consumers close to the rival. Firms are willing to price as low as the marginal cost in order to serve an additional consumer, and they are left only with the possibility to extract the saving in transportation cost over both dimensions. The standard Bertrand logic therefore applies also in two dimensional models.

Firm 1 has a transportation cost advantage when serving consumers with $x+y < 1$, whereas firm 2 has an advantage over those with $x+y > 1$. Therefore firms set their tailored offers accordingly to $p_1(x,y) = \max\{t(1-2x)+t(1-2y), 0\}$ and $p_2(x,y) = \max\{t(2x-1)+t(2y-1), 0\}$. Equilibrium personalized prices are

$$p_1^*(x,y) = 2t(1-(x+y)) \qquad \text{if} \quad x+y<1$$

$$p_2^*(x,y) = 2t((x+y)-1) \qquad \text{if} \quad x+y>1$$

and equal to the marginal cost otherwise. The market boundary is the same as in the full privacy benchmark, so that efficiency is achieved again. Notice that only consumers which are located along the bisector of the unit square effectively pay a price equal to the marginal cost. All other consumers pay a positive price that reflects the advantage in transportation cost from going to the closest retailer. Provided that the model is two dimensional, some consumers end up being charged the highest feasible price: consumers located precisely at firms' locations get an offer equal to $2t$ and are fully exploited. Price dispersion is maximal.

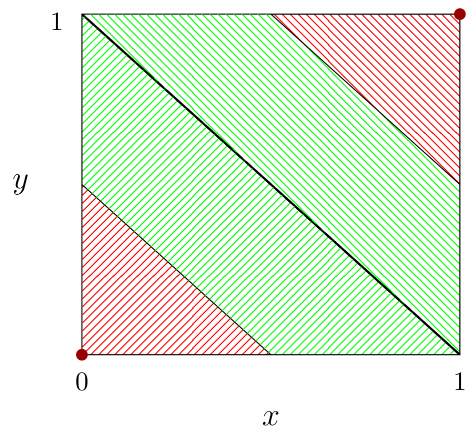


Figure 2.3: Winners (—) and losers (—) under no privacy

Nevertheless, even though some prices are larger than in the full privacy benchmark, the competition effect still prevails on the rent extraction effect. Firm 1's profit is

$$\pi_1^{np} = \int_0^{1-x} \int_0^1 p_1^*(x,y) f(x) f(y) dx dy = \frac{t}{3}$$

and symmetrically for the rival[20]. Whenever firms have symmetric access to all the information available, they would be better-off by committing not to price discriminate. Consumer surplus is larger and equal to $CS^{np} = v - \frac{4t}{3}$ so that total welfare is left unchanged with respect to the full privacy benchmark. However, there are winners and losers among consumers[21]. In particular, consumers with $0 \leq x < \frac{1}{2}$ and $0 \leq y < \frac{1}{2} - x$ are charged a personalized price larger than p_i^* at outlet 1, whereas

[20]Notice that, under all-out competition in one dimensional models, firms' profit is equal to $\frac{t}{4}$ (Thisse and Vives, 1988). The transition to a two dimensional model induces by construction a larger degree of product differentiation, which is reflected in an increase in profits with respect to a setting à la Thisse and Vives (1988).

[21]When customers are individually targeted, the standard one dimensional literature on price discrimination suggests that *all* consumers are better-off.

those with $\frac{1}{2} < x \leq 1$ and $\frac{3}{2} - x < y \leq 1$ at outlet 2. The mass of consumers paying a lower price is larger and therefore profits decrease.

2.2 Partial privacy

So far privacy was either fully enforced or not enforced at all: firms had no information or they accessed data about each consumer's willingness to pay. In online markets consumers are likely to be partially targeted. Here we assume that firms eventually know something about consumers but not everything or, equivalently, that there exists a privacy policy that bans individual targeting of consumers. Our focus is on the effect of partial information on competition and consumer privacy.

2.2.1 Symmetric partial information

Suppose that both firms have information on locations x and are able to tailor prices $p_i(x)$ to targeted consumer groups. Consumer (x, y) in group x accepts the personalized price of firm 1 if and only if

$$p_1(x) + tx + ty \leq p_2(x) + t(1 - x) + t(1 - y).$$

The expression for the indifferent consumer writes as in (2.1), except for the fact that prices are tailored on the realized value of x observed by both firms. Notice that for each x such prices take a specific value p_1 and p_2, which allows to write the demand of firm 1 at each location x as

$$D_1(p_1, p_2; x) = F(y \leq \tilde{y}(x)) = \frac{p_2 - p_1 + 2t(1 - x)}{2t} \tag{2.2}$$

while $D_2(p_1, p_2; x) = 1 - D_1(p_1, p_2; x)$. Similarly to the uniform pricing game, both firms have positive demand whenever $|p_1(x) - p_2(x)| < 2t$, for any value of x. Firms maximization problems yield asymmetric best responses $b_1(p_2) = (p_2 + 2t(1 - x))/2$ and $b_2(p_1) = (p_1 + 2tx)/2$ that depend on x. Firm 1's best response is strictly monotone decreasing in x, whereas firm 2's best response behaves the other way around.

Equilibrium personalized prices lie on the segment joining E_0 to E_1 in Figure 2.4. This segment is drawn by the translation of the best responses in the space $p_1(x) \times p_2(x)$, moving from $x = 0$ to $x = 1$: $b_1(p_2)$ shifts inwards whereas $b_2(p_1)$

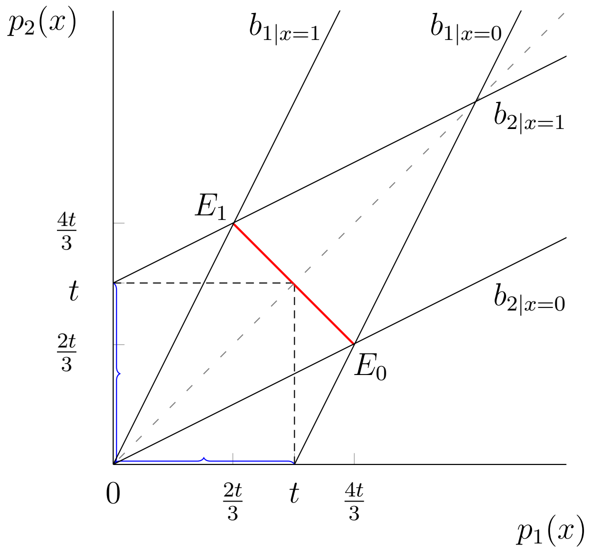


Figure 2.4: Best responses for $x = 0$ and $x = 1$ and locus of equilibria in personalized prices for any value of x. The red line represents the discriminatory prices in the two-dimensional model. The blue braces show the price range in a standard one-dimensional model.

shifts upwards as x increases. The resulting equilibrium prices are

$$p_1^*(x) = t + \frac{t}{3}(1 - 2x)$$
$$p_2^*(x) = t + \frac{t}{3}(2x - 1).$$

Each firm charges a maximum price of $\frac{4t}{3}$ to the closest consumers and a minimum price of $\frac{2t}{3}$ to the farthest consumers[22]. Group $x = \frac{1}{2}$ is charged a personalized price that is equal to the uniform Nash equilibrium price[23].

From (2.2) define the demand elasticity as $E_{D_i} = -\frac{\partial D_i}{\partial p_i}\frac{p_i}{D_i}$, which can be written as

$$E_{D_1} = \frac{p_1}{p_2 - p_1 + 2t(1 - x)} \qquad \text{and} \qquad E_{D_2} = \frac{p_2}{p_1 - p_2 + 2tx},$$

where E_{D_i} is the elasticity of consumers in group x with respect to firm i's price.

[22]When both firms discriminate among consumers the average price schedule is equal to the Nash equilibrium uniform price. This is not true in one dimensional models characterized by best-response asymmetry, where the average price decreases for both firms.

[23]Notice that the optimal uniform price lies between the highest and the lowest discriminatory prices, which is a well established feature of pricing games characterized by best-response *symmetry*. Stole (2003) argues that without such symmetry this clear-cut conclusion about prices does not exist. Instead, here this feature of equilibrium prices emerges also in presence of best-response *asymmetry*.

In comparison to the full privacy regime, it turns out that consumers in the neighbourhood of firm 1 have a more inelastic demand for firm 1's product, while far away consumers have a relatively more elastic demand for it. A symmetric argument applies to firm 2. The market is therefore divided in two regions: (i) for $0 < x < \frac{1}{2}$, firm 1 faces an inelastic demand whereas firm 2 has an elastic demand; (ii) for $\frac{1}{2} < x < 1$, the reverse holds true. Competing firms rank consumers in an opposite way with respect to the elasticity of demand, so that they have opposite incentives when setting prices conditional on the same information, as intuitively shown in Figure 2.5. In other words, when firms acquire the same consumer data, best response asymmetry holds.

However, it is not the case that all discriminatory prices need to be above or below

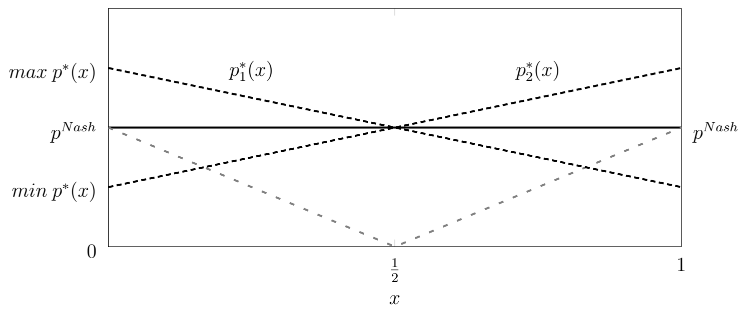


Figure 2.5: Personalized price schedules in two dimensional (densely dashed) and one dimensional models (loosely dashed).

the Nash equilibrium uniform price when best response asymmetry holds, as one would expect in one dimensional spatial models. Indeed, Corts (1998) clarifies that such asymmetry is a necessary - but not a sufficient - condition for having an increase or a decrease in all prices. In contrast to a one dimensional setup, price dispersion decreases when firms have symmetric access to information: the range of personalized prices is $p_i(x) \in \left[\frac{2t}{3}, \frac{4t}{3}\right]$ whereas standard models show that $p_i(x) \in [0, t]$, a range that is one-third wider, which suggests that prices are less sensitive to distance from the firms given the uncertainty in the unobservable dimension.

In equilibrium the market boundary is $y^*(x) = \frac{2-x}{3}$ and it exhibits an anticlockwise rotation with respect to the no information subgame. Each firm serves always a positive fraction of y realizations for all values of x. Compared to uniform pricing, price discrimination allows firm i to serve even the most loyal consumer of the rival firm in the observable dimension. Two major implications follow: (i) a subset of buyers is charged a personalized price that is larger than $p = t$ (i.e. $x < \frac{1}{2}$ at firm 1 and $x > \frac{1}{2}$ at firm 2), the Nash equilibrium price in the no information benchmark,

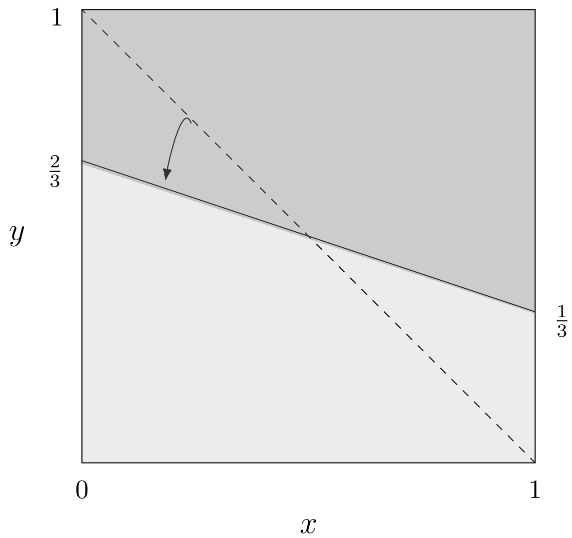


Figure 2.6: Rotation of the market boundary.

whereas other consumers get a price lower than t (i.e. $x > \frac{1}{2}$ at firm 1 and $x < \frac{1}{2}$ at firm 2); (ii) the allocation of consumers among the two outlets is inefficient, given that with respect to the baseline case some consumers switch between firms in order to benefit from a discounted personalized price offered by the rival, causing a net increase in overall transportation costs. Indeed, the anticlockwise rotation of the market boundary is driven by consumers with a realized x close to the ideal product of their chosen retailer in the no information benchmark but with an unobservable realized y close to the product characteristics of the rival: given that each firm increases its price precisely to these consumers, they benefit from switching to the low pricing rival, trading-off a better match in the dimension y with an increased mismatch in the dimension x.

Firm's 1 profit[24] is

$$\pi_1(x,x) = \int_0^{\frac{2-x}{3}} \int_0^1 \left(t + \frac{t}{3}(1-2x)\right) f(x)f(y)dxdy = \frac{14t}{27}$$

and symmetrically for firm 2. Each firm makes a larger profit when both competitors obtain access to consumer information. Symmetric partial information makes firms less aggressive when setting prices. In turn, even tough consumer data is symmetrically held by firms, less privacy is not beneficial to consumers or, at least, not

[24]Slightly extending the notation, in the following we denote with $\pi_i(a,b)$ the equilibrium value of firm i's profit when i has information set a and j has b, with a and b taking values in $\{0,x,y\}$. For aggregate outcomes, a (b) refers to firm 1 (2).

to all of them, differently from what is suggested in the one dimensional literature. Indeed, consumer surplus can be computed as follows

$$CS(x,x) = 2\int_0^{\frac{2-x}{3}} \int_0^1 (v - p_1^*(x) - t(x+y))\, f(x)f(y)dxdy = v - \frac{47t}{27}$$

and it can be easily shown that $CS(x,x) < CS^{fp}$ for any value of t. More rents are transferred from consumers to firms even though some buyers acquire the product at a lower price. The decrease in price for low valuation consumers in the x dimension is more than offset by the increase in price offered to high valuation consumers. The only consumer group indifferent between the two subgames is the mass of realizations y located at $x = \frac{1}{2}$. Consumers are overall worse-off, but the impact on individual net utilities is ambiguous. Instead, both firms are strictly better-off when allowed to simultaneously price discriminate among consumers on the basis of only one dimension of product differentiation. The uncertainty about the other dimension is crucial for the ability of firms to extract rents from consumers.

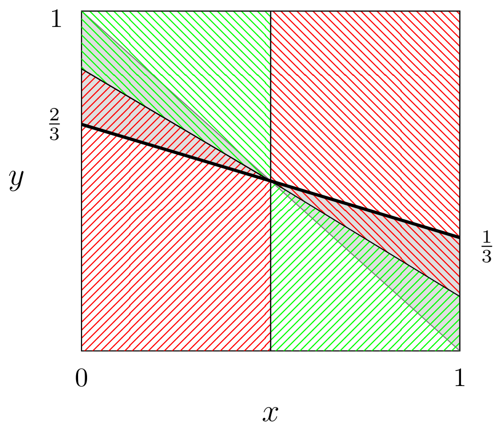


Figure 2.7: Winners (—) and losers (—) under symmetric partial information

Finally, total welfare is equal to $v - \frac{19t}{27}$ and it is lower than in both benchmarks[25]. The distortion in the allocation of consumers among competing firms leads to a redistribution of rents between the agents but some surplus is lost due to the net increase in transportation costs.

[25] In one dimensional models there is a redistribution of rents from firms to consumers, but overall welfare is unchanged. See the characterization of the equilibria with and without privacy in the linear city model of Taylor and Wagman (2014). Moreover, recall that the allocation of consumers in those two cases is efficient and the location of the marginal consumer does not change.

2.2.2 Asymmetric partial information

Suppose that both firms access partial consumer data but on different dimensions. The data allocation is asymmetric, with the two firms targeting consumers respectively on dimensions x and y. When the allocation is reversed the analysis is similar. Firm 1 sets a discriminatory price $p_1(x)$ and firm 2 simultaneously sets $p_2(y)$.

Lemma 2. *When firm 1 has partial information on x and firm 2 has partial information on y personalized prices are $p_1^*(x) = t(\frac{3}{2} - x)$ and $p_2^*(y) = t(\frac{1}{2} + y)$. Each firm's profit is equal to $\pi_i = \frac{7t}{12}$.*

Proof. See the Appendix 2.4. ■

Average price schedules are equal to t and the market boundary coincides with the bisector. Asymmetric access to partial information restores efficiency. However, firms are able to extract more surplus from consumers than in the symmetric case. Firm makes a profit equal to $\pi_1(x, y) = \pi_2(y, x) = \frac{7t}{12}$, but consumer surplus is driven down to $CS(x, y) = v - \frac{11t}{6}$, the lower bound across all subgames. Total welfare is maximized but efficiency is achieved at the expenses of consumer privacy.

Recall that under symmetric partial information the market was divided in two regions accordingly to elasticities, and one firm's equilibrium price was mirroring the schedule set by the rival at each x (see Figure 2.5). Here the elasticities of demand are equal to

$$E_{D_1} = \frac{p_1}{p_2 - p_1 + 2t(1 - x)} \qquad \text{and} \qquad E_{D_2} = \frac{p_2}{p_1 - p_2 + 2ty}$$

with E_{D_1} increasing in x and E_{D_2} decreasing in y, with the market divided in four regions. Firms now have symmetric incentives when setting prices, in particular in the two regions located along the negatively sloped diagonal, but contrasting incentives in the other two regions. In other words, along the diagonal, best response asymmetry fails to hold, and firms rank consumers similarly even tough they discriminate on different pieces of private information. When consumer privacy is partially enforced and firms have access to asymmetric information, there is a mixture of best responses symmetry and asymmetry. Profits increase even more, harming consumer privacy.

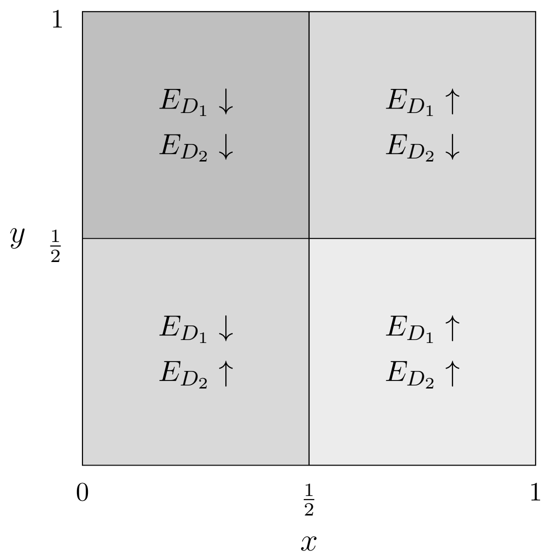


Figure 2.8: Price elasticity of demand under asymmetric partial information.

2.2.3 Exclusive partial information

Suppose that firm 1 acquires exclusively consumer information[26]. Consumer (x, y) buys from the informed firm at the personalized price if and only if

$$p_1(x) + tx + ty \leq p_2 + t(1-x) + t(1-y)$$

which yields the following expression for the locus of indifferent consumers:

$$y(x) = \frac{p_2 - p_1(x) + 2t(1-x)}{2t}.$$

The best reply of the informed player is given by $b_1(p_2) = (p_2 + 2t(1-x))/2$, which is defined for all values of x. However, the uninformed player is not able to optimally respond at each x to the schedule posted by the rival. Instead, firm 2 can set a unique price that "on average" is a best reply to the rival's optimal strategy: our result is that in equilibrium the candidate uniform price of firm 2 must be equal to the average candidate price schedule of firm 1.

Lemma 3. *When only firm 1 has partial information the uniform price is* $p_2^* = t$ *while the personalized price is* $p_1^*(x) = t\left(\frac{3}{2} - x\right)$. *Profits are equal to* $\pi_2 = \frac{t}{2}$ *and* $\pi_1 = \frac{13t}{24}$.

[26]Given that firms are horizontally differentiated, the analysis is symmetrical when it is firm 2 to receive exclusive information.

Proof. See Appendix 2.4. ■

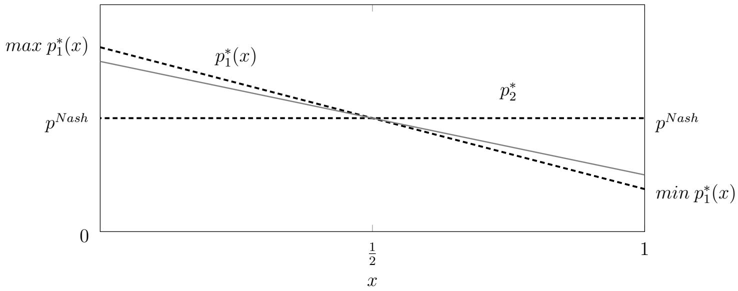


Figure 2.9: Rotation of the personalized price of the exclusively informed firm (densely dashed) with respect to the symmetric partial information game (not dashed).

Notice that the average value of $p_1^*(x)$ is exactly t, the optimal uniform price of firm 2. Firm 1's schedule $p_1^*(x) \in \left[\frac{t}{2}, \frac{3t}{2}\right]$ and price dispersion increases. In contrast to the symmetric subgame with price discrimination, the informed firm sets an even higher personalized price in its strong market, but it is also forced to price more aggressively to realizations of x close to the uninformed firm's location. As a result, the equilibrium market boundary is given by $y^*(x) = \frac{3-2x}{4}$, and interestingly it rotates clockwise with respect to the symmetric information game. In the partially exclusive information regime, the inefficiency is partially mitigated: more consumers buy their preferred product. Provided that $\pi_i(x,0) > \pi_i(0,x)$, exclusive information gives to the informed player a competitive advantage. However, exclusive access to data is not detrimental to the uninformed player: firm 2 is able to secure the same profit level as under full privacy, given that there is no negative externality arising from holding information exclusively. In other words, the asymmetry in data allocation does not result in a "too large" difference between the payoffs of the two players. Finally, consumer surplus is equal to $CS(x,0) = v - \frac{83t}{48}$. The level of consumer surplus is slightly higher than $CS(x,x)$ but it ranks below the benchmarks. Overall, exclusivity does not harm neither the uninformed player nor final consumers but rather it partially restores efficiency.

2.2.4 Welfare analysis

Effects on profits

I show that the standard ranking of profits of the literature on price discrimination and privacy does not hold when consumer private information is identified by the pair (x, y) and firms partially observe consumers' willingness to pay. Recall that in the one dimensional literature: (i) an exclusively informed firm has the largest payoff whereas the uninformed rival has the lowest payoff (below both π_i^{fp} and π_i^{np}) across all types of games, and (ii) when both firms have consumer data, they are always worse-off with price discrimination. Thus, total duopoly profits with information are below exclusive industry profits, and more surplus is extracted in the downstream market when data is allocated to a single competitor.

The first contribution of this paper is to show that the introduction of a slightly more rich information structure can lead to a ranking of individual profits that partially reshuffles to

$$\pi_i(x, y) > \pi_i(x, 0) > \pi_i(x, x) > \pi_i^{fp} = \pi_i(0, x) > \pi_i^{np}.$$

Firms can be better-off with price discrimination, but only if they are not able to target consumers individually: partial consumer data always lessens competition. Several considerations follow: (i) the uninformed firm is not harmed by partial exclusivity, (ii) the payoff of the exclusively informed firm dominates the symmetric payoff with information, but (iii) industry profits are maximized when both price discriminate but on different dimensions. Therefore, industry profits satisfy the inequality

$$\pi(x, y) > \pi(x, 0) > \pi(x, x),$$

and are always larger than π^{fp}. When two dimensions of product differentiation are considered, it is irrelevant for a firm without access to data whether its rival acquires or not partial information. This ranking will have a direct implication for the optimal selling strategy of a monopolistic data broker in the third chapter.

Consumer privacy

The literature on economics of privacy and price discrimination has widely shown that less privacy is better for all consumers when information is allocated to both competing firms since discriminatory prices are efficient, meaning that firms can extract through personalized prices at each location only the value of the reduction in transportation costs when buying the product from the closest outlet, and that all these prices are weakly below the full privacy Nash equilibrium price. From a consumer protection point of view, no privacy at all would be optimal. Instead, consumers are collectively worse-off under exclusivity.

Common wisdom suggests that more information in the market should benefit consumers (i.e. CS^{fp} should be a lower bound on surplus). I find instead that overall consumer surplus is reduced when there is partial information in the market, showing that in aggregate

$$CS^{np} > CS^{fp} > CS(x,0) > CS(x,x) > CS(x,y).$$

Consumers are overall worse-off when there is non exclusive access to partially informative and different data about their preferences. From a privacy protection perspective, consumers like extreme cases. They would prefer either full privacy or, if it is inevitable to be targeted, they would opt for no privacy at all, leading to individual targeting and fierce competition between firms. However, despite the clear policy conclusion that this result may suggest, totally banning or completely not regulating the use of information may not be beneficial for all consumers: under partial information, it is true that some of them are exploited with higher personalized prices, but others receive a tailored price that is truly a discount with respect to the Nash equilibrium uniform price. Indeed, when firms can partially price discriminate along one dimension only, some personalized prices are above t whereas other prices are below t. In the symmetric case, some consumers strategically but inefficiently switch between outlets when the information regime changes, accordingly to the realization of their type y, in order to benefit from tailored discounts. In the asymmetric case, efficiency holds and some consumers receive a discount as well. Moreover notice that under full information, which provides an upper bound on surplus, some consumers are in the worst possible scenario, given that efficient prices in two dimensions approach $2t$ for increasingly captive consumers. There are winners and losers in each scenario, which makes it hard to draw an unambiguous policy conclusion regarding consumer privacy.

The most interesting comparison is among the full privacy and the symmetric partial information case, which shows the largest inefficiency. All other comparisons directly follow from what is shown here. Recall that without information the market boundary is $y^*(x) = 1 - x$ and transportation costs are minimized. When the information regime changes, the market boundary $y^*(x) = \frac{2-x}{3}$ rotates anticlockwise. As shown in Figure 2.10, in order to benefit from a relatively lower tailored price, some consumers are willing to incur in a larger transportation cost as they go to the farthest outlet. Partial information generates a misallocation of consumers among the duopolists: some customers buy the "wrong" product. Therefore, under symmetric partial price discrimination, in contrast to full privacy, we can identify

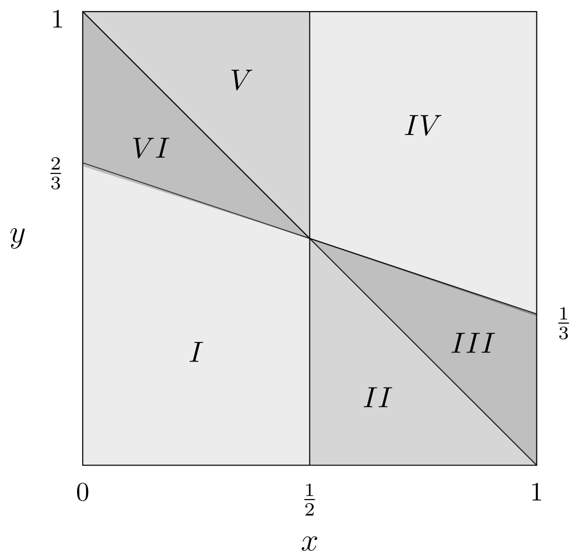


Figure 2.10: Consumers' allocation among firms: from full privacy to symmetric partial information.

three types of consumers:

1. Consumers in the sets $I = \left\{(x,y) : 0 \leq x < \frac{1}{2}, 0 \leq y < \frac{2-x}{3}\right\}$ and $IV = \left\{(x,y) : \frac{1}{2} \leq x \leq 1, \frac{2-x}{3} \leq y \leq 1\right\}$ are strictly worse-off but buy from the same firm;

2. Consumers in the sets $II = \left\{(x,y) : \frac{1}{2} \leq x \leq 1, 0 \leq y < 1-x\right\}$ and $V = \left\{(x,y) : 0 \leq x < \frac{1}{2}, 1-x \leq y \leq 1\right\}$ are strictly better-off and buy from the same firm;

3. Consumers in the sets $III = \left\{(x,y) : \frac{1}{2} \leq x \leq 1, 1-x \leq y < \frac{2-x}{3}\right\}$ and $VI = \left\{(x,y) : 0 \leq x < \frac{1}{2}, \frac{2-x}{3} \leq y < 1-x\right\}$ switch between firms and buy a mismatched product.

There is an inefficient flow of consumers between outlets. The switchers avoid the high personalized price of the nearest competing firm, and prefer to get the product from the farthest firm at a low tailored price. In aggregate, the positive effect on the switchers' net utility coming from the discount is perfectly offset by the increase in transportation costs, so that the overall surplus of switchers does not vary. However, as shown in Figure 2.11, only switchers located along the line $\frac{5-4x}{6}$ are really indifferent between the two information regimes: switchers located relatively far from the newly chosen retailer incur in an additional transportation cost that outweighs the discount; only the others actually have a net benefit from a discounted tailored price. Figure 2.11 gives a graphical intuition for the net increase in profits: the fraction of losers is clearly larger than the area of winners.

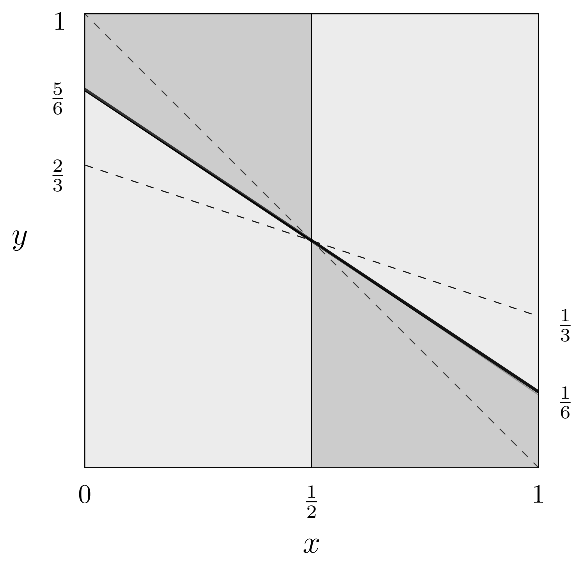


Figure 2.11: Indifferent consumers under symmetric partial information.

Firms' interests to acquire and use personal data are partially aligned with consumers' interests. A fraction of consumers would agree with the disclosure of personal information whereas others would prefer to conceal information. In the one dimensional literature instead these interests are always misaligned, as prices move in one direction only when both firms are informed.

Finally, it is worth to briefly analyze what happens in the asymmetric and exclusive data regimes, with respect to the symmetric case. Let us consider the exclusive case first. Inefficiency is partially mitigated, given that the market boundary rotates clockwise. More consumers buy the right product but, provided that price dispersion increases, some consumers are served by the informed firm at an even lower price, while other customers are charged more. The efficiency gain is sufficient to have a slight increase in consumer surplus. Exactly the opposite holds in the asymmetric partial information case. The market boundary is efficient, but consumer surplus reaches a lower bound. Price dispersion is the same as in the exclusive case, with the additional feature that now two differently informed firms target consumers with both high and low personalized prices. Efficiency ensures that total welfare is the same as in the two benchmarks, but the ability of firms to extract surplus from consumers is maximized. From a consumer privacy perspective, having two competing

	Gross profits	Consumer surplus	Total welfare
Full privacy	t	$v - \frac{5t}{3}$	$v - \frac{2t}{3} \uparrow$
No privacy	$\frac{2t}{3} \downarrow$	$v - \frac{4t}{3} \uparrow$	$v - \frac{2t}{3} \uparrow$
Symmetric partial info	$\frac{28t}{27}$	$v - \frac{47t}{27}$	$v - \frac{19t}{27}$
Asymmetric partial info	$\frac{7t}{6} \uparrow$	$v - \frac{11t}{6} \downarrow$	$v - \frac{2t}{3} \uparrow$
Exclusive partial info	$\frac{25t}{24}$	$v - \frac{83t}{48}$	$v - \frac{11t}{16}$

Table 2.1: Equilibrium outcomes (the arrows identify the maximum and minimum of each column).

firms endowed with different dimensions of private information generates the worst outcome possible. For completeness, we report in Table 2.1 a summary of all the equilibrium outcomes characterized so far. Total welfare is the same when efficiency is achieved.

2.3 Extensions

2.3.1 Exclusive full information

Suppose that firm 1 has access to a dataset containing full information on each single consumer's willingness to pay while firm 2 is uninformed. When the exclusively informed firm is able to set a different price for each consumer it will optimally set the individual price accordingly to

$$p_1(x, y) = \max\{0, p_2 + t(1 - 2x) + t(1 - 2y)\} \tag{2.3}$$

which directly follows from the indifference condition at prices $p_1(x, y)$ and p_2. The intuition is that the informed firm, having the exclusive advantage of being able to identify individual locations, makes each consumer just indifferent between the two products. Then it is possible to show the following result.

Lemma 4. *When only firm 1 has full information the uniform price is* $p_2^* = \frac{2t}{3}$ *and the personalized price is* $p_1^*(x,y) = 2t\left(\frac{4}{3} - (x+y)\right)$*. Profits are equal to* $\pi_2 = \frac{4t}{27}$ *and* $\pi_1 = \frac{62t}{81}$*.*

Proof. See Appendix 2.4. ■

Profits are equal to

$$\pi_2(0, xy) = \int_{\frac{4}{3}-x}^{1} \int_{\frac{1}{3}}^{1} p_2^* \, dxdy = \frac{4t}{27}$$

$$\pi_1(xy, 0) = \int_0^1 \int_0^{\frac{1}{3}} p_1^*(x,y) \, dxdy + \int_0^{\frac{4}{3}-x} \int_{\frac{1}{3}}^{1} p_1^*(x,y) \, dxdy = \frac{62t}{81}.$$

The informed firm is not only better-off with price discrimination but there is also a *business stealing* effect: firm 2 is more aggressive but it serves less consumers. This implies that "full" exclusivity is detrimental to the uninformed firm, differently from the case of "partial" exclusivity characterized in Section 2.2.3. When only one firm has information on a unique dimension, an exclusive allocation does not impose a negative externality on the uninformed firm. Instead, when only one firm has information on both dimensions, such negative externality plays again a role, as it is standard in the literature on selling data to competing firms.

2.3.2 Asymmetric dimensions

So far the model was perfectly symmetric in the degree of product differentiation across the two dimensions. Here we generalize the differentiation parameters while keeping the two dimensions of consumer information symmetric. Therefore, suppose that x and y remain orthogonal and both uniformly distributed on $[0,1]^2$. However, we assume that t_x and t_y are different. In the full privacy benchmark, the uniform Nash equilibrium price is $p_i^* = t_y$ for $t_x < t_y$ and $p_i^* = t_x$ for $t_x \geq t_y$, with equilibrium profits equal to $\pi_i^{fp} = \frac{t_y}{2}$ and $\pi_i^{fp} = \frac{t_x}{2}$ respectively. The efficient market boundary is $y^*(x) = \frac{1}{2} + \frac{t_x}{t_y}\left(\frac{1}{2} - x\right)$.

Symmetric partial information

Suppose that only dimension x is observable to both firms. The locus of indifferent consumers now writes

$$y(x) = \frac{1}{2}\left(1 + \frac{t_x}{t_y}\right) + \frac{p_2 - p_1}{2t_y} - \frac{t_x}{t_y}x. \tag{2.4}$$

The condition on prices so to have both firms active in the market trivially becomes $|\Delta_p| < t_x + t_y$. By solving for the firms' first order conditions, and for strictly positive values of t_x and t_y, equilibrium price schedules are

$$p_1^*(x) = t_y + \frac{t_x}{3}(1 - 2x)$$
$$p_2^*(x) = t_y + \frac{t_x}{3}(2x - 1)$$

These prices are positive for any value of x when $0 < t_x \leq 3t_y$ holds. In this range of the parameters the equilibrium market boundary is $y^*(x) = \frac{1}{2} + \frac{t_x}{6t_y}(1-2x)$ and it is interior for all $x \in [0, 1]$. Firm 1, and symmetrically firm 2, makes a profit equal to

$$\pi_1^I(x, x) = \int_0^{y^*(x)} \int_0^1 p_1^*(x) f(x) f(y) dx dy = \frac{t_x^2}{54t_y} + \frac{t_y}{2}.$$

Instead, for $t_x > 3t_y$ equilibrium prices write

$$p_1^*(x) = \begin{cases} t_y + \frac{t_x}{3}(1 - 2x) & \text{if} \quad 0 \leq x < \frac{1}{2} + \frac{3t_y}{2t_x} \\ 0 & \text{if} \quad \frac{1}{2} + \frac{3t_y}{2t_x} \leq x \leq 1 \end{cases}$$

and

$$p_2^*(x) = \begin{cases} 0 & \text{if} \quad 0 \leq x \leq \frac{1}{2} - \frac{3t_y}{2t_x} \\ t_y + \frac{t_x}{3}(2x - 1) & \text{if} \quad \frac{1}{2} - \frac{3t_y}{2t_x} < x \leq 1. \end{cases}$$

where $\bar{x} = \frac{1}{2} + \frac{3t_y}{2t_x}$ and $\underline{x} = \frac{1}{2} - \frac{3t_y}{2t_x}$ are the intercepts of the equilibrium market boundary with the lower and upper side of the unit square, respectively. Thus, in this case firm 1's profit is given by

$$\pi_1^{II}(x, x) = \int_0^1 \int_0^{\underline{x}} p_1^*(x) f(x) f(y) dx dy + \int_0^{y^*(x)} \int_{\underline{x}}^{\bar{x}} p_1^*(x) f(x) f(y) dx dy = \frac{t_x^2 + 6t_x t_y - 3t_y^2}{12t_x}.$$

Before moving to the next subgames, it is instructive to analyze what happens for limiting values of the parameters. First of all, it is useful to establish the following equivalence: to study $t_x \to \infty$ is equivalent to study $t_y \to 0$ and viceversa. The intuition is that as t_x grows indefinitely large, consumers can be viewed as increasingly homogeneous in dimension y (i.e. $t_y \to 0$), provided that any finite heterogeneity in that dimension would be negligible. For convenience, let us rearrange (2.4) in terms of x, which yields

$$x(y) = \frac{t_x + t_y - p_1 + p_2 - 2t_y y}{2t_x}. \tag{2.5}$$

As t_y approaches zero, only the observable dimension x matters for the consumers' choice between the two outlets. Indeed, the marginal consumer simplifies to $x^* = (t_x + p_2 - p_1/2t_x)$ and the personalized price schedules coincide with the one dimensional tailored prices à la Thisse and Vives (1988), namely $p_1^*(x) = \max\{t_x(1-2x), 0\}$ and $p_2^*(x) = \max\{t_x(2x-1), 0\}$. When both firms have access to information, the unit square is split in half accordingly to the vertical line $x^* = 1/2$ (i.e. $\underline{x} = \bar{x} = 1/2$). It is intuitive that when the unobservable dimension plays no role in consumer choices, the model is *de facto* one dimensional. In terms of data sales, the implication is that an hypothetical data broker would necessarily prefer to grant exclusive access to data. It is less straightforward to analyze the case in which $t_y \to \infty$. Consumers are homogeneous in the observable dimension and they select their preferred product accordingly to the distance in the (unobservable) y dimension. The candidate price equilibrium must be in uniform prices[27]. By noticing that this scenario is equivalent to $t_x \to 0$, in a symmetric equilibrium it must be that consumers, if they buy from one of the two outlets, split between the two firms accordingly to the horizontal line $y^* = 1/2$. When $t_x = 0$, consumers buy the product from the closest firm in the y dimension at a price $p_i = t_y$ as long as their net utility is positive.

The additional intuition that this generalization delivers is that, when firms cannot observe one of the two dimensions, they effectively "separate" them in the price-setting problem: the equilibrium price schedules are the sum of two terms. The uncertainty in the dimension y is captured by the uniform component t_y, whereas competition in the observable dimension x generates the personalized component that appears in the price schedules. The solution from the firms' perspective is simple: when t_x and t_y are asymmetric, it is more profitable to jointly identify consumers along the *less differentiated* dimension. In other words, it is optimal to condition prices on x when $t_x < t_y$.

[27]If information on y had been observable as well, then price schedules would have been à la Thisse and Vives (1988) also in this second case: $p_1^*(y) = \max\{t_y(1-2y), 0\}$ and $p_2^*(y) = \max\{t_y(2y-1), 0\}$.

Asymmetric partial information

Now we turn to the case in which not only the differentiation parameters are asymmetric, but so it is also the type of partial information held by firms. Suppose, as in Section 2.2.2, that firm 1 has x and sets a price $p_1(x)$ whereas firm 2 has y and sets $p_2(y)$. By standard procedures[28] we derive the equilibrium prices

$$p_1^*(x) = \frac{2t_y}{3} + \frac{t_x}{6}(5-6x)$$
$$p_2^*(y) = \frac{2t_x}{3} + \frac{t_y}{6}(6y-1).$$

As noted previously, each price schedule can be interpreted as the sum of two terms: a first term related to the unobserved dimension and a personalized component. When firms have asymmetric partial information, this decomposition is even more evident than in the previous subgame. Notice also that only for $\frac{t_y}{4} < t_x < 4t_y$ personalized prices are simultaneously positive[29]. The equilibrium market boundary in this case is drawn by the line $y^*(x) = \frac{5t_x+t_y}{6t_y} - \frac{t_x}{t_y}x$ (or equivalently $x^*(y) = 1-y(x)$), which is interior for $\frac{t_y}{4} < t_x < t_y$. Otherwise, firms are able to monopolize close market segments. Thus, firm 1's profit for $\frac{t_y}{4} < t_x \leq t_y$ is

$$\pi_1^I(x,y) = \int_0^{y^*(x)}\int_0^1 p_1^*(x)f(x)f(y)dxdy = \frac{5t_x}{18} + \frac{7t_x^2}{36t_y} + \frac{t_y}{9}$$

Instead, when $t_y < t_x < 4t_y$, the profit equals

$$\pi_1^{II}(x,y) = \int_0^1\int_0^{\underline{x}} p_1^*(x)dxdy + \int_0^{y^*(x)}\int_{\underline{x}}^{\bar{x}} p_1^*(x)dxdy = \frac{1}{72}\left(25t_x + 40t_y - \frac{23t_y^2}{t_x}\right)$$

where $\underline{x} = \frac{5(t_x-t_y)}{6t_x}$ and $\bar{x} = \frac{5t_x+t_y}{6t_x}$ are the intercepts of $y^*(x)$ with the upper and lower side of the unit square. We can write the profit of the rival symmetrically, switching t_x with t_y in the above expressions and noticing that, from the point of view of firm 2, $x^*(y)$ is interior for all y in the second case, so that

$$\pi_2^I(y,x) = \frac{1}{72}\left(25t_y + 40t_x - \frac{23t_x^2}{t_y}\right) \qquad \pi_2^{II}(y,x) = \frac{5t_y}{18} + \frac{7t_y^2}{36t_x} + \frac{t_x}{9}.$$

[28] The argument is similar to the proof of Lemma 2 in Section 2.2.2.

[29] For external values of the parameters one of the two firms prefers to deviate to a uniform price, given that p_i^* is always larger than the average price schedule. The intuition goes back to the decomposition of the optimal price schedule: when the differentiation parameters are "too asymmetric", the uniform component (i.e. the intercept of the schedule) would be excessively low.

What is reassuring is that all these quantities converge to $\frac{7t}{12}$ when the two dimensions of differentiation become symmetric. Moreover, it can be easily shown that $\pi_1^I(x,y) < \pi_2^I(y,x)$ but $\pi_1^{II}(x,y) > \pi_2^{II}(y,x)$. Differently from a scenario in which firms are partially informed on the same dimension, here players prefer to have partial information about the *more differentiated* dimension. When the rival firm has access to information on y, it is optimal to condition prices on x when $t_x > t_y$.

Exclusive partial information

Let us consider the subgame in which there is partial exclusivity. Suppose that only firm 1 has information, so that its maximization problem is the standard one, yielding a best response equal to $b_1(p_2) = \frac{t_y + t_x(1-x) + p_2}{2}$. Firm 2's objective function instead writes

$$\pi_2 = p_2 \left(1 - \int_{x \in [0,1]} y(x)dx\right)$$

where $y(x)$ is defined in (2.4). By standard procedures[30] we characterize the optimal uniform price starting from the average best response of the uninformed firm $b_2(\bar{p}) = \frac{t_y + \bar{p}}{2}$, where $\bar{p}$ is the average of the best response of the informed firm. Notice that only the differentiation parameter attached to the unobserved dimension of consumer heterogeneity appears in firm 2's average best response: it is possible to show that $p_2^* = t_y$ and $p_1^*(x) = t_y + \frac{t_x}{2}(1-2x)$. The equilibrium market boundary $y(x) = \frac{1}{2} + \frac{t_x}{4t_y}(1-2x)$ is interior for $0 < t_x \leq 2t_y$, which is a narrower range than in the symmetric partial information case: when only one firm has information and the two dimensions of heterogeneity are asymmetric, it is easier to monopolize close consumer groups. Profits are equal to $\pi_1(x,0) = \frac{t_y}{2} + \frac{t_x^2}{24t_y}$ and $\pi_2(0,x) = \frac{t_y}{2}$. Similarly to the main part of the paper, the uninformed firm gets the full privacy payoff.

2.4 Main highlights

Motivated by the huge collection and trade of consumer data in online markets, I study the impact of information on competition when price discrimination is a feasible pricing strategy, providing a complete characterization of the effects of different privacy regimes on profits and consumer surplus, in a model in which the information structure is slightly more complex and realistic. The major implication of this chapter is that partial information always relaxes price competition and, moreover,

[30]See the proof of Lemma 3 in the Appendix 2.4 for a similar argument.

partial exclusivity does not harm the uninformed firm, so that firms are better-off with price discrimination for any type of information structure but, with respect to consumer privacy, there are winners and losers among consumers in all the studied scenarios, implying that from a consumer protection point of view the policy recommendation remains unclear.

When the structure of consumer information is two dimensional (x, y) and firms observe only one dimension, the *rent extraction* effect is strengthened: each firm obtains a larger profit than in the full privacy benchmark either when the rival has the same data or different data. In terms of individual payoffs, the latter case dominates the first one. Not surprisingly, an exclusively informed firm is better-off with price discrimination, so that information acquisition is always a dominant strategy; however, an exclusive data allocation does not impose a negative externality on the uninformed firm: when only one firm gets the ability to partially target consumers while the rival is forced to price uniformly, the uninformed firm can still get the full privacy payoff. In terms of industry profits, no privacy strengthens the *competition* effect, so that firms are worse-off in the extreme cases (i.e. full privacy or no privacy), and an inverse U-shaped relationship between profits and quantity of data available to firms arises when partial information is introduced. Absent an upstream data seller, in order to be able to condition prices, firms would have to invest autonomously in tracking and data collection capabilities. In turn, it is possible to argue that firms would prefer to avoid the development of a perfectly accurate technology but rather to employ a partially accurate one. Crucially, firms would not face the usual coordination issue of information acquisition games, as they have aligned incentives to exploit partially informative consumer data. It is important to stress that the inability to observe a portion of relevant consumer information is crucial for the increase in industry profits. When information is partial the standard Bertrand competition argument at each location breaks down, and firms find it optimal to increase prices to close consumers in the observable dimension, even though this strategy triggers a mismatch of consumers in the unobserved dimension. Far away consumers are induced to inefficiently switch to the rival firm. This mechanism works only if the model is two dimensional. Indeed, when the unobserved dimension becomes irrelevant, in a symmetric equilibrium with price discrimination the closest outlet charges at each x a personalized price that equals the saving in transportation costs of that consumer and makes a profit equal to $\frac{t}{4}$ (Thisse and Vives, 1988).

The impact of information on consumer privacy is instead ambiguous: in comparison to the full privacy benchmark, while some consumers would prefer to hide from price discriminating firms, others prefer to receive a tailored offer, irrespective of partial or full information. Therefore, to ban completely the use of consumer information would not benefit all consumers. Nor it would be beneficial to allow individual targeting, as instead is predicted by one dimensional models, where price discrimination under best response asymmetry causes *all* consumers to pay less. Here, even though it is true that aggregate consumer surplus reaches an upper bound in the no privacy benchmark, still *some* customers are fully exploited and pay the highest admissible price. Obviously this result has to be taken cautiously as it depends on the information structure of the model and the type of consumer data exogenously available to the firms. The ambiguity of the effects of price discrimination on consumer surplus at the individual level seems to call for a more nuanced approach to privacy protection. Different types would have diverging preferences when deciding whether to opt for privacy, either complete or partial, or for information disclosure. A regulation that allows each consumer to make an informed choice about disclosure (or concealment) of personal information in on-line markets seems to point in the right direction. Indeed, this is the standard adopted in the General Data Protection Regulation (EU2016/679) which assigns to data subjects the right to consent with personal data collection and exploitation. The GDPR is centered around this empowerment of data subjects. To some extent, privacy is granted by default and individuals hold the right to directly enforce their personal privacy if needed. Data brokers have to obtain a clear and affirmative consent from users prior to collection of their personal information. However, many on-line services require such consent as a condition *sine qua non* for accessing the service itself, which implies that the choice to avoid giving consent is not really a viable option for users when close substitutes are not at hand. The service terms of many platforms are shown on a take-it-or-leave-it basis, so that the user does not have a real choice.

Appendix

A.1 Monopoly

Here we briefly present another benchmark, in addition to the full and no privacy benchmarks established under competition. Without loss of generality let us assume that firm 1, located in $(0,0)$, is a monopolist in the market. Consumers get a net utility equal to

$$u = v - p - tx - ty$$

when buying from firm 1 at price p. The choke price of demand is equal to v whereas at a price $p = v - 2t$ all consumers would buy the product. We keep the assumption that $v \geq 2t$ in order to be consistent with the main body of the paper.

Full privacy. The optimal monopoly price must lie in the interval $(v - 2t, v)$. The location of the indifferent consumers is given by $v - p - tx - ty = 0$, which can be rewritten as

$$y(x) = \frac{v-p}{t} - x.$$

When $p > v - t$ the indifference line is below the $1 - x$; otherwise it lies above the bisector of the unit square as shown in Figure 2.12.

Case I: $p > v - t$. Demand can be derived as

$$D^I = \int_0^{\hat{x}} \left(\frac{v-p}{t} - x \right) dx = \frac{(v-p)^2}{2t}$$

where $\hat{x} = \frac{v-p}{t}$. Solving for the first order conditions we obtain $p = v$ and $p = \frac{v}{3}$. The first solution implies zero profits and the second one violates the condition on prices.

Case II: $p \leq v - t$. Demand is characterized as

$$D^{II} = \frac{v-p-t}{t} + \int_{\tilde{x}}^1 \left(\frac{v-p}{t} - x \right) dx = 1 - \frac{(v-p-2t)^2}{2t^2}.$$

Solving for the first order condition we get $p = \frac{1}{3}(2v - 4t \pm z)$ where $z = \sqrt{v^2 - 4tv + 10t^2}$. Given that prices must be nonnegative, the equilibrium uniform price is $p^{fp} = \frac{1}{3}(2v - 4t + z)$. The uniform price converges to $\sqrt{\frac{2}{3}}t$ as $v \to 2t$. The equilibrium market boundary is $y^*(x) = \frac{v+4t-z}{3t} - x$ and it is equal to $\frac{(6-\sqrt{6})}{3} - x$ as $v \to 2t$. Unless

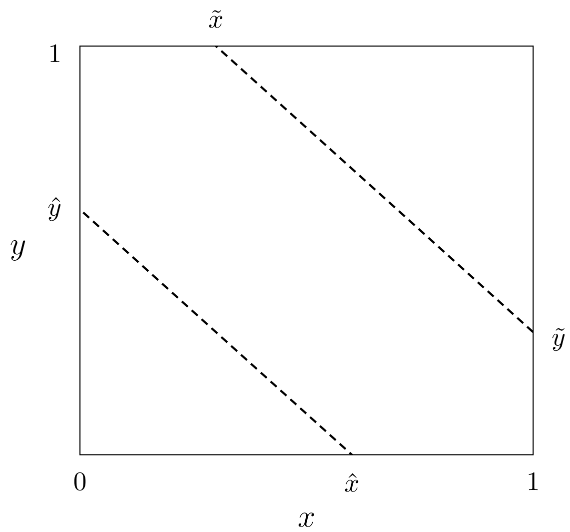


Figure 2.12: Demand structures for a monopolist located in $(0, 0)$.

v becomes extremely large the market is not entirely covered and the monopolist makes a profit equal to

$$\pi^{fp} = \frac{2tv(3v - 2z) - v^2(v - z) + 2t^2(3v + 5z) - 28t^3}{27t^2}.$$

Partial information. Suppose that the monopolist is able to segment consumers into groups accordingly to x. The objective function is

$$\pi(x) = p\left(\frac{v - p}{t} - x\right)$$

for each x. Taking the first order condition yields $p^*(x) = \frac{v-tx}{2}$. The market boundary is given by $y^*(x) = \frac{v}{2t} - \frac{x}{2}$ so that not all consumers are served in equilibrium. Therefore monopoly profit with partial information is equal to

$$\pi(x) = \int_0^1 \int_0^{\bar{x}} \left(\frac{v - tx}{2}\right) dxdy + \int_0^{\frac{v}{2t} - \frac{x}{2}} \int_{\bar{x}}^1 \left(\frac{v - tx}{2}\right) dxdy = \frac{6tv^2 - v^3 - 3t^2v - 3t^3}{12t^2}.$$

where $\bar{x} = \frac{v-2t}{t}$. Profit converges to $\frac{7t}{12}$ as $v \to 2t$.

No privacy. The monopolist can extract the entire surplus from each consumer. Intuitively, the personalized price is equal to $p^*(x, y) = v - t(x+y)$ for each consumer (x, y) and the entire market is covered. The monopoly profit under no privacy is maximized and it is equal to $\pi^{np} = v - t$ while consumer surplus is zero.

More consumers are increasingly served by the monopolist as we move from full privacy to no privacy. The monopolist is better off under no privacy, differently from competing firms. At the same time, from a consumer protection point of view, each consumer is fully exploited when the monopolist makes individual price offers. Not surprisingly, no privacy is detrimental to consumers under monopoly.

A.2 Proof of Lemmas

Proof of Lemma 1. Define $\Delta p = p_1 - p_2$. The fundamental equation for the analysis is

$$\tilde{y}(x) = (1 - x) - \frac{\Delta p}{2t} \tag{2.6}$$

In the above expression x is unknown to the players. Consider the extremes of the distribution of x, which is known to firms: the types with the lowest realization of x buy from firm 1 if $y < \tilde{y}(0)$, whereas those with the highest realization of x acquire 1's product if $y < \tilde{y}(1)$, with $\tilde{y}(1) < \tilde{y}(0)$. Since a low realization of x implies a preference for product 1 in the x dimension, relatively more consumers in the y dimension prefer firm 1 when x tends to zero, as captured by the negative unitary slope of $\tilde{y}(x)$. In particular: (i) when $\Delta p > 0$, $\tilde{y}(x)$ lies below $1 - x$, (ii) when $\Delta p < 0$, $\tilde{y}(x)$ lies above $1 - x$, whereas (iii) when $\Delta p = 0$ the indifferent consumers are located along the bisector of the unit square.

Consider firm 1 and fix p_2. Demand of firm 1 is necessarily zero whenever $p_1 \geq p_2 + 2t$ (i.e. $\Delta p \geq 2t$), whereas firm 1 captures the total mass of consumers for any $p_1 \leq p_2 - 2t$ (i.e. $-\Delta p \geq 2t$). Let us focus on interior cases ($-2t < \Delta p < 2t$) in which both firms have positive demand[31]. Moreover, prices are restricted to be non negative (i.e. above or at least equal to the marginal cost). The sign of Δp gives rise to two distinct segments of the demand function (notice that demand is continuous at $\Delta p = 0$, as shown later). We look for a symmetric equilibrium in uniform prices.

Case I: $\Delta p > 0$. The locus of indifferent consumers lies below the bisector. Thus, the intercepts with the axis are respectively on the left y-axis and the bottom x-axis. The coordinates are:

$$(0, \hat{y}) \;\Leftrightarrow\; \hat{y} = \frac{2t + p_2 - p_1}{2t} = \frac{2t - \Delta p}{2t}$$

[31] Therefore I simply denote with $\Delta p > 0$ cases in which $0 < \Delta p < 2t$ and with $\Delta p < 0$ cases in which $-2t < \Delta p < 0$.

$$(\hat{x}, 0) \;\; \Leftrightarrow \;\; \hat{x} = \frac{2t + p_2 - p_1}{2t} = \frac{2t - \Delta p}{2t}.$$

When firm 1 is pricing above the rival, the demand of firm 1 corresponds to the area of the triangle determined by these coordinates. Therefore

$$D_1^I = \int_0^{\hat{x}} F(y \leq \tilde{y}(x)) f(x) dx = \frac{(2t - \Delta p)^2}{8t^2}$$

whereas firm 2's demand is just the complement to one

$$D_2^I = 1 - D_1^I = \frac{4t^2 + 4t\Delta p - (\Delta p)^2}{8t^2}.$$

It is easy to show that for both firms $\partial D_i / \partial p_i = -\frac{1}{2t}\left(\frac{2t - \Delta p}{2t}\right) < 0$. However, notice that

$$\frac{\partial^2 D_1}{\partial p_1 \partial p_1} \geq 0 \qquad\qquad \frac{\partial^2 D_2}{\partial p_2 \partial p_2} < 0.$$

When $\Delta p > 0$ firm 1 is on the convex segment of its demand, whereas demand of firm 2 is concave (see Figure 2.13).

Case II: $\Delta p < 0$. The locus of indifferent consumers lies above $1 - x$. The intercepts with the axis are respectively on the right y-axis and the top x-axis. The coordinates are:

$$(1, \bar{y}) \;\; \Leftrightarrow \;\; \bar{y} = \frac{p_2 - p_1}{2t} = \frac{-\Delta p}{2t}$$

$$(\bar{x}, 1) \;\; \Leftrightarrow \;\; \bar{x} = \frac{p_2 - p_1}{2t} = \frac{-\Delta p}{2t}.$$

Notice that in this case it is easier to firstly derive the demand of firm 2 as the area of the triangle

$$D_2^{II} = \int_{\bar{x}}^1 \left(1 - F(y \leq \tilde{y}(x))\right) f(x) dx = \frac{(2t + \Delta p)^2}{8t^2}$$

and then the demand of firm 1 as

$$D_1^{II} = 1 - D_2^{II} = \frac{4t^2 - 4t\Delta p - (\Delta p)^2}{8t^2}.$$

For both firms we find again that $\partial D_i / \partial p_i = -\frac{1}{2t}\left(\frac{2t + \Delta p}{2t}\right) < 0$, whereas now

$$\frac{\partial^2 D_1}{\partial p_1 \partial p_1} < 0 \qquad\qquad \frac{\partial^2 D_2}{\partial p_2 \partial p_2} \geq 0$$

and the reverse holds true with respect to case I.

As anticipated before, notice that demand is continuous at the inflection point $p_1 = p_2$. Consider firm 1 and take the right and left limits of its price over the two demand segments characterized above:

$$\lim_{p_1 \to p_2^+} D_1^I(p_1, p_2) = \frac{1}{2} \qquad \lim_{p_1 \to p_2^-} D_1^{II}(p_1, p_2) = \frac{1}{2}.$$

Therefore, demand of firm i holding fixed the price of the rival j can be written as

$$D_i(p_i, p_j) = \begin{cases} 0 & p_i \geq p_j + 2t \\ \frac{(2t+p_j-p_i)^2}{8t^2} & p_j < p_i < p_j + 2t \\ \frac{1}{2} & p_i = p_j \\ \frac{4t^2+4t(p_j-p_i)-(p_j-p_i)^2}{8t^2} & p_j - 2t < p_i < p_j \\ 1 & p_i \leq p_j - 2t. \end{cases}$$

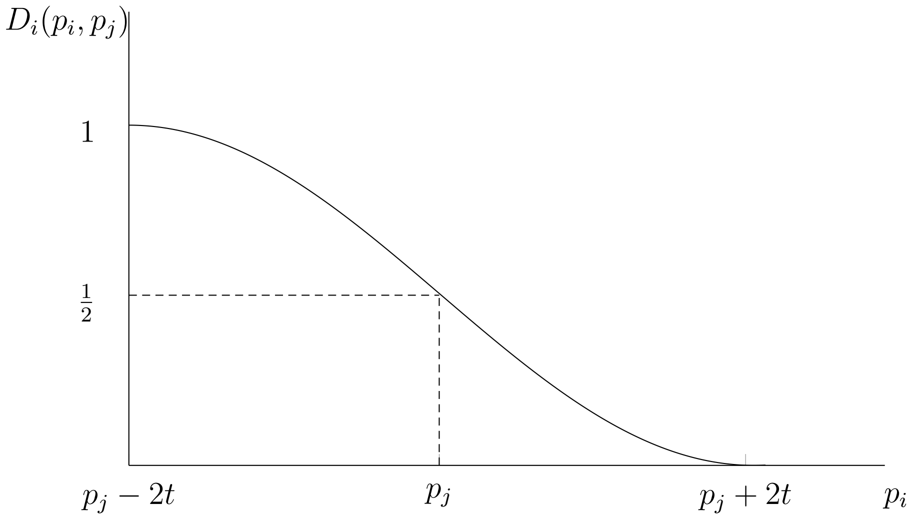


Figure 2.13: Demand of firm i as a function of p_i holding fixed p_j.

Finally, it remains to show that the solution to firms' maximization problems is the same under both structures of demand. Consider case I (case II is symmetric), in which firm 1 faces a convex demand whereas firm 2 faces a concave demand. First order conditions are quadratic in prices, and taking them equal to zero yields two best replies for each firm. Recalling that prices are restricted to be non-negative and that pricing above the rival's price by more than $2t$ leads firms out the market, we can disregard degenerate best replies that violates these conditions. Consider firm 1. Taking the first order conditions with respect to p_1 yields

$$b_1(p_2) = \frac{p_2 + 2t}{3} \qquad b_1(p_2) = p_2 + 2t.$$

It is immediate to see that the second response leads firm 1 out of the market, given that the firm sets a uniform price such that zero consumers are willing to buy the product for any price of the rival firm. Consequently, this best reply is eliminated. Similarly, it is possible to show that the unique best response of firm 2 satisfying the conditions on prices is $b_2(p_1) = (2p_1 - 4t + z)/3$ where $z = \sqrt{p_1^2 - 4tp_1 + 28t^2}$ (notice that the second computed response of firm 2 lies entirely in the quadrant $(p_1(+), p_2(-))$, and it does not even appear in Figure (2.14).

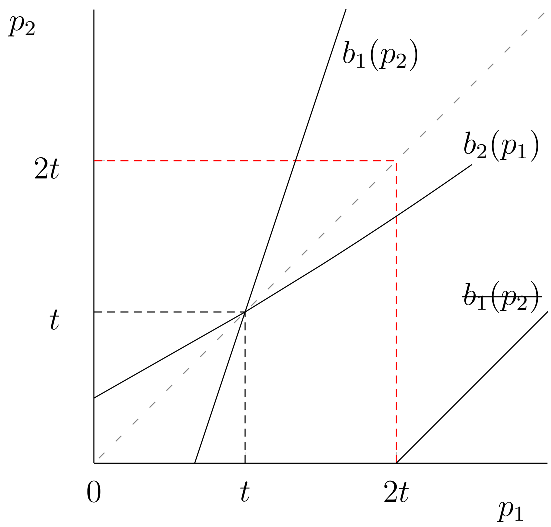


Figure 2.14: Best responses - not violating conditions on prices - in *Case I* ($\Delta_p > 0$).

Solving the system of best responses yields a unique equilibrium in positive prices: $p_1^* = p_2^* = t$.

■

Proof of Lemma 2. When firm 1 charges $p_1(x)$ and firm 2 sets $p_2(y)$ the indifference condition writes

$$p_1(x) + tx + ty = p_2(y) + t(1 - x) + t(1 - y).$$

Given that firms target consumers asymmetrically, firm 1 considers as indifferent consumers those located along

$$y(x) = \frac{p_2(y) - p_1(x)}{2t} + (1 - x)$$

and firm 2 considers the line

$$x(y) = \frac{p_2(y) - p_1(x)}{2t} + (1 - y).$$

Notice that $y(x)$ and $x(y)$ draw the same line within the unit square, so that $D_1 = y(x)$ and $D_2 = 1 - x(y)$. Existence and uniqueness of a discriminatory price equilibrium is proved in steps.

Part 1. Consider firm 1. For each group x, the rival is setting a continuum of prices in the y dimension. We will show that what matters for the optimization problem of a firm is only the average price of the rival. Therefore, we firstly solve for competition in average prices; the requirement is that in equilibrium the optimal price schedules must be equal to the average price derived in this part of the proof

$$\int_0^1 p_1(\tilde{x})dF(\tilde{x}) = \bar{p}_1 \qquad\qquad \int_0^1 p_2(\tilde{y})dF(\tilde{y}) = \bar{p}_2.$$

Firm 1 maximizes

$$\tilde{\pi}_1 = \int_0^1 p_1(\tilde{x})D_1(p_1(\tilde{x}), \bar{p}_2, \tilde{x})dF(\tilde{x})$$

and firm 2 maximizes

$$\tilde{\pi}_2 = \int_0^1 p_2(\tilde{y})D_2(p_2(\tilde{y}), \bar{p}_1, \tilde{y})dF(\tilde{y}).$$

Taking the first order conditions we get $\bar{p}_1 = \frac{\bar{p}_2+t}{2}$ and $\bar{p}_2 = \frac{\bar{p}_1+t}{2}$. The average price schedules are $\bar{p}_1 = \bar{p}_2 = t$.

Part 2. Now we derive the unique profit maximizing schedule satisfying the above constraint, taking into account the average discriminatory schedule of the rival[32].

[32]This procedure is equivalent to guessing linear schedules $p_1(x) = a - bx$ and $p_2(y) = \alpha + \beta y$, and plugging them into the optimization problem of the rival firm, taking the integral with respect to the information unobserved to that firm. The equilibrium values of $(a^*, b^*, \alpha^*, \beta^*)$ yields the same schedules derived in the two-step proof.

Firms' final objective functions therefore are

$$\pi_1 = p_1(x)\left(\frac{\bar{p}_2 - p_1(x)}{2t} + (1-x)\right), \forall x$$

$$\pi_2 = p_2(y)\left(1 - \left(\frac{p_2(y) - \bar{p}_1}{2t} + (1-y)\right)\right), \forall y.$$

Equilibrium prices are $p_1^*(x) = t(\frac{3}{2} - x)$ and $p_2^*(y) = t(\frac{1}{2} + y)$, with average schedules indeed equal to the transportation cost. The market boundary is $y^*(x) = 1 - x$ (or equivalently $x^*(y) = 1 - y$). Profits are

$$\pi_1 = \int_0^{1-x}\int_0^1 p_1^*(x)\ f(x)f(y)dxdy = \frac{7t}{12}.$$

$$\pi_2 = \int_{1-y}^1\int_0^1 p_2^*(y)\ f(y)f(x)dydx = \frac{7t}{12}.$$

Consumer surplus is equal to

$$\begin{aligned}CS = &\int_0^{1-x}\int_0^1 (v - p_1^*(x) - tx - ty)\ f(x)f(y)dxdy + \\ &\int_{1-y}^1\int_0^1 (v - p_2^*(y) - t(1-x) - t(1-y))\ f(y)f(x)dydx = v - \frac{11t}{6}.\end{aligned}$$

■

Proof of Lemma 3. When firm 1 charges a personalized price $p_1(x)$ and firm 2 a uniform price, the expression for the indifference line modifies to

$$y(x) = \frac{p_2 - p_1(x) + 2t(1-x)}{2t}.$$

As long as $|p_1(x) - p_2| < 2t$ for all x, the market boundary is interior, and the payoffs of the players are continuous. Players set prices simultaneously but player 2 is able to best reply only "on average" to the personalized price of player 1.

The objective functions of the two players are

$$\pi_1 = p_1(x)\left(\frac{p_2 - p_1(x)}{2t} + (1-x)\right), \forall x$$

and

$$\pi_2 = p_2\left(1 - \int_{x\in[0,1]}\left(\frac{p_2 - \bar{p}}{2t} + (1-x)\right)dx\right)$$

where $\bar{p}$ is the average of $p_1(x)$ over $x \in [0,1]$. Solving for the first order conditions we get

$$b_1(p_2, x) = \frac{p_2 + 2t(1-x)}{2} \qquad \text{and} \qquad b_2(\bar{p}) = \frac{\bar{p} + t}{2}$$

In order to get the optimal uniform price we plug the integral (i.e. the average) of the informed firm's best response into the above equation

$$p_2 = \frac{1}{2}\left(\int_{x\in[0,1]} \left(\frac{p_2 + 2t(1-x)}{2}\right) dx + t\right) dx$$

which yields $p_2^* = t$. The informed firm is always best responding by setting $p_1^*(x) = t(\frac{3}{2} - x)$ and the uninformed firm is best responding "on average". The equilibrium market boundary is equal to $y^*(x) = \frac{3-2x}{4}$ and final payoffs of the players are

$$\pi_2 = \int_{\frac{3-2x}{4}}^{1} \int_0^1 p_2^* \, f(x)f(y)dxdy = \frac{t}{2}$$

$$\pi_1 = \int_0^{\frac{3-2x}{4}} \int_0^1 p_1^*(x) \, f(x)f(y)dxdy = \frac{13t}{24}.$$

Consumer surplus is equal to

$$\begin{aligned} CS^{I,NI} = &\int_0^{\frac{3-2x}{4}} \int_0^1 (v - p_1^*(x) - tx - ty) \; f(x)f(y)dxdy + \\ &\int_{\frac{3-2x}{4}}^1 \int_0^1 (v - p_2^* - t(1-x) - t(1-y)) \; f(x)f(y)dxdy = v - \frac{83t}{48}. \end{aligned}$$

■

Proof of Lemma 4. Notice that the difference in transportation costs $t(1-2x) + t(1-2y)$ is negative for $x + y > 1$. Firstly we show that $p_2 = 0$ cannot be the equilibrium uniform price. When $p_2 = 0$, the price of firm 1 is given by eq. (2.3) and the informed firm serves all consumers with $x + y \leq 1$, whereas the other half of the market buys from firm 2. The market boundary is $y(x) = 1 - x$, the bisector of the unit square. However, this cannot be an equilibrium provided that $\pi_2 = 0$, which implies that firm 2 has an incentive to deviate and to set a price larger than zero, serving less consumers but making a positive profit.

Consider therefore a candidate equilibrium $p_2 > 0$. Let us firstly provide a geometric argument that simplifies the problem. When $p_2 > 0$ the market boundary must necessarily lie above the bisector: the informed firm is now able to match the net utility guaranteed by firm 2 also for some consumers with $x+y > 1$. The intuition is that in this region a positive price p_2 offsets the negative term $t(1-2x)+t(1-2y)$

that appears in eq.(2.3), and the more p_2 increases, the more the market boundary switches in the north-east direction. Indeed, the boundary is always identified by $p_1(x,y) = 0$ because, as long as the price of the informed firm is nonnegative, firm 1 has the advantage of "winning" the consumers at the margin by just matching the rival's offer. Therefore, the problem of characterizing the optimal uniform price reduces to the characterization of the set of locations (x, y) at which $p_1(x,y) = 0$ when $p_2 > 0$.

Moreover, notice that for each line that is parallel to the market boundary, the coordinates (x, y) are such that the transportation costs are the same along the entire line: these are the isocost lines within the unit square. Given that $p_1(x, y)$ depends only on x, y and p_2, which is the same for all consumers, it must necessarily be that along the isocost lines the price set by the informed firm is the same[33]. We can therefore further simplify the problem by focusing on locations along the curve $y = x$. Thus, set $x = y = z$ where $\frac{1}{2} < z < 1$. Then, for all $x + y = 2z$, the price of the informed firm will be the same by construction[34]. From the indifference condition we can write

$$p_2 = p_1(z, z) + t(4z - 2). \tag{2.7}$$

The marginal consumers are such that $p_1(z, z) = 0$ and the demand of firm 2 is then geometrically characterized in Figure (2.15) as $D(z) = \frac{(2(1-z))^2}{2}$.

The objective function of the uninformed firm is

$$\pi_2 = t(4z - 2)\frac{(2(1-z))^2}{2}.$$

Formally the equilibrium is characterized by the following equations:

$$\begin{cases} x = y = z \\ p_1(z, z) = 0 \\ 8t(1-z)^2 - 4t(1-z)(2-4z) = 0 \end{cases}$$

Solving with respect to z we get $z^* = \frac{2}{3}$. Substituting back into eq.(2.7) yields $p_2^* = \frac{2t}{3}$. From eq.(2.3) it follows that $p_1^*(x, y) = 2t\left(\frac{4}{3} - (x + y)\right)$, which is equal to zero along $y^*(x) = \frac{4}{3} - x$. Indeed, it is easy to verify that $x + y = 2\left(\frac{2}{3}\right)$: along

[33] In a three dimensional space, the personalized price of the informed firm can be represented as a negatively sloped plane with domain $[0, 1]^2$: if we cut the plane along the isocost lines, then we find the same price level for all (x, y) along each line.

[34] Basically, condition $x + y = 2z$ is equivalent to $x = y = z$.

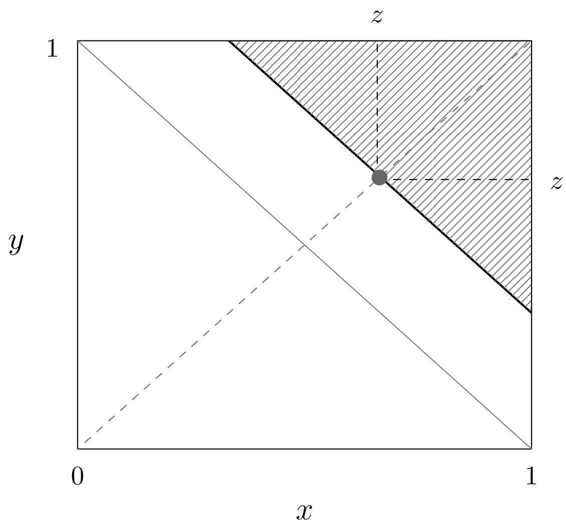


Figure 2.15: Demand of firm 2 when $p_2 > 0$.

this line consumers are indifferent between the two outlets at prices p_2^* and $p_1^*(x, y)$. Profits are equal to

$$\pi_2 = \int_{\frac{4}{3}-x}^{1} \int_{\frac{1}{3}}^{1} p_2^* \, dxdy = \frac{4t}{27}$$

$$\pi_1 = \int_0^1 \int_0^{\frac{1}{3}} p_1^*(x, y) \, dxdy + \int_0^{\frac{4}{3}-x} \int_{\frac{1}{3}}^{1} p_1^*(x, y) \, dxdy = \frac{62t}{81}.$$

■

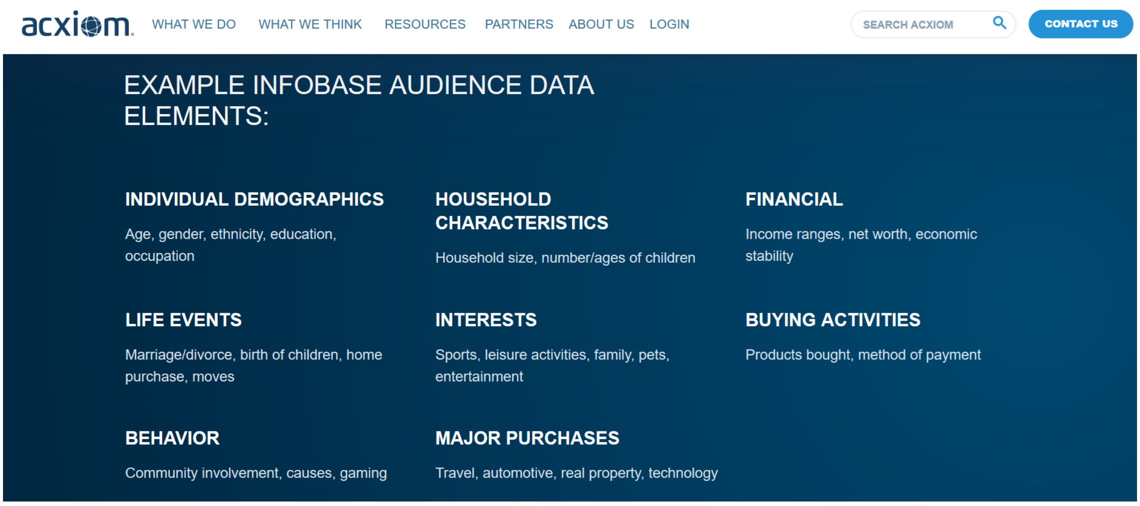
Chapter 3

Selling Information

The role of data collectors and data sellers, generally named data brokers, is far-reaching in the markets for information. Access to consumer data increases the ability of firms to segment and reach with personalized offers online users, and data sellers may find it optimal to discriminate among data buyers on the basis of their willingness to pay for consumer data (Pancras and Sudhir, 2007). For instance, an incentive to grant exclusive access to valuable consumer data to certain partners while foreclosing other firms emerged clearly in the recent Facebook case[35]. Ultimately, strategic behavior of data collectors and sellers and exclusive data access may harm consumers and the society as a whole (Duch-Brown et al., 2017). Economic literature on privacy and price discrimination has widely analyzed data brokers' incentives to sell (one dimensional) information when data buyers can exploit acquired information to make targeted price offers to consumers. Montes et al. (2018) show that in a duopoly characterized by horizontal product differentiation the seller has an incentive to induce maximal asymmetry from the point of view of downstream access to consumer data, therefore selling data exclusively to one competitor. Price competition is relaxed and retailers have a strong incentive to become the exclusive winner of consumer data. As a result, the data seller can extract the highest price for information. Kim et al. (2018) confirm such exclusivity result even when the downstream market is a triopoly. Clavorà Braulin and Valletti (2016) arrive to the same conclusion in a vertical differentiation duopoly, formalizing the conditions under which it is the high (low) quality firm to receive consumer data exclusively. A key feature of this literature on data brokers and consumer privacy is perfect price discrimination: the dataset can be sold either exclusively or not exclusively, but

[35] https://www.nytimes.com/2018/12/05/technology/facebook-documents-uk-parliament.html

only entirely, implying that firms can perfectly identify consumers tracked by the seller. This is clearly a great limitation of this strand of the literature. Other contributions assume that the data seller can adopt slightly more sophisticated strategies when it comes to sell information to competing firms. When the data broker has the possibility to partition the compiled dataset prior to the sale stage the exclusive equilibrium may not arise. Indeed, Bounie et al. (2018) show that the data broker sells information to both competitors. When the dataset put on sale can be optimally partitioned, the seller has an incentive to sell symmetric but not overlapping subsets of consumer data to firms in order to soften downstream competition. In a market for a homogeneous product, Belleflamme et al. (2019) show that a data broker always has an incentive to allocate information to both downstream firms, but only under vertical data differentiation: the quality of the dataset sold to firms must be different, implying that the duopolists can identify consumers with asymmetric but correlated technologies.



Source: `https://www.acxiom.com/what-we-do/infobase`

Figure 3.1: Data brokers collect multiple attributes about consumers.

The data market is populated by an enormous amount of firms that track consumers, collect and merge data from a variety of different sources, and create comprehensive lists of profiled consumers (Bergemann et al., 2018; Bonatti and Cisternas, 2018). However, this market is very opaque, in the sense that many connections between different firms active in the data market are quite shady (Brill et al., 2014). Yet, what clearly emerges is that data brokers collect data on several consumer attributes which concur to determine the consumer's type or profile. It seems therefore reasonable to investigate data sales in a framework, described in the second chapter, in which the willingness to pay of customers is determined by at least two characteristics. This setup is a minimal working example that allows me to capture the multi-dimensional nature of information sold about consumers in the data market. For instance, as it is possible to see in Figure 3.1, data brokers such as Acxiom have access not only to individual demographics but they gather many data pieces about consumers, ranging from past buying activities to interested and financial data.

In this chapter we analyze a sequential game in which information exchange or acquisition is feasible. Firstly we briefly discuss flows of consumer data between competing firms, when firms are already informed and the initial allocation of information is exogenously given. Then, in Section 2, we introduce a third party data seller which is exogenously endowed with or can collect consumer data and sells it to competing firms that do not have access to information on preferences. As highlighted in the literature on data sales, when the downstream data allocation is induced by an upstream seller, information is usually awarded exclusively in order to preserve the supra-competitive profit of the informed firm and, more importantly, to maximize the difference in profits between the informed firm and the uninformed one. In a two dimensional model, when partial privacy is enforced, exclusivity still holds when the data broker has information on a single dimension. However, if the dataset held by the seller is fully informative, then the data broker has an incentive to sell information non exclusively, awarding different subsets of consumer data to different competitors. This final result is somehow complementary to Bounie et al. (2018) and Belleflamme et al. (2019), even though the theoretical models built in their papers are different from the setup proposed here.

In Section 3 we will discuss a scenario in which the data market is populated by more data sellers, investigating the effects on the price and allocation of consumer data of a competitive upstream structure. A non exclusive downstream data allocation always arises given that information is potentially available from different sources, but conditionally on the type of data that each upstream seller holds, that allocation will be induced via different types of deals, that is, resorting to both exclusive or non exclusive contracts.

Finally, in Section 4, we relax the assumption that information can be collected and used only as partial (i.e. x or y) or full (i.e. x and y). The idea is that finer forms of data partitioning could be feasible and we further exploit the possibility of sectioning data about consumers' willingness to pay with less constraints in order to investigate whether consumers can be made further better-off. This analysis will be informative about the possibilities for a social planner to collectively improve the welfare of consumers beyond the no privacy outcome.

3.1 Data sales between competing fims

We consider a two dimensional model of horizontal product differentiation in which two firms exploit partially or fully informative data on consumer preferences to price discriminate (Section 2.1). Our primary focus in this introductory part of the chapter is to discuss a simple question: why should firms have an incentive to share or sell information with their rivals? Indeed, common wisdom and large part of the literature suggest that having exclusive access to consumer data gives a clear competitive advantage over rival firms, similarly to what would be an incumbency advantage as dicussed by Biglaiser et al. (2018). However, there may be instances in which a firm finds it profitable either to sell proprietary data to an uninformed competitor or to exchange data with another informed firm[36]. The relevant initial scenarios that we study therefore are: (i) one firm is partially informed about consumer preferences and the rival does not have any data, (ii) both firms are partially informed but they have access to different type of data[37], and (iii) only one firm has full information while the rival is not informed. It is possible to interpret this game as a two period model in which the first period allocations and outcomes were characterized in the second chapter. Then, at the beginning of the second period, firms can eventually trade or exchange data.

3.1.1 Partial but exclusive information

This case resembles the analysis in Section 2.2.3 and profits are reported in Lemma 3. We maintain the convention that the more informed firm, if any, is player 1 and that partial exclusivity implies that, for exogenous reasons such as data collection costs, only the x is collectible: for instance, the geographical location can be inferred when the consumer is browsing the web with active location tracking whereas more private information remains hidden to the competing firms.

It is immediate to see that trade occurs only if the condition $\pi_1(x, 0) < \pi_1(x, x)+$

[36] See the first chapter for an application without price discrimination but with product customization

[37] Clearly, the analysis of a symmetric partial information scenario has no relevance in this context.

$(\pi_2(x,x) - \pi_2(0,x))$ holds[38], which is equivalent to require that symmetric industry profits raise after the trade: $\pi_1(x,0) + \pi_2(0,x) < 2 \times \pi_i(x,x)$. Given the equilibrium payoffs previously characterized our condition is never met, even when the informed player has enough bargaining power to rip off the entire extra profit enjoyed by the counterpart, which seems to be a tenable assumption on the selling mechanism in this context. In our model $\pi(x,0) < \pi(x,x)$ holds and information is not shared.

This argument can be interestingly reversed to claim that if we start with a symmetric scenario, then one of the two firms is willing to sell its data, committing *de facto* to compete in uniform prices. At the same time, the rival is not acquiring data about dimension x *per se*, but it is buying the competitive advantage of remaining the unique firm to exploit this resource. The intuition is the same: partial exclusivity creates more value to the entire industry. The solely data owner gets an incremental profit equal to $\pi_1(x,0) - \pi_1(x,x) = \frac{5t}{216}$ which more than compensates the loss of the firm without data, that is measured by $\pi_2(x,x) - \pi_2(0,2) = \frac{t}{54}$. Firm 1 is therefore willing to pay a price P for information that is not *per se* valuable such that $P \in [\frac{t}{54}, \frac{5t}{216}]$: (i) if the price is equal to lower bound, then firm 2 is indifferent and *ex-post* only the data holder is better-off, whereas (ii) if the price hits the upper bound then the data owner is indifferent and *ex-post* the firm waiving away its data benefits from the trade. For any price within the two bounds both firms benefit when moving from a symmetric scenario to an exclusive one. To pin down such price we would need to additionally assume a profit sharing rule or how the bargaining power is eventually distributed across the two players.

3.1.2 Partial and different information

We now turn to the case in which different firms are able to identify consumers along different dimensions, building upon the analysis carried out in Section 2.2.2. Each player has data that is valuable to the rival. Suppose that firm 2, which knows only dimension y, is willing to sell data to the rival.

If the seller can credibly commit to a uniform price in the subsequent pricing game, then the relevant condition boils down to the comparison between $\pi_1(xy,0) - \pi_1(x,y) > 0$, firm 1's maximum willingness to pay for data about dimension y, and the loss incurred by the seller - firm 2 - which amounts to $\pi_2(y,x) - \pi_2(0,xy)$. Provided that industry profits under full exclusivity fall below the full privacy benchmark, it is clear that the even the highest price offered by firm 1 can never compen-

[38] The term in parenthesis which appears on the left hand side of the inequality is simply the price for data that firm 1 can get from the rival.

sate firm 2 for a commitment to compete in uniform prices, which is equivalent to claim that $\pi_1(xy, 0) + \pi_2(0, xy) < \pi_1(x, y) + \pi_2(y, x)$.

Consider now a scenario in which the seller cannot commit to a uniform price but rather, upon information sale, it will continue to condition prices along one dimension only. Firm 2 is not selling data but it is sharing data with the rival. If this is the case, firm 2 is able to limit its loss to the payoff $\pi_2(y, xy) > \pi_2(0, xy)$, whereas interestingly firm 1 does not benefit at all from the trade[39]. The willingness to pay for data without commitment is equal to zero. In general, when firms have access to asymmetric partial information data trade cannot arise in equilibrium, and the intuition is fairly straightforward: under this type of data allocation industry profits under competition reach an upper bound and any other allocation is therefore dominated. This argument is going to be relevant for the next subsection.

3.1.3 Fully exclusive information

Consider a big market player with the ability to target consumers individually along all data dimensions. An uninformed firm could be for instance a newly entered firm or an entrant not yet active in the market. Allowing data access could be seen as an accommodating strategy while blocking any type of data flow between firms could resemble an aggressive response to entry.

The outcomes of a scenario in which only one firm has an exclusive access to complete consumer data were characterized in Section 2.3.1. The distance between firms' payoffs is the largest one but industry profits fall below the no information benchmark. The trade-off for the data owner is clear: on the one hand, exclusive access to all the data is undoubtedly a major competitive advantage that maximizes the individual payoff; on the other hand, industry profits are not maximized and there is room for rent extraction. Across all the possible subgames, asymmetric partial information delivers the largest improvement in industry profits. Two conditions must be met in order to have a data flow between competing firms: (i) the data owner must commit to not use *ex-post* the part of information that is sold to the rival in order to condition prices, and (ii) the inequality $\pi_1(xy, 0) < \pi_1(x, y) + (\pi_2(y, x) - \pi_2(0, xy))$ must hold[40], where $P = \pi_2(y, x) - \pi_2(0, xy)$ represents as usual the price for data that makes the buyer indifferent between accepting the offer and rejecting it. Condition (i) is easily satisfied in the sense that individual targeting does not further

[39]This argument is developed later. Please see the proof of Lemma 5.

[40]We are assuming that firm 1 is selling data on dimension y. The analysis symmetrically holds for the other case.

improve the individual payoff of the seller[41] once the rival gets partially informed.

The interesting conclusion is that not only an exclusively informed firm is willing to dilute its informational advantage in favor of a competitor but also that it is committing to personalize prices in a less sophisticated way, moving from individual to group price targeting. Several policy considerations follow from this result. Firstly, within the debate surrounding dominant positions in data access that can undermine the flow of information in the data market, ultimately damaging competition, situations of "de facto" data ownership rise policy makers' concerns. Exclusive data ownership, as discussed in the first chapter, is considered to be one of the main threats to the free flow of information in the market. Our results, even though the model is a particular one, show that exclusive access to data is not an issue, provided that competition leads to a non exclusive allocation. Therefore intervention policies conceived around mandatory data sharing would be ineffective in this context, because the seller is effectively granting data access to the rival. Obviously, we have shown that the interest of the data holder is to induce in equilibrium a non exclusive allocation of partial data, while the policy maker would prefer a non exclusive allocation of full information. But once there is a data flow between firms, the regulator can hardly verify whether a partially informative or a fully informative dataset has been shared. Secondly, the regulatory interventions, such as the GDPR in EU, increasingly stress the need to scrutinize more severely situations in which consumers can be identified, in real time or ex-post, on an individual basis. In the context of our model, this would imply that a firm knows both x and y. There are requirements to anonymize data upon collection and before sharing them with other parties and, indeed, whereas individually personalized prices are looked at with suspicion, group price discrimination is well established in the market. In the light of the model predictions, the punchline therefore is twofold: a data seller sharing partial information with a rival firm simultaneously escapes potential concerns about its dominant data position and it complies with the prevailing approach of this regulation. However, we have also shown that price discrimination based on asymmetric partial information determines an overall reduction in aggregate surplus, which should be a warning that competition on almost equal grounds does not necessarily delivers an outcome favorable to consumers, even though, as extensively stressed in the second chapter, there are winners and losers among consumers.

[41] See the proof of Lemma 5 for the formal argument.

3.2 Monopolistic data broker

Here we employ again the same model of Section 2.1. The novelty is the presence of a data seller who observes consumer types either in the x or in the y dimension or both. The data seller gathers information at zero cost into datasets X, Y or XY. For the moment, the available consumer data is exogenously given. When data is acquired, the buyer receives the dataset entirely[42]. The timing of the game is changed to

1. The data seller observes consumer types and puts on sale consumer data.
2. If firm i does not acquire data, it sets a uniform price p_i; otherwise, it sets a personalized price conditional on the acquired information structure.
3. Consumers buy the product.

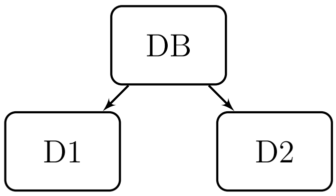


Figure 3.2: Monopolistic data broker.

We focus on the analysis of the first stage, in order to characterize the subgame perfect Nash equilibrium of the game.

3.2.1 Exogenous consumer data

Here we study the incentives of a data broker to sell partial information to downstream firms. The data seller posts a take-it-or-leave-it offer for the available information at the second stage. Payments are made at this stage and the information allocation among downstream firms becomes common knowledge before the pricing game. The seller is assumed to hold all the bargaining power. The selling mechanism exploited in the literature resembles an auction with downstream externalities (Jehiel and Moldovanu, 2000): when the dataset is exclusively offered to a competing

[42]The possibility to optimally partition the dataset before selling it is not considered in this paper. Bounie et al. (2018) investigate this issue building on the work of Liu and Serfes (2004), and they show that both firms receive information, although different and not overlapping partitions of the original dataset.

firm which does not buy consumer information, then the rival has the chance to acquire it. However, in a two dimensional model, partial exclusivity does not harm the uninformed player, differently from standard information acquisition games. This peculiarity of the outside option plays an important role.

In the following, the data broker is said to hold a partially informative dataset when he has collected only x (or equivalently, only y); the dataset is fully informative when both dimensions are collected. We keep the assumption that competing firms are allowed to know something about consumers but not everything or, in other words, that individual targeting is banned by a privacy policy. This is equivalent to say that the seller cannot grant full exclusivity (i.e. about both x and y) to a unique firm[43]. Two scenarios, depending on the type of information structure held by the seller, can arise.

Proposition 7. *When the dataset is partially informative, the seller sells partial information to only one firm at a price* $P^{EX} = \frac{t}{24}$.

	0	x
0	$t/2,\ t/2$	$t/2,\ 13t/24$
x	$13t/24,\ t/2$	$14t/27,\ 14t/27$

Table 3.1: Pricing game when the dataset is partially informative.

Proof. First of all, notice that the exclusive price[44] is equal to

$$P^{EX} = \pi_i(x, 0) - \pi_i(0, x) = \frac{t}{24}.$$

Suppose that the seller posts a take it or leave it offer $P_i^{NE} > 0$, where P_i^{NE} is the price at which both firms can acquire consumer data. The non exclusive price[45]

[43]The reader may wonder whether the seller could not be better-off precisely when granting full exclusivity to a single firm: in section 3.2.2 we prove that actually this is not the case.

[44]Bounie et al. (2018) consider a selling mechanism different from an auction with downstream externalities and write the exclusive price as the difference in profits $\pi_i(x, 0) - \pi_i^{fp}$, where the outside option is the standard Hotelling profit (i.e. our full privacy benchmark). They assume that the seller commits *ex-ante* to an exclusive deal with only one firm, without the possibility to offer the dataset to the rival in case of a rejection. Here these exclusive prices are equivalent, and therefore the difference between the selling mechanism of Bounie et al. (2018) and Montes et al. (2018) is immaterial.

[45]A more general expression for the price for information would be $P_i = \alpha \times (\pi_i(x, x) - \pi_i(0, x))$ and similarly for the exclusive price. The bargaining power measured by α is set equal to one in the proof.

writes

$$P_i^{NE} = \pi_i(x,x) - \pi_i(0,x) = \frac{t}{54}.$$

The seller therefore compares

$$\underbrace{2 \times (\pi_i(x,x) - \pi_i(0,x))}_{P^{NE}=\frac{t}{27}} < \underbrace{(\pi_i(x,0) - \pi_i(0,x))}_{P^{EX}=\frac{t}{24}}.$$

In the subgame perfect Nash equilibrium of the partially informative game the data seller grants information on a single dimension exclusively to one downstream competitor. ■

From the point of view of the seller, when data is only partially informative about consumer preferences, an exclusive sale is optimal. Now we move to the case in which the dataset is fully informative about consumer preferences.

Proposition 8. *When the dataset is fully informative, the seller sells partial but different information to both firms at a price* $P_i^{NE} = \frac{t}{12}$.

Proof. Firstly we establish that if information is allocated to both downstream firms, it must be that each firm receives different partial information. Recall that the non exclusive price when players acquire data on the same dimension was $P_i^{NE} = \pi_i(x,x) - \pi_i(0,x) = \frac{t}{54}$. By selling x to firm 1 and y to firm 2, the seller can get

$$P_1^{NE} = \pi_1(x,y) - \pi_1(0,y) = \frac{t}{12} \qquad \text{and} \qquad P_2^{NE} = \pi_2(y,x) - \pi_2(0,x) = \frac{t}{12}$$

The relevant pricing game is represented in Table 3.2. Clearly, the exclusive price

	0	y
0	$t/2,\ t/2$	$t/2,\ 13t/24$
x	$13t/24,\ t/2$	$7t/12,\ 7t/12$

Table 3.2: Pricing game when the dataset is fully informative.

does not change, and therefore the seller compares

$$\underbrace{(\pi_1(x,y) - \pi_1(0,y)) + (\pi_2(y,x) - \pi_2(0,x))}_{P^{NE}=\frac{t}{6}} > \underbrace{(\pi_i(x,0) - \pi_i(0,x))}_{P^{EX}=\frac{t}{24}}.$$

In the subgame perfect Nash equilibrium of the fully informative game, the data seller grants partial information on different dimensions to both competitors. ■

In terms of revenues, the straightforward implication is that a non exclusive sale dominates an exclusive sale. Given the optimal selling strategy, as long as data collection does not entail any cost, the following result holds.

Corollary. A monopolistic data seller has an incentive to collect a fully informative dataset.

In order to be able to sell data to both firms, the seller must hold information on both dimensions. If data collection is instead costly, it could be that for some cost configurations the gain from a non exclusive sale is outweighed by the increase in collection costs. If it is profitable to gather data on a single dimension, then an exclusive downstream allocation necessarily arises.

3.2.2 Endogenous data collection

So far the underlying assumption on data collection has been that the type of information observable by the seller was exogenously given. This assumption justified the separate focus on partial and full information. We also relax the assumption that firms may know something but not everything about consumers, and we allow for the sale of information that permit individual price targeting. Here we check not only the robustness of our previous findings when the data collection choice is made endogenous but we also characterize the data collection incentives: the question is whether the data seller decides to gather partial or full information about consumers and how he decides to optimally sell collected information. The goal of this section is to show that non exclusivity is robust.

Zero data collection costs

Suppose that data collection entails no additional costs for the data broker. This assumption implies that the data broker can freely select the optimal information structure to put on sale. In other words, the characterization of the seller's data collection incentives is equivalent to the analysis of how much information provide to firms. Absent any ex-ante costs, the seller can always dispose of all the available information without constraints. In order to maximize the price of information, the seller always induces a negative externality on the eventual loser of the data auction, leaving the data buyers with the worst outside option. When the decision on how much data to sell is made endogenous, the seller can resort to more complex selling mechanisms. In other words, if one player does not accept the initial take-it-or-leave-it data offer (either for partial or full information), then it is optimal from the point of view of the data seller to grant full exclusivity to only one firm, so that the outside option of the information acquisition game is always the less favorable payoff $\pi_i(0, xy) = \frac{4t}{27}$. Does the seller always have an incentive to induce an exclusive downstream allocation? Notice that when the seller is able to threaten firms with this exclusive allocation, then he has a strict incentive to award partial information non exclusively, provided that $P^{EX} = \pi_i(xy, 0) - \pi_i(0, xy) < 2 \times P^{NE} = 2 \times (\pi_i(x, y) - \pi_i(0, xy)) = \frac{47t}{54}$. This inequality reminds Proposition 8, but the selling price is different.

More formally, the data broker has access to x and y and can decide to allow individual targeting. Competing firms know that they can either receive partial or full information offers. However, it is common knowledge that (asymmetric) partial information can be eventually awarded only non exclusively, whereas full information - dimensions x and y tied together - is always sold exclusively. The relevant payoff matrix for the analysis is

	NI	I
NI	$t/2,\ t/2$	$4t/27,\ 62t/81$
I	$62t/81,\ 4t/27$	$7t/12,\ 7t/12$

Table 3.3: Payoff matrix when the type of dataset put on sale is endogenously determined.

where in the subgame $\{I, I\}$ firms receive asymmetric partial information and in the subgame $\{I, NI\}$ one firm is awarded full exclusivity. The crucial intuition is that, as long as the data broker can freely decide how much information to sell, the outside option of buyers is represented by the payoff of being the uninformed firm when the rival is informed under *full* information. Indeed, provided that data can be collected and added to the information structure put on sale at no cost, the threat of an exclusive sale that heavily disadvantages the losing firm is credible. In turn, the data broker extracts even higher rents from competing firms. Before stating the main result of this section, it is necessary to present the outcome of another type of subgame that was not characterized yet.

Lemma 5. *Suppose that both firms observe x but only firm 1 has access also to y. Then prices are equal to $p_2^*(x) = tx$ and $p_1^*(x, y) = t(2 - x - 2y)$ with profits $\pi_2 = \frac{t}{6}$ and $\pi_1 = \frac{7t}{12}$.*

Proof. See Appendix 3.2. ■

The following proposition establishes a *non exclusivity* result. Indeed, through a non exclusive sale, the seller is able to maximize its revenues from data trade.

Proposition 9. *If full information is collected at no cost, then the data seller sells asymmetric partial information to both firms at a price* $P_i^{NE} = \frac{47t}{108}$.

Proof. An offer from the seller includes the type of information put on sale, an individual price for it and, implicitly, whether information is offered exclusively or non exclusively. The sale of full information implies exclusivity, provided that symmetric full information lowers industry profits. The *ex-ante* exclusive selling price is

$$P^{EX} = \pi_i(xy, 0) - \pi_i(0, xy) = \frac{62t}{81} - \frac{4t}{27} = \frac{50t}{81}.$$

On the other hand, the individual non exclusive selling price for partial information now becomes

$$P_i^{NE} = \pi_i(x, y) - \pi_i(0, xy) = \pi_i(y, x) - \pi_i(0, xy) = \frac{7t}{12} - \frac{4t}{27} = \frac{47t}{108}.$$

The data broker therefore compares

$$2 \times P_i^{NE} = \frac{47t}{54} > P^{EX} \tag{3.1}$$

which implies that the maximum revenue from information sale is secured by awarding partial information to both downstream competitors. We claim that the optimal selling strategy of the data broker is structured as follows: (1) a non exclusive contract regarding partial information, with x sold to one firm and y to the rival; (2) an exclusive contract regarding full information (x and y) in case contract (1) is not accepted by all buyers.

The proof is articulated in two parts: in the first part, we check that the data broker has no incentives to deviate from the non exclusive contract proposed above, which also proves that Table 3.3 is indeed the relevant payoff matrix for the analysis; in the second part, we check that the buyers have an unilateral incentive to accept the offer P_i^{NE}.

Part 1. Suppose that a non exclusive contract has been accepted by firms at some positive price P_i possibly different from P_i^{NE}. Payments are made just before the pricing game, once the final data allocation becomes common knowledge. Suppose that the data broker has the option to offer exclusively x to the buyer that has acquired partial information on y. Suppose also that firm 1 had initially access to y, and it gets data on x and y whereas firm 2 is left with access to x only. Then, in the light of Lemma 5, the new prices would be $p_2^*(x) = tx$ and $p_1^*(x, y) = t(2 - x - 2y)$, with respective profits given by $\pi_2(x, xy) = \frac{t}{6}$ and $\pi_1(xy, x) = \frac{7t}{12}$. The firm that receives the after-sale exclusive offer gets exactly the same payoff, given that the rival is already partially informed. There is no incentive to acquire additional information about consumer preferences. In turn, the data broker does not have an incentive to deviate from the initial non exclusive contract.

Part 2. We show that a firm cannot gain by unilaterally deviating from the non exclusive contract proposed above, which also implicitly provides the rationale behind the choice of the payoff $\pi_i(0, xy)$ as the outside option in the expression for the relevant price. Suppose that firm 2 does not accept the non exclusive contract about partial information x. The data broker can then simply offer information on x along with that on y to firm 1. In other words, an exclusive contract is now offered to the not deviating firm. In case of a further rejection of the offer both firms obtain the no information payoff. However, if the exclusive offer is accepted, the deviating firm obtains exactly the outside option that appears in P_i^{NE}. Indeed, firm 1 would have an incentive to acquire full information at a lower *ex-post* exclusive price, which is equal to

$$P^{EX} = \pi_i(xy, 0) - \pi^{fp} = \frac{62t}{81} - \frac{t}{2} = \frac{43t}{162}$$

and firm 2 gets the payoff of the uninformed firm under full information. Therefore, each firm has an unilateral incentive to accept the non exclusive contract which maximizes the seller's revenue (i.e. $\bar{R} = \frac{47t}{54}$) at the first stage. As usual, given the standard assumption on bargaining power, the price charged for consumer data makes the buyer exactly indifferent between acquiring information and the outside option. When the collection choice of the seller is made endogenous, the outside option of the firms is less favorable, and therefore seller's revenue is larger than the one characterized in Section 3.2.1.

■

The possibility to offer full information allows the data broker to credibly threaten firms with an exclusive downstream data allocation. The threat is credible considering that it is impossible for the duopolists to coordinate and deviate together from the proposed non exclusive contract, as each competitor has an incentive to wait for the deviation of the rival and then get the exclusive offer. Notice also that this result holds independently of whether contracts are offered before or after the actual collection of consumer data. Zero collection costs imply that the seller can always propose a new offer in case of a deviation by one player (i.e. immediately acquire the additional information that is needed to propose the exclusive data package), which makes the outside option of data buyers less favorable.

Positive data collection costs

Finally, we interestingly show that the non exclusivity result holds also when data collection costs are positive but not too large. Suppose that the data broker incurs in the fixed costs $k_x \geq 0$ and $k_y \geq 0$ when collecting information on dimensions x and y respectively. We only require that $k_x + k_y \leq \frac{47t}{54}$, otherwise the data broker would not have enough resources to threaten firms with an exclusive offer, which is the key mechanism used by the seller to extract more rents from the data buyers. Put it differently, the seller has an incentive to actively gather full consumer data as long as the cost of investing in the tracking technology does not exceed the maximum revenue from information sale. The key intuition is that in presence of costly information acquisition, only the cost of collecting an extra dimension really matters for the data broker's incentives once the constraint is satisfied.

Proposition 10. *If $k_x + k_y \leq \frac{47t}{54}$, then the data broker collects full information but it sells asymmetric partial information to both firms at a price $P_i^{NE} = \frac{47t}{108}$.*

Proof. See the text. ■

As long as the constraint on costs is satisfied, the data broker collects full information and obtains the maximum revenue $\bar{R}$ when selling asymmetric partial information: the threat of an exclusive deal is credible since the seller has a fully informative dataset, and such threat is leveraged to extract more rents from firms through the non exclusive contract. What is left to investigate is what happens when the constraint on costs is not satisfied: the data broker cannot collect full information. This observation has an immediate impact on the outside option of buyers. Provided that it is common knowledge that a fully exclusive contract cannot be proposed when $k_x + k_y > \frac{47t}{54}$, competing firms are aware of the fact that an unilateral deviation from the contract designed for partial consumer data now yields $\pi_i(0, x) = \frac{t}{2}$, if we assume that x is collected first in the case of a partially informative dataset. The data broker then prefers to sell partial information exclusively, as long as the cost of collecting a single dimension of consumer data does not exceed the exclusive price characterized in Proposition 7. It is therefore straightforward to show the following result.

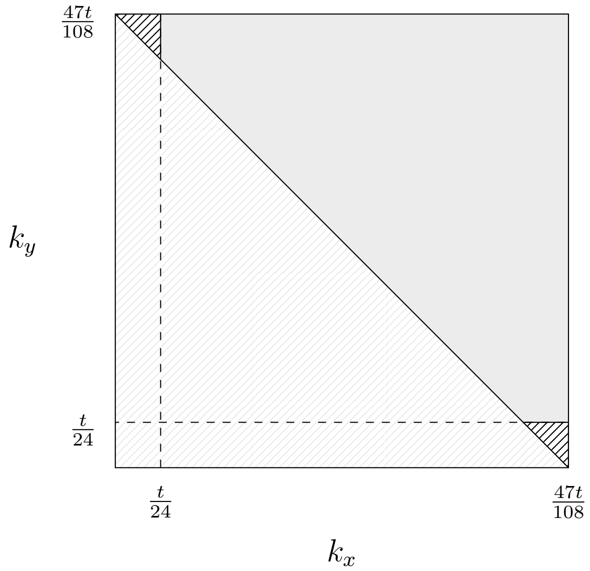


Figure 3.3: Selling strategy of the data broker when collection costs are positive.

Proposition 11. *Suppose that $k_x + k_y > \frac{47t}{54}$. If k_x (k_y) $\leq \frac{t}{24}$, then the data broker collects partial information on dimension x (y) and sells it exclusively to one firm at a price $P^{EX} = \frac{t}{24}$; otherwise, no information is collected.*

Proof. See the text. ■

The punchline is that positive costs of data collection, when the decision to collect such information is endogenous, determine which type of information structure is in the hands of the seller; in turn, availability of full or partial information determines whether a negative externality is imposed or not on the loser of the data auction. Accordingly to the prevalent outside option, which is $\pi_i(0, xy) = \frac{4}{27}$ under full information and $\pi_i(0, x) = \pi_i^{fp}$ under partial information, the seller is able to extract more or less rents. The optimal selling strategy is to differentiate the type of consumer data that each buyer can acquire, under the threat of an exclusive sale of full information, whenever possible.

3.3 Competitive data brokers

A quick look at the data brokerage industry returns the snapshot of a very populated yet opaque market. The next building block of this chapter is therefore an extension that aims to relax the assumption of a monopolistic data seller as the unique source of consumer data in the market. It is clear that the main objective of a single data broker would be that of maximizing industry profits in order to extract the largest revenue from data trade, even though we have shown that, when information is multidimensional, it is not obvious under which type of data allocation firms' payoffs are maximized. This argument may not hold anymore when introducing a more competitive upstream market structure. Moreover, such analysis delivers additional predictions which enlarge the scope for a discussion of possible policy interventions in the data market.

We assume that firms $i = 1, 2$ can now acquire data from two different upstream firms $k = A, B$. In this extension of the baseline monopolistic case, investment costs are ignored. Our focus is mainly on partially informed sellers. In other words, we will investigate two simple scenarios in which the sellers hold exogenously given data: (i) both upstream firms have partial information on one dimension only (i.e. both hold x or both y), and (ii) both have partial information but on different dimensions. Notice that in order to develop a meaningful analysis of the effects of upstream competition on the final downstream data allocation and on the price at which information is traded it is reasonable to consider scenarios in which both data brokers have "something" to sell. Otherwise, the analysis would simply boil down to the insights developed in Section 3.2.

In line with the previous section, with some necessary modifications, the salient aspects of the relatively simple contracting environment assumed here can be summarized as follows:

1. There is common knowledge about the type of data controlled by each upstream firm;

2. Firms make public offers for the data in the form of take-it-or-leave-it offers;

3. Firms can post either a non exclusive price at which whoever is interested can acquire the attached data, or they can set exclusive prices that are targeted to only one of the downstream buyers;

4. Differently from the mechanism exploited in the monopolistic case, which involved a negative downstream externality when an exclusive offer was turned down, here it is neither necessary nor tenable to impose the possibility of offering exclusive data to a second buyer; rather, we assume that an exclusive offer is truly targeted to a specific firm, without reselling possibilities.

Each buyer holds the conjecture that the rival will always buy information at least from one upstream seller, as it is individually rational to do so. These assumptions ensure that it is relatively easy to detect whether the game has a unique equilibrium or multiple equilibria, as it is easier to identify the relevant outside option for the data buyers. Rather than developing a comprehensive treatment of the effects of introducing upstream competition, this section aims to deliver a bunch of simple insights that complement the monopolistic data broker case.

3.3.1 Both sellers have x

Each seller k can either exclusively sell dataset X to firm 1 or firm 2, or it can post a non exclusive price at which both firms can eventually buy partial consumer information. The expected payoffs of the two competing data brokers are summarized in the 3×3 payoff matrix[46] depicted in Table 3.4. Notice that each payoff should be multiplied by $\alpha \geq 0$, which is a parameter that measures the bargaining power of the sellers. Under monopoly it was assumed equal to one, provided that it is quite natural to proceed with $\alpha = 1$ in that scenario. This assumption instead becomes quite untenable in a competitive setting; yet, in order to avoid abuses of notation, the parameter does not appear in the payoff matrix. We can still derive meaningful results under the assumption that α is constant across all cells.

This game possesses several interesting aspects generated by competition upstream. First of all, notice that when the data brokers offer exactly the same deal to the same buyer, data trade takes place with a probability equal to one half. Secondly, the sellers maximize the aggregate expected revenues from data sales when they simultaneously offer data exclusively to only one and the same downstream retailer. Thirdly, when different brokers make an offer to different downstream firms, this *ex-ante* exclusivity necessarily generates *ex-post* non exclusivity under the conjecture that the rival always finds it profitable to buy information. This observation justifies the presence of the profit $\pi_i(x, x)$ in the "exclusivity" payoff of the sellers. Finally, the key intuition is that when both sellers hold the same type of partial information, each broker finds it unilaterally optimal to sell to both firms when the rival makes exclusive offers: as a result, in the unique Nash equilibrium of this game neither firm is willing to promote exclusive deals.

Proposition 12. *If there are two data sellers endowed with the same partial information, they offer non exclusive deals.*

[46] The convention adopted is that seller A is the row player whereas seller B is the column player but clearly everything is symmetric.

	x to 1	x to 2	x to both
x to 1	$\frac{\pi_1(x,0)-\pi_1^{fp}}{2} = \frac{t}{48}$ $\frac{\pi_1(x,0)-\pi_1^{fp}}{2} = \frac{t}{48}$	$\pi_1(x,x) - \pi_1(0,x) = \frac{t}{54}$ $\pi_2(x,x) - \pi_2(0,x) = \frac{t}{54}$	$\frac{\pi_1(x,x)-\pi_1(0,x)}{2} = \frac{t}{108}$ $\frac{\pi_1(x,x)-\pi_1(0,x)}{2} + \pi_2(x,x) - \pi_2(0,x) = \frac{t}{36}$
x to 2	$\pi_2(x,x) - \pi_2(0,x) = \frac{t}{54}$ $\pi_1(x,x) - \pi_1(0,x) = \frac{t}{54}$	$\frac{\pi_2(x,0)-\pi_2^{fp}}{2} = \frac{t}{48}$ $\frac{\pi_2(x,0)-\pi_2^{fp}}{2} = \frac{t}{48}$	$\frac{\pi_2(x,x)-\pi_2(0,x)}{2} = \frac{t}{108}$ $\pi_1(x,x) - \pi_1(0,x) + \frac{\pi_2(x,x)-\pi_2(0,x)}{2} = \frac{t}{36}$
x to both	$\frac{\pi_1(x,x)-\pi_1(0,x)}{2} + \pi_2(x,x) - \pi_2(0,x) = \frac{t}{36}$ $\frac{\pi_1(x,x)-\pi_1(0,x)}{2} = \frac{t}{108}$	$\pi_1(x,x) - \pi_1(0,x) + \frac{\pi_2(x,x)-\pi_2(0,x)}{2} = \frac{t}{36}$ $\frac{\pi_2(x,x)-\pi_2(0,x)}{2} = \frac{t}{108}$	$\frac{2\times\pi_i(x,x)-\pi_i(0,x)}{2} = \frac{t}{54}$ $\frac{2\times\pi_i(x,x)-\pi_i(0,x)}{2} = \frac{t}{54}$

Table 3.4: Both data brokers sell information on dimension x.

When only one dimension of consumer private information is collectible, upstream competition is sufficient to ensure that no downstream buyer is foreclosed.

3.3.2 One seller has x while the other one has y

Suppose that seller A has collected data on x while the competing seller B holds data on y. The expected payoffs appear in Table 3.5. An important assumption will be that there is "one-stop shopping": each buyer can at most acquire data from one upstream firm. This is a simplifying assumption, but the intuition is that, provided that the game is simultaneous, once a buyer visits one data broker it is impossible to subsequently visit also the other broker. More substantively, this assumption rules out the possibility of having cases in which the full exclusivity payoff appears downstream (i.e. $\pi_i(xy, \cdot)$).

Prior to the analysis of the matrix, it is useful to establish an important result that simplifies the game played by the upstream firms.

Remark. When the upstream sellers hold different partial information, selling information to both downstream firms (i.e. to set a price that induces with some positive probability both firms to buy) is always a strictly dominated strategy.

With this idea in mind, notice the payoff matrix will be simply a 2×2 matrix.

	y to 1	y to 2
x to 1	$\frac{\pi_1(x,0)-\pi_1^{fp}}{2} = \frac{t}{48}$ $\frac{\pi_1(y,0)-\pi_1^{fp}}{2} = \frac{t}{48}$	$\pi_1(x,y) - \pi_1(0,y) = \frac{t}{12}$ $\pi_2(y,x) - \pi_2(0,x) = \frac{t}{12}$
x to 2	$\pi_2(x,y) - \pi_2(0,y) = \frac{t}{12}$ $\pi_1(y,x) - \pi_1(0,x) = \frac{t}{12}$	$\frac{\pi_2(x,0)-\pi_2^{fp}}{2} = \frac{t}{48}$ $\frac{\pi_2(y,0)-\pi_2^{fp}}{2} = \frac{t}{48}$

Table 3.5: Data brokers hold partial information but on different dimensions.

The argument is relatively simple. In any case there is the possibility for at least one downstream firm to get partial information (i.e. either x from A or y from B) at any of the upstream firms. When both firms offer non exclusivity they will trade data with a probability equal to one half. If we fix the non exclusive strategy for one of the sellers, it is easy to show that the best response of the rival is to always sell information exclusively. Obviously, the downstream firm that is not targeted via the exclusive offer will always have an incentive to buy partial information from the other seller. But, if this is the case, the exclusive seller is able to ensure that his offer is accepted with probability one by making the exclusively targeted buyer just indifferent between the non exclusive price and his price. As a result, when one seller offers non exclusivity, the competing broker will always target either firm 1 or firm 2 with an exclusive offer. Provided that the argument is symmetric for the other upstream player, it is easy to show that, when endowed with different partial data about consumers, the firms will always play exclusivity. Notice that this argument builds also on the indifference result derived in Lemma 5. Namely, when one downstream firm already has partial information (for instance, about dimension x) there is no reason for the rival to spend additional resources to acquire full information if dimension y can be easily acquired from a competing seller. In turn, this implies that in the upstream market non exclusive contracts are not proposed.

The game in Table 3.5 has multiple Nash equilibria: there are two equilibria in which the data brokers target different downstream firms with exclusive deals.

Proposition 13. *If there are two data sellers endowed with different partial information, they offer exclusive deals to different downstream buyers.*

This multiplicity of equilibria is only apparent, as the two scenarios are clearly symmetric in terms of market outcomes. What is interesting to notice is that "asymmetric" competition upstream - meaning that the firms are endowed with information about different dimensions - always triggers a non exclusive downstream allocation induced via exclusive contracts: both retailers can target prices on the basis of partial information but they have access to different dimensions of consumers' willingness to pay. It would be clearly interesting in future research to investigate the incentives of these competitors to vertically integrate.

3.4 Data partitioning: Beyond no privacy

In both the second and the third chapter, firms had access to two dimensional information about consumer willingness to pay. Yet, under both data sharing between firms and data brokerage by a third party, the somehow restrictive assumption was that consumer data was sold either as a bundle (i.e. full information) or as a single dimension (i.e. partial information). In the latter case it has been shown that competition is relaxed and firms are better-off whereas consumers collectively prefer a scenario in which full information is allocated to all firms in the market. Yet, this no privacy regime is not the upper bound on aggregate consumer surplus in the model: adopting the view of a benevolent social planner, more sophisticated data management strategies may lead to an outcome even more favorable to customers (i.e. $CS \to W^*$) than no privacy. Recall that a no privacy regime would overall benefit final consumers, with the crucial caveat that there would be winners and losers. The obvious question is whether it is possible to design a privacy regime that not only collectively benefits consumers but also makes them all at least weakly better-off than under full privacy. In this section we characterize a "consumer friendly" equilibrium. In doing so, we firstly check whether a first best outcome is feasible (i.e. $CS = W^*$); otherwise, we proceed to characterize a data partitioning strategy that leads to a second best outcome[47]. Finally, we discuss possible policy implications of this refinement of the analysis, contrasting and complementing the results of this chapter with those of the previous one.

Consumers keep all of their surplus only if at each location $(x, y) \in [0, 1]^2$ both firms set prices equal to the marginal cost. The following lemma formalizes an immediate observation.

Lemma 6. *Consumer first best never arises as an equilibrium of the game.*

Proof. Suppose that both firms are setting a price $p_i = 0$. If we consider a consumer with $x + y < 1$ patronizing firm 1 it is possible to see that a switch between outlets does not occur as long as

$$v - p_1 - tx - ty \geq v - 2t + tx + ty \qquad \Rightarrow \qquad p_1 \leq 2t(1 - x - y)$$

which implies that in the neighborhood of the most contestable consumers (i.e. $x + y = 1$) there exist a location such that a buyer trades off a positive but low enough price at outlet 1 against the larger disutility at outlet 2 in favor of the first firm. Such positive price always exists and a zero price equilibrium cannot be sustained. ■

The intuition is straightforward: it would be possible to sustain an equilibrium with prices equal to the marginal cost if and only if the degree of perceived product

[47]In this analysis potential data externalities are not considered, in order to maintain ourselves within the same framework previously employed; yet, we acknowledge that they could play an important role in this context and they should be considered in future research.

differentiation was driven down to zero (i.e. $t \to 0$). Trivially, each consumer regards the two products as perfect substitutes and price competition à la Bertrand kicks in. The question therefore is whether a particular data partitioning could improve over the no privacy benchmark, which is the scenario that collectively benefits consumers. The first observation is that when both firms address with targeted prices each consumer there are winners and losers with respect to the full privacy equilibrium in which all customers are charged a price equal to the differentiation parameter. Personalized prices indeed are equal to $p_1^{np}(x,y) = \max\{2t(1-(x+y)),0\}$ and $p_2^{np}(x,y) = \max\{2t((x+y)-1),0\}$. It is useful to define the set of efficiently served consumers in the following way:

$$E_1 = \left\{(x,y) \in [0,1]^2 : x+y \leq 1\right\}$$

$$E_2 = \left\{(x,y) \in [0,1]^2 : x+y \geq 1\right\}$$

and identify among these consumers the "losers" under no privacy in comparison to the no information benchmark:

$$L_1 = \{(x,y) \in E_1 : p_1^{np}(x,y) > t\}$$

$$L_2 = \{(x,y) \in E_2 : p_2^{np}(x,y) > t\}.$$

Intuitively the "winners" are the consumers located close to the margin whereas types which are more captive are exploited by the firms. Thus, two other observations follow: (i) the knowledge of precise data about the rival's efficient consumers has a clear procompetitive effect which drives down prices towards the marginal cost, whereas (ii) the targeting of own captive consumers is detrimental to high willingness to pay buyers. A consumer friendly partitioning of information should therefore point in the direction of promoting the first effect while avoiding the second one, and at the same time it should also maintain an efficient allocation in order to minimize overall transportation costs. In other words, each firm i must be informed about consumers belonging to the set E_j while remaining uninformed about the portion of its most valuable consumers (i.e. those in the set L_i). In this way non exclusive access to consumer data is preserved as much as possible and exclusivity is given only over consumers that must be targeted with very aggressive discounts. In turn, the ability of firms to extract higher rents from close consumers is eroded via tough competition by the more informed rival.

Our candidate disclosure rule can be formalized as follows:

- Firm 1 is informed about all consumers in E_2 and only those in $E_1 \backslash L_1$;
- Firm 2 is informed about all consumers in E_1 and only those in $E_2 \backslash L_2$.

When the consumer data available to the competing firms resemble this "almost" fully informative allocation the following result can be stated[48].

Lemma 7. *When firms are almost fully informed but they cannot identify own captive consumers profits are equal to $\pi_i = \frac{7t}{24}$, lying below the no privacy level, and the equilibrium is efficient.*

Proof. Consider firstly the consumers labeled as winners under the no privacy regime (i.e. $\frac{1}{2} - x \leq y \leq \frac{3}{2} - x$). Provided that they are identified simultaneously by both firms, it is straightforward to see that the personalized prices charged to them are equal to the full information tailored prices: $p_1^*(x, y) = \max\{2t(1 - (x + y)), 0\}$ and $p_2^*(x, y) = \max\{2t((x + y) - 1), 0\}$.
The second step of the proof is the characterization of the equilibrium prices in the regions of captive consumers. Consider region L_1 and recall that the boundary of this area is determined by the line $y(x) = \frac{1}{2} - x$. In particular, along this line, firm 1's prices in both benchmarks satisfied $p_1^{fp} = p_1^{np}(x, y) = t$ while we have shown that the price of firm 2 was equal to the marginal cost in E_1 under no privacy. However, given that firm 1 can now set only a uniform price to its captive consumers, a trade-off emerges: (i) to increase the basic price in order to extract more surplus, which however allows competition via targeted discounts by the informed rival to kick in, or (ii) to maintain a low enough basic price such that it is unprofitable for firm 2 to compete (i.e. firm 2 continues to charge a price equal to the marginal cost as in the no privacy benchmark). The symmetric argument holds for firm 2 when facing its (unidentified) captive consumers in region L_2.
The ability to target individually all consumers in region L_1 allows firm 2 to match any uniform price that is above the price level making competition unprofitable. For each small increase in the uniform price, firm 2 steals an increasing fraction of captive consumers from firm 1. The problem therefore boils down to the characterization

[48]Notice that an even more extreme disclosure rule induces even lower profits for the retailers than the outcome characterized in the lemma; namely, an *ex-post* scenario in which each player has full information only about all consumers in the efficient market of the rival (i.e. firm i gets precise location data on consumers in E_j). This information allocation triggers even more intense competition and each retailer gets a profit equal to $\pi_i = \frac{20t}{81} < \frac{7t}{24}$. However it has a crucial drawback: the allocation of consumers among the two outlets induced in equilibrium is inefficient. The implication is that overall consumer surplus would be lower than in the equilibrium proposed as "consumer friendly".

of the uniform price that do not induce firm 2 to have positive sales in this region. The best response of the informed firm is $\tilde{p}_2(x,y,p_1) = p_1 + t(2x-1) + t(2y-1)$. We claim that the lowest possible price for firm 1 is $p_1 = t$. Indeed, if this is the candidate uniform price, $\tilde{p}_2(x,y,t) < 0$ for all $x+y<\frac{1}{2}$. For any other price $p_1 > t$ firm 2 would be able to make positive profits in L_1, frustrating the increase in price. Finally, it remains to check that all consumers in $E_1 \backslash L_1$ prefer to buy the product from outlet 1 at price $p_1^*(x,y)$ rather than from outlet 2 at price $\tilde{p}_2(x,y,t)$, which is easily verified. Efficiency therefore holds and captive consumers pay at most the same price as in the full privacy equilibrium. ■

Profit of firm 1 is

$$\begin{aligned}\pi_1 = &\int_0^{\frac{1}{2}-x}\int_0^{\frac{1}{2}} t\,dxdy + \int_{\frac{1}{2}-x}^{1-x}\int_0^{\frac{1}{2}} 2t(1-(x+y))\,dxdy \\ &+\int_0^{1-x}\int_{\frac{1}{2}}^{1} 2t(1-(x+y))\,dxdy = \frac{7t}{24} \quad < \pi_i^{np}\end{aligned}$$

and symmetrically for firm 2. Total consumer surplus is equal to $CS = v - \frac{5t}{4} > CS^{np}$ and $W = W^*$.

We have therefore characterized a possible data allocation which favors consumers not only in aggregate but also - at least weakly - on an individual basis. The main feature of this scenario is to have simultaneously each firm uninformed about the own high willingness to pay consumers but informed about the rival's same type of customers, and both firms equally informed about the marginal consumers in order to preserve intense price competition for the consumers with relatively less polarized valuations. The risk of aggressive competition by the rival on the subset of the market populated by high willingness to pay consumers will prevent each firm from setting a too high uniform price. The described allocation of information among competing firms could arise, for instance, under a data portability regime. If in the *status quo* we assume that each firm is likely to be more informed about its close consumers (i.e. firm 1 about consumers with $x+y<1$ and firm 2 about consumers with $x+y>1$), data portability should ensure that each data subject can make available to the rival firm his or her personal data. However, this is not sufficient. Consumers should be sophisticated enough, and the data portability right should be accordingly designed in order to to ensure that the high willingness to pay consumers not only port their data to the rival but also definitively erase personal information held by the original data controller. This second part of the argument is clearly less realistic, as data portability under the GDPR (2018) ensures the right to port data but it is more opaque about the duties of the original firm. It is therefore likely that a switch from a no privacy regime to a "consumer-friendly" equilibrium would be hard to realize even with portability rights.

3.5 Main highlights and policy recommendations

Within competition policy circles, justified by an extensive literature on foreclosure and exclusive contracts, the common view is that exclusive data contracts pose great antitrust concerns. In this chapter I focus on data flows between firms, investigating what type of data allocation will emerge in a variety of settings based on the market outcomes of the second chapter. First of all, I briefly discuss the incentives of competing firms to exchange data about consumer preferences in the context of a two dimensional model of product differentiation. Then I introduce a monopolistic data broker that collects and sells information to competing firms and I further extend the analysis to an upstream competitive data market. Provided that the variety of the scenarios investigated is quite rich, heterogeneous results emerge: exclusivity is not only the less likely outcome, but under some conditions data exclusivity may be even preferred.

More substantively, I consider a monopolistic seller that exogenously holds either a partially informative dataset or a fully informative one. In the former case, only one downstream firm gets access to partial consumer data. In the latter case, in the light of the previous literature, we should expect the seller to induce an even stronger asymmetry in the downstream market by selling full information to a single retailer; instead, I show that the seller prefers to induce both firms to acquire partial but different information about consumers. The intuition is straightforward: industry profits under asymmetric partial information dominate profits in a fully exclusive scenario. As a result, no firm is foreclosed and, moreover, the non exclusive allocation is efficient: however, consumers would prefer exclusivity, at least in aggregate terms. When the seller's choice to collect consumer data is made endogenous and possibly costly, the prediction is that, if data collection costs are relatively low, the seller gathers all the possible data about consumers and the optimal selling mechanism is a non exclusive one; otherwise, if such costs are in aggregate too large but sufficiently asymmetric so that it is feasible to create at least one of the two datasets, only one attribute is collected, which triggers an exclusive data allocation. When only one dimension is collectible, one of the downstream firms is foreclosed from access to data.

The introduction of competition in the upstream data sector solves the exclusivity problem, with the important remark that consumers may collectively prefer precisely an exclusive outcome under some conditions. However, from a competition policy point of view, exclusive data access does not seem to be an issue for downstream firms. Two cases are investigated: when the data brokers have the same partial information both makes non exclusive offers to all downstream market participants whereas, when the data brokers hold different partial information, there is a multiplicity of equilibria in which the upstream players target with exclusive offers different downstream retailers. In either case, the prediction is that the final data allocation contemplates all firms to be partially informed about consumers' preferences. Exclusivity does not arise as an equilibrium of the game when the upstream sector is competitive.

However, with respect to privacy, competition alone is not sufficient, meaning that it does not waive away the privacy concerns when we evaluate the market outcomes through the lens of consumer protection. Rather, exclusivity is not *per se* bad for consumers when compared with other possible data allocations. Still, the variety of scenarios analyzed in this chapter delivers important conclusions with respect to the suspicion that surrounds exclusive data access from an antitrust point of view. Exclusivity is a real competition concern only in one situation: in the data market there is a unique data seller which knows something but not necessarily everything about consumers. In other words, the monopolistic broker is not able to offer conclusive inference about consumers' valuations for the products. In all the other cases, either when the monopolistic seller is fully informed or when there is competition upstream, both downstream firms will have access to *some* consumer data. Therefore, what emerges from this analysis is that we should not ban exclusivity *a priori*; rather, the competition authorities should invest more resources in investigating the underlying characteristics of the market for data and what type of information is collected and traded. Clearly, if we believe that information may be available from several sellers we should not be concerned about exclusive access to data. Similarly, this issue does not arise if we believe that, even though there is a monopoly position in the data market, the seller is a quasi-omniscient firm that knows everything about consumers. Yet, this is hardly the case: it is more likely to see in different (sub)markets firms that are specialized in collecting particular attributes about consumers which however are not conclusive about the type of the potential buyers. This is why our main concerns should be raised not about the type of contracts but about the type of data held by the data controllers. If each broker has a monopoly position with respect to a particular type of consumer data then exclusivity arises and the standard competition policy concern is indeed there; otherwise, all firms get some information about consumers, which is not necessarily good for them.

Appendix

A.1 Proof of Lemmas

Proof of Lemma 5. Consider a scenario in which both firms are partially informed. For instance, firm 1 initially has only y and firm 2 observes y. Suppose now that the seller can further offer data on y to firm 1, which gets full information whereas the rival remains only partially informed. Consumer (x, y) is always served by firm 1 when

$$p_1(x, y) + tx + ty = p_2(x) + t(1 - x) + t(1 - y)$$

and the informed firm optimally sets its personalized price at each location in order to match the rival's price

$$p_1(x, y) = p_2(x) + t(1 - 2x) + t(1 - 2y).$$

The objective function of firm 2 is defined for each x as

$$\pi_2 = p_2 \left(1 - \left(\frac{p_2}{2t} + (1 - x)\right)\right)$$

and taking the first order condition yields the personalized price $p^*(x) = tx$. In turn, firm 1 sets $p_1^*(x, y) = t(2 - x - 2y)$. The market boundary is interior ($y^*(x) = \frac{2-x}{2}$) and profits are equal to

$$\pi_1 = \int_0^{\frac{2-x}{2}} \int_0^1 p_1^*(x, y)\ f(x)f(y)dxdy = \frac{7t}{12}$$

$$\pi_2 = \int_{\frac{2-x}{2}}^1 \int_0^1 p_2^*(x)\ f(x)f(y)dxdy = \frac{t}{6}$$

■

Conclusions

In this thesis were investigated some policy relevant questions on digital markets . The policy debate, supported by the economic but also the law literature, revolves around several competition policy issues and consumer protection concerns posed by developments in information technologies and digital markets, such as the rise of dominant platforms, pervasive data collection and its consequences on consumer privacy. In this thesis, the leading approach has been to ultimately consider many of these concerns as competition issues. Indeed, we know from the literature that when an equal level playing field is ensured then dominance may not be a concern anymore. Moreover, if competition properly works in the market, privacy is not an issue provided that consumers, even if information about them is disclosed, get always the best deal. Problems arise when competition is lessened. In particular, one of the main channels through which the intensity of competition may be varied is the collection and exploitation of data about consumer preferences.

The red thread in this work is information about consumers. The three chapters deal with applications related to exploitation of data about the willingness to paof consumers in digital market, ranging from product customization to price discrimination. Another crucial and strictly related topic is also investigated; namely, the concerns posed by data exchanges or data sales among firms. Indeed, conditionally on these flows of information between firms active in digital markets, more or less firms are able to implement sophisticated strategies, with not obvious consequences on competition and, ultimately, consumer privacy. In the first chapter I show that a proprietary informational asset that grants a competitive advantage, in the form of exclusive access to data or a better customizing technology, may be traded with a rival firm under some conditions. In particular, when customization induces more product differentiation in the market, then firms find mutually beneficial to compete with a technology characterized by a similar precision in targeting customers. In turn, it is this symmetry that ultimately allows firms to coordinate on higher prices charged to final consumers. This finding is in contrast to the standard argument that more symmetric competition should benefit consumers: this is not necessarily true in digital markets. where the informational asset exploited by firms are very particular inputs. This argument fails to hold also in the second chapter, in which I present a two dimensional model of horizontal product differentiation. This information structure allows to consider both scenarios in which the players are precisely informed about users and cases in which the firms may know something but not necessarily everything about consumers. In other words, the chapter contrasts regulatory regimes in which privacy is or is not enforced at all against regimes in which there is only partial privacy. Being not perfectly informed about consumers' willingness to pay relaxes price competition, and a hump-shaped relationship between information and industry profits emerge. We know from the policy debate and from the economic literature that, in general, more information in the market should benefit final consumers, at least when it is available to all the players without problems of foreclosure. Instead, in this model, when moving from full privacy to partial privacy, firms win and consumers lose. Under no privacy, instead, the standard argument of Bertrand competition applies again and the reverse holds true, with consumers collectively winning and firm losing. The ambiguity with respect to consumer privacy arises at the individual level, provided that it is always possible to find a subset of consumers that is strictly worse-off even when consumer surplus rises. Consumers should be therefore empowered with more control over personal information, which implies that the regulatory approach to consumer privacy should be more nuanced. Finally, in the third chapter, I investigate how prevalent are exclusive sales of consumer information in the data market, showing that, differently from what is commonly highlighted in the policy debate, exclusivity is not a competition concern in all scenarios. Rather, the type of data allocation among competing firms depends on the type of information in the hands of the data seller. The main contribution is to show that both exclusive and non exclusive sales of consumer data can be an equilibrium in the data market, conditionally on whether the seller holds a partially or a fully informative dataset. Competition upstream is sufficient to induce always a non exclusive allocation of data among downstream firms, so that no competing firm is foreclosed.

Bibliography

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*.

Anderson, S., De Palma, A., and Thisse, n. J.-F. (1989). Demand for differentiated products, discrete choice models, and the characteristics approach. *The Review of Economic Studies*, 56(1):21–35.

Armstrong, M. (1999). Competition in the pay-tv market. *Journal of the Japanese and International Economies*, 13(4):257 – 280.

Armstrong, M. (2006). Price discrimination. *MPRA Paper*.

Armstrong, M. and Vickers, J. (2001). Competitive price discrimination. *RAND Journal of Economics*, 32(4):1–27.

Baye, I., Reiz, T., and Sapi, G. (2018). Customer recognition and mobile geo-targeting. *DICE Discussion Paper*, (285).

Baye, I. and Sapi, G. (2019). Should mobile marketers collect data other than geo-location? *The Scandinavian Journal of Economics*, 121(2):647–675.

Belleflamme, P., Lam, W. M. W., and Vergote, W. (2019). Competitive imperfect price discrimination and market power.

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

Bergemann, D., Bonatti, A., and Smolin, A. (2018). The design and price of information. *American Economic Review*, 108(1):1–48.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–957.

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95.

Bhaskar, V. and To, T. (2004). Is perfect price discrimination really efficient? An analysis of free entry. *RAND Journal of Economics*, 35(4):762–776.

Biglaiser, G., Calvano, E., and Crémer, J. (2018). Incumbency advantage and its value. *Journal of Economics and Management Strategy*, 28:41–48.

Bimpikis, K., Crapis, D., and Tahbaz-Salehi, A. (2019). Information sale and competition. *Management Science, Articles in Advance*, pages 1–19.

Bonatti, A. and Cisternas, G. (2018). Consumer scores and price discrimination. *CEPR Discussion Paper Series*.

Bounie, D., Dubus, A., and Waelbroeck, P. (2018). Selling strategic information in digital competitive markets.

Brill, J., McSweeny, T., Ohlhausen, M., Ramirez, E., and Wright, J. (2014). Data brokers: A call for transparency and accountability. Technical report, Federal Trade Commission.

Caplin, A. and Nalebuff, B. (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, 59(1):25–59.

Chen, Y., Li, X., and Sun, M. (2019). Competitive mobile geo targeting. *Management Science*, 36(5):666–682.

Chen, Y. and Riordan, M. H. (2008). Price-increasing competition. *The RAND Journal of Economics*, 39(4):1042–1058.

Clavorà Braulin, F. and Valletti, T. (2016). Selling customer information to competing firms. *Economics Letters*, 149:10–14.

Corts, K. S. (1998). Third-degree price discrimination in oligopoly: All-out competition and strategic commitment. *RAND Journal of Economics*, 29(2):306–323.

Crémer, J., de Montjoye, Y.-A., and Schweitzer, H. (2019). Competition policy for the digital era. Technical report, European Commission, Directorate-General for Competition.

D'Aspremont, C., Gabszewicz, J., and Thisse, J.-F. (1979). On hotelling's "stability in competition". *Econometrica*, 47(5):1145–1150.

De Corniere, A. and Montes, R. (2017). Consumer privacy and the incentives to price-discriminate in online markets. *Review of Network Economics*, 16(3):291–305.

Drexl, J., Hilty, R. M., Desaunettes, L., Greiner, F., Kim, D., Richter, H., Surblyte, G., and Wiedemann, K. (2016). Data ownership and access to data. *Max Planck Institute for Innovation and Competition.*

Dube, J., Fang, Z., Fong, N., and Luo, X. (2017). Competitive price targeting with smartphone coupons. *Marketing Science*, 36(6):944–975.

Dube, J. and Misra, S. (2017). Scalable price targeting. Discussion paper, University of Chicago Booth School of Business.

Duch-Brown, N., Martens, B., and Mueller-Lang, F. (2017). The economics of ownership, access and trade in digital data. Technical report, Digital Economy Working Paper 2017-01; JRC Technical Reports.

Economides, N. (1986). Minimal and maximal product differentiation in hotelling's duopoly. *Economics Letters*, 21:67–71.

Ezrachi, A. and Stucke, M. E. (2016). *Virtual Competition. The Promise and Perils of the Algorithm-Driven Economy.* Harvard University Press.

Gans, J. (2018). Enhancing competition with data and identity portability. The Hamilton Project.

Ganuza, J. and Llobet, G. (2018). Personalized prices in the digital economy. In Ganuza, J. and Llobet, G., editors, *Economic analysis of the digital revolution*, pages 117–142. FUNCAS Social and Economic Studies.

Goldfarb, A. and Tucker, C. (2019). Digital economics. *The Journal of Economics Literature*, 57(1):3–43.

Graef, I., Wahyuningtyas, S. Y., and Valcke, P. (2015). Assessing data access issues in online platforms. *Telecommunications Policy*, 39(5):375 – 387.

Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. (2014). Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*, pages 305–318. ACM.

Harbord, D. and Ottaviani, M. (2002). Contracts and competition in the pay-tv market. Industrial Organization 0203005, University Library of Munich, Germany.

Hindermann, C. M. (2018). Price discrimination in online retail. ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.

Holmes, T. (1989). The effects of third-degree price discrimination in oligopoly. *American Economic Review*, 79(1):244–250.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153):41–57.

Hupperich, T., Tatang, D., Wilkop, N., and Holz, T. (2018). An empirical study on online price differentiation. In *Proceedings of the Eight ACM Conference on Data and Application Security and Privacy*, pages 76–83.

Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, 110(2):569–595.

Jehiel, P. and Moldovanu, B. (2000). Auctions with downstream interaction among buyers. *RAND Journal of Economics*, 31(4):768–791.

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

Johnson, J. and Myatt, D. (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review*, 96(3):756–784.

Kamien, M. I. and Tauman, Y. (1986). Fees versus royalties and the private value of a patent. *The Quarterly Journal of Economics*, 101(3):471–491.

Katz, M. (1984). Price discrimination and monopolistic competition. *Econometrica*, 52(6):1453–1471.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75(3):424–440.

Kim, J.-H., Wagman, L., and Wickelgren, A. L. (2018). The impact of access to consumer data on the competitive effects of horizontal mergers.

Kox, H., Straathof, B., and Zwart, G. (2017). Tareted advertising, platform competition, and privacy. *Journal of Economics and Management Strategy*, 26:557–570.

Lambrecht, A. and Tucker, C. E. (2017). Can big data protect a firm from competition? *CPI Antitrust Chronicle January 2017*.

Lederer, P. and Hurter, A. (1986). Competition of firms: Discriminatory pricing and location. *Econometrica*, 54(3):623–640.

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

Liu, Q. and Serfes, K. (2005). Imperfect price discrimination, market structure, and efficiency. *Canadian Journal of Economics*, 38(4):1191–1203.

Liu, Q. and Shuai, J. (2013). Multi-dimensional price discrimination. *International Journal of Industrial Organization*, 31:417–428.

MacLeoad, W., Norman, G., and Thisse, J.-F. (1988). Price discrimination and equilibrium in monopolistic competition. *International Journal of Industrial Organization*, 6:429–446.

Matsumura, T. and Matsushima, N. (2015). Should firms employ personalized pricing? *Journal of Economics and Management Strategy*, 24(4):887–903.

Mendelson, H. and Parlaktürk, A. K. (2008). Competitive customization. *Manufacturing & Service Operations Management*, 10(3):377–390.

Montes, R., Sand-Zantman, W., and Valletti, T. (2018). The value of personal information in markets with endogenous privacy. *Management Science*, Articles in Advance:1–21.

Myatt, D. and Wallace, C. (2016). Information use and acquisition in price-setting oligopolies. *The Economic Journal*, 128(609):845–886.

Neven, D. (1986). On Hotelling's competition with non-uniform customer distributins. *Economics Letters*, 21:121–126.

Neven, D. and Thisse, J.-F. (1987). Combining horizontal and vertical differentiation: The principle of max-min differentiation.

Pancras, J. and Sudhir, K. (2007). Optimal marketing strategies for a customer data intermediary. *Journal of Marketing Research*, 44(4):560–578.

Prufer, J. and Dengler, S. (2018). Consumers' privacy choices in the era of big data. *CentER Discussion Paper*, 12.

Prufer, J. and Schottmüller, C. (2017). Competition with big data. *CentER Discussion Paper*, 7.

Segal, I. (1999). Contracting with externalities. *The Quarterly Journal of Economics*, 114(2):337–388.

Sen, D. and Tauman, Y. (2007). General licensing schemes for a cost-reducing innovation. *Games and Economic Behavior*, 59(1):163 – 186.

Shilony, Y. (1981). Hotelling's competition with general customer distributions. *Economics Letters*, 8:39–45.

Shy, O., Stenbacka, R., and Zhang, D. (2016). History-based versus uniform pricing in growing and declining markets. *International Journal of Industrial Organization*, 48:88–117.

Steppe, R. (2017). Online price discrimination and personal data: A General Data Protection Regulation perspective. *Computer Law and Security Review*, 33:768–785.

Stole, L. (2003). Price discrimination and imperfect competition.

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

Stucke, M. E. (2018). Should we be concerned about data-opolies? *Geo. L. Tech. Rev.*, 275.

Syam, N. B. and Kumar, N. (2006). On customized goods, standard goods, and competition. *Marketing Science*, 25(5):525–537.

Tabuchi, T. and Thisse, J.-F. (1995). Asymmetric equilibria in spatial competition. *International Journal of Industrial Organization*, 13:213–227.

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

Thisse, J.-F. and Vives, X. (1988). On the strategic choice of spatial price policy. *American Economic Review*, 78(1):12–137.

Townley, C. P., Yeung, K., and Morrison, E. (2017). Big data and personalized price discrimination in EU.

Tucker, D. S. and Wellford, H. (2014). Big mistakes regarding big data. *Antitrust Source, American Bar Association*.

Varian, H. (1989). *Handbook of Industrial Organization, Volume I, Price Discrimination*, pages 598–654. Elsevier Science Publishers B.V.

Vogel, J. (2011). Spatial price discrimination with heterogeneous firms. *The Journal of Industrial Economics*, 59(4):661–676.

Wattal, S., Telang, R., and Mukhopadhyay, T. (2009). Information personalization in a two dimensional product differentiation model: Impact of market structure and the quality-fit ratio. *2009 42nd Hawaii International Conference on System Sciences*, pages 1–10.

Weeds, H. (2016). Tv wars: Exclusive content and platform competition in pay tv. *The Economic Journal*, 126(594):1600–1633.

Whinston, M. (1990). Tying, foreclosure and exclusion. *American Economic Review*, 80(4):837–859.