The value of personal information in markets with endogenous privacy*

Rodrigo Montes[†], Wilfried Sand-Zantman[‡], and Tommaso Valletti[§]

May 27, 2016

Abstract

We investigate the effects of price discrimination on prices, profits, and consumer surplus when at least one competing firm can use consumers' private information to price discriminate yet consumers can prevent this by paying a "privacy cost". Unlike a monopolist, competing duopolists do not always benefit from higher privacy costs because each firm's profit decreases—while consumer surplus increases—with the privacy cost. We also show that, under such competition, the optimal strategy for an owner of consumer data is selling to only one firm, thereby maximizing the stakes for rival buyers. We argue that the resulting inefficiencies imply that policy makers should devote more attention to discouraging exclusivity deals and less to ensuring that consumers can easily protect their privacy.

1 Introduction

This paper studies how customer information and privacy affect the price targeting behavior of a firm facing imperfect competition. We investigate the effects of price discrimination on prices, profits, and consumer surplus when firms can use consumers' private information to discriminate and when consumers are given the means to prevent that discrimination. This analysis leads to some interesting conclusions regarding the value of customer information—that is, the willingness to pay of firms that seek data about potential customers.

In the past several years, price discrimination on the Internet has been documented many times.[1] An especially (in)famous case occurred in 2000 when a customer complained that, after erasing the "cookies" from his computer's browser, he observed a lower price for a particular DVD on Amazon.com.[2] More

*We thank Florian Baumann, Paul Belleflamme, Steffen Hoernig, Morten Hviid, Bruno Jullien, Marc Lebourges, Yassine Lefouili, Martin Peitz, Oz Shy, Olga Slivkó, Rune Stenbacka, Chia-Yu Tsai, Catherine Tucker, J. Miguel Villas-Boas, and Liad Wagman as well as seminar participants at DICE, Florence, Liège, Mannheim, Marburg, Munich, Paris, and UEA. Sand-Zantman gratefully acknowledges Orange for financial support under the IDEI/Orange convention. Montes and Sand-Zantman also acknowledge support from the Agence Nationale de la Recherche, grant ANR-12-BSH1-0009.

†Toulouse School of Economics, rodrigomontesm@gmail.com

‡Toulouse School of Economics and ESSEC Business School, wsandz@tse-fr.eu

§Imperial College London and University of Rome II, t.valletti@imperial.ac.uk

[1] See, for example, Mikians et al. (2012, 2013) for a systematically collected data set and Shaw and Vulkan (2012) for an experimental approach.

[2] `https://en.wikipedia.org/wiki/Amazon.com_controversies#Differential_pricing`

recently, the *Wall Street Journal* (2012a) reported that the travel agency Orbitz Worldwide was showing more expensive hotel offers to Mac users than to PC users. A similar practice has been employed by Staples.com: the *Wall Street Journal* (2012b) revealed that this site displayed different prices once the potential buyers' locations had been identified. Other firms using customers' browsing history and geolocation to vary the offers and products were identified by the newspaper: Discover Financial Services, Rosetta Stone, Home Depot and Office Depot.

The use of data is not a new phenomenon. Maintaining a customer database and conducting market research have long been staples of every business activity. However, technical progress and the digitalization of the economy have expanded the nature (e.g., real-time location data fed by smartphones), sources (e.g., cross-device tracking of a user's web journey), applications (e.g., machine learning) and volume of data. The widespread use of personal data in marketing has thus created an immense demand for consumers' personal information, which in turn has spawned a *data brokers* industry generating an estimated $156 billion (US) annually. Some of these brokers are large data aggregators; others specialize in specific types of consumers (Pasquale, 2014). Examples of such data intermediaries include Acxiom, BlueKai (Oracle), Experian, or Teradata. These firms collect data from a variety of sources, including their own data collection technology, arrangements with website owners allowing the intermediary to implement user tracking technologies (such as cookies and pixels), public information (phone numbers, information available on social networks), data from public authorities and third-party companies (websites, banks, other data brokers).

Our aim in this paper is to study the role of information and privacy in markets. More precisely, we would like to understand the value of consumer information—both for the sellers of this information and for the firms that are buying it. We also study whether consumers should be given more or less control over the information that can be used by firms to adjust their commercial offers. In order to set grounds for our research question and modeling approach, imagine two differentiated retailers that are competing against each other to supply sport shoes to customers in a certain geographic area. Call these companies Adidas and Nike. Adidas and Nike can both approach a data broker that holds unique information about some customers interested in shoes in that area. The several (and linked) questions that we aim to answer are as follows: a) How would Nike price its shoes once it has bought some detailed profiles from the data supplier?; b) How does Nike's pricing strategy depend on whether also Adidas possesses detailed profiles about some customers? c) In the anticipation that the ability to conduct price targeting depends on having access to customers profiles, how do customers protect their information? d) Ultimately, will the data supplier prefer to sell the database exclusively to only one retailer, or will it be more profitable to obtain the fee for data both from Adidas and Nike? After answering these questions, we will also be able to conduct a welfare analysis related to relevant policy questions: e) How are the various parties affected when it is made easier or more difficult for customers to conceal their information (privacy costs)? f) Should data exclusivity contracts be allowed or challenged by the relevant authorities?

1.1 Preview of the model

Towards these ends, we use a model in which firms can first acquire from an intermediary information on consumers' tastes and then offer their products at a set of different personalized prices. We assume that there are two broad groups of consumers, that we call "new" and "old" consumers. For the first group, individual information cannot be extracted (as in the standard Hotelling model of differentiated products with uniform prices). For the second group, information may be available—that is, for purchase from a data supplier. In our setting, the first group can be thought of as offline consumers and online consumers who are not yet active because they are newcomers; the second group is comprised of consumers who are active, leave many traces of their activities which can be collected by the data broker, and so need to engage *ex post* in costly actions to erase those traces. The way active consumers can protect their privacy is by paying a cost to "disappear" from the firm's database. We can interpret this *privacy cost* as the effort consumers spend to conceal their actions online. This means that, if a firm does not know a consumer's type even after paying for information, then the firm cannot tell whether this consumer is a less active consumer or rather an active one who erased her browsing history.

As a benchmark, we consider the case where privacy is not allowed and so information on *all* old consumer preferences is available on a data exchange market. A monopolist will fully exploit this situation and extract all surplus from each old consumer while behaving as a standard uninformed monopolist for new consumers. Under duopoly, the outcome depends on the informational structure—that is, on which firms acquire information about old consumers. We show that, in equilibrium, the data broker chooses to sell the data about consumers to only one firm. Then the uninformed firm sets a price lower than the Hotelling price while the informed firm pursues new consumers less aggressively. As regards those old consumers, the informed firm will match what they could gain by buying from the other firm but will not capture the entire market.

Next we study the case where consumers can pay for privacy and thus avoid being in the database. In this case also there are two de facto markets: one with the new consumers and all the old consumers who paid the privacy cost, and one consisting only of old consumers that chose not to pay the privacy cost. On the first or *anonymous* market, a monopolist sets a price that is increasing in the number of old consumers in this market or, equivalently, that is decreasing in the cost of privacy. On the second market, this firm captures all of the consumer surplus. A higher privacy cost benefits new consumers but makes old consumers in the anonymous market worse-off; however, the monopolist's profit always increases with the privacy cost. In the duopoly case, we show that the data supplier still chooses to deal exclusively with one firm and that the price of information is U-shaped with respect to the privacy cost. Investigating market prices in this case reveals that, provided the privacy cost is not too great, prices on the anonymous market will be higher than in the Hotelling case and will decline as the privacy cost increases. Indeed, for small privacy cost, most old consumers choose to buy on the anonymous market. The uninformed firm can therefore obtain a large profit by focusing on this market and setting a high price. If the privacy cost increases then the size and the taste characteristics of consumers on the anonymous market change, which makes both firms more aggressive. In this setting, larger privacy costs increase competition; hence both firms' individual profits are decreasing in the privacy cost and consumer surplus is increasing in that cost.

We then extend our analysis in three directions. First, we relax the assumption that the data supplier can commit *ex ante* to a particular selling strategy. Relaxing the data seller's commitment requires that consumers form beliefs on the information structure *ex post*. Despite this change, we still find that the data broker chooses to sell exclusively to one firm. The second extension involves the case where the market sizes of old and new consumers differ. We show that varying these relative sizes changes firms' pricing strategies, with greater competition evident when the (relative) mass of new consumers is small. We also show that our result of exclusive dealing on the data market is robust to changing the relative masses of consumers. In the last extension the data seller can offer a recognition technology that allows firms to differentiate the groups of new from old consumers, in addition to offering personal data. In this context we find that the data seller still allocates personal data exclusively but refrains from offering the recognition technology to either firm.

1.2 Literature

Privacy is not a new topic in economics. In the early 1980s, Stigler (1980) and Posner (1981) argued that privacy could reduce economic efficiency by allowing individuals to hide some characteristics. The disclosure of information, and the price discrimination that follows (as shown by Thisse and Vives, 1988), are known to have different effects depending on the perspective taken. Disclosure can increase welfare in that the firm can then sell to consumers with lower valuations (the welfare effect of price discrimination, see Tirole, 1988). *Ex ante*, however, early information disclosure renders insurance impossible and thus hinders efficient risksharing (Hirshleifer, 1971). We shall assume that the market is always fully covered and so price discrimination affects only the distribution of surplus among the agents—that is, it has no direct effect on efficiency. We can still conduct a meaningful welfare analysis because our model incorporates both transportation costs and privacy costs; hence the privacy factors of interest still generate tangible economic trade-offs.

Given the last decade's rise of online markets and Big Data, privacy research has been revived in three main directions.[3] The first one is the study of behavior-based price discrimination (see, e.g. Villas-Boas

[3]For a comprehensive review of the literature on the economics of privacy, see Acquisti et al. (2015). See also Lane et al. (2014) for an accessible approach to policy issues. For empirical work on privacy, see Goldfarb and Tucker (2011, 2012).

1999; Fudenberg and Tirole, 2000; Villas-Boas, 2004; Esteves, 2010; Fudenberg and Villas-Boas, 2012).[4] These papers explore dynamic pricing situations where competition is repeated and where firms use the past behavior of consumers to infer their tastes and then price accordingly. When each consumer is biased toward one particular firm, the first-period choices tell who is a fan of which firm. This revelation leads firms to set high prices for their fans but low prices for the fans of the other firms.

Our model differs from the literature on behavior-based price discrimination in that we do not assume a period in which there is no information. Our model is new and technically simpler compared to the dynamic pricing literature. The fact that we abstract from the process of creating information allows us to extend the research questions in several directions, focusing first on the privacy actions undertaken by consumers and then on how this information is sold to retail firms. Also, we explicitly have in mind situations where retail firms, despite having repeated interactions with their customers, do not have the capability to infer much about them directly so they need to buy additional information and analytics services from a data seller. In turn, a data seller might have collected information about customers in the past and matched both on-line and off-line information, in order to create a mapping of consumer preferences about all those "old" consumers that left digital traces but did not protect their identities.[5]

The second strand of this literature has tackled the privacy issue more directly. Taylor (2004) and Acquisti and Varian (2005) consider repeated sales by a monopolist in each period while assuming intertemporal correlation for tastes. Both papers show that using past information can benefit a firm with myopic but not with rational consumers. Following the Coase (1972) conjecture argument, using past information deters consumers from consuming in the first place and thus reduces the firm's profit. So far there has been only limited analysis of information disclosure in a more competitive setting; a notable exception is Taylor and Wagman (2014), who show that full privacy can be detrimental to consumer welfare, depending on the details of the model used.[6] We endogenize the consumer privacy choice and also insist that the data owner follow its optimal selling strategy, a combination that has not been analyzed before in the privacy literature.[7]

The third strand of the literature shares with us the idea that consumers may decide how much information to reveal. In Conitzer et al. (2012) and Koh et al. (2015), consumers face a monopoly and may chose whether or not to reveal themselves. Conitzer et al. find that the monopoly cannot commit to forgoing the use of past information, as in the Coase conjecture. When the cost of protecting their privacy is high, consumers refrain from paying this cost but also from consuming in the first place. Thus lowering the cost of anonymity has similar effects to increasing the monopolist's commitment power, as both increase the firm's

[4]The marketing literature on couponing, market segmentation, and consumer addressability is closely related (see Chen et al., 2001; Chen and Iyer, 2002).

[5]Another example could be an Internet Service Provider, such as Verizon, capable of putting together a complete picture of a consumer by tracking the sites the consumer visits using its network. Similarly, a broadband service like Google Fiber could blend consumer information with search, maps and email habits of a consumer. The US Federal Communications Commission is now considering whether to ban such practices without permission from a consumer (see Kang, 2016).

[6]See also Liu and Serfes (2006), Chen and Zhang (2009), and Shy and Stenbacka (2016).

[7]A related literature is that on intermediary gatekeepers; see Wathieu (2002), Pancras and Sudhir (2007), and Weeds (2015).

profit. Our paper differs in that the cost of privacy reduces only the monopolist's ability to tailor prices and does *not* affect commitment. Koh et al. focus on voluntary profiling, where refusals to participate lead the monopolist to propose high prices. Voluntary profiling reduces the consumer's search costs of finding the ideal product but generates unsolicited advertisements. Both of these papers cover only the monopoly case and so do not discuss an information seller's strategy. In Casadesus-Masanell and Hervas-Drane (2015), consumers can choose the amount of personal data they reveal to competing platform firms and product quality is increasing in the amount of this information. Their model also features firms that derive revenue from selling consumer data and consumers who value privacy. In contrast, we separate the market for goods and the market for information; we also model differently the manner in which consumers choose their privacy. Chen et al. (2001) shed some light on the market for information. In their model, firms compete to attract nonloyal customers. The authors find that a data broker may be incentivized to sell data nonexclusively if targetability (the firm's ability to distinguish loyal from nonloyal consumers) is sufficiently low. This result runs counter to ours and will be discussed later (see Section 4.2).

The rest of the paper is organized as follows. Section 2 presents the model, and in section 3 we solve the case where consumers are not allowed to have privacy. Section 4 introduces endogenous privacy, and Section 5 details three extensions of the model. We conclude the paper in Section 6 by discussing our results and offering suggestions for future research.

2 The model

There are three different types of agents: consumers, two competing retail firms, and a data supplier that can collect information about consumers. We start by describing preferences of consumers and the two competing firms. There is a continuum of customers each of them with unit demand. A customer receives a utility $v > 0$ from buying the product but nothing otherwise. Consumers have horizontal preferences θ that are uniformly distributed on $[0, 1]$. Let p_i be the price charged by firm i for its product ($i = A, B$). Each firm's marginal cost is normalized to zero. Additionally, consumers must pay a linear transportation cost $t > 0$ in order to buy the good. So if firm i is located at point $x \in [0, 1]$, then consumer θ derives net utility $v - p_i - |x - \theta|t$. When two firms compete on the market, consumers buy from the one that gives them the greatest utility. Throughout the paper we assume that $v \geq 2t$; thus we ensure that the gross utility is high enough that all consumers would consume even if the price were set by a monopolist seller.[8]

Suppose there are two sets of consumers (like the set just described) that differ in terms of the information on consumer preferences obtainable by firms. We refer to the first set as the *new consumers*. Because there is no way to obtain detailed information on them, firms can offer only basic or uniform prices that do

[8]This assumption simplifies the analysis by reducing the number of cases to be studied without loss of economic insight.

not depend on θ. The second set consists of what we call *old consumers*. A firm may be able to obtain information on these consumers, which would allow it to make tailored offers that depend on θ in a way that we will specify. The total mass of each set is normalized to unity.[9]

To flesh out these notions, consider an online retailer that supplies two distinct customer groups. On the one hand, there are newcomers: those about whom there is not yet enough data. These consumers are not necessarily young, only new to the Internet. Yet because their preferences could differ considerably from what the retailer offers, information about the (eventual) online activity of these consumers may not be useful. Suppose, for example, that the retailer sells sporting goods; then even the most detailed data about consumers who never visited any sports websites and never shopped for sport-related items will be of no use to the retailer. At the same time, there are old consumers—that is, those who have actively used the Internet: visiting websites, shopping, leaving reviews, and so forth. These data are likely to be informative about consumer preferences for the goods or services sold by the retailer.

The detection (via Internet activity) of preferences can be evaded by consumers who are willing to pay a cost. We call this the *privacy cost* and denote it by $c > 0$. This cost is associated to the actions consumers take (or the payments they make) to prevent any firm or third party from holding personal data about them. One example of this cost is the difficulty of erasing browser cookies after shopping on the Internet or visiting a website. Regulators could reduce that cost by imposing a full-disclosure policy on the use of cookies.[10] The privacy cost could increase if firms were allowed to trade customer data—so that a consumer might need to request many websites to erase her data.[11] There is also evidence that some consumers are willing to incur a monetary cost to protect their privacy. For example, Reputation.com charges individuals $9.95 per month to remove personal data from online data markets. Another company, Private Internet Access, charges $3.33 per month for a VPN (virtual private network) connection.

Throughout the paper we set $c \leq t$, which generates the economically more interesting cases (it can be shown that, when $c > t$, no consumer would ever pay the privacy cost). We assume that c has a direct effect on utility: if a consumer θ buys the good at p_i from a firm located at x and also pays the privacy cost, then her utility is $v - p_i - |x - \theta|t - c$. For the rest of the paper, c will not depend on the consumer type; hence we can explore more generally the effects of policies that facilitate or impede privacy on the Internet.

Firm i can price its product differently for the two groups. A basic price will be offered to all consumers that i does not recognize; call this the *anonymous market*. This group consists of all new consumers plus those old consumers who paid the privacy cost and thus cannot be detected. The basic price will be based on the firm's belief about the average willingness to pay in the anonymous market.

[9] In an extension, we discuss the effects of changing the relative size of these two sets (see Section 5.2).

[10] As in Europe, for example (see Directive 2002/58/EC).

[11] Firms could also facilitate privacy voluntarily. For example the Network Advertising Initiative (an association of advertisers) encourages its members to offer opt-out provisions so that consumers can more easily avoid being tracked and targeted. However, there is evidence that some participants do not respect consumers' opt-out choices (Mayer, 2012).

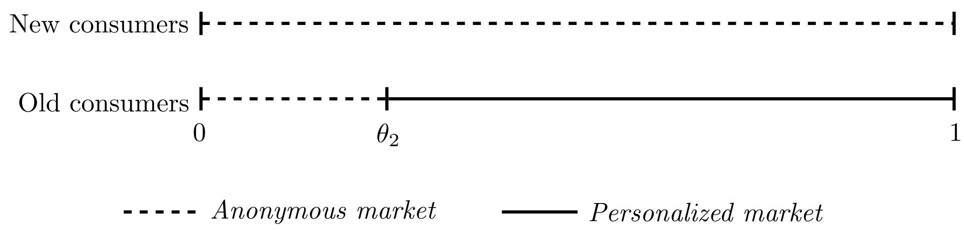

Figure 1: Example of the anonymous and personalized markets

For consumers it does recognize, the firm will make tailored offers based on each consumer's location θ. The consumers on this *personalized market* are simply the old consumers who chose not to pay the privacy cost. Figure 1 represents a case where old consumers with locations between 0 and θ_2 paid the privacy cost and thus cannot be recognized. Therefore, the anonymous market is comprised of these consumers *plus* every new consumer; whereas the personalized market is comprised of old consumers to the right of θ_2.[12]

The last agent in our model is the broker that collects and sells data about consumer preferences. This information can be viewed as personal taste, willingness to pay, brand fidelity, and so on. We assume that the data broker is a monopolist, in that it possesses unique information related to specific customer profiles. In this paper, we do not model how the broker *acquires* its supply of data; we are more interested in how it chooses to *sell* the data and how the data are used by retailers. Data brokers collect and store a vast amount of data on almost every household and commercial transaction in the US. The Federal Trade Commission (FTC) recently surveyed the US landscape, and found a concentrated market structure, which is possibly expected given the economies of scale associated with Big Data analytics. According to the FTC, one data broker's database has information on 1.4 billion consumer transactions and over 700 billion aggregated data elements; another data broker's database covers one trillion dollars in consumer transactions; and yet another data broker adds three billion new records each month to its databases (FTC, 2014). This information is then sold to retail firms. For instance, BlueKai supplies centralized information (on consumers' past experience and observed characteristics) that is then auctioned to retailers (BlueKai, 2015).

However, consumers are not passive and can take actions to avoid appearing in the database. There are many reasons to do so. First, privacy is a good in itself, and many consider it unethical for a company to compile personal information. Second, some people may wish to conceal criminal or shameful activities. Third, consumers might be aware that firms can set different prices for different customers based on personal data. It is this aspect of such information that our model incorporates.[13]

[12] In section 5.3 we study an extension where the data seller has a further recognition technology to differentiate the market of old consumers from the market of new consumers.

[13] We emphasize that the only reason for paying the privacy cost in our model is because a tailored price is expected to be higher than the basic price for a particular consumer θ; thus we do not include any intrinsic benefits of privacy. However, c could also be viewed as a *net* privacy cost that includes an intrinsic value for privacy. The comparative statics are unchanged in that case.

3 No privacy

As a benchmark, we first study the case where consumers cannot pay for privacy. This is equivalent to saying that the cost c is extremely high. First we analyze the monopoly case, followed by the case where two firms compete. In this section, the set of new consumers is equivalent to the anonymous market and the set of old consumers is equivalent to the personalized market; hence we use the terms interchangeably.[14] Throughout the paper we solve for perfect Bayesian equilibria in pure strategies.

3.1 Monopoly

Consider a monopolist seller (denoted as firm i) located at 0 with respect to both sets of consumers (old and new).[15] The model's timing proceeds as follows:[16]

Stage 1. The data supplier posts a price T for the data.

Stage 2. Firm i decides whether or not to purchase the data.

Stage 3. Firm i determines its basic price p_i.

Stage 4. Firm i can offer tailored prices.

Stage 5. Consumers buy and consume.

Let us assume that the monopolist bought the information from the data broker at stage 2. The monopolist's market share among new consumers is then all the consumers $\theta \in [0, \theta_1]$ such that $v - p - \theta_1 t = 0$. This means that θ_1 represents a consumer who is indifferent between buying the good and taking the outside option, which he values at zero:

$$\theta_1 = \min\left\{\frac{v-p}{t}, 1\right\}.$$

For the old consumers, the firm offers prices $p(\theta)$ tailored to each θ. Since the outside option is assumed equal to zero, it follows that the tailored prices can be as high as needed to capture all the surplus. Therefore,

$$p(\theta) = v - t\theta;$$

here $v \geq 2t > t$ guarantees that $p(\theta) > 0$. Hence, the monopolist solves

$$\max_{p \geq 0} \quad \int_0^{\theta_1} p \, d\theta + \int_0^1 p(\theta) \, d\theta$$

[14]This will not be the case in Section 4, which incorporates privacy.

[15]The results of this section would be qualitatively similar if the firm were located in the middle of each market.

[16]Merging stages 3 and 4 would make no difference in this case, but we maintain this distinction for consistency with the competition case to be analyzed in Section 3.2. Also note that we denote the monopolist more generally as firm i in order to use the same timing in the later sections with competition.

$$\text{s.t.} \quad 0 \le \theta_1 \le 1.$$

Under the mentioned assumption $v \geq 2t$, we obtain $p^* = v - t$ and $\theta_1 = 1$. The price p^* is the equilibrium basic price and is offered only to new consumers. Old consumers are offered the tailored price $p(\theta)$ and are left with no surplus. These results imply that both markets are fully covered.

Total profits π are given by the sum of the profits π^O obtained from selling to old consumers and the profits π^N from selling to new consumers. The respective equilibrium values are

$$\pi^O = \int_0^1 (v - t\theta)\, d\theta = v - \frac{t}{2} \quad \text{and} \quad \pi^N = \frac{v - p^*}{t} p^* = v - t.$$

Hence total profits are $\pi = 2v - 3t/2$. Note that *without* information on consumers the monopolist's profit is $2(v - t)$—that is, twice π^N. As a result, the maximum price T^* that a retailer will pay for this data set is

$$T^* = \pi - 2(v - t) = \frac{t}{2}.$$

The consumer surplus for old and new consumers are given, respectively, by

$$\text{CS}^O = \int_0^1 (v - (v - t\theta) - t\theta)\, d\theta = 0 \quad \text{and} \quad \text{CS}^N = \int_0^1 (v - (v - t) - t\theta)\, d\theta = \frac{t}{2}.$$

These expressions imply that total consumer surplus is simply $\text{CS} = \text{CS}^N = t/2$.

3.2 Competition

We now study the case where two firms, A and B, compete. We assume that their locations are fixed with respect to all consumers, with firm A located at $\theta = 0$ and firm B at $\theta = 1$. We look at different cases depending on which firms have information about consumers. As in Section 3.1, information is available only about old consumers and no information can be obtained about new consumers. The timing in the competitive case is the same as in Section 3.1, for firm $i = \{A, B\}$.[17]

In stage 2, firms can buy information from an upstream data supplier. We assume that this is the only stage at which information can be sold. In other words, the data supplier's stage 1 commitment to T is *ex ante*—prior to the retailers' purchase decisions.[18]

As a benchmark, consider the case in which neither firm has information. This is equivalent to solving

[17] In this section, sequentiality in the choice of basic and tailored prices is necessary for there to be an equilibrium in pure strategies. Firm A will tailor a price that is optimally a function of B's basic price, so B can always reduce its price and thereby increase its market share. Sequentiality helps ensure that a pure-strategy equilibrium will exist because the optimal tailored price will always exist. We prove in Appendix A that a pure-strategy equilibrium does not exist if prices are chosen simultaneously.

[18] We will relax this assumption in an extension where interim sales are allowed; that is, firms can buy information at later stages (see Section 5.1).

a Hotelling model in two identical markets. In each market, the prices p_A and p_B are both equal to the transportation cost t; this implies that the market is evenly split between firms.[19] Then each firm's profits from selling to old and new consumers are $t/2$ while total profits for each firm are $\pi_A = \pi_B = t$.

Finally, for each consumer type the surplus is given by

$$\text{CS}^O = \text{CS}^N = \int_0^{1/2} (v - t - t\theta)\, d\theta + \int_{1/2}^{1} (v - t - t(1-\theta))\, d\theta = v - \frac{5t}{4}.$$

Therefore, total consumer surplus $\text{CS} = 2v - 5t/2$.

3.2.1 When firm A and firm B both have information

The case in which both firms have purchased consumer information is studied by Taylor and Wagman (2014). The only difference is that we consider two sets of consumers and two types of prices.

Proposition 1. *Assume that both firms A and B buy the information and that consumers cannot pay for privacy. Then the basic prices are*

$$p_A = t \quad \textit{and} \quad p_B = t$$

while the tailored prices are

$$p_A(\theta) = \max\{t(1-2\theta), 0\} \quad \textit{and} \quad p_B(\theta) = \max\{t(2\theta-1), 0\}.$$

Profits are $\pi_A = \pi_B = 3t/4$ *and consumer surplus is given by* $\text{CS} = 2(v - t)$.

Proof. See Taylor and Wagman (2014). □

As compared with the no-information case, profits decrease while consumers' surplus increases. The reason is that competition reduces the potential for rent extraction when both firms have information about consumers; it is as if both firms were competing à la Bertrand for each consumer. Because the goods are differentiated, both prices are equal to marginal cost (here equal to zero) only for the consumer equidistant from both firms. For all other consumers, one of the two firms is preferred; hence that firm can set a strictly positive price and still sell its products.

3.2.2 When only firm A has information

In this section, we focus on the case where information is sold to one firm only and assume, without loss of generality, that it is firm A. Note that in this case the consumer's outside option is not zero;[20] rather, it is

[19] See Belleflamme and Peitz (2010).

[20] This nonzero outside option is guaranteed by our assuming $v \geq 2t$, so that even consumers closer to A would derive strictly positive surplus if they were to buy from B.

the utility associated with buying from B instead of from A. Buying from A leads to the utility level

$$v - t\theta - p_A$$

whereas buying from B leads to

$$v - t(1-\theta) - p_B;$$

here p_A and p_B are the prices that firm A and firm B charge for their respective products. In the *anonymous* market, the market shares $[0, \theta_1]$ and $[\theta_1, 1]$ are determined by the indifferent consumer's choice of between buying from A or B:

$$\theta_1(p_A, p_B) = \frac{1}{2} + \frac{p_B - p_A}{2t}.$$

In the *personalized* market, firm A offers (at stage 4) a tailored price $p_A(\theta)$ that leaves consumers indifferent between accepting and buying B's product.[21] As a consequence,

$$p_A(\theta) = p_B + (1 - 2\theta)t.$$

We define the last consumer buying from A by θ_2^0, where θ_2^0 is such that $p_A(\theta) = 0$. Using the expression just displayed for $p_A(\theta)$, we obtain

$$\theta_2^0(p_B) = \frac{p_B + t}{2t}.$$

Next we consider the game's third stage where both firms choose their basic prices. Firm A's profits are given by

$$\pi_A = \int_0^{\theta_1} p_A \, d\theta + \int_0^{\min\{\theta_2^0, 1\}} p_A(\theta) \, d\theta,$$

and B's profits are

$$\pi_B = \int_{\theta_1}^1 p_B \, d\theta + \int_{\min\{\theta_2^0, 1\}}^1 p_B \, d\theta.$$

There are two possible cases. In the first, B's price p_B is low enough to attract some old consumers; this corresponds to the case where $\theta_2^0 < 1$. If instead $\theta_2^0 \geq 1$, then firm B does not operate in the personalized market. In either case, A's reaction function is the same because $p_A(\theta)$ does not depend on p_A. The implication is that firm B must determine the game's outcome based on its own profits. Firm B can choose either to give away every customer in the personalized market by setting a high price or to lower its price and thus keep some consumers in this market.

Our next proposition shows that firm B finds it optimal to operate in the personalized market and therefore chooses a low basic price.

[21] Under indifference, consumers buy from the informed firm (which we have assumed to be firm A).

Proposition 2. *Assume that only firm A buys the information and that consumers cannot pay for privacy. Then the equilibrium prices on the anonymous market are*

$$p_A = \frac{6t}{7} \quad \text{and} \quad p_B = \frac{5t}{7};$$

profits are

$$\pi_A = \frac{54t}{49} \quad \text{and} \quad \pi_B = \frac{25t}{49};$$

and consumer surplus is given by $\mathrm{CS} = 2v - 110t/49$.

Proof. We verify that firm B finds it profitable to set a price such that it operates in the personalized market ($\theta_2^0 < 1$). The reaction of firm A is simply $p_A = \frac{p_B+t}{2}$. If B decides to compete in the personalized market then its reaction function will be the solution of

$$\arg\max_{p_B} \, [(1 - \theta_2^0(p_B)) + (1 - \theta_1(p_A, p_B))]p_B,$$

which leads to $p_B(p_A) = \frac{2t+p_A}{4}$. Combining this with A's reaction function yields $p_A = 6t/7$ and $p_B = 5t/7$ as well as respective profits for A and B of $\pi_A = 54t/49$ and $\pi_B = 25t/49$. If firm B does *not* operate in the personalized market, then its program can be written as

$$\arg\max_{p_B} \, [1 - \theta_1(p_A, p_B)]p_B.$$

Firm B's reaction function is then given by the standard Hotelling reaction function, $p_B(p_A) = \frac{t+p_A}{2}$, which results in a price $p_B = p_A = t$ and in profits of $\pi_B = t/2$. Because this latter profit is smaller than the former, B chooses a price to compete both in the personalized market and in the anonymous market. □

We find that prices are lower here than in the no-information case. As a result, consumer surplus is at its highest when both firms have consumer information and at its lowest when neither firm has such information. We find that the informed firm is more profitable than an uninformed competitor. Moreover, the informed firm makes higher profits in this case than in any of the previous cases.

3.2.3 The price of information

We can now analyze how the value of information under the various setups studied in Section 3.2 should affect the data broker's selling strategy. Here we suppose that this data supplier, denoted DS, owns the information about old consumers and can post a price T for it and sell it at stage 2, and only at that stage. The DS can choose different selling strategies by setting different prices, which in equilibrium will induce either one or both firms to buy the data. We denote T_i the price paid by each firm when a number i of firms buys the information. The allocation and payments happen at stage 2, and we further assume that this trade is common knowledge.

If we suppose that DS chooses a strategy to maximize its profit, then this profit is given by

$$\pi_{\text{DS}} = \max\{T_1, 2T_2\}.$$

To characterize the prices T_1 and T_2, we assume that DS has all the bargaining power; a natural assumption given that both firms A and B are competing. The setting is similar to an auction with externalities, as in Jehiel and Moldovanu (2000). Indeed, suppose first that DS sells to only one firm. Then the maximum price T_1 that DS can set is the difference between winning and losing the auction for one firm's profits—that is, the firm's profits when it has consumer information *minus* the firm's profits when its rival has that information. In this first case we obtain

$$T_1 = \frac{54t}{49} - \frac{25t}{49} = \frac{29t}{49}.$$

Analogously, if DS sells to both firms then T_2 represents the difference in (say) firm A's profit between the case when both firms can use consumer information and that when only firm B can use this information. Plugging in the profit values derived previously, we obtain

$$T_2 = \frac{3t}{4} - \frac{25t}{49} = \frac{47t}{196}.$$

Therefore, DS chooses to sell the information only to one firm because $T_1 > 2T_2$. The next proposition formalizes this section's main result.

Proposition 3. *Information is sold to only one firm at* $T_1 = 29t/49$ *and the firms' net profits are* $\pi_A = \pi_B = 25t/49$.

Proof. See the text. □

Proposition 3 establishes that it is optimal for the owner of information to grant only exclusive rights over the database. The reason is that competition is too intense when both firms have information and the potential for rent extraction is minimized. That being said, an informed firm can extract more rents from consumers if its competitor must set a single basic price. The exclusive allocation can be implemented via an auction under which the DS *commits* to selling the data to only one firm.

4 Privacy

We now turn to the case where consumers can pay for privacy. As noted in the description of our model, consumers can incur a cost c to not appear in the database. We assume that consumers can observe which firm(s) purchased this information and can form rational expectations about prices.[22]

4.1 Monopoly

We continue to assume that the monopolist is located at $\theta = 0$ with respect to both new and old consumers. The timing is the same one as in Section 3.1 *except* for the addition of a stage at which consumers can protect their privacy. This new stage, 2.5, occurs just after the firm decides whether or not to buy the data.

Stage 2.5 Consumers make their privacy choice—that is, they decide whether or not to pay the cost c.

As in the case with no privacy, any consumer located at θ for whom the firm has information will be charged the tailored price $p(\theta) = v - t\theta > 0$ and will derive zero utility. This group is called the *personalized market*, and it consists solely of every old consumer who did not opt for privacy. All other consumers make up the anonymous market. Because each consumer's privacy choice is made at stage 2.5 (i.e., before any price has been set), we need to define an anticipated price $p^a \geq 0$. This price is the basic price that old consumers expect to pay if they are not in the firm's database. In effect, for these consumers the basic price is now $p^a + c$. As consumers anticipate a higher price, fewer consumers would pay to avoid detection. If $p^a + c$ remains small, however, then consumers would be incentivized to opt for privacy at stage 2.5. A consumer located at θ will pay the privacy cost when

$$v - t\theta - p^a - c \geq 0 \iff \theta \leq \theta_2 = \frac{v - p^a - c}{t}.$$

To solve the model, assume that the firm buys consumer information at stage 2. Then, at stage 3, the monopolist maximizes profits $\pi(p)$ given the anticipated p^a:

$$\pi(p) = \int_0^{\theta_1} p\, d\theta + \int_0^{\min\{\theta_1,\theta_2\}} p\, d\theta + \int_{\theta_2}^1 p(\theta)\, d\theta. \tag{1}$$

The first term on the right-hand side (RHS) represents the profits from selling to new consumers, and the second term represents the profits from the old consumers who have paid the privacy cost. Together, these two terms capture the firm's profits on the anonymous market. Observe that the firm could set a high enough price p that the actual market share θ_1 is exceeded by the proportion of consumers who are willing to pay the privacy cost. This is why the upper bound in the second integral is $\min\{\theta_1, \theta_2\}$. The last term on

[22] In Section 5.1, we investigate the case where consumers make their privacy decision *before* information is bought.

the RHS represents the personalized market; consumers on this market are subject to price discrimination and pay the tailored price $p(\theta)$.

The firm then maximizes the profits expressed in equation (1) for $p \geq 0$ subject to the following constraint:

$$1 \geq \theta_1 \geq 0. \tag{2}$$

Furthermore, in any equilibrium we require that the proportion of consumers who pay the privacy cost remains bounded:

$$1 \geq \theta_2 \geq 0. \tag{3}$$

Proposition 4. *Suppose the firm buys the consumer information.*

- *If* $c \leq \max\{3t - v, 0\}$, *then not all new consumers buy the product, with*

$$p^* = \frac{2v - c}{3} \quad \textit{and} \quad \theta_2^* = \frac{v - 2c}{3t}$$

while profits and consumer surplus are, respectively,

$$\pi = \frac{2c^2 + v^2}{6t} + \frac{1}{2}(2v - t) \quad \textit{and} \quad \text{CS} = \frac{5c^2 - 2cv + 2v^2}{18t}.$$

- *If* $t \geq c > 3t - v$, *then all new consumers buy the product, with*

$$p^* = v - t \quad \textit{and} \quad \theta_2^* = 1 - \frac{c}{t}$$

while profits and consumer surplus are, respectively,

$$\pi = \frac{c^2}{2t} + 2(v - t) \quad \textit{and} \quad \text{CS} = t + \frac{c^2}{2t} - c.$$

Proof. Note that $\theta_2 > \theta_1$ cannot be an equilibrium because consumers at stage 2.5 correctly anticipate the prices chosen by the monopolist at stage 3 and so $p^a = p$. Then in any equilibrium we must have $\theta_2 \leq \theta_1$. Under this condition, the firm takes the quantity of consumers who pay the privacy cost θ_2 as given. If we neglect conditions (2) and (3), then the first-order condition $\pi'(p) = 0$ leads to

$$p = \frac{-p^a - c + 2v}{2}.$$

We finally apply rational expectations, under which $p^a = p$, and solve for p. We remark that if $p^a = p$ then $\theta_1 \leq 1$ implies $\theta_2 \leq 1$. In this case, the solution requires only that $\theta_1 \leq 1$ be interior. The second solution is the corner solution when $\theta_1 = 1$. □

The basic price p^* is now obtained by assessing the average willingness to pay in the anonymous market, which corresponds to the weighted average between the willingness to pay of new consumers and the willingness to pay of old consumers who have already paid a privacy cost c. Note also that the average willingness to pay of those who paid the privacy cost is greater than the average willingness to pay of new consumers—given that consumers who pay that cost are located closer to the firm. Therefore, we found that when c increases p^* decreases because fewer consumers choose privacy, which reduces the demand for the firm's product due to a lower average valuation.

Corollary 1. *Assume that the firm buys the information. Then profits are always increasing in c. Additionally, if $c \leq 3t - v$ then consumer surplus is U-shaped with respect to c and is minimized at $v/5$; if $c > 3t - v$ then consumer surplus is always decreasing in c.*

Proof. The claims follow directly from the relevant expressions in Proposition 4. □

Corollary 1 states that profits are increasing in the privacy cost. Profits increase in c because, notwithstanding the lower basic price, a smaller proportion of consumers pay the privacy cost. Then higher profits in the personalized market compensate the firm for its losses in the anonymous market.

Consider first the interior case $c \leq 3t - v$, where consumers' surplus is U-shaped in c. This pattern reflects that new consumers always gain with a larger c whereas old consumers always lose. For new customers, a larger cost is associated with a larger market share and a lower price; the opposite holds for old customers. This can be seen by observing that the effective basic price paid by old consumers is $p^* + c$. Corollary 1 states that the losses from old consumers (resp., the gains from new consumers) dominate when $c \leq v/5$ (resp., when $c > v/5$).

If instead $c > 3t - v$, then increasing c has no effect on the set of new consumers because full coverage is achieved and so the firm no longer lowers its price on the anonymous market. Then, when c increases, the surplus of old consumers declines as fewer consumers pay for privacy and hence must pay their full valuation.[23]

As in the case without privacy, the data supplier can extract any profit the monopolist makes in excess of $2(v - t)$, which are the profits the monopolist could make without having any information at all.

4.2 Competition

Now we consider the duopoly case where consumers can pay for privacy. The presentation here is similar to that in Section 3.2, although we incorporate (as in Section 4.1) an intermediate stage 2.5 in which consumers

[23] When $c = t$, Proposition 4 leads to the no-privacy result with $\theta_2^* = 0$.

can choose to pay for privacy. So the timing in this section likewise changes to accommodate consumers' actions.

Remark that if neither firm buys the information then the ability of consumers to pay for privacy does not alter the benchmark result derived in Section 3.2. Hence if consumers observe that no firm has information then no consumer will pay the cost c, and the resulting prices will be the standard Hotelling prices.

4.2.1 When firm A and firm B both have information

Consider the case where, at stage 2, both firms buy the consumer information. In this case, increased competition leads to tailored prices low enough that it makes no economic sense for *any* consumer to pay for privacy.

Lemma 1. *If consumers observe that both firms purchased the information, then no consumer pays the privacy cost.*

Proof. Take any subset of old consumers who did not pay the privacy cost. Competition in this subset is for every consumer, so the tailored prices are equivalent to those given in Proposition 1. The implication is that any consumer who prefers privacy does so for the resulting opportunity to pay only the basic price offered by the nearest firm — because the alternative (the tailored price) already matches the competitor's offer. The rest of the proof relies on the following two arguments.

1. If a consumer $\hat{\theta} \in [0, \frac{1}{2}]$ pays the privacy cost, then every consumer $\theta \in [0, \hat{\theta}]$ pays the privacy cost because utility is decreasing in the transportation cost $t\theta$. By symmetry, a similar statement is true in $(\frac{1}{2}, 1]$.

2. If a strictly positive mass of consumers pays the privacy cost, then p_A and p_B should be greater than t. Suppose consumers $\theta \in [0, \theta_A]$ pay the privacy cost to hide from firm A. Then A solves

$$\max_{p_A} \int_0^{\theta_A} p_A \, d\theta + \int_0^{\frac{1}{2}+\frac{p_B - p_A}{2t}} p_A \, d\theta.$$

Firm B solves an analogous problem (where consumers $\theta \in [\theta_B, 1]$ pay the privacy cost); then prices are

$$p_A = t + \frac{2t(2\theta_A + 1 - \theta_B)}{3} \quad \text{and} \quad p_B = t + \frac{2t(\theta_A + 2(1 - \theta_B))}{3}.$$

The first argument implies that consumers to the left of $\frac{1}{2}$ would pay the privacy cost if and only if $p_A + c < t(1 - 2\theta)$, and the second argument implies that $p_A > t$. These two conditions cannot hold for $c > 0$, from which it follows that no consumer would pay the privacy cost. □

In the personalized market, both firms will compete for every consumer and the tailored prices will be low. Therefore, consumers will pay the privacy cost only if basic prices are even lower than tailored prices (after accounting for the privacy cost c); but such basic prices are not profitable.

Note that, from the perspective of total welfare, we achieve a first-best allocation in this case: transportation costs are minimized (since market shares along both Hotelling lines are symmetric) and no consumer engages in wasteful activities (since no one pays the privacy cost).

In sum: if consumers *can* pay for privacy and if information is distributed symmetrically between firms, then in equilibrium consumers *do not* pay for privacy. This means consumers' entitlement to privacy has no effect over the outcome of the model when both firms have information. Therefore, Proposition 1 still holds with Hotelling prices offered to new consumers and lower personalized prices offered to old consumers.

4.2.2 When only firm A has information

As in Section 3.2.2, we assume that firm A buys the information and B does not. Some of the old consumers might pay the privacy cost to avoid the tailored price. The proportion of such customers — who prefer to pay c and to buy from firm A at the basic price, rather than the tailored price — is given by

$$v - p_A^a - \theta t - c \geq v - p_A^a(\theta) - \theta t = v - t(1-\theta) - p_B^a.$$

As in Section 4.1, this share must be based on *anticipated* prices because consumers make their privacy choices at stage 2.5 yet prices are not set until stages 3 and 4. In other words, by the time prices are set, firm A treats the proportion of consumers who paid the privacy cost as fixed. In this case, the indifferent consumer type is given by

$$\theta_2 = \frac{1}{2} + \frac{p_B^a - p_A^a - c}{2t}. \tag{4}$$

The equilibrium is given in the following result.

Proposition 5. *Assume that only firm A buys the information and that consumers can pay for privacy. Then there exists a threshold $\bar{c} < t$ such that, if $c \leq \bar{c}$, then the equilibrium consists of*

$$p_A = t + \frac{t-c}{2}, \quad p_B = t + \frac{t-c}{4}, \quad \text{and} \quad \theta_2 = \frac{3(t-c)}{8t};$$

if $c > \bar{c}$, then the equilibrium is as described in Proposition 2.

Proof. First we verify that firm B chooses a price such that it does not sell to old consumers ($\theta_2^0 \geq 1$). Under this setup, A's reaction function is always given by

$$\arg\max_{p_A} \, [\theta_1(p_A, p_B) + \theta_2(p_A^a, p_B^a)]p_A.$$

If B does *not* sell to old consumers, then its reaction function is the same as in the no-privacy setup:

$$\arg\max_{p_B} \, [1 - \theta_1(p_A, p_B)]p_B.$$

These two functions determine $p_A(\theta_2)$ and $p_B(\theta_2)$. Finally, because in equilibrium $p_A = p_A^a$ and $p_B = p_B^a$ (which implies $\theta_2 < \theta_1$), we can solve for θ_2 and find $\theta_2 = \frac{3(t-c)}{8t}$. The resulting prices lead to $\pi_B = \frac{(5t-c)^2}{32t}$. If B *does* sell to old consumers, then its reaction function is again

$$\arg\max_{p_B} \, [(1 - \theta_2^0(p_B)) + (1 - \theta_1(p_A, p_B))]p_B.$$

Note that the value if θ_2^0 is a direct consequence of $p_A(\theta) \geq 0$. Since $p_A(\theta) = p_B + (1-2\theta)t$ is always chosen by firm A at stage 4, it follows that θ_2^0 is a function of p_B and not of p_B^a.

Following the same procedure as before, we obtain $\theta_2 = \frac{6t-7c}{20t}$ and $\pi_B = \frac{(8t-c)^2}{100t}$. Then, to keep $\theta_2 \geq 0$ we must have $c \leq \frac{6t}{7}$. For a larger c, no consumer pays for privacy and prices are the same as in Proposition 2. It is therefore sufficient to establish for $c \leq \frac{6t}{7}$ that $\frac{(5t-c)^2}{32t} \geq \frac{(8t-c)^2}{100t}$. Then there exists a $\bar{c} = 5t - \frac{20}{7}\sqrt{2}t \approx 0.96t$ such that, for $c \leq \bar{c}$, we have $\frac{(5t-c)^2}{32t} \geq \frac{25t}{49}$. □

We discover that prices under this scenario are higher than Hotelling prices for $c \leq \bar{c}$. From firm A's perspective, the average willingness to pay in the segment $[0, \theta_2]$ is greater than the average willingness to pay of new consumers because the former are closer to A. It follows that the average willingness to pay in the anonymous market is greater than the average willingness to pay in a set comprising new consumers only;[24] hence $p_A \geq t$. It is worth noting that the inequality $p_A \geq p_B$ results from the difference in the average willingness to pay for each product, given that no consumer in $[0, \theta_2]$ buys from firm B.

That the price p_B is never less than t stems from B's finding it unprofitable to compete for old consumers; in other words, firm A supplies every old consumer. This means that $\theta_2^0 \geq 1$, which in turn implies that $p_B \geq t$. The change of regime at $\bar{c}$ is given by B's decision to sell to old consumers,[25] after which B's equilibrium profits are continuous at $\bar{c}$. Yet, θ_2 is discontinuous at $\bar{c}$, where it drops to 0 (i.e., no consumer pays for privacy). We will therefore focus on $c \leq \bar{c}$ in order to study how the equilibrium is affected by privacy costs.

[24]To see this, let v^N be the average willingness to pay of new consumers and let v^O be the average willingness to pay in the segment $[0, \theta_2]$ of old consumers. Since $v^O > v^N$, it follows that $\frac{v^N + \theta_2 v^O}{1+\theta_2} > v^N$.

[25]Details are given in the proof of Proposition 5.

Just as in the monopoly case, a change in c has a negative impact on firm A's basic price. Indeed, as the privacy cost increases, fewer consumers pay for privacy. This fact leads to a reduction in the average willingness to pay for A's product in the anonymous market, that in turn reduces the demand for its good. Finally, due to price complementarity, B's price is also decreasing in c.

Corollary 2. *Assume that only firm A buys the information and that consumers can pay for privacy. Then the profits of both firms are decreasing in c whereas consumer surplus is increasing in c.*

Proof. If $c \leq \bar{c}$, then profits and consumer surplus can be written as follows:

$$\pi_A = \int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^{1} (p_B + (1 - 2\theta)t) \, d\theta = \frac{11c^2 - 22ct + 107t^2}{64t},$$

$$\pi_B = \int_{\frac{p_B - p_A + t}{2t}}^{1} p_B \, d\theta = \frac{(5t - c)^2}{32t};$$

$$\text{CS} = \int_0^{\frac{p_B - p_A + t}{2t}} (v - p_A - \theta t) \, d\theta + \int_0^{\theta_2} (v - p_A - \theta t - c) \, d\theta + \int_{\theta_2}^{1} (v - (p_B + (1 - 2\theta)t) - \theta t) \, d\theta$$

$$+ \int_{\frac{p_B - p_A + t}{2t}}^{1} (v - p_B - (1 - \theta)t) \, d\theta = \frac{10ct + 5c^2 - 103t^2}{32t} + 2v.$$

If $c > \bar{c}$, then the equilibrium is the same as the one found in Proposition 2. We can show that $\pi'_A(c), \pi'_B(c) < 0$ because $c < t$ and $\text{CS}'(c) > 0$. In the proof of Proposition 5 we established that there is a threshold $\bar{c}$ above which firm B poaches some old consumers from firm A, which leads to basic prices low enough that consumers do not pay for privacy; hence the equilibrium in Proposition 2 applies. Then profits and consumer surplus are linear in $c > \bar{c}$. It follows that Corollary 2 holds for any $c \leq t$. □

Observe that, as c increases, a smaller proportion of consumers pay the privacy cost. This dynamics reduces the average willingness to pay for A's basic price. In response, A reduces its price and then B does the same (to remain in competition for the new consumers). Because p_B decreases, the tailored price should also decrease because A must leave more surplus to the consumers—hence again they will have no incentive to buy from firm B. Thus the profits of both firms decline, which contrasts sharply with the monopoly case.

Of course, consumers benefit from this competition. From Corollary 2 it follows that consumer surplus is always increasing in c. This is because when c increases, all prices fall and the number of consumers who pay the privacy cost also declines. These two effects are more than enough to compensate for the consumers who pay a higher c. Furthermore, it can be seen that consumers surplus exhibits an upward jump at $\bar{c}$ because the firms compete more aggressively when B supplies old consumers as well.

We can also compute the privacy gains for old consumers close to A as the difference in utility from paying for privacy *minus* the utility from not paying:

$$(v - p_A - t\theta - c) - (v - p_A(\theta) - t\theta) = \frac{1}{4}(t(3 - 8\theta) - 3c),$$

which is decreasing in c. Indeed, as c increases these consumers benefit from a lower price but suffer from the direct effect of an increased privacy cost—and that direct negative effect dominates the positive, price-mediated effect. Therefore, consumers who protect their privacy are worse-off when privacy is more costly while all other consumers are better-off in that case.

4.2.3 The price of information

The price of information when consumers can pay for privacy is calculated in the same way as in Section 3.2.3. The data supplier DS compares what it can obtain from selling to one firm, T_1, with what it can obtain from selling to both firms, $2T_2$. Proposition 6 shows that, even when consumers can pay for privacy, the DS finds it optimal to sell its consumer information to one firm only.

Proposition 6. *If $c \leq \bar{c}$, then information is sold to one firm at a price $T = \frac{9c^2 - 2ct + 57t^2}{64t}$ and the net profits are $\pi_A = \pi_B = \frac{(5t-c)^2}{32t}$. If $c > \bar{c}$, then Proposition 3 holds.*

Proof. The case where $c > \bar{c}$ has already been solved. For $c \leq \bar{c}$ we use the proof of Corollary 2 to check that $T_1 > 2T_2$, where $T_1 = \pi_A - \pi_B$ and $T_2 = \frac{3t}{4} - \pi_B$. Indeed,

$$\pi_A - \pi_B \geq 2\left(\frac{3t}{4} - \pi_B\right)$$

and so $T = T_1$. □

Thus our previous finding that the data supplier prefers to sell its consumer information to just one firm is robust to introducing privacy. Remarkably, the price that DS can set for the information is higher when the privacy cost is not too high. In fact, the informed firm (A) knows that if privacy is allowed (or if c is not too high) then it can charge higher prices because its customers have a higher average valuation (since only those consumers close to the informed firm will pay for privacy). Therefore, competition is less intense and there are larger gains to be made when acquiring information exclusively. Hence the possibility of privacy reinforces the result that information is allocated to only one firm.[26]

Proposition 6 is at odds with Chen et al. (2001), who postulate that a nonexclusive allocation can be optimal for the data seller. That could occur because price competition is not strong when targetability is sufficiently low, and firms set higher prices the more they mistake loyal customers from searches. The

[26] We are able to show that the exclusive data sales result is robust to introducing vertically differentiated retail firms instead of horizontally differentiated.

authors define privacy as the state that prevails when targetability is imperfect and firms cannot properly recognize their loyal customers; in this model, mistargeting occurs uniformly across consumers. In contrast, privacy (and hence targetability) in our model depends on the information structure and on competition. That explains why we do not encounter cases with both a high level of privacy (low targetability) and nonexclusive sales of data.

Finally, given that T in Proposition 6 is U-shaped with respect to c, we can show that the corner $c = \bar{c}$ yields the highest revenue for the DS.[27] Furthermore, since T is minimized at $c = t/9$, it follows that the value of information is generally increasing in c because, when $c > t/9$, the outside option of not buying π_B decreases more rapidly than profits are generated from buying the information π_A.

To see why this is the case, note that, since p_A decreases faster in c than does p_B, it follows that B's market share among new consumers decreases with c. This means that B's profits must decline when c increases. Firm A, however, faces two effects in the new consumer segment: its market share increases with c even as its price decreases with c. Among old consumers, a larger privacy cost means that fewer consumers will pay for it, which increases the revenue from tailored prices even though $p_A(\theta)$ decreases. At the same time, fewer of the old consumers pay the basic price. This makes A's (gross) profit less sensitive to c than is B's profit.

5 Extensions and robustness checks

5.1 Interim sales of information

So far we have assumed that the data supplier DS sells the information *ex ante*, at a price T that was set at stage 1. This selling strategy presumes some strong commitment ability on the side of the DS, that instead may want to keep some pricing flexibility when selling its dataset to the retail firms. For this reason, we now consider that data sales could be made at a later stage of the game. More precisely, we change the timing so that buying information occurs *after* consumers have decided whether or not to pay the privacy cost (but still before the pricing game). We remark that if the sale of information occurs after basic prices are set, then all consumers can benefit from the basic price at no cost and so consumer information has no value.[28] In this extension, we concentrate on the range of privacy costs $c \leq \bar{c}$ to avoid any change of regime.

The timing now still reflects the basic timing as in Section 3.1. But, in contrast to Section 4.2, consumers make their privacy decision at the very beginning of the game, at a stage that we denote as 0.5, so that the information dataset is not sold *ex ante* anymore:

[27] Setting the cost of privacy above $\bar{c}$ would result in a change of regime whereby no consumer opts for privacy (since it is too costly), although information's competitive value would then be lower.

[28] If the basic prices are set before the information is bought by the firm(s) *but not publicly announced* to the anonymous market until after the information is bought, then the equilibrium is the same as the one derived in this section.

Stage 0.5 Consumers make their privacy choice—that is, they decide whether or not to pay the cost c.

In this section, consumers must decide whether or not to pay the privacy cost *before* knowing which firm will be able to make targeted offers. Therefore, not only firms A and B, but also—and more importantly—the data supplier, may wish to deviate from the actions anticipated by consumers. In light of this possibility, the next lemma shows that information cannot be sold to both firms simultaneously and also that the structure of exclusive sales should differ from the structure characterized in Section 4.2.3.

Lemma 2. *If consumers must decide about paying the privacy cost before knowing which firm will be informed, then there is no equilibrium in which information is sold to only one firm with certainty or in which information is sold to both firms with certainty.*

Proof. We first show that there is no equilibrium when information is sold to only one firm with certainty. By contradiction, consider a candidate equilibrium in which consumers anticipate that only firm A buys the data. In this case, the outcome of the game is the same as that in Proposition 5. However, the data supplier could deviate from selling consumer information exclusively to firm A and sell it to firm B. If it does, then the DS can charge a price equivalent to the profits that B generates from old consumers. Note that, after the DS deviates at stage 1, both firms set the same basic prices at stage 3 as in Proposition 5; from stage 0.5 onward θ_2 is fixed (and is also defined in Proposition 5). Therefore,

$$\pi_B^O = \int_{\frac{3(t-c)}{8t}}^{1} \left(\frac{t-c}{2} + t(2\theta - 1) + t \right) d\theta = \frac{10ct - 21c^2 + 75t^2}{64t}.$$

We can easily show that π_B^O is strictly greater than the price of information given in Proposition 6. So if consumers anticipate that the information will be bought by A (resp., by B), then the data supplier has an incentive to sell to B (resp., to A). As a result, there cannot be an equilibrium in which the data seller necessarily allocates data to a single firm.

Next we establish that there is no equilibrium when information is sold to both firms with probability 1. Consider a candidate equilibrium under which that certainty prevails. In this case, as shown in Section 3, consumers do not pay the privacy cost. But then the DS would be incentivized to sell the information to only one firm (and we are back to the scenario covered in Section 3). So consumers should rationally anticipate this deviation. Therefore, this cannot occur in equilibrium. □

An equilibrium could arise if the DS randomized its selected buyer of information and charged a price such that the chosen firm is indifferent between buying or not buying.

Proposition 7. *There exists an equilibrium where information is sold exclusively to one firm with probability*

$\frac{1}{2}$ *and where the price of that information is*

$$T = \begin{cases} 4c - 4c^2/t & \textit{if } c \leq t/2, \\ t & \textit{if } c > t/2. \end{cases}$$

Proof. Suppose that consumers anticipate that each firm will have information with probability $\frac{1}{2}$. Consider a symmetric equilibrium in which a fraction θ_A pays for privacy (to hide from A) and a fraction $1-\theta_B$ pays for privacy (to hide from B), with $\theta_A = 1-\theta_B$.

If a consumer close to A pays for privacy, then she anticipates a utility of $-p_A^a - c - t\theta_A + v$. Yet, if the same consumer does not pay for privacy, then her expected utility is

$$\frac{1}{2}(-p_A^a - t(2\theta_A - 1) - t(1-\theta_A) + v) + \frac{1}{2}(-p_B^a - t\theta_A - t(1-2\theta_A) + v).$$

The first term represents the utility derived by this consumer from paying a tailored price (as calculated before) to A; the second term represents the utility from paying a tailored price to B. Each term is multiplied by the probability of the respective event. The choice of an indifferent consumer θ_A is then

$$\theta_A = \frac{p_B^a - p_A^a + t - 2c}{2t}.$$

Given θ_A, firm A's profits can be written as

$$\frac{1}{2}\left(\theta_1(p_A, p_B)p_A + \theta_A p_A + \int_{\theta_A}^{\theta_B} [p_B + (1-2\theta)t] \ dx\right) + \frac{1}{2}\left(\theta_1(p_A, p_B)p_A + \theta_A p_A\right),$$

where $\theta_1(p_A, p_B)$ represents A's market share among new consumers. The first term represents profits when the firm has data, and the second term when it does not (but B does). By symmetry, firms A and B solve (respectively)

$$\arg\max_{p_A} \ [\theta_1(p_A, p_B) + \theta_A]p_A,$$

$$\arg\max_{p_B} \ [(1-\theta_1(p_A, p_B)) + \theta_A]p_B.$$

The equalities $p_B = p_B^a$ and $p_A = p_A^a$ (consumers' rational expectations) lead to equilibrium prices $p_A = p_B = 2(t-c)$ and $\theta_A = \frac{t-2c}{2t}$. Then, for such a solution to be feasible, the condition $c \leq \frac{t}{2}$ must be satisfied. In this solution, the data supplier can set a price T equal to the profits in the personalized market (i.e., $[\theta_A, \theta_B]$). This is because these profits are in addition to the profits the firm would make without data. Therefore,

$$T = \int_{\theta_A}^{\theta_B} [2(t-c) + (1-2\theta)t] \ dx = 4c - \frac{4c^2}{t}.$$

When $c > t/2$, consumers do not pay for privacy and the personalized market consists of all the old consumers. Since $\theta_A = 0$, it follows that $p_A = p_B = t$. Then the extra profits enabled by having information (equivalently, the price of that information) is equal to:

$$T = \int_0^1 p_A(\theta)\, d\theta = t.$$

By construction of the price T, neither firm A nor firm B deviates at stage 2. Finally, it can be shown that the data broker does not deviate and sells to both firms because more revenue is always generated by selling to one firm only. Formally, such deviation would yield the revenue:

$$T^D = 2 \int_{\theta_A}^{1/2} t(1 - 2\theta) d\theta < T.$$

We can rule out price deviations by A and B at stage 3 by observing that, because both firms are playing their best response in the anonymous market, changing basic prices could only generate losses. Furthermore, the data supplier has no incentive to change the allocation rule because DS is indifferent between selling the data to A or B—and this indifference is consistent with consumer expectations. $\square$

The result concerning exclusivity of purchased information is therefore robust to changes in the baseline model's timing if we consider *ex post* rather than *ex ante* exclusivity. Furthermore, we can demonstrate that the data supplier would like to commit *ex ante* for small values of c but would rather sell later for a large c (see Figure 2 where we plot the price of information with and without commitment).

To understand this result, recall that without commitment and for low c, nearly all old consumers will pay the privacy cost and so the data supplier has almost nothing to sell. With commitment, however, some consumers choose not to pay even a low privacy cost; firms compete strongly for these customers because in this case the basic price is high (and so their outside options are limited). Now suppose the privacy cost is high. Without commitment, no consumer decides to pay this cost because his preferred firm will have some information about him only with probability $\frac{1}{2}$; hence the competition for information is intense and so the data supplier's price is high. With commitment, there may be more consumers paying for privacy as well as less competition for information.

5.2 Markets of different sizes

Until now we have assumed that there were as many old as new consumers. We have shown that, in the asymmetric scenarios when only one firm has information about customers, the other firm would prefer not to compete in the market for old consumers, which is left entirely to its informed rival. One may therefore wonder the extent to which this result depends on having two markets of identical sizes. In this section we

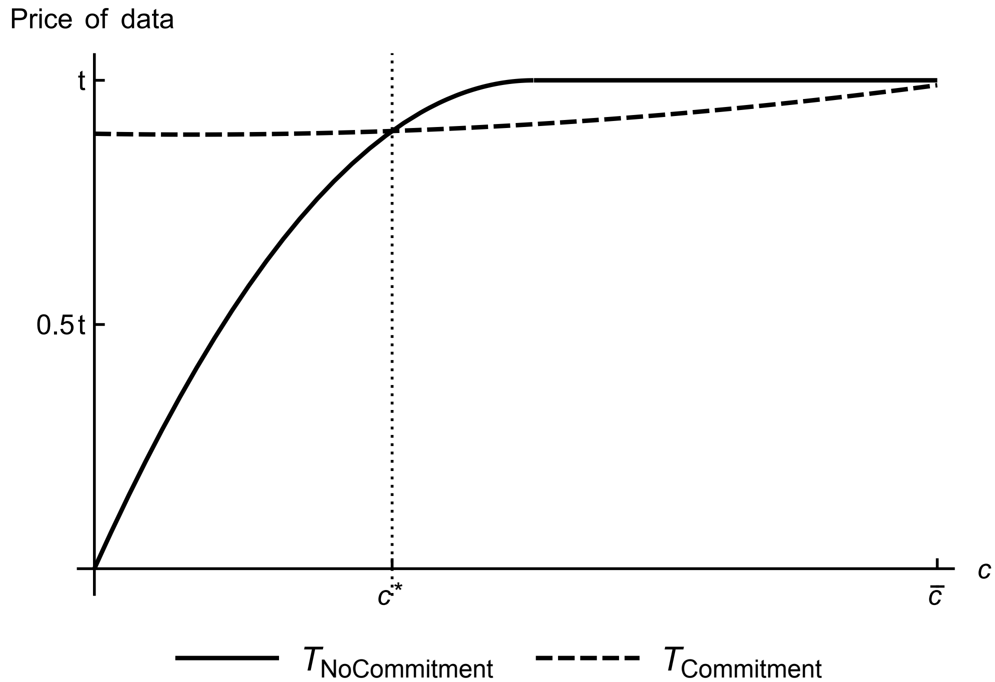

Figure 2: Information prices with and without commitment

generalize the analysis by allowing the mass of new consumers to be of any size $m > 0$ (while still assuming a unit mass of old consumers). Thus we explore the effect of changing the relative market size on the incentives to acquire information in the presence of privacy. In this endeavor we assume that $c \leq \frac{t(3m+3)}{3m+4} < t$ so that some consumers pay for privacy in every case (as shown below, this condition ensures that $\theta_2 \geq 0$).

The main effect on the equilibrium of varying m is to change firm B's incentives to compete for old consumers.[29] Recall that doing so requires B to set a low price (smaller than t) because it is disadvantaged when competing with A, which has purchased information about consumers. Furthermore, B can set only one price for both markets. So if m is large, then it is more profitable for B to set a higher price and compete only for new consumers. Conversely, if m is small then forsaking old consumers will not be profitable.

Lemma 3. *Assume that only firm A buys the information and that consumers can pay for privacy. Then there are two types of equilibria, as follows.*

(i) *If B does not compete for old consumers, then*

$$p_A = t + \frac{2(t-c)}{1+3m}, \quad p_B = t + \frac{t-c}{1+3m}; \quad \theta_2 = \frac{3m(t-c)}{2t(1+3m)}.$$

[29]Firm B's choice is always independent of firm A's actions.

(ii) *If B does compete for old consumers, then*

$$p_A = t + \frac{t-2c}{2+3m}, \quad p_B = t - \frac{cm+(1+m)t}{(1+m)(2+3m)}; \quad \theta_2 = \frac{3m(m+1)(t-c)-cm}{2(m+1)(3m+2)t}.$$

Proof. In equilibrium we must have $p_A = p_A^a$ and $p_B = p_B^a$, in which case $\theta_2 < \theta_1$. However, reaction functions should be calculated by taking consumer actions as given. In case (i) B's reaction function is given by

$$\arg\max_{p_B} m[1-\theta_1(p_A,p_B)]p_B,$$

while A's reaction function is

$$\arg\max_{p_A} [m\theta_1(p_A,p_B)+\theta_2(p_A^a,p_B^a)]p_A.$$

These two functions allow us to find $p_A(\theta_2)$ and $p_B(\theta_2)$. Finally, because $p_A = p_A^a$ and $p_B = p_B^a$ in equilibrium, we can plug these reaction functions into equation (4); doing so, leads to the expression for θ_2.

In case (ii) we solve in the same manner but change B's reaction function to account for selling also to old consumers:

$$\arg\max_{p_B} [(1-\theta_2^0(p_B))+m(1-\theta_1(p_A,p_B))]p_B.$$

Recall that θ_2^0 (B's market share among old consumers) depends on p_B but not on p_B^a; the reason is that θ_2^0 is set by $p_A(\theta_2^0)=0$. And since $p_A(\theta) = p_B + (1-2\theta)t$ is always chosen by A at stage 4, it follows that θ_2^0 is a function of p_B and not of p_B^a. Therefore, B optimizes in both markets at the same time. The assumption $c \leq \frac{t(3m+3)}{3m+4}$ is necessary to ensure a positive proportion of consumers who pay the privacy cost. □

Prices are decreasing in m when firm B does not compete for old consumers. As m continues to increase, prices converge to the Hotelling prices. Indeed, as the market for new consumers becomes relatively more important, firms focus increasingly on that market and the situation is akin to a standard Hotelling competition. When B competes for old consumers, the effect of the relative size m on prices is more ambiguous.

Lemma 4. *Suppose once again that only firm A buys the information and that consumers can pay for privacy. Under these conditions there exists an $m^*(c) > 0$, such that B does not compete for old consumers if and only if $m \geq m^*(c)$.*

Proof. Compute the difference in B's profits when B does not sell to old consumers (case (i) in Lemma 3) and when B does (case (ii) in Lemma 3):

$$\Delta\pi = \frac{m(c-(3m+2)t)^2}{2(3m+1)^2 t} - \frac{\left(-cm+3m^2t+4mt+t\right)^2}{2(m+1)(3m+2)^2 t}.$$

The first term on the RHS represents B's profits when it does not sell to old consumers; the second term is B's profits when it does. It can be verified that that $\Delta\pi \to t/6$ as $m \to \infty$. Moreover, $\Delta\pi \to -t/8$ as $m \to 0$. It can also be proved that $\Delta\pi$ is always increasing. Therefore, a threshold $m^* > 0$ exists. □

We can now predict the data supplier's behavior for any $m > 0$. Lemma 4 gives us two cases to analyze based on whether m is greater or less than $m^*(c)$. By Lemma 1 (which can be generalized to any $m > 0$), if both firms buy information then consumers do not pay for privacy. It follows from Proposition 1 that firms make equal profits from selling to either set of consumers: $mt/2$ for new consumers and $t/4$ for old consumers.

Proposition 8. *For any $m > 0$, it is always more profitable for the data supplier to sell the information to only one firm.*

Proof. If $m \geq m^*(c)$, then B does not compete for old consumers. So just as in the proof of Proposition 6, it suffices to show that $\pi_A - \pi_B \geq 2\left(\frac{t(2m+1)}{4} - \pi_B\right)$ for

$$\pi_A = m\int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^1 (p_B + (1-2\theta)t)\, d\theta$$

and

$$\pi_B = m\int_{\frac{p_B - p_A + t}{2t}}^1 p_B \, d\theta,$$

where p_A, p_B, and θ_2 are as given in Lemma 3 (i).

If $m < m^*(c)$, then B does supply old consumers. Again as in the proof of Proposition 6, it suffices to show that $\pi_A - \pi_B \geq 2\left(\frac{t(2m+1)}{4} - \pi_B\right)$ for

$$\pi_A = m\int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^{\frac{p_B + t}{2t}} (p_B + (1-2\theta)t)\, d\theta$$

and

$$\pi_B = m\int_{\frac{p_B - p_A + t}{2t}}^1 p_B \, d\theta + \int_{\frac{p_B + t}{2t}}^1 p_B \, d\theta,$$

where p_A, p_B, and θ_2 are as given in Lemma 3 (ii). □

We can therefore conclude that, while the decision of firm B to compete for old consumers does depend on the relevant size of the two markets, the DS's optimal selling strategy does not. Proposition 8 implies that the data seller's preference for exclusivity in the allocation of information is robust to any relative market size—provided that information is sold *ex ante*. A similar result also arises in a different but related extension, whereby the tracking technology that reveals information about consumers in not perfect, as we have assumed so far, but imperfect. As the tracking technology becomes more imperfect, the profits become more and more dependent on the new market for which information is not available, an effect similar to an increase in the mass of new consumers. Exclusivity also arises in this case.[30]

5.3 Recognition technology

In this section we discuss alternative ways in which the data seller might contract with the retail firms, and alternative ways to think about personalized information. In the baseline model, we assumed that firms could not distinguish between new consumers (for whom no information is available) and old consumers who paid the privacy cost (because they canceled any available trace of their activities). While this is realistic, one could argue that some minimal traces of the old consumers would always be left, even if they paid for the privacy cost. In terms of Figure 1, this would allow to distinguish the aggregate market of new consumers from the aggregate market of old consumers who protected themselves. We now introduce a *recognition technology* that allows firms to identify whether a consumer is old or new. The data broker can sell this technology to either firm *in addition* to the data set containing consumers' tastes analyzed in the baseline model. For instance, if firm i has the technology and the data set, it can charge personalized prices $p_i(\theta)$ to those consumers about whom it has personalized data and two different basic prices, p_i^N for new consumers and p_i^O for old consumers.[31]

In this setting multiple cases must be analyzed. For instance, a case where one firm has only the consumer data and the competitor has only the recognition technology. We consider all the possible combinations generating different pricing games and profits in Appendix B. The less trivial part comes when we analyze the selling strategy of the DS. In the baseline model, the DS looks for the most profitable way to allocate consumer data. If the DS offers the data to one firm then there is only one object for sale (the data set). On the other hand, if the DS offers the data to both firms, there are two objects for sale, even though they are identical. In this extension we follow a similar procedure as we assume that the DS can offer up to two objects for sale.

However, the difference with the baseline model is that the objects offered by the DS need not be identical. The DS can offer up to two goods from the following set: consumer data, the recognition technology, or a bundle of both. In this way, if any of the retail firm refuses to buy the good allocated to it, the competitor gets both objects put on the market. For example, in the baseline model we studied a case where the DS does not sell the recognition technology. So if a firm had chosen not to buy the data it would have made the profits of a firm without *both* the technology and the data set, which competes against a firm that is only in possession of the data set. Moreover, if the DS wants to sell the bundle exclusively to firm A and the recognition technology to firm B and firm A does not buy the bundle, the bundle will be offered to

[30] See Appendix C for details where we make this informal argument precise.

[31] See the discussion in Shy and Stenbacka (2016) on pricing instruments and consumer segmentation.

firm B in addition to the recognition technology (even if the second "copy" of the technology is valueless). This procedure is needed to obtain the price of information, in analogy with the previous sections where we discussed how this is related to the difference between winning and losing the auction for the objects put on sale.

We are able to show that any selling strategy that involves the recognition technology is not preferred by the DS, provided that the privacy cost c is not too large. Proposition 9 establishes this section's main result.

Proposition 9. *The DS always sells the data set exclusively. Additionally, if $c \leq \bar{c}$, then the DS does not sell the recognition technology to either firm.*

Proof. See Appendix B. □

In the Proposition we show that consumer data are still sold exclusively. The reason is that the availability of the recognition technology does not alter the broker's incentives to limit competition between the firms. Furthermore, we show that, as long as the privacy cost is low, the DS does not sell the recognition technology. The reasoning for this is as follows. On the one hand, if a firm has consumer data in addition to the technology, such firm can deter privacy by setting a high basic price for old consumers. This allows the firm to set a lower price to new consumers and in turn the competitor prices more aggressively as well, which leads to lower tailored prices. On the other hand, if the DS sells the technology to the competitor, the latter can set a low price to old consumers independently from the price set for new consumers. Then the firm with the data set is forced to match a lower price for old consumers and hence makes lower profits. Overall, the recognition technology increases competition between the firms and so it reduces the broker's ability to benefit from the information asymmetry created by selling consumer data exclusively. Our main results are thus quite robust to more elaborate consumer segmentation strategies, in that these will not arise along the equilibrium path: the DS does not sell the recognition technology at all despite this being available at a zero cost to the DS. Therefore, the equilibrium will be identical to that of the baseline model.

6 Main Highlights and Final Remarks

Big Data enable firms to approach consumers via three types of targeted offers: targeted advertisements, customized products, and tailored prices. In this paper, we have focused on prices.[32] The novelty of our approach is in allowing consumers to respond to the possibility of being targeted. We have also explored the previously unstudied aspect of the data owner's selling strategy. Our model can be used to answer two

[32] We could easily have modeled product customization by making the firm offer type-θ consumers a good that generates gross utility $v + \theta t$ at a cost to the firm of θx, where $0 < x < t$. In this case, the data supplier would appropriate any efficiency gains and still prefer exclusivity; neither the uninformed firm's profits nor consumer surplus would change. In addition, if one firm is more efficient than its rival at tailoring goods then the DS always sells to the more efficient firm.

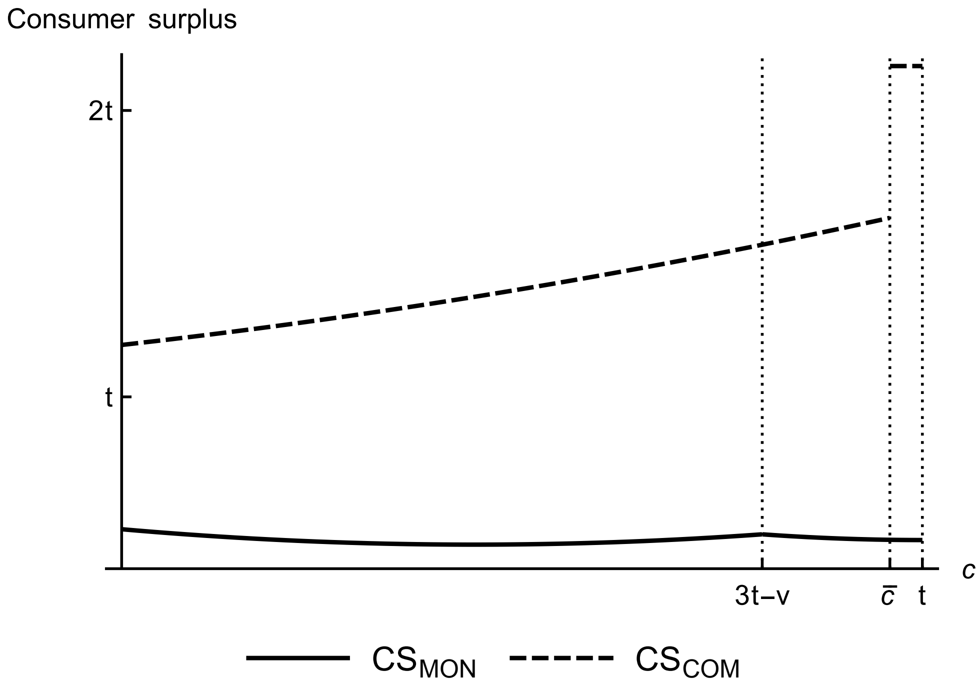

Figure 3: Consumer Surplus for $t = 1$ and $v = 2.2$

important policy questions that we anticipated in the introduction, one related to the value of privacy for consumers and one to the value of information for firms.

With respect to privacy, a current debate concerns the effect of restricting how commercial firms use information. If increasing restrictions are reasonably proxied by reductions in the privacy cost, then our study indicates that one must distinguish between the case of monopoly and the case of more competitive market structures. In the monopoly case, making it easier for consumers to protect their privacy has an ambiguous effect on consumers: some consumers gain while others lose. Using a utilitarian social welfare function that weighs consumer surplus and firm profits equally also leads to ambiguous results—even if there is a stronger presumption that society is worse-off when privacy costs are low. The results are less ambiguous under duopoly: total consumer surplus increases with the privacy cost and total welfare is maximized when consumers *cannot* protect their data. However, if consumers do protect their data (i.e., cases with $c < \bar{c}$), welfare is U-shaped with respect to c. So from a social welfare perspective, it would make sense to promote either a little or a lot of protection because intermediate levels are suboptimal.[33] Figures 3 and 4 illustrate these results; the solid lines plot the monopoly case and the dashed lines plot the duopoly case (the subscript MON refers to "monopoly" and COM signifies "competition").

With regard to the second key question, one may be interested in computing the value of information

[33] Recall that efficiency in our model is reached when market shares along both Hotelling lines are shared equally by the firms (under competition) and when consumers do not spend their resources on privacy. The other variables in the model (such as consumer prices and the cost of data) are simply transfers that affect individual payoffs but not overall welfare.

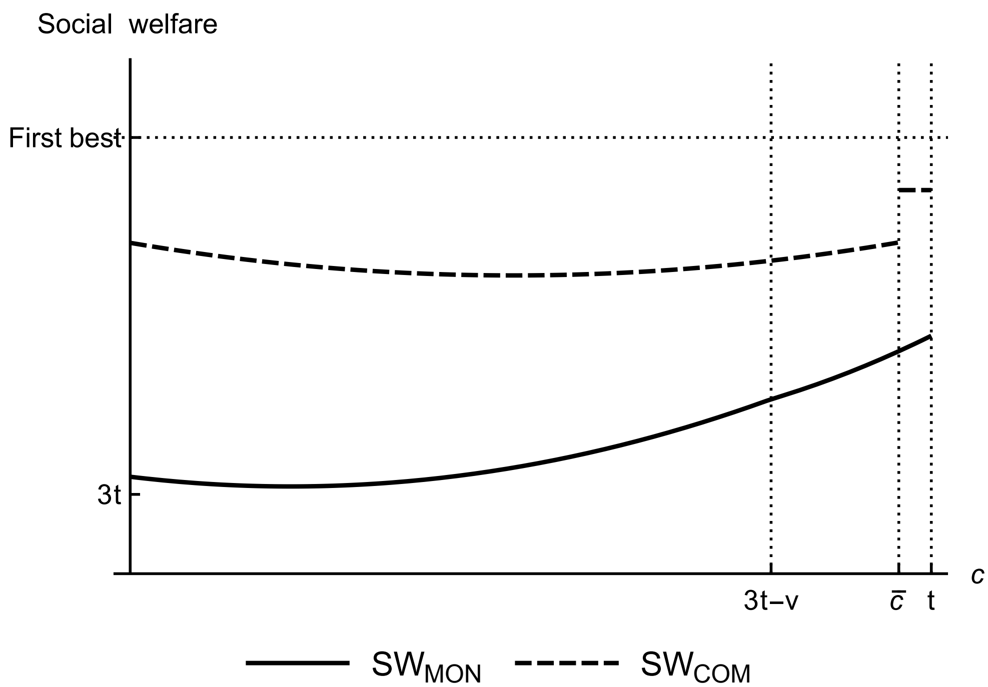

Figure 4: Social Welfare for $t = 1$ and $v = 2.2$

for the retail firm purchasing the consumer data or, equivalently, for the data supplier. This value depends on the strategy chosen by the DS, so the question could alternatively be phrased in terms of a study on the data supplier's optimal selling strategy. We establish that this optimal selling strategy consists of creating strong competition between the firms to acquire information. As mentioned previously, the competition for information can be analyzed as an auction, with externalities, in which each firm's payoff is affected by the final informational structure. The best way to generate high information value is to increase the difference between the firm's profits when only it has the information and when it has no information. Toward that end, the information should be sold in an exclusive way. Figure 5 depicts the price for data in the monopoly case (when the exclusivity strategy is not available) and also in the duopoly case when the data supplier sells to only one of the competing firms. Even when profits are greater in the monopoly than in the duopoly case, for every c we observe that the price of data under a competitive duopoly (T_{COM}) is higher than its price under a single-firm monopoly (T_{MON}). These prices are computed while considering the differential effects of information on profit; as shown in Figure 6, the impact is greater in the duopoly case. This result illustrates the notion that, for the data supplier it is optimal to maximize the competition between firms and so create a winner-takes-all situation.

A first investigation of the strategies of firms in the market for data seems to confirm our findings. Indeed, BlueKai auctions off information about consumers. The BlueKai Exchange offers access to "data on more than 300 million users offering more than 30,000 data attributes; it processes more than 750 million data

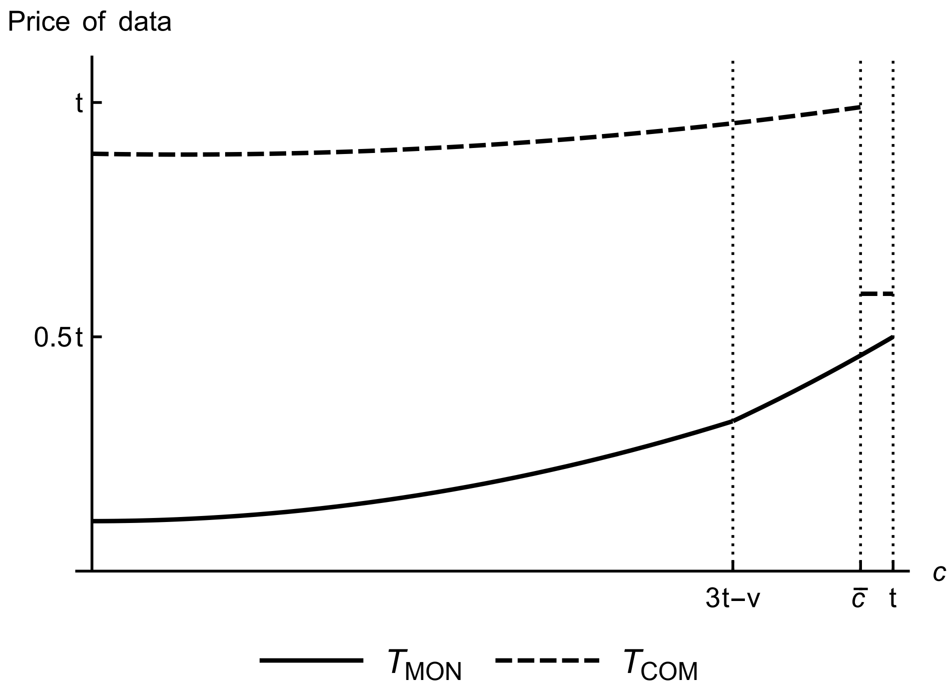

Figure 5: Price of data for $t = 1$ and $v = 2.2$

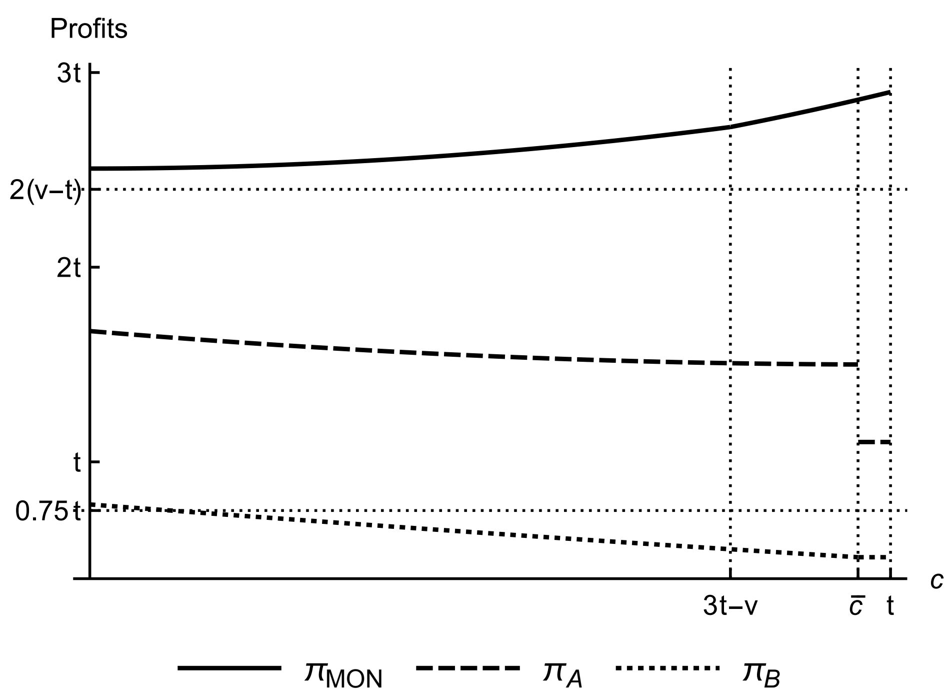

Figure 6: Profits for $t = 1$ and $v = 2.2$

events and transacts over 75 million auctions for personal information a day" (OECD, 2013, p. 15). Bidders are typically firms engaged in a marketing campaign that requires information to tailor offers. According to this data broker's website, its segmentation of "audiences" makes it unlikely for two firms to be sold data on the same consumers—even if those firms are in the same industry (BlueKai, 2015).

Exclusivity is often observed for premium rights on media platforms (e.g., sport rights or premium movies). In this paper we underscore the analogy to the Big Data context, even in the absence of network effects. It would be interesting to know more about exclusivity deals offered by data suppliers, although many details about such transactions are relatively less visible to researchers. Our model predicts that exclusivity should be offered to one firm in a competing market—for example, one local restaurant or one manufacturer of training shoes. This empirical implication of our results could be tested with suitable data.

Our finding on the robustness of exclusivity deals has also important policy implications. Recently, the FTC published a report that discusses the benefits and risks created by the use of big data analytics (FTC, 2016). This is part of a conversation that the agency is having with experts to test hypothetical harms with economic reasoning and empirical evidence. While everyone acknowledges the great potential for efficiency gains that can be brought about by the Big Data revolution, it is also fair to say that the size of the players involved, as well as the insufficient transparency of many deals, is a potential source of concerns (see also FTC, 2014). Influential antitrust scholars have started discussing how significant effects of conduct or transactions (such as the exclusivity deals we considered in this paper) may occur in the market for data rather than in the market for final goods and services. A focus on customer data, therefore, can reveal competitive effects with consequences on final consumers (Shelanski, 2013).[34]

Finally, we draw a clear policy recommendation concerning privacy and price discrimination. In our model, policy makers have two tools: a fairly standard one that makes privacy more or less costly; and the oversight of exclusive data arrangements. We find that regulators should promote a symmetric allocation of consumer data across competing firms but perhaps should *not* advocate for easier privacy when only one firm has consumer data.[35] Furthermore, an asymmetric allocation of information is doubly inefficient in our model because it induces not only inefficient consumption patterns (oversupply from firm A in both markets) but also wasteful privacy expenditures by consumers. So even though regulation was initially concerned with price discrimination, its focus should be redirected toward how information is transacted—but not toward facilitating consumer privacy.

[34]As a further example of the active policy interest in this area, the Competition Authorities of Germany (Bundeskartellamt) and of France (Autorit de la Concurrence) published in May 2016 a joint report on Big Data looking at the potential effects of the collection of large sets of personal users data on competition in digital markets. See Competition Law and Data, available at `http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big\%20Data\%20Papier.pdf`.

[35]For example, there is currently some debate about whether the bankrupt Radioshack should be allowed to sell its customer data collected on a prospective basis many years ago (Thielman, 2015). This conundrum illustrates how consumers may be myopic when assessing the disclosure of personal data. This issue, which is related to the concerns addressed by Gabaix and Laibson (2006), is left for future research.

References

Acquisti, A., C. Taylor and L. Wagman (2015). "The Economics of Privacy", *Journal of Economic Literature* (forthcoming).

Acquisti, A. and H. Varian (2005). "Conditioning Prices on Purchases History", *Marketing Science*, Vol. 14(3), 367-381.

Belleflamme, P. (2015). "Monopoly Price Discrimination and Privacy: The Hidden Cost of Hiding", *Mimeo*.

Belleflamme, P. and M. Peitz (2010). *Industrial Organization, Markets and Strategies*, Cambridge University Press.

BlueKai, Oracle (2015). "Frequently Asked Questions", `http://www.bluekai.com`

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

Chen, Y. and G. Iyer (2002). "Consumer addressability and customized pricing", *Marketing Science*, Vol. 22(2), 197-208.

Chen, Y., C. Narasimhan, and Z. J. Zhang (2001). "Individual marketing with imperfect targetability", *Marketing Science*, Vol. 20(1), 23-41.

Chen, Y. and T. Zhang (2011). Equilibrium price dispersion with heterogeneous searchers, *International Journal of Industrial Organization*, Elsevier, Vol. 29(6), pages 645-654.

Coase R. (1972). "Durability and Monopoly", *Journal of Law and Economics*, Vol. 15(1), 143-149.

Conitzer, V., C. Taylor and L. Wagman (2012). "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases", *Marketing Science*, Vol. 31(2): 277-292.

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, 2002 O.J. L 201/37.

Esteves, R. (2010). "Pricing with customer recognition", *International Journal of Industrial Organization*, Vol. 28, 669-681.

Federal Trade Commission (2014). "Data Brokers: A Call for Transparency and Accountability", *FTC Report, Washington, D.C.*

Federal Trade Commission (2016). "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues", *FTC Report, Washington, D.C.*

Fudenberg, D. and J. Tirole (2000). "Customer Poaching and Brand Switching", *RAND Journal of Economics*, Vol. 31(4), 634-657.

Fudenberg, D. and J. Villas-Boas (2012). "Price Discrimination in the Digital Economy", in M. Peitz and J. Waldgofel (Eds.),*The Oxford Handbook of the Digital Economy*.

Gabaix, X. and D. Laibson (2006). "Shrouded attributes, consumer myopia, and information suppression in competitive markets", *The Quarterly Journal of Economics*, Vol. 121 (2), 505-540.

Goldfarb, A. and C. Tucker (2011). "Privacy regulation and online advertising", *Management Science*, Vol. 57(1), 57-71.

Goldfarb, A. and C. Tucker (2012). "Shifts in privacy concerns", *American Economic Review: Papers and Proceedings*, Vol. 102(3), 349-353.

Hirshleifer, J. (1971). "The Private and Social Value of Information and the Reward to Inventive Activity", *American Economic Review*, Vol. 61(4), 561-574.

Jehiel, P. and B. Moldovanu (2000). "Auctions with Downstream Interaction among Buyers", *RAND Journal of Economics*, Vol. 31(4), 768-791.

Kang, C. (2016). "F.C.C. Proposes Privacy Rules for Internet Providers", *The New York Times*, 10/03/16, `http://www.nytimes.com/2016/03/11/technology/fcc-proposes-privacy-rules-for-internet-providers.html?_r=0`

Koh, B., S. Raghunathan and B. Nault (2015). "Is voluntary profiling welfare enhancing?", *MIS Quarterly* (forthcoming).

Lane, J., V. Stodden, S. Bender, and H. Nissenbaum (2014). *Privacy, Big Data, and the Public Good: Frameworks for Engagement.* Cambridge University Press.

Liu, Q. and K. Serfes (2009). Customer information sharing among rival firms, *European Economic Review*, Vol. 50(6), 1571-1600.

Mayer, J. (2012). "Tracking the Trackers", The Center for Internet and Society at Stanford Law School, retrieved on 16/04/2015, `http://cyberlaw.stanford.edu/blog/2011/07/tracking-trackers-early-results`

Mikians, J., L. Gyarmati, V. Erramilli and N. Laoutaris (2012). "Detecting price and search discrimination on the internet", *Hotnets'12, October 29-30, 2012, Seattle, WA, USA.*, 79-84.

Mikians, J., L. Gyarmati, V. Erramilli and N. Laoutaris (2013). "Crowd-assisted Search for Price Discrimination in E-Commerce: First results", *Mimeo.*

OECD (2013). "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", OECD Digital Economy Papers, No. 220, OECD Publishing.

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

Pasquale, F. (2014). "The Dark Market for Personal Data", *The New York Times*, 16/10/14, `http://www.nytimes.com/2014/10/17/opinion/the-dark-market-for-personal-data.html?_r=1`

Posner, R. (1981). "The Economics of Privacy", *American Economic Review*, Vol. 71(2), 405-409.

Shaw, I. and N. Vulkan (2012). "Competitive Personalized Pricing: An Experimental Investigation", *Mimeo.*

Shelanski, H. (2013). "Information, Innovation, and Competition Policy for the Internet", *University of Pennsylvania Law Review*, Vol. 161, 1663-1705.

Shy, O. and R. Stenbacka (2016). "Customer Privacy and Competition", *Journal of Economics & Management Strategy*, (forthcoming).

Stigler, G. (1980). "An Introduction to Privacy in Economics and Politics". *The Journal of Legal Studies*, Vol 9(4), 623-644.

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

Taylor, C. and L. Wagman (2014). "Customer privacy in oligopolistic markets: winners, losers, and welfare", *International Journal of Industrial Organization*, Vol. 34, 80-84.

Thielman, S. (2015). "Bankrupt RadioShack aims to sell 67m customer names and addresses", *The Guardian*, 18/05/2015, `http://www.theguardian.com/business/2015/may/18/bankrupt-radioshack-selling-customer-data-names-addresses`

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

Tirole, J. (1988). *The Theory of Industrial Organization*, MIT Press.

Villas-Boas, J. (1999). "Dynamic competition with customer recognition", *RAND Journal of Economics*, Vol. 30(4), 604-631.

Villas-Boas, J. (2004). "Price cycles in markets with customer recognition", *RAND Journal of Economics*, Vol. 35(3), 486-501.

Wall Street Journal (2012a). "On Orbitz, Mac Users Steered to Pricier Hotels", 23/08/2012, `http://online.wsj.com/news/articles/SB10001424052702304458604577488822667325882?mg=reno64-wsj\&url=http\%3A\%2F\%2Fonline.wsj.com\%2Farticle\%2FSB10001424052702304458604577488822667325882.html`

Wall Street Journal (2012b). "Websites Vary Prices, Deals Based on Users' Information", 24/12/2012, `http://online.wsj.com/news/articles/SB10001424127887323777204578189391813881534?mg=reno64-wsj\&url=http\%3A\%2F\%2Fonline.wsj.com\%2Farticle\%2FSB10001424127887323777204578189391813881534.html`

Wathieu, L. (2002). "Privacy, exposure and price discrimination", Technical Report 03-015, Harvard Business School.

Weeds, H. (2015). "TV Wars: Exclusive Content and Platform Competition in Pay TV", *Economic Journal* (forthcoming).

Appendix A

Here we present an interesting stand-alone result that justifies our sequential setting for the choice of basic and tailored prices. The alternative timing of simultaneous choices has the disadvantage of not yielding equilibria in pure strategies.

Lemma A1. *If only one firm has information, and all prices (basic and tailored) are chosen simultaneously, there is no equilibrium in pure strategies.*

Proof. Assume w.l.o.g. that A has information about some consumers. Note that for any p_B, firm A can tailor a price $p_A(\theta) \in [0, \max\{0, p^I(\theta, p_B)\}]$; here $p^I(\theta, p_B)$ is the price that makes consumer θ indifferent between buying from A or from B at p_B. It is then a dominant strategy for A to set $p_A(\theta) = \max\{0, p^I(\theta, p_B)\}$. Consider any equilibrium candidate with $p_B > 0$. For any $p^I(\theta, p_B)$, firm B could deviate and set $p_B - \varepsilon$. Then, by assumption, old consumers are no longer indifferent and now buy from B. The term ε needs to be small enough that B's losses from selling to new consumers do not offset its gains from deviating, and such an ε always exists.

Consider now a candidate equilibrium with $p_B = 0$, which implies that B's profit is zero. In that case, $p_A = (2t - c)/3$. Yet this cannot be an equilibrium because B would then rather choose $p_B > 0$ so as to maximize its profit on the anonymous market only. Firm B's best reaction is thus to set $p_B = (5t - c)/6$; in which case it makes a strictly positive profit. Hence for any candidate pure-strategy equilibrium there exists a profitable deviation for firm B. We conclude that there can be no equilibrium in pure strategies when prices are chosen simultaneously. □

Appendix B

Proof of Proposition 9. We begin the proof with a result that characterizes the consumers' privacy behavior when at least one firm has both the technology and the data set.

Result 1. *Assume at least one firm has both the technology and the data set. There exists no equilibrium in which consumers pay for privacy.*

Proof. Assume some consumer θ_2 is indifferent between paying $c > 0$ for privacy and not paying for it. Then, private consumers must be located in $[0, \theta_2]$ because the total transportation cost paid by consumers is increasing in θ. At stage 3, firm A can set either the monopoly price or a price such that consumer θ_2 is *ex post* indifferent to buying from A or B at the basic prices. In the former case, some consumers will be excluded by A and then buy from B or not at all, so those consumers would deviate at stage 2.5 (i.e., not pay for privacy).

In the latter case, consumer θ_2 faces a price $p_A = t + p_B - 2t\theta_2$. Then it can be checked that consumer θ_2 will always deviate at stage 2.5 and buy from B later on, given that

$$v - p_A - \theta_2 t - c < v - p_B - (1 - \theta_2)t.$$

Hence, in any equilibrium $\theta_2 = 0$. $\square$

Next we can study the different cases that can arise. To do so, we introduce new notation to denote each case. Let us call d the consumer data set and τ the recognition technology, while $\emptyset$ denotes the lack of both. For example, the profits when a firm has both the data and the technology whereas its competitor has only the technology are written as $\pi^{d\tau,\tau}$.[36]

1. Both firms have the data and the technology; or both firms have the data and one firm has the technology

   By Result 1, these two cases are equivalent because consumers do not pay for privacy. This implies that in equilibrium only one basic price is used by each firm. Therefore these cases are identical to the one studied in section 3.2.1 and Proposition 1 holds.

2. Firm A has the data and the technology while B has neither

   According to Result 1 consumers do not pay for privacy. This implies that in equilibrium only one basic price is used (i.e., $p_A^N = p_A$). Therefore this case is identical to the one studied in section 3.2.2 and Proposition 2 holds. Hence $\pi^{d\tau,\emptyset} = 54t/49$ and $\pi^{\emptyset,d\tau} = 25t/49$.

3. Firms A and B have the technology but neither firm has the data

   In this case consumers have no reason to pay for privacy because no firm has customer data. Then the solution is that both firms charge the Hotelling prices to every consumer, which leads to $\pi^{\tau,\tau} = t$.

4. Only firm A has the technology but neither firm has the data

   This case is identical to case 3 because $p_A^N = p_A^O$ given that B's reaction function is the same in both markets.

5. Both firms have the technology but only firm A has the data

   According to Result 1 consumers do not pay for privacy. In this case, firm A sets the same tailored price $p_A(\theta)$ as in the baseline model and sets a basic price only for new consumers.

   B can now set two prices, one for new consumers and one to old consumers. To do so, it solves

   $$\arg\max_{p_B^N, p_B^O} \, (1 - \theta_2^0(p_B^O))p_B^O + (1 - \theta_1(p_A, p_B^N))p_B^N,$$

   where θ_2^0 represents the consumer type who is offered the tailored $p_A(\theta) = 0$. Then $p_B^O = t/2$ while $\theta_2^0 = 3/4$. In turn, A uses its basic price to maximize profits from selling to new consumers. In this case it can be checked that $\pi^{d\tau,\tau} = 17t/16$ and $\pi^{\tau,d\tau} = 5t/8$.

[36]The cases where the data broker sells only the data set were covered in the baseline model.

6. Firm A has the data and firm B has the technology

   In this case B faces the same maximization problem as in the previous case. However, A's problem changes because some consumers will pay for privacy. A solves

$$\arg\max_{p_A} [\theta_1(p_A, p_B^N) + \theta_2(p_A^a, p_B^{O^a})]p_A.$$

   Next, consumer θ pays for privacy if

$$v - p_A^a - \theta t - c \geq v - p_A^a(\theta) - \theta t = v - t(1-\theta) - p_B^{O^a},$$

   which implies that for $c > 0$ consumers who pay for privacy buy from A in equilibrium, as in the baseline model. Finally, it can be verified that if $c > t/2$, no consumer pays for privacy. Then profits in this case are

   - If $c \leq t/2$
   $$\pi^{d,\tau} = \frac{17c^2 - 12ct + 108t^2}{100t} \quad \text{and} \quad \pi^{\tau,d} = \frac{2c^2 - 22ct + 73t^2}{100t}.$$
   - If $c > t/2$, $\pi^{d,\tau} = \pi^{d\tau,\tau}$ and $\pi^{\tau,d} = \pi^{\tau,d\tau}$.

It remains to verify in which of these cases the DS generates more revenue. First remark that according to Proposition 3, the data broker prefers case 2 to case(s) 1, so the latter can be eliminated. In case 2 the DS can charge for the data and the technology $T^{d\tau,\emptyset} = \pi^{d\tau,\emptyset} - \pi^{\emptyset,d\tau} = 29t/49 \approx 0.59t$.

Next, note that the DS makes zero revenue in cases 3 and 4, because firms are always indifferent between buying or not buying the technology (regardless of whether the competitor has the technology or not).

In case 5 the DS can set 2 prices, T_A and T_B, where

$$T_A = \pi^{d\tau,\tau} - \pi^{\emptyset,d\tau} \approx 0.55t \text{ and } T_B = \pi^{\tau,d\tau} - \pi^{\emptyset,d\tau} \approx 0.12t,$$

which leads to total revenue of $T = T_A + T_B \approx 0.67t$. This implies that the DS prefers case 5 to case 2. We can compare this result with the revenue the DS makes in the baseline model (Proposition 6) and we verify that the broker prefers not to sell the technology for $c \leq \bar{c}$.

In case 6 the DS can set[37]

$$T_A = \pi^{d,\tau} - \pi^{\emptyset,d\tau} = \frac{833c^2 - 588ct + 2792t^2}{4900t} \text{ and } T_B = \pi^{\tau,d} - \pi^{\emptyset,d\tau} = \frac{98c^2 - 1078ct + 1077t^2}{4900t},$$

[37] Only for $c \leq t/2$ as the case where $c > t/2$ is identical to case 5.

then

$$T = T_A + T_B = \frac{931c^2 - 1666ct + 3869t^2}{4900t}.$$

We are able to show that the revenue in Proposition 6 is larger than that of case 6.

Finally, for $c > \bar{c}$, the revenue in cases 5 or 6 (which are the same) is larger than what the DS makes in the baseline model. In the case, the DS strictly prefers to sell the technology, but it is indifferent between selling it to both firms (case 5) or only to the firm that has no data (case 6).

Appendix C

We consider the case of an imperfect tracking technology. More precisely, following Belleflamme (2015), we assume that absent any protection, a consumer's characteristics will be known by the data broker with probability $\lambda \in [0, 1]$. We will show that our main results, in particular the exclusivity result, are preserved in this modified setting. We focus here on the analysis of the duopoly case.

We consider first the case where the two firms have acquired the information and show that the result of Lemma 1 stating that the privacy cost is never paid by consumers still holds. We know that for consumers whose information has been acquired by both firms, the prices proposed by firms A and B will be given by Proposition 1. But now, these prices only concern a share λ of the consumers that have not paid the privacy cost. The other share $1 - \lambda$ can be pooled with the new consumers in the eyes of the firms. Keeping the notation of Lemma 1, the set of old consumers having paid the privacy cost have types in $[0, \theta_A]$ or $[\theta_B, 1]$.

When choosing its basic price, A solves $\max_{p_A} \lambda \int_0^{\theta_A} p_A \, d\theta + (2 - \lambda) \int_0^{\frac{1}{2} + \frac{p_B - p_A}{2t}} p_A \, d\theta$. Firm B solves an analogous problem and the prices are now given by $p_A = t + \frac{2t}{3} \frac{(2\theta_A + 1 - \theta_B)}{2 - \lambda}$ and $p_B = t + \frac{2t}{3} \frac{(\theta_A + 2(1 - \theta_B))}{2 - \lambda}$.

The condition under which consumers pay for privacy slightly differ from the one used in the main model, because not paying leads to type revelation only with probability λ. For the people to the left of the Hotelling line, the condition writes $v - p_A - \theta t - c \geq \lambda[v - p_A(\theta) - \theta t] + (1 - \lambda)[v - p_A(\theta) - \theta t]$ or, using $p_A(\theta) = (1 - 2\theta)t$, equivalently $\theta \leq \theta_2 = \frac{t - p_A - \frac{c}{\lambda}}{2t}$.

Since $p_A = t + \frac{2t}{3} \frac{(2\theta_A + 1 - \theta_B)}{2 - \lambda} > t$, no consumers with positive c wants to pay the privacy cost in this case. Therefore, a firm buying the data from the broker will be fully informed on the share λ of old consumers. For the other group of consumers, of mass $(2 - \lambda)$ and uniformly distributed, the firms will compete à la Hotelling. The profit of the two firms will then be $\pi_A = \pi_B = (2 - \lambda)t - \lambda \frac{t}{4}$. Let us define $M = (2 - \lambda)/\lambda$. The profit is thus $\lambda[M \frac{t}{2} + \frac{1}{4}]$, that is proportional to the profit found in Section 5.2 where markets have different sizes. The imperfection of the tracking technology tends to increase the mass of consumers for which no information is available, an effect similar to an increase in the mass of new consumers.

We consider now the case where only firm A has acquired the information. Remark that whether firm B has chosen or not to compete for consumers firm A has some information about does not change the way firm A chooses its basic price. It is the solution of the program $\max_{p_A}[(2-\lambda)\theta_1(p_A,p_B)+\lambda\theta_2]$ or equivalently $\max_{p_A}\lambda[M\theta_1(p_A,p_B)+\theta_2]p_A$, with $M=(2-\lambda)/\lambda$. This leads to a reaction function given by $p_A=\frac{t+p_B}{2}+\frac{t\theta_2}{M}$. In this case, the personalized price chosen by A is given by $p_A(\theta)=p_B+(1-2\theta)t$, when it is positive. Therefore, the condition under which consumers pay the privacy cost is given by $\theta\leq\theta_2=\frac{t+p_B-p_A-\frac{c}{\lambda}}{2t}$.

When firm B does not compete for consumers firm A has some information about, its basic price can attract new consumers and a share $1-\lambda$ of old consumers. Firm B's objective writes as $\max_{p_B}(2-\lambda)(1-\theta_1(p_A,p_B))p_B$ and the reaction function is $p_B=\frac{p_A+t}{2}$. Still denoting $M=(2-\lambda)/\lambda$, the equilibrium basic prices are given by the expressions of Lemma 3 (i) with c/λ instead of c and M instead of m.

Suppose instead that firm B tries to compete for consumers firm A has some information about. It means that a share of the share λ of old consumers that has not pay the privacy cost will be served by firm B. These consumers have a type $\theta\in[\theta_2^0,1]$ with $\theta_2^0=\frac{p_B+t}{2t}$. So firm's B objective writes as

$$\max_{p_B}(2-\lambda)(1-\theta_1(p_A,p_B))p_B+\lambda(1-\theta_2^0)p_B$$

and the reaction function is $p_B=\frac{p_A+t(1+M)}{4}$. Here, the equilibrium basic prices are given by the expressions of Lemma 3 (ii) with c/λ instead of c and M instead of m.

This extension with an imperfect tracking technology is therefore similar to the one studied in Section 5.2. A change in the value of λ has two impacts. First it changes the value of the privacy cost, but this change is the same for all the expressions of the profit. Second, it modifies the relative size of the two markets, exactly as studied in Section 5.2. This means that, for a given value of λ (hence of c), the ranking of the profits derived in Section 5.2 is unaffected. The result of Proposition 8 on exclusivity, which states that the data broker sells consumer data to only one firm, regardless of c and m, is also valid for any value of λ.