WORKING PAPERS

N° 583

Revised May 2017

"The value of personal information in markets with endogenous privacy"

Rodrigo MONTES, Wilfried SAND-ZANTMAN
and Tommaso VALLETTI

Toulouse School of Economics

The value of personal information in markets with endogenous privacy *

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

May 19, 2017

Abstract

We investigate the effects of price discrimination on prices, profits, and consumer surplus when (a) at least one competing firm can use consumers' private information to price discriminate yet (b) consumers can prevent such use by paying a "privacy cost". Unlike a monopolist, competing duopolists do not always benefit from a higher privacy cost because each firm's profit decreases—and consumer surplus increases—with that cost. 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. 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 under imperfect competition. We investigate the effects of price discrimination on prices, profits, and consumer surplus when firms can use consumers' private information to price-discriminate and when consumers are given the means to prevent that discrimination. Our analysis yields some interesting conclusions regarding the value of customer information—in other words, 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"

*We thank Florian Baumann, Paul Belleflamme, Alessandro Bonatti, 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, Northwestern Searle Center Conference, and University of East Anglia. 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] For a systematically collected data set, see Mikians et al. (2012, 2013); for an experimental approach, see Shaw and Vulkan (2012).

from his computer's browser, he observed a lower price for a particular DVD on Amazon.com.[2] More 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 in marketing 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, and 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 buy 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 in this area who are interested in shoes. There are several questions we aim to answer: 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 Adidas also possesses detailed profiles about some customers? c) How do customers, given that price targeting requires access to customers' profiles, protect their information? d) Will the data supplier be better-off selling the database to only one retailer or to both Adidas and Nike? After answering these questions, we will be able to conduct a welfare

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

analysis pertaining to these related policy questions: e) How are the various parties affected when it is made easier or more difficult for customers to conceal their information (i.e., when privacy costs vary)? f) Should data exclusivity contracts be allowed or instead be challenged by the relevant authorities?

1.1 Preview of the model

To answer these questions, we build 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, which we call "new" and "old" consumers. For the first group, individual information cannot be extracted (we use here the standard Hotelling model of differentiated products). 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 consists of consumers who are active, leave many traces of their activities that can be collected by the data broker, and so must engage *ex post* in costly actions to erase those traces. Active consumers can protect their privacy by paying a cost to "disappear" from the firm's database. We interpret this *privacy cost* as the effort consumers spend to conceal their actions online. So, if a firm does not know a consumer's type even after paying for information, then it cannot tell whether this consumer is a less active consumer or rather an active one who erased his 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 and the informed firm pursues new consumers less aggressively. With regard to 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.

We then study the case where consumers can pay for privacy and thus avoid being in the database. Here, too, there exist two *de facto* markets: one with the new consumers and all the old consumers who paid the privacy cost, and one containing only old consumers who 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 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 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.

Finally, we extend our analysis in different directions, related to timing, selling strategies, markets size. First, we show that our main analysis is robust to a change of timing where consumers can pay for privacy after prices have been set. In this case, consumers pay the same total price (including the privacy cost) as in the benchmark case but firms are able to increase their profit by de facto capturing the privacy cost, since no one hides. Second, and still related to timing, we relax the assumption that the data supplier can commit *ex ante* to a particular selling strategy. Despite this change, we still find that the data broker chooses to sell information exclusively to one firm.

In the second main extension, we allow the data supplier to use more subtle selling strategies by considering two alternative options. We initially discuss the possibility of splitting the data set into two parts, and selling each part to a different firm. We show that this way to partition the data set may indeed increase the profit of the data supplier. Next, we consider the possibility that, on top of selling the full data set to one firm, the data supplier sells a less precise version to the rival firm. We show that, if the data seller can choose the level of precision of the second data set, it should optimally be as imprecise as possible, strengthening the result that the full data set should be sold exclusively.

The final extension investigates 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, but our result of exclusive dealing on the data market is still robust.

1.2 Literature

Privacy is a long-standing topic in economics. Several decades ago, 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. 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—in other words, 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 continue to generate tangible economic trade-offs.

Given the recent rise of online markets and Big Data, privacy research has been revived along three main directions.[3] The first of these is research addressing behavior-based price discrimination (see e.g., Villas-Boas 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. If each consumer is biased toward one particular firm, then consumers' first-period choices show 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. 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, are unable 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 online and offline information, in order to create a mapping of consumer preferences about all those "old" consumers who left digital traces but did not protect their identities.[5] An unprotected old consumer will therefore face a targeted price and will not have access to the offers that the same firm may have for other consumers. This assumption is at the very core of the idea of price-discrimination and has been used in many contexts.[6] In the behavior-based price discrimination literature, consumers make implicit privacy choices when they decide whether or not to buy from a firm at an initial stage. Our paper makes this choice and its related cost explicit. In both cases, the firm is able to make a personalized offer to consumers who already bought the product and a potentially distinct offer to consumers who did not.

[3]For a comprehensive review of the literature on the economics of privacy, see Acquisti et al. (2016). See also Lane et al. (2014) for an accessible approach to policy issues. For empirical work on privacy, see Goldfarb and Tucker (2011, 2012).

[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, that is capable of assembling a complete picture of any consumer by using its network to track the sites that the consumer visits. 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 firms from engaging in such practices without the consumer's permission (see Kang, 2016).

[6]For example, in Chen (1999) local shoppers always buy from the same retailer and do not compare prices; in Fudenberg and Tirole (2000) firms offer promotional prices to attract new consumers, which are unaccessible to loyal shoppers; in Villas-Boas (2004) the seller chooses a price for returning shoppers, who do not have access to other prices; in Acquisti and Varian (2005) consumers observe a discriminatory price if the seller identifies a cookie in the customer's browser; in Armstrong (2006) firms can use marketing data to set higher prices to consumers who prefer their brand. Additionally, see Stole (2007) for examples of price discrimination identified in the literature.

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 that tastes are intertemporally correlated. 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—depending on the details of the model used— full privacy can be detrimental to consumer welfare.[7] 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.[8]

The third strand of the literature shares with our paper 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 choose 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. Hence lowering the cost of anonymity has similar effects to increasing the monopolist's commitment power, as both increase the firm's 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 firms and product quality is increasing in the amount of this information. Their model also features firms that derive revenue from selling consumer data as well as 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. These 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).

Finally, our paper is also related to the literature on information acquisition in oligopolistic markets. At an early stage, this literature focused mainly on aggregate information, such as common demand parameters (see Ponssard, 1979; Vives, 1984; Gal-Or, 1985; Li et al., 1987). The literature later shifted towards consumers'

[7]See also Liu and Serfes (2006), Chen and Zhang (2011), and Shy and Stenbacka (2016).

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

information acquisition through repeated interactions (as in the research previously cited), or by testing individual consumers (Burke et al., 2012). In contrast, we allow for individual information to be acquired from a third party. Moreover, consumers can individually protect their personal data; this results in an endogenous information structure and hence in an endogenous object that can be traded.

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 the preferences of consumers and the two competing firms. There is a continuum of customers each of whom has unit demand. A customer receives 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 0. In addition, 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. We assume throughout 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.[9]

Suppose there are two sets of consumers (each like the set just described) but differing 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 about them, firms can offer only basic or uniform prices that do 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.

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

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

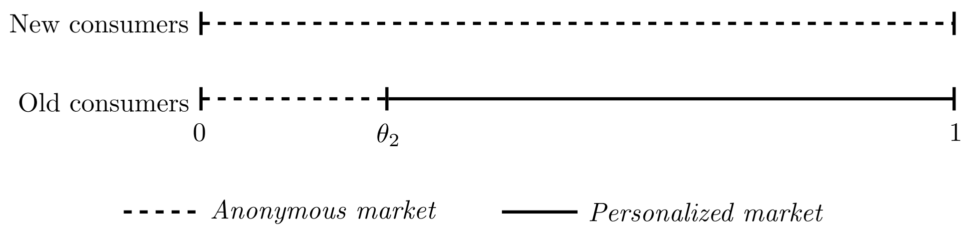

Figure 1: Example of anonymous and personalized markets

use to this retailer. At the same time there are old consumers 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 $c > 0$. That cost is associated with the actions consumers take (or the payments they make) to prevent any firm or third party from holding personal data about them.[10] 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.[11] 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 his data.[12] There is 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 virtual private network connection.

We set $c \leq t$ throughout, which generates the economically more interesting cases (since it can be easily 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 his utility is $v - p_i - |x - \theta|t - c$. Hereafter, 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 that the *anonymous market*. This group consists of all new consumers plus those old consumers who paid the privacy cost and so cannot be detected. The basic price will be based on the firm's belief about the average willingness to pay in the anonymous market.

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

[10]The same analysis could be made with variable costs, that would depend on the level of effort chosen by every consumer as a function of the gain they expect from being private.

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

[12]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).

cost. Figure 1 represents a case where old consumers with locations between 0 and θ_2 paid the privacy cost and therefore cannot be recognized. Hence, the anonymous market comprises these consumers *plus* every new consumer, whereas the personalized market consists of old consumers to the right of θ_2.

The last agent in our model is the data supplier, a broker who 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 supplier is a monopolist, in that it possesses unique information related to specific customer profiles. We do not model how it acquires its supply of data because 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 not surprising 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 broker's database covers $1 trillion in consumer transactions; and yet another adds 3 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]

3 No privacy

As a benchmark, we first study the case where consumers cannot pay for privacy. This is identical 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 to the personalized market; hence we use the respective terms interchangeably.[14] Throughout, we solve for perfect Bayesian equilibria in pure strategies.

[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.

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

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.

We 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 0, 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

$$\max_{p \geq 0} \int_0^{\theta_1} p \, d\theta + \int_0^1 p(\theta) \, d\theta \quad \text{s.t. } 0 \leq \theta_1 \leq 1.$$

Given our 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

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

[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 keep them distinct for consistency with the competition case to be analyzed in Section 3.2.

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 (CS) for old consumers and new consumers is given by

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

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

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 the price T is *ex ante*—that is, prior to the retailers' purchase decisions.

As a benchmark, consider the case in which neither firm has information. This is equivalent to solving 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.[18] 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

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

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

[17] In this section, sequentiality in the choice of basic and tailored prices is necessary for there to be an equilibrium in pure strategies (as in related papers; see e.g., Thisse and Vives, 1988). 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] See Belleflamme and Peitz (2010).

3.2.1 Both firm A and firm B 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* $\mathrm{CS} = 2(v-t)$.

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 set at the marginal cost (here equal to 0) only for the consumers who is 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 Only firm A has information

Here, we focus on the case where information is sold to one firm only; we assume, without loss of generality, that it is firm A. In this case the consumer's outside option is not zero;[19] rather, it is 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 between

[19] 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 bought from B.

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.[20] As a consequence,

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

We define the last consumer buying from A by θ_2^0, which is such that $p_A(\theta_2^0) = 0$. 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 at which 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 when $\theta_2^0 < 1$. 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. In other words, firm B can choose either to lose 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, in this benchmark case with equal market sizes,[21] firm B finds it optimal to operate in the personalized market and therefore chooses a low basic price.

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*

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

profits are

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

[20]Under indifference, consumers buy from the informed firm (which we are assuming is firm A).

[21]In Section 5.2, we study the case where the market sizes can differ.

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

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

it follows that $p_B(p_A) = \frac{2t+p_A}{4}$. 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. $\square$

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 (DS) owns the information about old consumers and can post a price T for it and sell it at stage 2, but 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 use T_k to denote the price paid by each firm when a number k 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

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

To characterize the prices T_1 and T_2, we assume that DS has all the bargaining power; this is a natural assumption given that firms A and B are both 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, any consumer can incur a cost c so that he does 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]

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

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 here is the same 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 who is located at θ and 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 the old consumers who did not opt for privacy; all other consumers constitute 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 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 *total* price is now $p^a + c$. To the extent that consumers anticipate a higher price, fewer pay to avoid detection. However, if $p^a + c$ remains small, then consumers would have an incentive 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 the RHS represents the personalized market; consumers in 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 remain bounded:

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

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

(i) *If $c \leq 3t - v$, 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,

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

(ii) *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,

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

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. Observe 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. Thus we found that if c increases then p^* decreases (because fewer consumers choose privacy) which reduces the demand for the firm's product owing to a lower average valuation.

Corollary 1. *Assume that the firm buys the information. Then profits are always increasing in c. Moreover, 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 for the firm 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; hence 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 customer information at all.

4.2 Competition

Now we consider the duopoly case when 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 at which consumers can choose to pay for privacy. So the timing in this section likewise changes to accommodate consumers' actions.

We 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.

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

4.2.1 Both firm A and firm B 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 an efficient allocation in this case: transportation costs are minimized (since market shares along both Hotelling lines are symmetric) and no consumer engages in wasteful activities (i.e., 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 that consumers' entitlement to privacy has no effect on the outcome when both firms have information. Hence Proposition 1 still holds with Hotelling prices offered to new consumers and lower (personalized) prices offered to old consumers.

4.2.2 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 paying c so that they can buy from firm A at the basic 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. The indifferent consumer type is then 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*

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

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 that $\theta_2 < \theta_1$), 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 consequently enough 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 $\frac{(5t-c)^2}{32t} \geq \frac{25t}{49}$. $\square$

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 when 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).

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.*

[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 $(v^N + \theta_2 v^O)/(1 + \theta_2) > v^N$.

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

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

$$\begin{aligned}
\pi_A &= \int_0^{\frac{p_B - p_A + t}{2t}} p_A \, d\theta + \int_0^{\theta_2} p_A \, d\theta + \int_{\theta_2}^1 (p_B + (1-2\theta)t) \, d\theta = \frac{11c^2 - 22ct + 107t^2}{64t}, \\
\pi_B &= \int_{\frac{p_B - p_A + t}{2t}}^1 p_B \, d\theta = \frac{(5t-c)^2}{32t}; \\
\text{CS} &= \int_0^{\frac{p_B - p_A + t}{2t}} (v - p_A - \theta t) \, d\theta + \int_0^{\theta_2} (v - p_A - \theta t - c) \, d\theta \\
&\quad + \int_{\theta_2}^1 (v - (p_B + (1-2\theta)t) - \theta t) \, d\theta \\
&\quad + \int_{\frac{p_B - p_A + t}{2t}}^1 (v - p_B - (1-\theta)t) \, d\theta = \frac{10ct + 5c^2 - 103t^2}{32t} + 2v.
\end{aligned}$$

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 holds. Then profits and consumer surplus are linear in $c > \bar{c}$. It follows that Corollary 2 holds for any $c \le t$. $\square$

Note first that, even when c is small, some consumers choose not to protect themselves. Indeed, they anticipate an intense competition for them, and the way to benefit from it is to leave public the information about them.[26]

Observe now that, as c increases, a smaller proportion of consumers pay the privacy cost. This dynamics reduces the average willingness to pay for A's product in the anonymous market. In response, A lowers its price and then B does the same (to continue competing for the new consumers) owing to strategic complementarity. Since p_B decreases, it follows that 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. We deduce from Corollary 2 that consumer surplus is always increasing in c 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, consumer 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 utility from paying for privacy

[26]Even if this is not an option in our paper, these consumers would even be ready to pay for their information to be public.

minus the utility from not paying:

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

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. So when the cost of privacy is higher consumers who protect their privacy are worse-off whereas all other consumers are better-off.

4.2.3 The price of information

The price of information when consumers can pay for privacy is calculated just 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 $T_2 = \frac{3t}{4} - \pi_B$. We have $\pi_A - \pi_B \geq 2(\frac{3t}{4} - \pi_B)$ and so $T = T_1$. $\square$

Our previous finding—that the data supplier prefers to sell its consumer information to just one firm—is therefore robust to introducing privacy. It is remarkable that the DS can set a higher price for information when the privacy cost is not too great. 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). As a result, 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.[27]

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 authors define privacy as the state prevailing when targetability is imperfect and firms cannot properly recognize their loyal customers; in that model, mistargeting occurs uniformly across consumers. In contrast, privacy (and hence targetability) in our model depends on the information structure and also on competition. That

[27] We can show that this "exclusive data sales" result is robust to using vertically (rather than horizontally) differentiated retail firms.

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 solution $c = \bar{c}$ yields the highest revenue for the DS.[28] 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 do the profits π_A generated from buying information.

To explain further this result, note that, since p_A decreases more rapidly in c than does p_B, it follows that B's market share among new consumers decreases with c. This, in turn, means that B's profits must decline when c increases. However, firm A encounters 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 despite the associated decline in $p_A(\theta)$. At the same time, fewer of the old consumers pay the basic price; hence, A's (gross) profit is less sensitive to c than is B's profit.

5 Extensions and robustness checks

We now consider alternatives in the timing of the game, we envisioned different data selling strategies and we discuss the robustness of our results to changes in consumers' markets size.

5.1 Alternative timing

5.1.1 Ex-post privacy

We modify the timing of the game so that consumers can pay for privacy *after* both basic and tailored prices have been set by the firms. Therefore, consumers observe all prices before deciding whether to acquire privacy and prior to consumption. In turn, firms cannot alter their prices after consumers have made their privacy choices. This change of timing is relevant when consumers could have access to all prices, but at a cost. [29]

This extension shows that our main message carries through. To see why, first consider a case where only firm A has data and let $\tilde{p_A}(\theta)$ be the tailored price. Then consumers pay for privacy if

$$v - p_A - t\theta - c \geq v - \tilde{p_A}(\theta) - t\theta.$$

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

[29]For instance, for online grocery shopping, introductory offers are available to all consumers, old and new, but to benefit from them, old consumers would have to create a new profile – a time-consuming action – and they would lose the benefit of predefined personalized baskets.

Since firm A commits to its basic price p_A, it can always choose a tailored price p_A+c (minus a small rebate) to discourage privacy. Hence, the tailored price firm A chooses for old consumers is given by

$$\tilde{p_A}(\theta) = \min\{p_A + c, p_A(\theta)\},$$

where $p_A(\theta)$ was defined in Section 4.2.2. This implies that firm A sets a tailored price of p_A+c for consumers with types $\theta < \theta_2$, as defined in the baseline model. Therefore, A's profits are

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

In this version of the model, firm A can effectively condition three prices: p_A+c for high-value old customers, p_A for new consumers, and $p_A(\theta)$ for old consumers who may be attracted to the competitor's offer. It is immediate to show that Firm A's reaction function derived from π_A is the same as the one obtained in the main model.[30] As Firm B's pricing incentives remain unchanged in this new timing, we obtain the same reaction functions for both firms as in Section 4.2.2. Therefore, results are unchanged when the privacy decision is ex-post. The only difference is that no consumer pays for privacy in equilibrium and A's profits (as well as the price of the data) increase by $c \times \theta_2^*$.

5.1.2 Interim sales of information

So far we have assumed that the data supplier sells the information *ex ante* at a price T set at stage 1. This selling strategy presupposes some strong commitment ability on the part of the DS, which may actually prefer to retain some pricing flexibility when selling its dataset to the retail firms. For this reason, we now consider the possibility that data sales could be made at a later stage of the game. More precisely, we consider the case in which the firms buy information *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.[31] In this extension, we concentrate on the range of privacy costs $c \leq \bar{c}$ to avoid any change of regime.

The timing here still observes the basic sequence described in Section 3.1. But in contrast to Section 4.2, consumers make their privacy decision at the very start of the game—at a stage that we denote as 0.5, so that the information dataset is no longer sold *ex ante*:

[30] This can be understood by looking at the the expression for the profit π_A. We focus on the first two terms that reflect the old consumers' market. For these consumers, the net extensive margin effect of varying p_A cancels out because, when p_A increases, the same share of consumers who do not pay p_A+c start paying $p_A(\theta)$. The only effect that remains, as far as the old consumers are concerned, is the one on the intensive margin. And when added to the impact on the new consumers' market, it generates the same reaction function as in the main model.

[31] If the basic prices are set before the information is bought by the firm(s) *but are 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 about paying the privacy cost *before* they know which firm will be making the targeted offers. Therefore, not only firms A and B, but also—and more importantly—the data supplier, itself 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. See Appendix B □

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 in which information is sold exclusively to each firm with probability $\frac{1}{2}$ and where the price of that information is*

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

Proof. See Appendix B. □

The result concerning exclusivity of purchased information is therefore robust to changes in the baseline model's timing—in particular, if we consider *ex post* rather than *ex ante* exclusivity. Furthermore, we can observe by comparing the results of Proposition 6 and Proposition 7 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, and the threshold c^* below which the DS prefers to commit).

One can understand this result by recalling 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 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.

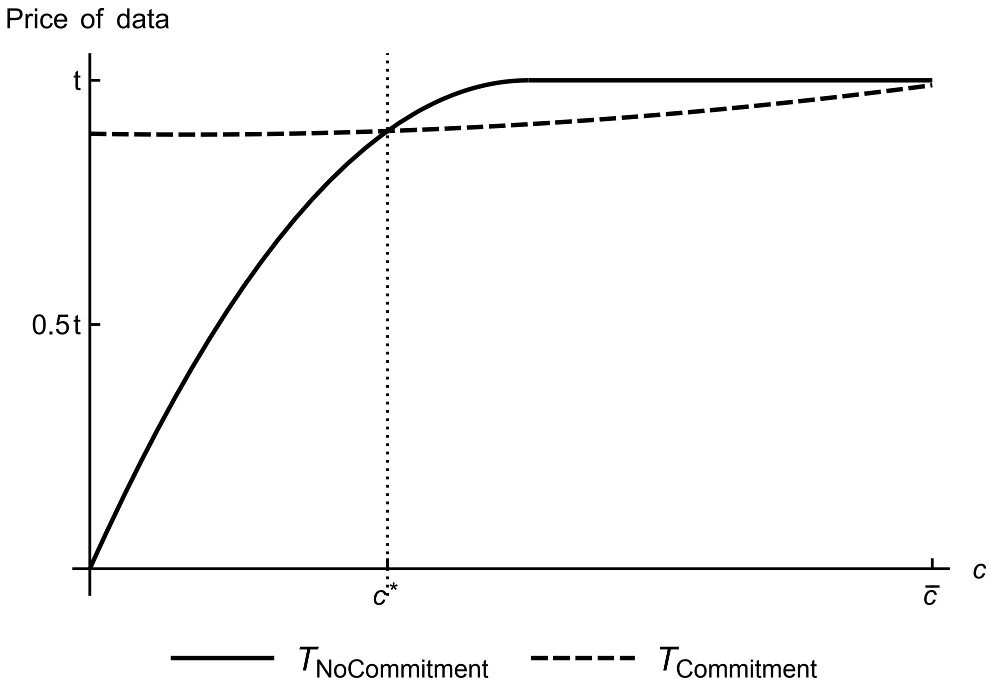

Figure 2: Information prices with and without commitment

5.2 Alternative data selling strategies

Here we discuss alternative ways in which the data supplier might contract with the retailers. First, we consider a case where the DS can create two independent data segments. By doing so, he can sell the whole data set while ensuring that firms do not propose personalized prices to the same consumers. In the second part of this section we study whether the DS can profit from reducing the quality of the data that are offered to the firms. More precisely, we assume that the DS can sell datasets with different qualities, one with perfect information and another that only gives a signal on whether the consumer's type is below or above a threshold.

5.2.1 Data segmentation

Suppose that the DS segments the data in two halves and offers the subset $[0, 1/2]$ to firm A and the remainder to firm B.[32] So when a firm recognizes a consumer it can propose a tailored price that leaves the consumer indifferent between paying such price and the competitor's basic price. Hence, tailored prices and the privacy shares are calculated as in section 4.2.2. If both firms buy the data, consumers close to firm A and firm B pay for privacy. Call the share of consumers who hide from firm i, $\theta_i \in [0, 1/2]$. Profits in this

[32] We focus here on the case where $c \leq \bar{c}$.

case are given by

$$\pi_i = p_i \left(\max\left\{ 1-\theta_j, \frac{-p_i+p_j+t}{2t} \right\} - (1-\theta_j) \right) + p_i \left(\frac{-p_i+p_j+t}{2t} + \theta_i \right) + \int_{\theta_i}^{\frac{1}{2}} (p_j + (1-2\theta)t)\, d\theta, \quad (5)$$

where the max{.} in (5) captures the fact that firm i could reduce its basic price and sell to some of the private consumers in its competitor's turf.

Given that the data is equally divided between the firms and old consumers at both ends of the distribution exhibit the same behavior, any equilibrium has to be symmetric, i.e, $\theta_i = \theta_j$. Therefore, it must be that $1-\theta_j > \frac{-p_i+p_j+t}{2t} > \theta_i$. Then, it can be shown that equilibrium prices are $p_i = p_j = t(1+2\theta_i)$, where θ_i is common to both firms and depends on the value of c. It remains to verify that firms do not deviate from this equilibrium during the pricing stage. Such deviation would require a sufficiently large reduction in price such that a firm can sell to private consumers located closer to its competitor. We establish in the next Lemma a condition under which that is not the case.

Lemma 3. *There exists a threshold $\tilde{c} = \frac{t}{2}(2-\sqrt{2}) < \bar{c}$ such that, for $c \geq \tilde{c}$, firms do not deviate.*

Proof. See Appendix B. □

This result can be interpreted as follows. When c is large there are fewer private consumers. If this is the case, a significant reduction in the basic price is required to sell to consumers closer to the rival, which is not profitable. Then, for $c \geq \tilde{c}$, it can be shown that equilibrium profits are

$$\pi_i = \frac{c^2}{4t} - c + 2t.$$

Let us assume that the DS can offer the *whole* data set to the rival when a firm refuses to buy its share of the data.[33] Therefore, a firm that refuses to buy the data makes profits π_B in Corollary 2 and the DS generates revenue

$$T = 2(\pi_i - \pi_B) = \frac{17c^2 - 22ct + 39t^2}{16t},$$

which exceeds the broker's revenue in the baseline model.

This example show that one has to qualify our main result on exclusivity as it may not apply to the whole dataset but only to part of it. At the same time, the result of this example hinges critically on the assumption that the DS can perfectly and at no cost determine which segment of the data is most profitable for each firm, which may not be realistic. One can also believe that, by splitting the dataset, the information each firm can extract from its part is reduced.

[33] Alternatively, we could assume the DS leaves one half of the data unsold if a firm refuses to buy its share. This case leads to qualitatively similar results.

5.2.2 Data quality

We consider now the possibility that the data supplier sells not only the fully informative dataset to firm A but also another dataset, less precise, to firm B. This additional dataset would give firm B a signal on whether the old consumer's type is above of below σ with $\sigma \in [1/2, 1]$.[34]

If we compare this situation with the benchmark case, B can now set three prices instead of two : p_B for the anonymous market; $\underline{p}_B$ for the old consumers with type less that σ; $\overline{p}_B$ for the old consumers with type greater than σ. As far as the old consumers are concerned, it is as if there were three submarkets: one with consumers in $[0, 1/2)$, another with consumers in $[1/2, \sigma)$ and the last with consumers in $[\sigma, 1]$. Regarding the first submarket, because consumers are closer to firm A than firm B, firm A will always be able to tailor its price sufficiently to serve all consumers. The last submarket (made of consumers with types in $[\sigma, 1]$) can be totally captured by firm B if σ is close enough to 1 but may also be contested by firm A when σ is not too large. This depends on whether σ is greater or smaller than 3/4. As for the intermediate submarket made of consumers with type between 1/2 and σ, it is divided between the two firms and the price paid by these consumers increases with σ.

This new informational structure also affects the old consumers' decision to pay the privacy cost. We can show that, for any value of the privacy cost or of the signal σ, no consumer close to firm B chooses to pay the privacy cost. Indeed, on the anonymous market, the price is at least equal to t, and because of the privacy cost, the total cost born is at least $t + c$. The price paid by the old consumers on the other market – when they do not hide – reflects competition between firm A and firm B. If firm B wants to get at least some old consumers, and be able to compete with the most aggressive offers from firm A, it must set a price less than t. So no consumers on firm B's turf will ever have any incentive to hide, a result that echoes Lemma 1. On the contrary, it may be in the interest of some consumers located to the left of the line to pay the privacy cost to hide. Indeed, since B has only an imprecise information about consumers located on firm A's turf, it cannot exert a strong competitive pressure on tailored price set by firm A. As these prices remain high, some consumers close to 0 may still find it profitable to hide from A and buy on the anonymous market, at least when the privacy is cost is not too large. In the Appendix, we build on these results to prove the following lemma.

Lemma 4. *Assume that the data supplier sells a dataset disclosing full information to firm A and a less precise dataset to firm B, who only learns if the type of each (old) individual is above or below $\sigma \in [1/2, 1]$. If the data supplier can choose the optimal level of precision, it will choose $\sigma = 1$ in order to maximize its revenues from the sale of information.*

Proof. See Appendix B. □

[34] In Appendix B we fully characterize the equilibrium in this case so here we will focus on explaining the main intuitions.

The revenues earned by the data supplier depend on the profits earned by the two firms. From firm A's point of view, the optimal level of σ is 1 since it is this value that preserves the most its comparative advantage.[35] From Firm B's point of view, the optimal precision should be between 1/2 and 1. Indeed, when $\sigma = 1/2$, the information learnt by firm B does not allow it to improve its offer on the consumers located to the left – Firm A's can always undercut any positive price with a profitable offer – and not much on the right – these consumers are heterogenous so the value of the information is low. Similarly, when $\sigma = 1$, the signal only tells firm B who is an old and who is a new consumer but this is not helpful to screen old consumers. On the contrary, when the signal is intermediate, it allows firms B to make specific offers to the two groups of consumers, depending on whether they are almost captive (located close to 1) or potentially attracted to firm A (when they are close to 1/2). But any information given to firm B affects the way competition occurs between the two firms. In particular, when σ increases, the price set for the old consumers increases so more people pay the privacy cost. This leads to a price increase on the anonymous market which benefits all firms, and allows the data supplier to increase the price at which the information is sold to both firms.

One can now compare the revenues earned by the data supplier when a second albeit imprecise dataset is sold to firm B to the revenues of the benchmark case when only firm A benefits from information about consumers. When firm B only knows who is old and new, and this is the case when it buys the least precise dataset, any offer it could make would be matched by a tailored offer made by firm A for any consumer with type between 0 and 1/2. Therefore firm B will only consider consumers with type greater than 1/2, which leads it to set a price equal to $t/2$. But there was no effective competition for these consumers in the benchmark case, they were all buying from A, so the revenues gained by firm A are greater than when firm B has also some information about consumers. This explains why the revenues the data supplier can generate are greater when the information is sold to firm A only than when some additional information is sold to the other firm.

Proposition 8. *The data seller's profits are greater when it sells the full dataset to one firm only rather than selling a additional imperfect dataset to the other firm competing on the same market.*

Proof. See Appendix B. □

The results of this subsection confirm that the data supplier has an interest in creating competitive differences between downstream firms. Conditional on selling the dataset to both firms, the DS tries to make one as noisy as possible but it is even more profitable to go one step further and grant full exclusivity.

[35] Note that $\sigma = 1$ corresponds to the idea that firm B only knows whether a consumer is old (in this case, it receives a signal) or new (no signal is received).

5.3 Markets of different sizes

Until now we have assumed that there are as many old consumers as new ones. 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 therefore 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 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 in what follows, this condition ensures that $\theta_2 \geq 0$).

With regards to the equilibrium, the main effect of varying m is to change firm B's incentives to compete for old consumers. Recall that doing so requires B to set a low price (lower than t) because it is disadvantaged when competing with A, which has purchased information about consumers. Moreover, 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 5. *Assume that only firm A buys the information and that consumers can pay for privacy. Then there are two types of equilibria, as described next.*

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

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

(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. See Appendix B. □

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 6. *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. See Appendix B. □

We can now predict the data supplier's behavior for any $m > 0$. Lemma 4 gives us two cases to analyze: one is which m is greater than and one in which it is 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—that is, $mt/2$ for new consumers and $t/4$ for old consumers.

Proposition 9. *For any $m > 0$, it is always more profitable for the data supplier to sell the information to only one firm.*

Proof. See Appendix B. □

Hence we may conclude that, whereas firm B's decision to compete for old consumers depends on the respective sizes of the two markets, the data supplier's optimal selling strategy does not. Proposition 8 implies that the DS's preference for exclusivity in the allocation of information is robust to any market size—provided that information is sold *ex ante*.[36]

A similar result also arises in a related extension, where the tracking technology that reveals information about consumers is not perfect (as we have, until now, assumed it to be). As the tracking technology becomes less perfect, profits become more dependent on the anonymous market (for which information is not available); the effect is similar to that of an increase in the mass of new consumers. Exclusivity also arises in this case.[37]

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.[38] 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 important policy and managerial questions, one related to the value of privacy for consumers and one to the value of information for firms.

With regard to the value of 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,

[36] When $m < m^*(c)$, the results of Corollary 2 are preserved for sufficiently small values of the privacy cost.

[37] See Appendix C for the details that make this informal argument more precise.

[38] 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.

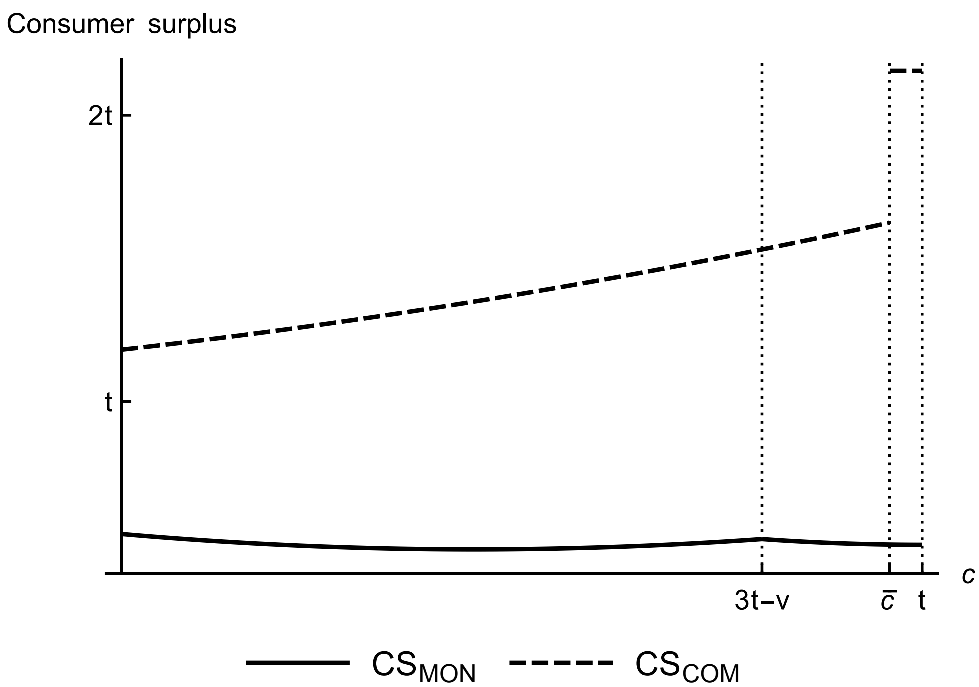

Figure 3: Consumer Surplus for $t = 1$ and $v = 2.2$

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 gain while others lose. Using a utilitarian social welfare function that weighs consumer surplus and firm profits equally also leads to ambiguous results—even though the presumption is then stronger 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., when with $c < \bar{c}$), then 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.[39]

Figure 3 and Figure 4 illustrate these results; the solid lines plot the monopoly case (subscript MON) and the dashed lines plot the case of duopoly (subscript COM).

As for the second key managerial question, one may be interested in computing the value of information—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 such strategy consists of creating strong competition between the firms to acquire information. As mentioned previously, competing for information can be

[39]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 (e.g., consumer prices, the information price) are simply transfers that affect individual payoffs but not overall welfare.

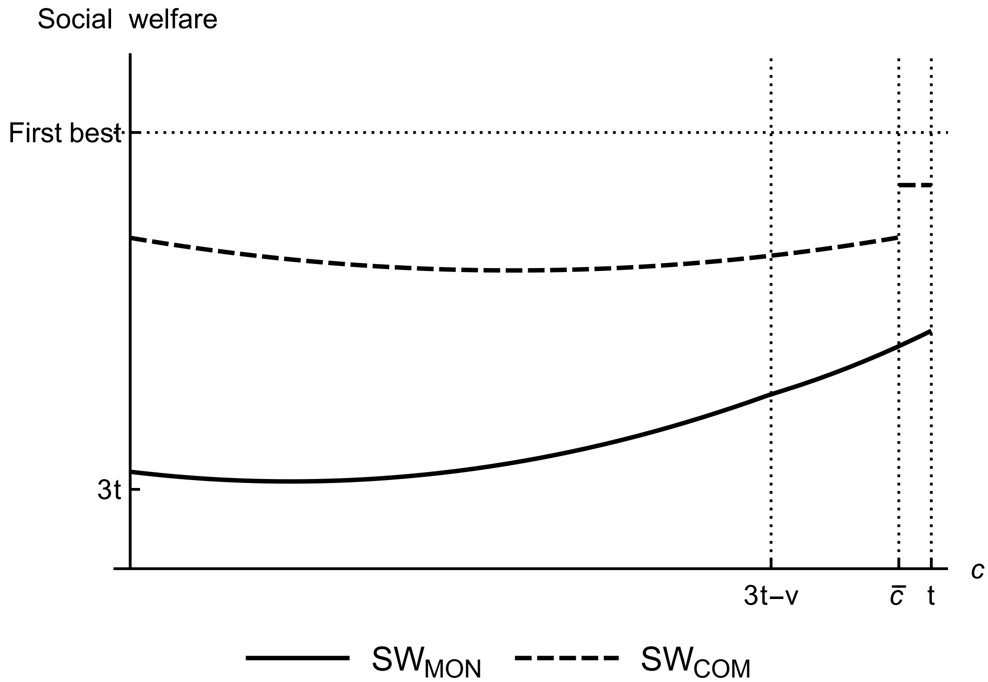

Figure 4: Social Welfare for $t = 1$ and $v = 2.2$

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. Achieving this goal requires that the information be sold to one exclusive buyer. Figure 5 plots 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.

It would be illuminating to know more about the exclusivity deals offered by data suppliers, but 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 about the robustness of exclusivity deals also bears important policy implications. The FTC (2016) recently published a report detailing the benefits and risks created by the use of Big Data analytics. This publication is part of a conversation that the agency pursues with experts to address, by way of economic

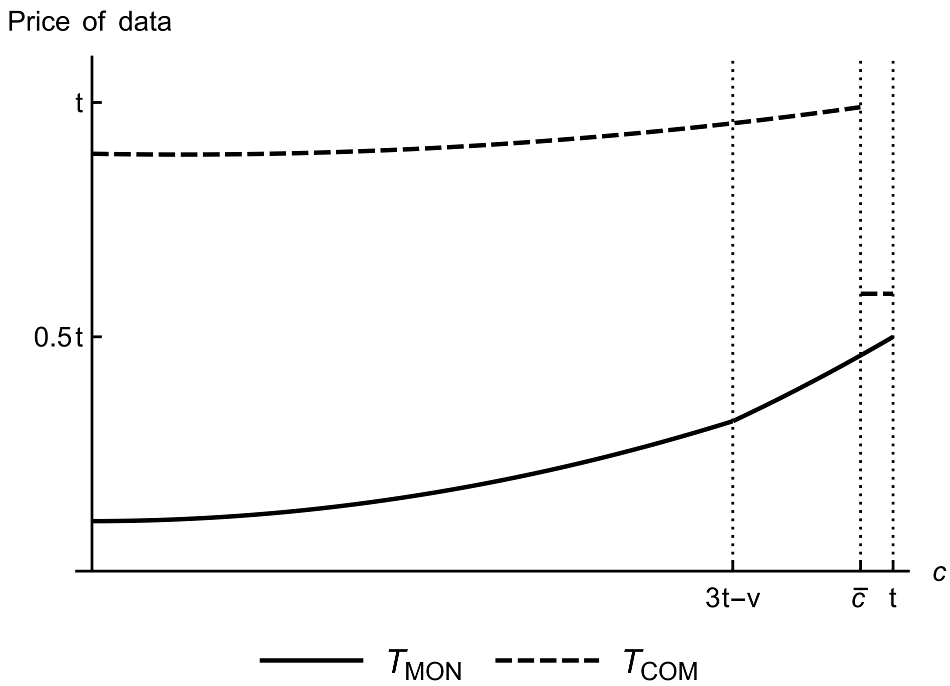

Figure 5: Price of data for $t = 1$ and $v = 2.2$

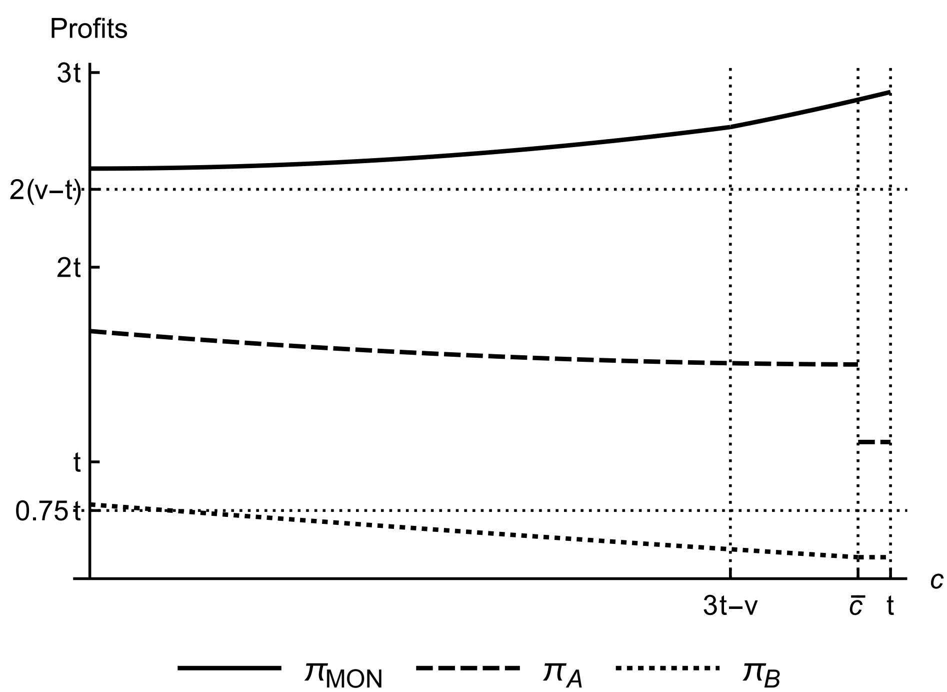

Figure 6: Profits for $t = 1$ and $v = 2.2$

reasoning and empirical evidence, the possible downsides of Big Data. Although most acknowledges the great potential for efficiency gains from the data revolution, areas of concern include the size of the players involved and the insufficient transparency of many deals (FTC, 2014). Influential antitrust scholars have begun to discuss how significant effects of particular conduct or transactions (e.g., exclusivity deals considered in this paper) may manifest in the market for data though not in the market for final goods and services. Therefore, a focus on customer data can reveal competitive effects that have consequences for final consumers (Shelanski, 2013).[40] As a related example of prominence of customer information, we mention the 2016 acquisition of LinkedIn – the largest provider of social networking services for professionals– by Microsoft for $26.2 in cash. Microsoft managed to beat Salesforce in a fierce bidding battle during the acquisition process. Both Microsoft and Salesforce are active in the market for customer relationship management (CRM). Both companies believed that LinkedIn database of profiles could be coupled with CRM systems. When Microsoft won, Salesforce complained that after the merger, Microsoft will have the ability and the incentive to refuse access to LinkedIn data to its CRM competitor. This example shows that LinkedIn chose to sell itself exclusively to a single buyer rather than sharing its dataset with multiple downstream firms.

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. However, if only one firm has consumer data then regulators should probably *not* advocate for easier privacy.[41] 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 despite regulation's initial concern with price discrimination, its focus should be re-directed toward how information is transacted—but not toward facilitating consumer privacy.

References

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

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

[40]As a further example of active policy interest in this area, the competition authorities of Germany (Bundeskartellamt) and of France (Autorité de la Concurrence) published in May 2016 a joint report on Big Data; this report looked at the potential effects of collecting large sets of users' personal data on competition in digital markets. See "Competition Law and Data", available at `http://www.autoritedelaconcurrence.fr/doc/reportcompetitionlawanddatafinal.pdf.`

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

Armstrong, M. (2006). "Recent Developments in the Economics of Price Discrimination", *In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. 2.*

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`

Burke, J., C. Taylor, and L. Wagman (2012). "Information Acquisition in Competitive Markets: An Application to the US Mortgage Market", *American Economic Journal: Microeconomics*, Vol. 4(4): 65-106.

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

Chen, Y. (1999). "Oligopoly price discrimination and resale price maintenance", *RAND Journal of Economics*, , Vol. 30(3), 441-455.

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

Federal Trade Commission [FTC] (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.

Gal-Or, E. (1985). “Information Sharing in Oligopoly”, *Econometrica*, Vol. 53(2), 329-343.

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.

Li, L., R. McKelvey and T. Page (1987). “Optimal Research for Cournot Oligopolists”, *Journal of Economic Theory*, Vol. 42(1), 140-166.

Liu, Q. and K. Serfes (2006). “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”. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, 79-84.

Mikians, J., L. Gyarmati, V. Erramilli, and N. Laoutaris (2013). “Crowd-assisted search for price discrimination in e-commerce: First results”. In *Proceedings of the 9th ACM Conference on Emerging Networking Experiments and Technologies*, 1-6.

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

Ponssard, J. (1979). "The Strategic Role of Information on The Demand Function in an Oligopolistic Market", *Management Science*, Vol. 25(3), 243-250.

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.

Stole, L. (2007). "Price Discrimination and Competition", *Handbook of Industrial Organization, Volume III, Edited by M. Armstrong and R. Porter.*

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.

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.

Vives, X. (1984). "Duopoly Information Equilibrium: Cournot and Bertrand", *Journal of Economic Theory*, Vol. 34(1), 71-94.

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 if all prices (basic and tailored) are chosen simultaneously, then 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, yet 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. Therefore, firm B's best reaction is 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: Proofs of the results stated in the extension section

Proof of Lemma 2

First we 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 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. Observe 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 consequence, there cannot be an equilibrium in which the data supplier necessarily allocates data to a single firm.

Next we establish that there is no equilibrium in which 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). Since consumers should rationally anticipate this deviation, it cannot occur in equilibrium.

Proof of Proposition 7

Suppose consumers anticipate that each firm will have information exclusively 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), where $\theta_A = 1 - \theta_B$.

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

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

here $\theta_1(p_A,p_B)$ denotes 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)

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

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:

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

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—by selling to *both* firms—because more revenue is always generated by selling to one firm only. Formally, such deviation (D) would yield the revenue

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

Price deviations by A and B at stage 3 can be ruled out because, since 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 it is indifferent between selling the data to A or B—and this indifference is consistent with consumer expectations.

Proof of Lemma 3

Let us consider, without loss of generality, a possible deviation from firm A. Its profit, when it tries to poach on firm B's turf, can be written as

$$\pi_A^D = \frac{1}{t}[p_B - p_A + 2t\theta_A]p_A + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta$$

Note that only the first term of the RHS is affected by the choice of firm A. Using the equilibrium price for firm B, $p_B = t(1+2\theta_A)$, and optimizing with respect to p_A leads to $p_A^D = \frac{t}{2}(4\theta_A + 1)$ and the optimal deviation profits are then

$$\pi_A^D = \frac{t}{4}(1+4\theta_A)^2 + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta,$$

The equilibrium profits being equal to $\pi_A = \frac{t}{2}(1+2\theta_A)^2 + \int_{\theta_A}^{\frac{1}{2}} (p_B + (1-2\theta)t)\, d\theta$, we find that the equilibrium profits are greater when $\theta_A < \sqrt{2}/4$. At last, we use the definition θ_A, given by the consumer θ indifferent between paying $p_A + c$ and $p_B + (1-2\theta)t$, to derive $\theta_A = \frac{t-c}{2t}$ and the result of the lemma. Note at last that $\tilde{c} = \frac{t}{2}(2-\sqrt{2}) \approx 0.29t < \bar{c} \approx 0.96t$.

Proof of Lemma 4

We assume that firm A has full information about old consumers – if they do not pay the privacy cost – whereas firm B has only a signal on whether the old consumer's type is above of below σ with $\sigma \in [1/2, 1]$. B will then set 3 prices: p_B for the anonymous market; $\underline{p}_B$ for the old consumers with type less that σ; $\overline{p}_B$ for the old consumers with type greater than σ. As before, we denote by θ_A the cut-off type of old consumers that hide from A (located to the left) and $1-\theta_B$ the cut-off type of old consumers that hide from B (located to the right).

We first characterize the standard price for the anonymous market, that is p_A and p_B. The price p_A is chosen to maximize

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

This leads to $p_A = \frac{p_B + t(1+2\theta_A)}{2}$. Similarly $p_B = \frac{p_A + t(1+2\theta_B)}{2}$. Combining the two equations leads to

$$p_A = t + \frac{2t}{3}[2\theta_A + \theta_B] \text{ and } p_B = t + \frac{2t}{3}[2\theta_B + \theta_A].$$

To go further, we now investigate the old consumers market. For any consumer with type $\theta \leq 1/2$, firm A can offer a price that matches any positive price set by firm B. For types $\sigma \geq \theta > 1/2$, Firm A can match B's offer if and only if $\theta \leq \frac{\underline{p}_B + t}{2t}$. Let us look at the choice of $\underline{p}_B$. One should consider two cases. If $\underline{p}_B \geq 2\sigma t - t$, then A will capture the whole market in $[1/2, \sigma]$. If $\underline{p}_B \leq 2\sigma t - t$, then B's program is

$\max_{\underline{p}_B} \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \underline{p}_B d\theta$ so $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

We now look at the choice of $\bar{p}_B$. For this, we assume (and will check the consistency *ex post*) that $\sigma < 1 - \theta_B$. If B can capture a share of the market $[\sigma, 1-\theta_B]$, its program is $\max_{\bar{p}_B} \int_{\frac{\bar{p}_B+t}{2t}}^{1-\theta_B} \bar{p}_B d\theta$ s.t. $\frac{\bar{p}_B+t}{2t} > \sigma$. This leads to $\bar{p}_B = t[\frac{1}{2} - \theta_B]$. We need $\frac{\bar{p}_B+t}{2t} \geq \sigma$, that is $\sigma \leq \frac{3-\theta_B}{4}$. And note that we also need $\frac{\bar{p}_B+t}{2t} < 1 - \theta_B$, that is $\theta_B < 1/2$.

Next we turn to the choice of privacy. Without any protection, the price paid by the agents located on the right of the line is $\bar{p}_B$. With protection, the price paid is $p_B + c = t + \frac{2t}{3}[2\theta_B + \theta_A] + c$. Since $\bar{p}_B \leq t/2$ and $p_B + c \geq t + c$, no consumer located to the right of the Hotelling line will choose to hide. Therefore, we will have $\theta_B = 0$ and $\bar{p}_B = t/2$. Remark that, with $\theta_B = 0$ and $\bar{p}_B = t/2$, we need $\sigma \leq \frac{3}{4}$.

Let us now compute θ_A, characterized by $p_A + c = p_A(\theta_A)$. We know that $p_A(\theta) = \underline{p}_B + (1-2\theta)t$. This means that θ_A is such that

$$p_A + c = \underline{p}_B + (1-2\theta_A)t \Leftrightarrow p_A = \frac{t}{2}\big[2\sigma + 1 - 4\theta_A\big] - c.$$

Moreover, since $\theta_B = 0$, we have $p_A = t + \frac{4t}{3}\theta_A$. Using these two relationships leads to $\theta_A = \frac{3}{10}\big[\sigma - \frac{1}{2} - \frac{c}{t}\big]$. This implies that Firm A's price on the anonymous market is $p_A = \frac{2t}{5}\big[\sigma + 2 - \frac{c}{t}\big]$ and that Firm B's price on the anonymous market is $p_B = \frac{t}{5}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big]$.

Let us summarize this case.

1. For $\sigma \leq 3/4$ and $c < (\sigma - 1/2)t$ (that is $\theta_A > 0$), then $\theta_A = \frac{3}{10}\big[\sigma - \frac{1}{2} - \frac{c}{t}\big]$, $\theta_B = 0$, $p_A = \frac{2t}{5}\big[\sigma + 2 - \frac{c}{t}\big]$, $p_B = \frac{t}{5}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big]$, $\bar{p}_B = t/2$ and $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

2. For $\sigma \leq 3/4$ and $c \geq (\sigma - 1/2)t$, we will have $\theta_A = 0$, $p_A = p_B = t$, $\bar{p}_B = t/2$ and $\underline{p}_B = t[\sigma - \frac{1}{2}]$.

Let us compute the profit in the first sub-case where $c < (\sigma - 1/2)t$:

$$\begin{aligned}
\pi_A &= \theta_A p_A + \int_0^{1/2+(p_B-p_A)/2t} p_A d\theta + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \Big(\underline{p}_B + (1-2\theta)t\Big)\, d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} \left(\bar{p}_B + (1-2\theta)t\right) d\theta \\
&= p_A\big[\theta_A + 1/2 + \frac{p_B - p_A}{2t}\big] + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \Big(\underline{p}_B + (1-2\theta)t\Big)\, d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} \left(\bar{p}_B + (1-2\theta)t\right) d\theta.
\end{aligned}$$

But we can show that $\frac{p_B-p_A}{2t} + \theta_A + 1/2 = \frac{2\theta_A}{3} + 1/2 = \frac{1}{5}\big[\sigma + 2 - \frac{c}{t}\big]$ and $\frac{\underline{p}_B+t}{2t} = \frac{\sigma+1/2}{2}$ and $\frac{\bar{p}_B+t}{2t} = 3/4$. Therefore

$$\pi_A = \frac{2t}{25}\big[\sigma + 2 - \frac{c}{t}\big]^2 + \frac{t}{100}\big[2\sigma + 4 + 3\frac{c}{t}\big]^2 + (\frac{3}{4} - \sigma)^2 t.$$

Similarly

$$\begin{aligned}
\pi_B &= \int_{1/2+\frac{p_B-p_A}{2t}}^{1} (p_B)\, d\theta + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \left(\underline{p}_B\right) d\theta + \int_{\frac{\bar{p}_B+t}{2t}}^{1} (\bar{p}_B)\, d\theta \\
&= \frac{t}{50}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big]^2 + \frac{t}{2}\big[\sigma - \frac{1}{2}\big]^2 + \frac{t}{8}.
\end{aligned}$$

Here, we need to define and characterize the revenues of the DS. To each firm, he can threaten to sell the whole set of information to the competing firm, which would lead to a profit equal to the value π_B derived in Corollary 2. Let us denote $\underline{\pi}_B$ this minimal value. The total revenue of the DS is therefore $R = \pi_A + \pi_B - 2\underline{\pi}_B$. We can now differentiate R with respect to σ:

$$\begin{aligned}
\frac{\partial R}{\partial \sigma} &= \frac{4t}{25}\big[\sigma + 2 - \frac{c}{t}\big] + \frac{t}{25}\big[2\sigma + 4 + 3\frac{c}{t}\big] - \frac{t}{2}\big[\frac{3}{4} - \sigma\big] \\
&\quad + \frac{t}{25}\big[\sigma + \frac{9}{2} - \frac{c}{t}\big] + t\big[\sigma - \frac{1}{2}\big]
\end{aligned}$$

Note first that $\frac{\partial^2 R}{\partial \sigma^2} > 0$. Note also that $\frac{\partial R}{\partial \sigma}\mid_{\sigma=1/2} = \frac{t}{25}\big[25 - 2\frac{c}{t}\big] - \frac{t}{8}$ which is positive (since $c < t$). So, for $c < (\sigma - 1/2)t$, the optimal σ for the DS in $3/4$.

We now look at the second subcase where $c \geq (\sigma - 1/2)t$. Then, Firm A's profit writes as

$$\begin{aligned}
\pi_A &= \frac{t}{2} + \int_0^{\frac{\underline{p}_B+t}{2t}} \left(\underline{p}_B + (1-2\theta)t\right) d\theta + \int_{\sigma}^{\frac{\bar{p}_B+t}{2t}} \big[\bar{p}_B + (1-2\theta)t\big] d\theta \\
&= \frac{t}{2} + \frac{t}{4}\left[\sigma + \frac{1}{2}\right]^2 + t\left[\frac{3}{4} - \sigma\right]^2.
\end{aligned}$$

Similarly, Firm B's profit are

$$\begin{aligned}
\pi_B &= \frac{t}{2} + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \underline{p}_B d\theta + \int_{\frac{\bar{p}_B+t}{2t}}^{1} \bar{p}_B d\theta \\
&= \frac{t}{2} + \frac{t}{2}[\sigma - \frac{1}{2}]^2 + \frac{t}{8}.
\end{aligned}$$

As before, the Data Seller's revenue is $R = \pi_A + \pi_B - 2\underline{\pi}_B$ and we differentiate R with respect to σ.

$$\frac{\partial R}{\partial \sigma} = t\big[\frac{5\sigma}{2} - \frac{5}{4}\big] + t[\sigma - \frac{1}{2}] > 0, \text{ for all } \sigma \in [1/2, 3/4].$$

Therefore, for any $c \in [0, t]$ and $\sigma \in [1/2, 3/4]$, the optimal level of precision is $\sigma = 3/4$.

We now look at the case where $\sigma \geq 3/4$. In such a case, firm A will not be able to compete with Firm B for the consumers too close to 1, i.e., for consumers with types greater than σ. Indeed, the information gained by Firm B about these captive consumers is more precise, so Firm B will set the maximal price such that

$\frac{\bar{p}_B+t}{2t} = \sigma$ so $\bar{p}_B = t[2\sigma - 1]$. We now consider a consumer with type $\theta \in [\sigma, 1]$ and his choice to pay the privacy cost or not. As before, he must compare $\bar{p}_B$ and $p_B + c$. But again, $\bar{p}_B < t$ and $p_B + c \geq t + c$ so no agent to the right of the line will choose to pay the privacy cost, i.e., $\theta_B = 0$ as in the first case studied. This implies that θ_A and the prices p_A and p_B are the same as before. We now compute the firms' profit in order to characterize the DS's revenue.

As far as firm A is concerned,

$$\begin{aligned}
\pi_A &= \theta_A p_A + \int_0^{1/2+(p_B-p_A)/2t} p_A d\theta + \int_{\theta_A}^{\frac{\underline{p}_B+t}{2t}} \left(\underline{p}_B + (1-2\theta)t\right) d\theta \\
&= \frac{2t}{25}\left[\sigma + 2 - \frac{c}{t}\right]^2 + \frac{t}{100}\left[2\sigma + 4 + 3\frac{c}{t}\right]^2.
\end{aligned}$$

Regarding firm B,

$$\begin{aligned}
\pi_B &= \int_{1/2+\frac{p_B-p_A}{2t}}^{1} (p_B)\, d\theta + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} \left(\underline{p}_B\right) d\theta + \int_{\sigma}^{1} (\bar{p}_B)\, d\theta \\
&= \frac{t}{50}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right]^2 + \frac{t}{2}(\sigma - \frac{1}{2})(\frac{7}{2} - 3\sigma).
\end{aligned}$$

As before, we differentiate the DS's revenue with respect to σ. This leads to

$$\begin{aligned}
\frac{\partial R}{\partial \sigma} &= \frac{t}{25}\left[6\sigma + 12 - \frac{c}{t}\right] + \frac{t}{25}\left[\sigma + \frac{9}{2} - \frac{c}{t}\right] + \frac{t}{2}[5 - 6\sigma] \\
&= \frac{t}{50}\left[158 - 136\sigma - 4\frac{c}{t}\right] > 0, \text{ for all } \sigma \in [3/4, 1].
\end{aligned}$$

Let us assume now that $c > (\sigma - 1/2)t$, that is $\theta_A = 0$. We then get $p_A = p_B = t$, $\underline{p}_B = t(\sigma - \frac{1}{2})$ and $\bar{p}_B = 2t(\sigma - \frac{1}{2})$. Then

$$\begin{aligned}
\pi_A &= \frac{t}{2} + \int_0^{\frac{\underline{p}_B+t}{2t}} \left(\underline{p}_B + (1-2\theta)t\right) d\theta \\
&= \frac{t}{2} + \frac{t}{4}(\sigma + 1/2)^2.
\end{aligned}$$

Similarly

$$\begin{aligned}
\pi_B &= \frac{t}{2} + \int_{\frac{\underline{p}_B+t}{2t}}^{\sigma} (\underline{p}_B) d\theta + \int_{\sigma}^{1} \bar{p}_B d\theta \\
&= \frac{t}{2} + \frac{t}{2}(2\sigma - 1)(\frac{7}{4} - \frac{3}{2}\sigma).
\end{aligned}$$

Still denoting R the DS's revenue, we differentiate R with respect to σ. This leads to

$$\frac{\partial R}{\partial \sigma} = \frac{t}{2}(\frac{11}{2} - 5\sigma) > 0, \text{ for all } \sigma \in [3/4, 1]$$

Hence, the DS's revenue is still optimized for $\sigma = 1$, which completes the proof of Lemma 4.

Proof of Proposition 8

As shown in Lemma 4, the optimal level of precision for the DS is to choose $\sigma = 1$. Using the expression of π_A and π_B derived above, and setting $\sigma = 1$, we can write the equilibrium profit of the DS as

$$\begin{aligned} R &= \frac{2t}{25}\left[3 - \frac{c}{t}\right]^2 + \frac{t}{100}\left[6 + 3\frac{c}{t}\right]^2 + \frac{t}{50}\left[\frac{11}{2} - \frac{c}{t}\right]^2 + \frac{t}{8} - 2\underline{\pi}_B \\ &= \frac{t}{100}\left[\frac{337}{2} - 36\frac{c}{t} + 19\left(\frac{c}{t}\right)^2\right] + \frac{t}{8} - 2\underline{\pi}_B \end{aligned}$$

where $\underline{\pi}_B = \frac{(5t-c)^2}{32t}$ is the profit of Firm B derived in Corollary 2. This must be compared to the DS's revenue in the benchmark case derived in Proposition 6, that is $T = \frac{9c^2-2ct+57t^2}{64t}$. Then

$$\begin{aligned} T \geq R \quad &\Leftrightarrow \quad \frac{9c^2 - 2ct + 57t^2}{64t} \geq \frac{t}{100}\left[\frac{337}{2} - 36\frac{c}{t} + 19\left(\frac{c}{t}\right)^2\right] + \frac{t}{8} - 2\frac{(5t-c)^2}{32t} \\ &\Leftrightarrow \quad \left(\frac{149}{64} - \frac{337}{200}\right)t + \left(\frac{9}{25} - \frac{21}{32}\right)c + \left(\frac{13}{64} - \frac{19}{100}\right)\frac{c^2}{t} \geq 0. \end{aligned}$$

Note first that the last bracket $\left(\frac{13}{64} - \frac{19}{100}\right)$ is positive. Moreover, since $c \leq t$, the LHS is larger than

$$\left(\frac{149}{64} - \frac{337}{200}\right)c + \left(\frac{9}{25} - \frac{21}{32}\right)c = \left(\frac{107}{64} - \frac{265}{200}\right)c > 0 \tag{6}$$

The LHS is therefore positive so the DS's revenue is higher in the benchmark case where only one firm can benefit from some information about consumers.

Proof of Lemma 5

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

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

and 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.

For 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 (i.e., 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. We must assume that $c \le \frac{t(3m+3)}{3m+4}$ in order to ensure that a positive proportion of consumers pays the privacy cost.

Proof of Lemma 6

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

Proof of Proposition 9

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

Appendix C

Here 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 under this modification. We focus here on the analysis of the duopoly case.

First we consider the case where the two firms have acquired the information and show that the result of Lemma 1—which states 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 as given in Proposition 1. But now, these prices are relevant to only a share λ of the consumers who have not paid the privacy cost. The other share $1 - \lambda$ can, from the firm's perspective, be pooled with new consumers. Retaining the notation of Lemma 1, old consumers who have paid the privacy cost are of 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, after which the prices are 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 differs slightly from the one used in the main model, because not paying now leads to type revelation only with probability λ. For consumers to the left of the Hotelling line, this modified condition is written as $v - p_A - \theta t - c \geq \lambda[v - p_A(\theta) - \theta t] + (1 - \lambda)[v - p_A - \theta t]$, or equivalently, as $\theta \leq \theta_2 = \frac{t - p_A - \frac{c}{\lambda}}{2t}$ (where we have used $p_A(\theta) = (1 - 2\theta)t$).

Because $p_A = t + \frac{2t}{3} \frac{(2\theta_A + 1 - \theta_B)}{2 - \lambda} > t$, 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 about 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 two firms' profit will then be $\pi_A = \pi_B = (2-\lambda)t - \lambda\frac{t}{4}$. Let $M = (2-\lambda)/\lambda$. Then, the profit is $\lambda[M\frac{t}{2} + \frac{t}{4}]$, which 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 that is similar to an increase in the mass of new consumers.

We now consider the case where only firm A has acquired the information. Firm B may or may not have decided to compete for the consumers about which firm A has some information about; however, that decision does not affect how A chooses its basic price. It is the solution of the program $\max_{p_A}[(2-\lambda)\theta_1(p_A,p_B)+\lambda\theta_2]p_A$ or equivalently $\max_{p_A}\lambda[M\theta_1(p_A,p_B)+\theta_2]p_A$, where $M=\frac{(2-\lambda)}{\lambda}$. 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 that A chooses is given by $p_A(\theta) = p_B + (1-2\theta)t$ (when that value is positive). The condition under which consumers pay the privacy cost is therefore $\theta \leq \theta_2 = \frac{t+p_B-p_A-\frac{c}{\lambda}}{2t}$.

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

Now suppose that firm B does try to compete for consumers about which firm A has some information. In this case, a portion of the share λ of old consumers who have not paid the privacy cost will be served by firm B. These consumers are of type $\theta \in [\theta_2^0, 1]$ where $\theta_2^0 = \frac{p_B+t}{2t}$. So firm B's 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 in Lemma 3(ii) but with c/λ and M replacing (respectively) c and 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 effects. First it changes the value of the privacy cost—although this change is the same for all expressions of the profit. Second, it modifies the relative size of the two markets exactly as described 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. Proposition 8's exclusivity result, whereby the broker sells consumer data to just one firm no matter what the values of c and m, is also valid for any value of λ.