Privacy, Competition, and Multi-Homing

Document de Travail
Working Paper
2021-15

Marc Bourreau
Jean-Marc Zogheib

Privacy, Competition, and Multi-Homing

Marc Bourreau[*], Jean-Marc Zogheib[†]

Abstract

Two firms compete in prices and information disclosure levels. Firms derive revenues from two possible channels, i.e., by selling their service to consumers and by exploiting user data, sold to a monopoly data broker. A consumer signing up to one firm's service decides on the amount of personal information to provide. In a single-homing framework, firms engage in either a strict privacy regime with no information disclosure and high prices or a flexible privacy regime with positive disclosure levels and low prices, depending on consumer valuations. With the possibility of multi-homing, firms face issues in the monetization of multi-homing user data, which affects privacy regimes. On top of consumer valuations, the incentives to multi-home and product differentiation also impact firms' strategies. Firms may even end up engaging in a zero-privacy regime with maximal disclosure levels if monetization issues on multi-homing user data are not too significant.

Keywords: competition, online privacy, information disclosure, multi-homing.

JEL codes: D11; D40; L21; L41.

1 Introduction

Online privacy has become a critical variable of competition between digital players. They serve online users who are subject to growing privacy concerns regarding the collection and use of their personal data. In a survey conducted June 3-17 2019 by the Pew Research

[*]Telecom Paris, Institut Polytechnique de Paris. Email: marc.bourreau@telecom-paris.fr

[†]Economix, UPL, University Paris Nanterre, CNRS, University Paris Nanterre and EconomiX-CNRS, Nanterre F92000, France. Email: jmzogheib@parisnanterre.fr

Center on 4,272 U.S. adults, it has been found that 79% of them are concerned about how much data is collected about them by companies. In the same survey, around half (52%) of U.S. adults said they decided recently not to use a product or service because they were worried about how much personal information would be collected about them.[1] Therefore, how a firm determines the privacy level of its own service, i.e., chooses the level of user information disclosure, may strongly affect its profitability and market position.

On top of that, the business model of digital firms is characterized by important specificities related to digital markets. Most of these markets are two-sided, that is, when a digital firm, or platform, serves two distinct groups of users present at each side of the market. On one side of the market, platforms can offer low-price or even free services to consumers. In this way, platforms are able to attract consumer attention in the form of a higher data collection. On the other side of the market, platforms can "monetize" consumer attention on a data market to, say, data brokers. A direct consequence is that two-sided markets are likely to affect the dynamics of price competition. Notably, the existence of free services is related to the presence of indirect network effects, whereby a higher number of consumers on one side of the market increases the value of joining the platform for data brokers on the other side of the market. In addition to the privacy dimension, the pricing framework is an essential competition parameter in digital markets.

Competition in both prices and privacy thus shapes firms' business models. For instance, Google supplies consumers with various digital services, e.g., search engine (Google Search), mail (Gmail), maps (Google Maps, Waze), at a zero price. Thanks to the exploitation of personal data provided by its users, it can profile them and charge advertisers to target them. Contrary to Google, Apple charges for its digital services, e.g., data storage (iCloud) or music (Apple Music), bundled with Apple products (iPad, iPhone, etc.). Meanwhile, it has a strict privacy policy which is a central aspect of Apple's brand image.

Digital firms' business models could be further impacted by user behavior. First, digital service users can single-home (signing up to the service of only one platform) or multi-home

[1]See Pew Research Center (June 3, 2019) (1) and (2).

(signing up to the service of two or more platforms). For example, a consumer may buy different variants of horizontally differentiated services, depending on the extent to which service functionalities overlap (e.g., Google Search vs. Bing, Netflix vs. AppleTV). Second, users who have privacy concerns may provide a lower amount of personal information to firms. For instance, they can make use of privacy-enhancing technologies. Among others, the "Facebook Container" on Mozilla Firefox browser, adblockers (e.g., AdBlock, Ghostery), or the App Tracking Transparency (2021) on Apple IPhones through which "users will be able to see which apps have requested permission to track, and make changes as they see fit".

Competition in prices and privacy brings up the consideration of evolving business models for digital players. Users have changing behavior when signing up to firms' services (single vs. multi-homing) and privacy concerns over the collection and use of their personal information. In this paper, we examine the impact of competition between two firms in prices and information disclosure levels. We build a two-sided market model where two firms supply a horizontally differentiated service to consumers on one side, and sell access to consumer information, or user data, to a monopoly data broker on the other side.[2] Consumers observe the level of disclosure to which firms engage and their price before deciding which service to patronize and how much personal information to provide. The level of information disclosure is an inverse measure of privacy. Consequently, the perceived quality of the firm's service for each consumer increases with information provision and decreases with the firm's disclosure level. Firms derive revenue from two sources: purchase revenues from the prices charged to consumers and information disclosure revenues from the exploitation of user data. The latter depend on the firm's disclosure level, the amount of user data, and the price for access to this data charged to the data broker.

In a single-homing (SH) framework, we find that firms tend to adopt two types of business strategies captured by a trade-off between disclosure levels, consumer information provision, and consumer valuations. If consumer valuations for the service are sufficiently

[2] As in Casadesus-Masanell and Hervas-Drane (2015) and Anderson, Foros, and Kind (2019), we call this framework as "two-sided" in that firms (platforms) can derive revenues not only from the demand side (consumers) but also from the supply side (data broker).

high, firms adopt a strict privacy regime whereby they do not engage in the disclosure of user data. Firms therefore rely exclusively on purchase revenues. However, if consumer valuations are low enough, firms adopt a flexible privacy regime whereby they engage in the disclosure of user data. If consumer valuations increase, firms charge lower disclosure levels, and consumers are willing to provide more information. Firms therefore rely on both purchase and disclosure revenues, and they all the more rely on disclosure revenues that consumer valuations decrease. We show that if consumer valuations are very low or if competition intensifies (through less product differentiation), firms ultimately choose to subsidize consumers by charging negative prices. Firms represent bottlenecks on the data from their users and charge the data broker a price that leaves her with zero surplus.

We then open up for the possibility of consumer multi-homing (MH), firms may face monetization issues on MH user data, on which they are no longer bottlenecks. We show that firms' business models are altered compared to SH. Indeed, firms' trade-off with respect to privacy is not only impacted by consumer valuations, but also by the value of purchasing a second service (i.e., the value of MH), the level of product differentiation, and the value of MH user data. Firms' adoption of privacy regimes thus depends on the interaction between these parameters. Three distinct scenarios emerge with the possibility of MH. There exists a first one where monetization issues on MH user data are significant. In such a case, firms decrease their disclosure levels as consumer valuations or the value of MH increases, which leads firms towards stricter privacy regimes. Higher product differentiation drives up disclosure levels, which leads firms towards more flexible privacy regimes when competition softens. Firms may even charge maximal disclosure level if product differentiation is sufficiently high and thus engage in zero-privacy regimes. The impact of the value of MH user data on firms' disclosure level is more complex and depends on the firms' trade-off between the extent of monetization issues on MH user data and marginal disclosure revenues raised on MH users. A second scenario happens when monetization issues on MH user data are small and a higher disclosure level induces a sharp decrease in consumer willingness to pay. We find the same results qualitatively as in the first scenario, but the second scenario is more likely to happen if monetization issues are not too small and consumer willingness to pay is sufficiently high. Moreover, there exists a third scenario where monetization issues are small and a higher disclosure level induces a slight decrease in consumer willingness to pay. In such a case, firms engage in a zero-privacy regime by charging maximal disclosure levels. Finally, we show that firms choose to subsidize consumers by trading-off between consumer willingness to pay and expected disclosure revenues per MH user.

All things being equal, we may therefore end up with extreme or intermediate outcomes in the presence of multi-homing compared to single-homing only.

In the next section, we discuss the related literature. Section 3 presents the model framework. In Section 4, we solve for firms' decisions when consumers can only single-home. In Section 5, we examine the case where multi-homing is possible. Section 6 concludes. All proofs are in the appendix.

2 Relation to literature

Our paper is related to the literature on the economics of privacy.

A strand of the literature investigates the link of privacy with allocation efficiencies and externalities (Stigler, 1980; Posner, 1981; Hermalin et Katz, 2006; Calzolari and Pavan, 2006; Hui and Png., 2006). Calzolari and Pavan (2006) consider information disclosure between two firms (principals) interested in discovering consumers' willingness to pay. In a model of sequential contracting, a common agent strategically decides whether to report her true type. They find that the transmission of personal data from one company to another may in some cases reduce information distortions and enhance social welfare.[3] We contribute to this literature by setting a framework where consumers strategically decide to provide some of their personal data to the firm they patronized. Consumers have privacy concerns over the disclosure of their data, while providing more information to firms is beneficial to consumers through more personalized service.

Closer to our paper, another strand of the literature study the link between privacy and

[3] See also Acquisti and Wagman (2016) for a comprehensive literature review on the economics of privacy.

competition (Noam, 1995a,1995b; Spulber, 2009; Taylor and Wagman, 2014; Casadesus-Masanell and Hervas-Drane, 2015; Shy et al., 2016; Montes et al., 2018; Choi et al., 2019; Lefouili and Toh, 2020, Kim, 2020; Argenziano and Bonatti, 2020; Ichihashi, 2020a). Casadesus-Masanell and Hervas-Drane (2015) analyze the effect of competition on consumer privacy. The authors examine a model where firms compete with privacy (i.e., information disclosure levels). Consumers voluntarily provide personal information to firms in order to obtain higher-quality products. Firms can disclose and sell some of this information to an outside firm, which negatively impacts consumers' utility. They determine that the market equilibrium involves vertical differentiation for the disclosure of consumer information. One firm positions itself as a high-quality (low-disclosure) provider, and the other firm as a low-quality (high-disclosure) provider. Moreover, they show that more intense competition implies more disclosure by the firms, i.e., lower consumer privacy. The policy implication of their framework is that one should expect a low level of privacy in a competitive marketplace, but this does not necessarily harm consumer welfare. We build on the framework of Casadesus-Masanell and Hervas-Drane (2015). However, our paper differs first in that we investigate a model of horizontal differentiation. Second, we study the impact of consumer multi-homing on top of single-homing. Third, we provide a more general utility function compared to the authors. Finally, we model the interaction between firms and the data broker, when the former intend to monetize user data by disclosing it to the latter.

Our paper is also related to the literature dealing with the impact of collection and use of user data on the competition between digital firms (Prufer and Schottmuller, 2017; Prat and Valletti, 2019; Kim et al., 2019; Belleflamme et al., 2020; De Cornière and Taylor, 2020; Ichihasi, 2020b).[4] Prat and Valletti (2019) study digital platforms as attention brokers with proprietary information about their users' product preferences and sell targeted ad space to retail product industries. They show that increased concentration among attention brokers may reduce entry in retail product industries, which ends up harming consumer welfare. They argue that a monopolistic attention broker has an incentive to

[4] Closer to targeted advertising, see Tag (2009), Anderson and Gans (2011), and De Cornière and Nijs (2014, 2016).

create an attention bottleneck by reducing the supply of targeted advertising. If fewer ads are sold, this will reduce the number of retail firms with access to consumers, thus increasing their market power. This translates into higher total profits for the retail industry, partially captured by the platform through higher total ad revenue. The authors finally evaluate that a merger between platforms can increase market power in the retail industry to the detriment of consumers. In our framework, firms are attention brokers in that they are bottlenecks for access to user data. However, they are no longer bottlenecks as there is consumer multi-homing, and the data broker is left with a positive surplus. Thus, firms may reduce access to user data, particularly if they engage in a strict-privacy regime.

Our paper is finally related to the seminal literature on two-sided markets and the impact of consumer single and multi-homing (Rochet and Tirole, 2003; Caillaud and Jullien, 2003; Armstrong, 2006; Choi, 2010; Belleflamme and Peitz, 2019; Anderson et al., 2019). Anderson et al. (2019) study the impact of consumer multi-homing on market equilibrium and performance. The authors show that equilibrium consumer prices are independent of the number of platforms when some but not all consumers multi-home. On the contrary, advertising prices decrease as the fraction of multi-homers increases. They conclude that compared to single-homing, multi-homing flips the side of the market on which platforms compete. In our framework, we find that due to monetization issues, multi-homing tends to decrease the access price of user data tends to decrease. This results from competition for selling user data of multi-homing consumers, whereas there were no competition with single-homing. In a companion paper, Anderson et al. (2017) examine in a one-sided market the characteristics of single-homing and multi-homing equilibrium. The authors assume that each product has its own specific part, while a common overlapping part belongs to both products. Without product overlap, allowing multi-homing should be better for the firms because they have greater demand and less fierce competition. With overlap, each firm cannot charge for the common part, as they compete à la Bertrand. Therefore, allowing multi-homing could make the firms worse off because of overlap and horizontal differentiation. As Anderson et al. (2017), we assume that choose to multi-home by considering the rate of overlap between the two services.

3 Model

We study a two-sided market where firms compete in prices and information disclosure levels. Two firms, A and B, are located on a Hotelling line and supply a service to consumers at zero marginal cost. Firms are located at the two extremes of a line of length 1, with firm A located at point 0 and firm B at point 1. Consumers are uniformly distributed along the line, and choose to sign up for a service from firm A, firm B, or both if this is possible.

A consumer purchasing only the service of firm $i = A, B$ (single-homing) decides on the level of personal information $y_i \in [0,1]$ to provide to this firm.[5] The net utility of the consumer, located at $x \in [0,1]$, when purchasing service i, for given price p_i and information disclosure $d_i \in [0,1]$, is

$$U_i = vq(y_i, d_i) - p_i - tx_i, \tag{1}$$

where $v > 0$ is a parameter which reflects the intrinsic benefit of the service, and $tx_i = tx$ is the transportation cost incurred when buying for firm $i = A$ and $tx_i = t(1-x)$ the transportation cost when buying from firm $i = B$.

We interpret term $q(y_i, d_i)$ in equation (1) as the quality of firm i's service, where quality is assumed to be increasing and concave in the level of information provision y_i (i.e., $\partial q/\partial y_i \geq 0$ and $\partial^2 q/\partial y_i^2 \leq 0$). It means that consumers benefit from providing information because it allows the firm to provide a personalized, higher quality service; but at the margin, this benefit is decreasing. Moreover, quality is assumed to be decreasing and concave in the level of information disclosure d_i, meaning that consumers incur disutility from the disclosure of their personal information (i.e., $\partial q/\partial d_i \leq 0$ and $\partial q^2/\partial d_i^2 \leq 0$). Finally, as d_i is higher, providing more information affects negatively consumer utility ($\partial q/\partial y_i \partial d_i \leq 0$). For the sake of exposition, let $q(y_i, d_i) \equiv q_i$.

[5]The framework where a consumer can purchase both services (multi-homing) is presented later in Section 5.

Firm i decides on a price $p_i \in \mathbb{R}$ and a disclosure level $d_i \in [0, 1]$. Firms derive revenues from purchases, and we allow them to set negative prices (i.e., to subsidize consumers). Firms can also derive revenues by disclosing user data on a data market. More precisely, firm i sells access to user data to a monopoly data broker, at a price r_i per user and piece of information. This means that firm i's revenue from selling access to the information of one user is $d_i y_i \times r_i$. Let $D_i(.)$ be the demand of firm i. The profit of firm i is then

$$\Pi_i = (p_i + d_i y_i r_i)\, D_i.$$

We now characterize the utility of buying access to user data for the data broker. Let $v^b > 0$ be the value placed by the data broker on the information of each user signing up to service i. When acquiring firm i's user data, the net utility of the data broker is thus given by $\left(v^b - r_i\right) D_i$.

We make the following stability assumption.

Assumption 1.

$$D_i \left| \frac{\partial^2 D_i}{\partial d_i^2} \right| > \left(\frac{\partial D_i}{\partial d_i} \right)^2 .$$

We consider the following sequence of events. At the first stage, firms simultaneously set their disclosure level.[6] At the second stage, firms simultaneously set the price of their service to consumers and the price for access to user data charged to the data broker. At the third stage, having observed prices and disclosure levels, consumers choose to purchase firm A's or firm B's service, or to stay out of the market. At the fourth stage, consumers decide on the level of information provision to the firm they have patronized.

We look for the subgame perfect equilibrium of this game.

In what follows, an individual signing up to the service of a firm is designated by either a consumer or a user.

[6] As in Casadesus-Masanell and Hervas-Drane (2015), firms commit to level of information disclosure announced in the first stage.

4 Single-homing

We start by considering single-homing, that is, a situation where consumers purchase one service or none.

In what follows, we determine the equilibrium under duopoly and monopoly. We restrict our attention to parameter values such that the market is covered in equilibrium.

4.1 Duopoly

We start by solving for the equilibrium with single-homing when firm A and B are active.

At the fourth stage, a consumer decides on the level of information to the firm she has patronized, let's say firm i. For a given disclosure level d_i and price p_i, the consumer chooses y_i by maximizing (1), which gives

$$y_i^c(d_i) \equiv \arg\max_{y_i} vq(y_i, d_i) - p_i - tx_i. \tag{2}$$

Using the implicit function theorem, we show that the higher firm i's disclosure level is, the less a consumer is willing to provide information

$$\frac{\partial y_i}{\partial d_i}_{|y_i=y_i^c(d_i)} = -\frac{\partial^2 q/\partial y_i \partial d_i}{\partial^2 q/\partial y_i^2}_{|y_i=y_i^c(d_i)} \leq 0.$$

Moreover, let us assume that $\partial^2 y_i^c(d_i)/\partial d_i^2 \leq 0$. To save notations, let $y_i^c(d_i) = y_i^c$.

At the third stage, we compute firms' demands by determining the location of the consumer who is indifferent between A and B. Replacing for y_i^c into (1), we find that the indifferent consumer is located at

$$x^* = \frac{1}{2} - \frac{p_A - p_B}{2t} + \frac{v}{2t}\left(q(y_A^c, d_A) - q(y_B^c, d_B)\right). \tag{3}$$

The demand of firm A is therefore $D_A(p_A, p_B) = x^*$, while the demand of firm B is $D_B(p_A, p_B) = 1 - x^*$.

At the second stage, firm i sets the price for access to user data to the data broker. Each firm makes a take-it-or-leave-it offer to the data broker because they represent a bottleneck for access to user data. The data broker is willing to access to firm i's user data as long as $r_i \leq v^b$. Firm i therefore charges the data broker $r_i = v^b$, and the data broker is left with zero surplus. For simplicity, we normalize v^b to 1, meaning that $r_i = 1$.

If the market is covered, firm i's profit is then given by

$$\Pi_i(p_A, p_B) = (p_i + d_i y_i^c) D_i(p_A, p_B). \tag{4}$$

From (4), we observe that firm i derives profits from selling its service at a price p_i, i.e., through purchase revenues, and from the exploitation of user data y_i^c at a price $r_i = 1$ and a disclosure level d_i, i.e., through disclosure revenues.

We now write the following lemma on firm i's disclosure revenue per user, $d_i y_i^c$.

Lemma 1. *Firm i's disclosure revenues per user $d_i y_i^c$ are concave and increasing in d_i if and only if $d_i \in (0, \hat{d}]$, where $\hat{d} \equiv \arg\max_{d_i} d_i y_i^c$ and $\hat{d} \in (0, 1]$.*

From Lemma 1, we find that firm i's disclosures revenues per user are maximized at $d_i = \hat{d}$, where $\hat{d} \equiv \arg\max_{d_i} d_i y_i^c$. This implies that it does not benefit from setting a disclosure level d_i higher than $\hat{d}$.

By examining the impact of a higher disclosure level d_i on firm i's disclosure revenues per user, we observe that there actually exists two opposite effects at work:

$$\frac{\partial(d_i y_i^c)}{\partial d_i}(d_i) = \underbrace{y_i^c}_{>0} + d_i \underbrace{\frac{\partial y_i^c}{\partial d_i}}_{\leq 0}.$$

On the one hand, there is a positive effect coming from the consumer providing personal information to firm i ($y_i^c > 0$). On the other hand, the consumer suffers from being disclosed as it has privacy concerns; it then provides less information to firm i at the margin ($d_i \partial y_i^c / \partial d_i \leq 0$). When solving Stage 1, we show that any $d_i > \hat{d}$ involves a corner solution where $d_i = 0$. We thus focus the analysis on the case where $d_i \in (0, \hat{d}]$, i.e., when firm i obtains an interior solution.

Each firm sets its price p_i to maximize its profit, which is given by equation (4), taking the rival's price p_j as given. Solving for the price reaction functions denoted by

$$p_i(p_j) = \frac{t + v(q_i - q_j) - d_i y_i^c + p_j}{2}, \quad i = A, B,$$

we obtain the equilibrium prices[7]

$$p_i^c(d_i, d_j) = t + \frac{v(q_i - q_j) - 2d_i y_i^c - d_j y_j^c}{3}.$$

We now study how disclosure levels d_i and d_j affect firm i's price if $d_i \in (0, \hat{d})$.

$$\frac{\partial p_i^c}{\partial d_i}(d_i) = \frac{v}{3}\frac{\partial q_i}{\partial d_i} - \frac{2}{3}\frac{\partial (d_i y_i^c)}{\partial d_i} \leq 0; \quad \frac{\partial p_i^c}{\partial d_j}(d_j) = \underbrace{-\frac{v}{3}\frac{\partial q_j}{\partial d_j}}_{>0} \underbrace{-\frac{1}{3}\frac{\partial (d_j y_j^c)}{\partial d_j}}_{<0}. \tag{5}$$

The first expression in (5) shows that the effect of a higher disclosure level d_i affects negatively firm i's price because of a lower perceived quality of service ($\partial q_i / \partial d_i \leq 0$), and the firm thus relies more on disclosure revenues. The second expression in (5) shows that a higher disclosure d_j has an ambiguous effect on firm i's price. On the one hand, a higher d_j affects positively firm i's price since consumers who subscribe to firm i's service benefit from a relatively better quality (a higher d_j lowers firm j's service quality), which induces firm i to increase its price. Firm i then relies more on purchase revenues. On the other hand, a higher d_j negatively affects p_i^c because it increases firm j's disclosure revenues, which induces firm i to decrease its price and to rely more on disclosure revenues as well.

Plugging equilibrium prices into the profit function (4), we now solve for firms' optimal disclosure levels at Stage 1. Firm i's profit can be rewrited

$$\Pi_i(d_i, d_j) = 2t \left(D_i(d_i, d_j)\right)^2 \quad \text{where} \quad D_i(d_i, d_j) = \frac{1}{2} + \frac{v(q_i - q_j) + d_i y_i^c - d_j y_j^c}{6t}.$$

Each firm sets its disclosure level d_i to maximize its profit, taking the rival's disclosure level d_j as given. Solving for the disclosure reaction functions, we obtain the equilibrium disclosure levels, d_A^c and d_B^c.[8]

[7] The second-order condition is always satisfied, as $\partial^2 \Pi_i / \partial p_i^2 = -1/t < 0$.

[8] As already mentionned, we look for an interior solution ans therefore consider the case where firm i's

The following proposition characterizes the equilibrium:

Proposition 1. *(i) In the duopoly equilibrium with a covered market, if* $0 < t \leq vq(y^c, d^c) + d^c y^c/2$*, consumers provide information* y^c *(*$i = A, B$*). Firms' optimal prices and disclosure levels are* $p^c = t - d^c y^c$ *and* d^c*.*

(ii) Firms' choice to charge positive disclosure levels depends on consumer valuations v*, where* $\partial d^c/\partial v \leq 0$*. It implies that if* $v \geq v^c$*,* $d^c = 0$*.*

From Proposition 1, we find that depending on consumer valuations v, there are two equilibria types.

If consumer valuations for the service are sufficiently high ($v \geq v^c$), firms do not engage in the disclosure of user data ($d^c = 0$) and charge positive prices ($p^c = t$). Firms rely exclusively on purchase revenues and then adopt a *strict privacy regime.*

If consumers have low valuations ($v < v^c$), firms engage in the disclosure of user data ($d^c > 0$). As consumer valuations (v) increases, the level of information disclosure decreases ($\partial d^c/\partial v \leq 0$) whereas consumers provides more information ($\partial y^c/\partial v \geq 0$) and prices increase ($\partial p^c/\partial v \geq 0$). As v decreases, firms rely comparatively more on revenues from the disclosure of user data than purchase revenues. In this respect, the equilibrium price p^c represents firm i's trade-off between purchase revenues (t) and disclosure revenues ($d^c y^c$) per user, as this can be seen on Figure 1. As consumer valuations v decrease, d^c increases and firm i charges a lower price p^c, which explains why it relies more on disclosure revenues. In sum, if $v < v^c$, firms adopt a *flexible-privacy regime.*

Notably, firms are bottlenecks for access to user data and extract all the surplus from the data broker.

Corollary 1. *There is negative pricing under a flexible-privacy regime if, for* $d^c > 0$*, product differentiation* t *or consumer valuations* v *are sufficiently low.*

From Corollary 1, we find that firm i may charge a negative price, that is, subsidizing consumers. This is possible if the level of product differentiation is not too high (i.e., low

disclosure revenue per user is increasing in d_i ($\partial(d_i y_i^c)/\partial d_i \geq 0$). Otherwise, we would obtain a corner solution with $d^c = 0$.

t), or if consumer valuations (v) are sufficiently low (Figure 1). In other words, when competition intensifies or consumer valuations for the service are low, firms may subsidize consumers.

Figure 2 illustrates the duopoly equilibrium. We observe that firms face a trade-off between the disclosure of user data (d^c), the price (p^c), and the level of information provision (y^c), depending on consumer valuations (v). A higher disclosure level decreases the level of information provision and prices. Consumers with low valuations provide less information, but this information is more exploited to generate larger disclosure revenues, while these consumers may be subsidized (flexible privacy regime). Consumers with high valuations provide more information, but firms generate more value through purchase revenues (strict privacy regime).

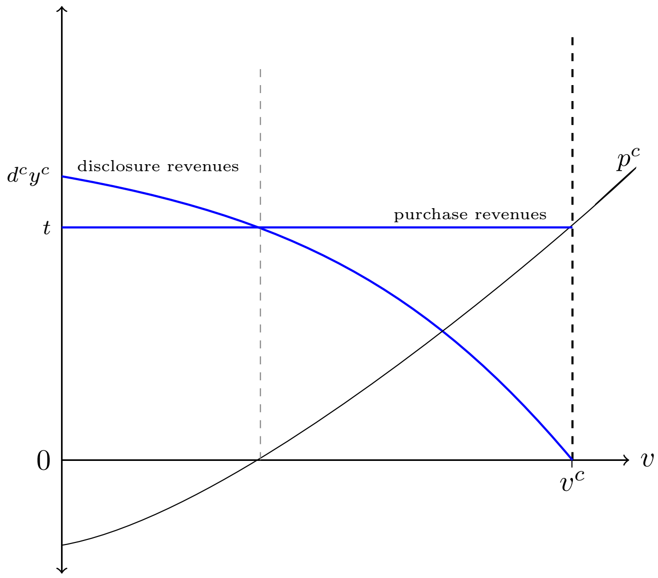

Figure 1: Pricing trade-off in flexible-privacy regime

4.2 Monopoly

We now study the equilibrium when A and B act as a monopoly, for example, after a merger. We consider a multi-product monopoly, which supplies services A and B at prices p_A and p_B, with disclosure levels d_A and d_B, respectively.

At Stage 4, the level of information provision y_i for service i is as given in (2): for a given d_i, $y_i = y_i^m(d_i) = y_i^c(d_i)$.

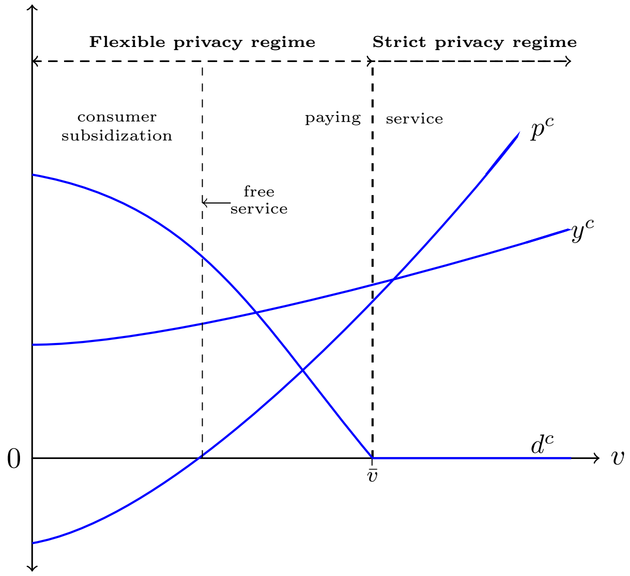

Figure 2: Equilibrium in single-homing duopoly

At Stage 3, if the market is covered, proceeding similarly as in the duopoly, we find that the demand for service A is $D_A(p_A, p_B) = x^*$ and the demand for service B is $D_B(p_A, p_B) = 1 - x^*$, where x^* is given by (3).

In a not covered market, we determine the location of the consumer indifferent between buying service i and staying out of the market, i.e., $U_i = 0$. The demand for service i in this case is:

$$D_i^u(p_i) = \frac{vq_i(y_i^m, d_i) - p_i}{t}, \quad i = A, B.$$

At Stage 2, we determine the access price of user data. As in the duopoly, the monopolist has monopoly power on user data (i.e., it represents a bottleneck on user data). Thus, the data broker obtains this data at a price $r_A = r_B = 1$.

We now consider the pricing problem of the monopoly. If the market is covered, the monopoly profit is given by:

$$\Pi^m(p_A, p_B) = \sum_{i=A,B} (p_i + d_i y_i^m) D_i(p_A, p_B). \tag{6}$$

At the optimum for the monopoly with a covered market, the indifferent consumer receives zero surplus, i.e., $vq_A - p_A - tx^* = 0$. Substituting for x^* in (3), we obtain the relation between prices that ensures market coverage:

$$p_B = v\,(q_A + q_B) - p_A - t.$$

We can thus express Π^m as a function of only p_A.[9] We solve for the first-order condition $\partial\Pi^m/\partial p_A = 0$ (the SOC holds as $\partial^2\Pi^m/\partial p_A^2 = -4/t < 0$). When the market is covered, we obtain the equilibrium prices denoted by

$$\begin{aligned} p_i^m(d_i, d_j) &= \frac{v(q_i+q_j)}{2} - \frac{d_i y_i^m - d_j y_j^m}{4} - \frac{t}{2}; \quad i \neq j = A, B, \\ \text{with} \quad D_i &= \frac{1}{2} + \frac{v(q_i - q_j)}{2t} + \frac{d_i y_i^m - d_j y_j^m}{4t}. \end{aligned}$$

If the market is not covered, the profit of the monopoly is given by

$$\Pi^u(p_A, p_B) = \sum_{i=A,B} (p_i + d_i y_i^m)\, D_i^u(p_i).$$

Profit maximization yields $p_i^u(d_i) = (vq_i - d_i y_i^m)/2$.

Consider now the first stage where the monopolist chooses the disclosure levels.

If the market is covered, the monopolist sets the disclosure levels to maximize its profit Π^m. Solving for the first-order conditions, we obtain the equilibrium disclosure levels, d_A^m and d_B^m.

We proceed in a similar way for the uncovered monopoly and obtain d_i^u.

The following proposition characterizes the monopoly outcome:

Proposition 2. *(i) In a monopoly with a covered market, if $0 < t < vq(y^m, d^m) + d^m y^m$, consumers provide information y^m $(i = A, B)$. Firm's optimal prices and disclosure levels are $p^m = vq^m - t/2$ and d^m, where $q^m \equiv q(y^m, d^m)$.*

(ii) The monopolist's choice to charge positive disclosure levels depends on consumer valuations v, where $\partial d^m/\partial v \leq 0$. If $v \geq v^m$, $d^m = 0$.

As in Proposition 1, there are two types of equilibria. If consumer valuations are

[9] Therefore, $D_A(p_A) = (vq_A - p_A)/t$ and $D_B(p_A) = 1 - (vq_A - p_A)/t$.

sufficiently high ($v \geq v^m$), the monopolist does not engage in the disclosure of user data ($d^m = 0$) and charges positive prices ($p^m = vq(y^m(0), 0) - t/2 > 0$). The monopolist therefore relies exclusively on purchase revenues and, as in the duopoly, the firm adopts a strict-privacy regime.

If consumers have low valuations, ($v < v^m$), the monopolist engages in the disclosure of user data ($d^m > 0$). As consumer valuations (v) increase, the level of information disclosure decreases ($\partial d^m / \partial v \leq 0$) while consumers provide more information ($\partial y^m / \partial v \geq 0$) and prices increase ($\partial p^m / \partial v > 0$). The firm here engages in a flexible-privacy regime.

The monopolist is a bottleneck for access to user data and extract all surplus from the data broker.

Corollary 2. *There is negative pricing under a flexible-privacy regime if for $d^m > 0$, consumer valuations v are sufficiently low and product differentiation t is sufficiently high.*

From Corollary 2, we observe that the monopolist may charge negative prices, that is, subsidize consumers: it may happen if product differentiation is sufficiently high (i.e., a high t) and if consumer valuations (v) are sufficiently low. It differs from the duopoly where a lower t, that is, a higher competitive intensity, drives prices below zero. This result is explained by the price maximization of the monopolist, through which it chooses the maximum price that extracts the surplus of the indifferent consumer.

5 Multi-homing

We now consider the possibility that some of the consumers purchase both services (i.e., multi-home).

Multi-homing choice. We first characterize how consumers value multi-homing (MH) over single-homing (SH). Let $\sigma \in [0, 1]$ represent the incremental value of signing up to a second service. If $\sigma = 1$, there is no overlap between the two services and consumers derive the full utility from the second service. On the other hand, if $\sigma = 0$, there is a large overlap and the consumption of the second service brings no incremental gross utility.

We examine a framework where some consumers choose to single-home, while others choose to multi-home. Figure 3 depicts a possible market outcome, where consumers located to the left of point x_B^0 purchase service A only, those on the right of x_A^0 purchase service B only, and finally consumers between x_B^0 and x_A^0 purchase both.

Utility functions. The utility of a consumer who purchases service j in addition to service i is

$$\begin{aligned} U_{i,j} &= U_i + \{\sigma v q(y_j, d_j) - p_j - t x_j\} \\ &= U_i + U_j - (1-\sigma) v q(y_j, d_j) \end{aligned} \tag{7}$$

If $\sigma = 0$, the consumer does not benefit from the consumption of a second service.[10] As σ increases, the consumer increasingly values the consumption of a second service, which implies that multi-homing is more valuable. If $\sigma = 1$, there is no overlap between services A and B and $U_{A,B} = U_{B,A}$. In equation (7), $(1-\sigma)vq(y_j, d_j)$ can be interpreted as the disutility from multi-homing by purchasing service j in addition to service i; this disutility increases as σ decreases.

Valuation of MH user data. We finally determine how the data broker may value MH user data. Let $\alpha v^b = \alpha$ (v^b has been normalized to 1) be the value of the information of each user when sold twice to the data broker, where $\alpha \in (0, 1]$. Since the information of a MH user is in possession of both firms, this characterization captures the idea that access to MH user data may be sold twice (i.e., by both firms A and B). As a result, it becomes less valuable for the data broker compared to access to SH user data, sold only once.[11] In other words, firms face a monetization issue over MH user data if $\alpha < 1$. Notably, α could be interpreted as an "expected" value placed by the data broker when user data is sold twice.[12]

In what follows, we first examine the duopoly framework. Then, we study a monopolistic market. As in the previous section, we restrict our attention to parameter values such that the market is covered.

[10]We assume that if $\sigma = 0$, there is no multi-homing, i.e., $vq(y_i, d_i) - p_A - p_B - t < 0$.

[11]Anderson et al. (2019) use this type of modelling with $\alpha \in [0, 1]$.

[12]Another possible interpretation is that each piece of information of MH users takes only two extreme values, 0 with probability $1-\alpha$ (i.e., the data is not valuable) or 1 with probability α (i.e., the data is fully valuable). In this case, the expected value of data per MH user for the data broker would be $\alpha \times 1 + (1-\alpha) \times 0 = \alpha$.

Note that for the timing at Stage 3, consumers now choose to purchase firm A's or firm B's service, both firms' services, or to stay out of the market.

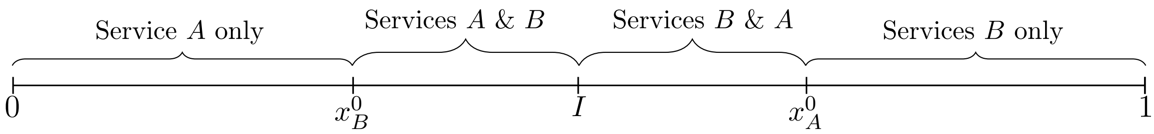

Figure 3: Demand composition with multi-homing

5.1 Duopoly

We solve for the equilibrium in the duopoly case.

At Stage 4, each consumer decides on the level of information to provide to the firm(s) he has patronized. If the consumer signs up to service i only, he chooses y_i by maximizing (1), and the level of information provision is then given by (2). If the consumer signs up to both services, he chooses y_A and y_B to maximize $U_{i,j}$, which gives $y_i^c(d_i) \equiv \arg\max_{y_i} U_{i,j}$.

Therefore, MH and SH consumers provide the same level of information when using service i.

At Stage 3, consumers choose which service(s) to patronize. With our specification, a consumer who purchases service i in addition to service j does not necessarily derive the same utility than if she purchases j in addition to i, i.e., we can have either $U_{AB} > U_{BA}$ or $U_{AB} < U_{BA}$. We therefore make the simplifying assumption that multi-homers to the left of $I = 1/2$ sign up to service B in addition to service A and the ones to the right of I sign up to service A in addition to service B (see Figure 3).[13] We justify this assumption insofar as multi-homers incur disutility from firstly signing up to a service far from their location; multi-homers to the left of 1/2 (to the right of 1/2) are therefore assumed to primarily sign up to service A (service B). This disutility from location is exogenous to this framework and is simply assumed.

The location of the consumer who is indifferent between purchasing service A and purchasing both services is then given by $U_A = U_{A,B}$ (location x_B^0 on Figure 3). Similarly,

[13] Anderson et al. (2017) specify a utility function where a consumer who purchases service i for its incremental value over service j does so depending on its location x.

the location of the consumer who is indifferent between purchasing service B and purchasing both services is given by $U_B = U_{B,A}$ (location x_A^0 on Figure 3).

Let $D_i^{sh}(.)$ be the single-homing demand of firm i and let $D^{mh}(.)$ be the multi-homing demand, common to firms A and B. The single-homing demand of firm i is then

$$D_i^{sh}(p_j) = 1 - \frac{\sigma v q_j - p_j}{t}, \quad i \neq j,$$

where $x_B^0 \equiv D_A^{sh}(p_B)$ and $1 - x_A^0 \equiv D_B^{sh}(p_A)$. The multi-homing demand is given by

$$D^{mh}(p_A, p_B) = \frac{\sigma v(q_A + q_B) - (p_A + p_B)}{t} - 1,$$

where $x_A^0 - x_B^0 \equiv D^{mh}(p_A, p_B)$.

Firm i's total demand is then given by $D_i = D_i^{sh} + D^{mh}$, that is,

$$D_i(p_i) = \frac{\sigma v q_i - p_i}{t}, \quad i = A, B. \tag{8}$$

Firm i's total demand therefore only depends on its own price and disclosure level. However, the composition of firm i's demand between SH and MH consumers depends on the prices and disclosure levels of the two firms.

At Stage 2, firm i determines the price for access to user data charged to the data broker. We consider the possibility that firms A and B choose to cooperate to solve the monetization issue from the double selling of MH user data. Let then $c_D \geq 0$ be the exogenous cost of data processing. If each firm incurs this cost, they may be able to split MH user data into two distinct datasets, thereby selling access to MH user data at one time to the data broker. Firm A would sell access to the data of MH users to the left of 1/2 only whereas firm B would sell access to the data of MH users to the right of 1/2 only, and by that, creating bottlenecks on MH user data.

We explore the scenario of cooperation on the sale of MH user data and write the following lemma.

Lemma 2. *(i) The non-cooperative outcome on the sale of MH user data by firms to* *the data broker is the only Nash equilibrium in pure strategies.*

(ii) The data broker obtains SH user data from firm i at a price $r_i = v^b = 1$ whereas she obtains MH user data from firm i at a price $r_i = \alpha$, where $\alpha \leq 1$.

From Lemma 2, we show that firms compete for selling access to MH user data ($r_i = \alpha$) rather than behaving strategically to sell access to this data once. Indeed, any deviation from a cooperative behavior is profitable for firm i and this is why firms do not cooperate at equilibrium.[14] Firms are therefore bottlenecks on SH user data only ($r_i = 1$).

Let $D_i^{sh}(p_j) + \alpha D^{mh}(p_i, p_j) = (1-\alpha)D_i^{sh}(p_j) + \alpha D_i(p_i)$. Firm i's profit can then be written as:

$$\Pi_i(p_i) = p_i D_i(p_i) + d_i y_i^c \left((1-\alpha)D_i^{sh}(p_j) + \alpha D_i(p_i)\right), \quad \alpha \in (0,1]. \tag{9}$$

The first term on the first line in (9) represents firm i's purchase revenues, while the second term represents firm i's disclosure revenues from the expected number of users from whom firm i can raise these revenues. Indeed, $(1-\alpha)D_i^{sh}(p_j) + \alpha D_i(p_i)$ can be interpreted this way since with probability $1-\alpha$, it generates disclosure revenues from SH users only, whereas with probability α, it generates them from all users, i.e., SH and MH users.

Each firm i sets its price p_i to maximize its profit given by (9), taking its rival's price p_j as given. Solving for the first-order conditions of profit maximization, we obtain the equilibrium price[15]

$$p_i^c(d_i) = \frac{\sigma v q_i - \alpha d_i y_i^c}{2}, \quad i = A, B. \tag{10}$$

We can divide firm i's price $p_i^c(d_i)$ in (10) into two terms. The first term ($\sigma v q_i$) represents consumers willingness to pay for service i: it depends on consumer valuations (v), the value of multi-homing (σ), and the quality of the service (q_i), which decrease with the disclosure level d_i and increases with consumer information provision y_i^c. The second

[14] The magnitude of the cost of data processing c_D does not change the outcome of the game; indeed, even if $c_D = 0$, firms would prefer not cooperating.

[15] The second-order conditions are always satisfied: $\partial^2 \Pi_i / \partial p_i^2 = -2/t < 0$.

term ($\alpha d_i y_i^c$) represents expected disclosure revenues of firm i per MH user. Since we look for an interior solution, this second term depends positively on the disclosure level d_i (see Lemma 1). Therefore, firm i charges a price to consumers by trading-off between consumer willingness to pay and expected disclosure revenues per MH user.

Firm i's demand now writes:

$$D_i(d_i) = \frac{\sigma v q_i + \alpha d_i y_i^c}{2t}, \quad D_i^{sh}(d_j) = 1 - D_j(d_j), \quad D^{mh}(d_i, d_j) = D_i(d_i) + D_j(d_j) - 1.$$

Plugging the equilibrium prices into the profit function (9), we now solve for the firms' equilibrium disclosure levels at Stage 1.

Let us first examine how firm i's profit $\Pi_i(p_i^c(d_i), p_j^c(d_j), d_i, d_j)$ is affected by a variation of d_i. Using the envelope theorem, we observe that the impact of a variation of d_i on firm i's profit is such that

$$\frac{d\Pi_i}{dd_i} = (p_i + \alpha d_i y_i^c)\frac{\partial D_i}{\partial d_i}(p_i) + \frac{\partial(d_i y_i^c)}{\partial d_i}(\alpha D_i(d_i) + (1-\alpha)D_i^{sh}(d_j)). \tag{11}$$

The first term on the right-hand side of (11) represents the negative impact of a higher disclosure d_i on firm i's demand $D_i(p_i)$: firm i's demand decreases since a higher disclosure level has a negative effect on the quality of service. The second term on the right-hand side of (11) represents the marginal disclosure revenue raised by firm i on the expected number of users over whom data is disclosed.

Note that dividing firm i's demand between SH and MH consumers ($D_i^{sh}(d_j)$ and $D^{mh}(d_i, d_j)$, respectively), we observe that as the incremental value of signing up to a second service (σ) increases, there are more (less) MH (SH) consumers. We observe the same pattern with consumer valuations (v). On the contrary, more product differentiation (i.e., a higher t) decreases (increases) the number of MH (SH) consumers. The intuition is that while a higher differentiation level makes multi-homing costlier, higher consumer intrinsic valuations (v), on top of a higher value of MH σ, encourage multi-homing. In addition, as MH user data is more valuable (i.e., a higher α), the number of MH (SH) consumers increases (decreases). Indeed, if MH user data is more valuable, firm i earns higher disclosure revenues and decreases its price (as seen in (10)), which encourages multi-homing.

Firm i's profit can be rewrited

$$\Pi_i(d_i) = t(D_i(d_i))^2 + (1-\alpha)d_i y_i^c D_i^{sh}(d_j)$$

Each firm sets its disclosure level d_i to maximize its profit, taking the rival disclosure level d_j as given. We then obtain equilibrium disclosure levels, d_A^c and d_B^c.

In the analysis that follows, let us write the following equation

$$\underbrace{d_i y_i^c \left(\frac{\partial D_i}{\partial d_i} + \frac{\alpha-1}{2t}\frac{\partial(d_i y_i^c)}{\partial d_i}\right)}_{A} + \underbrace{\frac{\partial(d_i y_i^c)}{\partial d_i} D^{mh}}_{B}. \tag{12}$$

In equation (12), term A represents the effect on disclosure revenues ($d_i y_i^c$) of the impact of a higher disclosure level d_i on (i) firm i's demand ($\partial D_i/\partial d_i$) and (i) disclosure revenues when firm i faces monetization issues on MH user data (i.e., if $(\alpha-1)/2t \times \partial(d_i y_i^c)/\partial d_i \leq 0$). Combining the two terms into brackets in term A, we obtain $\sigma v \partial q_i/\partial d_i + (2\alpha - 1)\partial(d_i y_i^c)/\partial d_i$.[16] Therefore, we deduce that if monetization issues are significant ($\alpha \leq 1/2$), the sign of A is negative ($A \leq 0$), whereas if monetization issues are small ($\alpha > 1/2$), the sign of A depends on the values of the parameters. Term B represents the effect of a higher disclosure level d_i on disclosure revenues raised on MH users.

The following proposition summarizes the results:

Proposition 3. *(i) In the duopoly equilibrium with a covered market, if $(\sigma v q^c + \alpha d^c y^c)/2 < t < \sigma v q^c + \alpha d^c y^c$, consumers provide information y^c. Firms' optimal prices and disclosure levels are $p^c = (\sigma v q^c - \alpha d^c y^c)/2$ and d^c.*

(ii) Firms' choice to charge positive disclosure levels depends on consumer valuations v, the value of MH σ, MH user data valuation α, and the level of product differentiation t.

[16]Term A thus becomes $\frac{d_i y_i^c}{2t}\left(\sigma v \frac{\partial q_i}{\partial d_i} + (2\alpha-1)\frac{\partial(d_i y_i^c)}{\partial d_i}\right)$.

(a) *If* $\alpha \leq 1/2$, $\partial d^c/\partial v \leq 0$, $\partial d^c/\partial \sigma \leq 0$, $\partial d^c/\partial t \geq 0$, *and*

$$\begin{cases} \partial d^c/\partial \alpha \leq 0 & \textit{if} \quad |A| \geq B, \\ \partial d^c/\partial \alpha > 0 & \textit{otherwise}. \end{cases}$$

(b) *If* $\alpha > 1/2$ *and* $A \leq 0$, *we find the same results as in (a);*

(c) *If* $\alpha > 1/2$ *and* $A > 0$, *we obtain a corner solution where* $d^c = 1$.

From Proposition 3, we observe different types of equilibria when we open up for the possibility of multi-homing. Firms' choice to charge disclosure levels are not only impacted by consumer valuations v as in the SH benchmark, but also by the value of MH σ (i.e., the incremental value of signing up to a second service), product differentiation t, and the value of MH user data α. In what follows, we examine the impact of each one of the parameters, given the others.

Proposition 3(ii)(a) describes a first scenario which occurs if the value of MH user data is low ($\alpha \leq 1/2$), i.e., if firm i faces significant monetization issues. Firm i engages in the disclosure of user data ($d^c > 0$) if consumer valuations v are sufficiently low, as in the SH benchmark ($\partial d^c/\partial v \leq 0$). In such a case, we find that as v increases, the price and the level of information provision increase ($\partial p^c/\partial v \leq 0$ and $\partial y^c/\partial v \leq 0$). Therefore, firm i engages in either a flexible or strict privacy regime.

The value of MH (σ) impacts firm i's disclosure level as well, and we find the same type of results as with consumer valuations v. A higher σ increases the value of consuming a second service and, in turn, consumer willingness to pay for the service. Firms thus tend to rely more on purchase revenues. Therefore, firm i engages in either a flexible or strict privacy regime.

We find that higher product differentiation t drives up firm i's disclosure level ($\partial d^c/\partial t \geq 0$). The intuition is that, for given disclosure levels, more product differentiation decreases the MH utility in (7), and therefore impact negatively the number of MH consumers ($\partial D^m/\partial t \leq 0$) (and hence increases the number of SH consumers). Firm i has more ability to monetize user data, which means that a higher t allows it to earn more disclosure revenues. In other words, a lower competitive intensity implies lower privacy ($\partial d^c/\partial t \geq 0$)

and prices ($\partial p^c/\partial t \leq 0$), and consumers provide less information ($\partial y^c/\partial t \leq 0$). Firms may even engage in maximal disclosure levels ($d^c = 1$) if t is sufficiently high, that is, a "zero-privacy" regime.[17]

Finally, firm i adjusts its disclosure level by considering the value of MH user data α. If the magnitude of the effect monetization issues on MH user data is stronger than the effect of higher marginal disclosure revenues raised on MH users, i.e., if $|A| \geq B$, firm i adjusts its disclosure level downward as α increases ($\partial d^c/\partial \alpha \leq 0$). However, if $|A| < B$, we find the opposite result ($\partial d^c/\partial \alpha > 0$). To determine the impact of a higher value of MH user data on its disclosure level, firm i trades-off the negative effect from monetization issues ($A < 0$ since $\alpha \leq 1/2$) and higher disclosure revenues raised on MH users at the margin. If $|A| \geq B$, firm i ultimately chooses to decrease its disclosure level when α increases. Then, there may exist a cutoff value $\tilde{\alpha} \in (0, 1/2]$ above which firm i engages in a strict-privacy regime ($d^c = 0$). If $|A| > B$, firm i increases its disclosure level with α, which implies that it may exists a cutoff value $\tilde{\alpha}' \in (0, 1/2]$ above which firm i may engage in a zero-privacy regime ($d^c = 1$).

Proposition 3(ii)(b) designates a second scenario which occurs if the value of MH user data is high ($\alpha > 1/2$), i.e., if firm i faces small monetization issues, and a higher disclosure level induces a strong decrease in consumer willingness to pay ($\sigma v|\partial q_i/\partial d_i| > (2\alpha - 1)\partial(d_i y_i^c)/\partial d_i$). We find the same variations as in Proposition 3(ii)(a). However, the second scenario is all the more likely to occur that monetization issues are not too small (even if $\alpha > 1/2$), and that σ and v are not too low, (i.e., the decrease in consumer willingness to pay is significant).

Moreover, the intuition behind the variations of firm i's disclosure level d^c with respect to α can be revisited. The case where $|A| \geq B$ should occur when the magnitude of monetization issues on MH user data and lower willingness to pay dominates the impact of higher disclosure revenues on MH users, so that firm i increases d^c when α increases ($\partial d^c/\partial \alpha \leq 0$). This is the contrary if $|A| < B$ and d^c increases with α ($\partial d^c/\partial \alpha > 0$).

Proposition 3(ii)(c) designates a third scenario which occurs if the value of MH user

[17]However, if t is sufficiently low, the firm engages in a strict-privacy regime.

data is high ($\alpha > 1/2$), i.e., if firm i faces small monetization issues, and a higher disclosure level induces a weak decrease in consumer willingness to pay ($A > 0$, or $\sigma v|\partial q_i/\partial d_i| > (2\alpha - 1)\partial(d_i y_i^c)/\partial d_i$). In such a case, we find a corner solution where firm i sets the highest disclosure level ($d^c = 1$),i.e., a zero-privacy regime. In fact, $A > 0$ implies that $\partial D_i/\partial d_i > 0$. This in turn means that the first-order conditions of firm i with respect to d_i are positive ($\partial \Pi_i/\partial d_i > 0$). As firm i's demand D_i perceives positively a higher disclosure level, firm i ultimately charges $d^c = 1$. The third scenario with a zero-privacy regime is likely to occur if monetization issues are not too significant (α closer to 1), and σ and v are not too high (i.e., the decrease in consumer willingness to pay is small).

While we examined the impact on price and disclosure levels of each parameter given the others, firms' incentives to engage in either privacy regime depends on the interaction between all the parameters (v, σ, t, α). All things being equal, we may therefore end up with extreme or intermediate outcomes in the presence of multi-homing compared to single-homing only.

We now study the impact of α on firm i's price p^c given by:

$$\frac{\partial p^c}{\partial \alpha} = \frac{1}{2}\left(\sigma v \frac{dq^c}{d\alpha} - \alpha \frac{d(d^c y^c)}{d\alpha} - d^c y^c\right), \tag{13}$$

and we write the followng corollary.

Corollary 3. *(i) If $\partial d^c/\partial \alpha \leq 0$, $\partial p^c/\partial \alpha \gtreqless 0$, whereas if $\partial d^c/\partial \alpha > 0$, $\partial p^c/\partial \alpha < 0$.*

(ii) There is price subzidization if for $d^c > 0$, consumers willingness to pay ($\sigma v q^c$) are lower than expected disclosure revenues per MH user ($\alpha d^c y^c$).

From Corollary 3(i), if the disclosure level d^c decreases as the value of MH user data α increases ($\partial d^c/\partial \alpha \leq 0$), the variation of firm i's price p^c with respect to α is unclear. On the one hand, a higher value of MH user data positively affects consumer willingness to pay since d^c decreases (first term on the right-hand side of (13)), thereby increasing firm i's price. Moreover, a higher α lowers disclosure revenues (second term on the right-hand side of (13)), which drives the price up. On the other hand, a higher α means that firm i

will suffer from fewer monetization issues on MH user data, which increases disclosure revenues and drives price p^c down (the third term on the right-hand side of (13)).

However, if the disclosure level d^c increases as the value of MH user data α increases ($\partial d^c/\partial\alpha > 0$), firm i reacts by decreasing its price ($\partial p^c/\partial\alpha < 0$). Indeed, a higher α lowers consumers willingness to pay (since d^c increases), and increases expected disclosure revenues per MH user ($d(\alpha d^c y^c)/dd^c \geq 0$). This is why firm i relies less on purchase revenues and decreases its price p^c.

From Corollary 3(ii), we show that firm i may subsidize its own consumers by trading-off between consumer willingness to pay and expected disclosure revenues per MH user. Consumer subsidization is possible as long as $\alpha > 0$. If, at an extreme, MH user data is valued $\alpha \to 0$, firm i would always charge a positive price. Figure 4 depicts firm i's pricing trade-off with respect to v. We observe that as consumer valuations increase, firm i relies progressively less on disclosure revenues, till the point where it no longer subsidizes consumers (i.e., $\sigma v q^c > \alpha d^c y^c$).

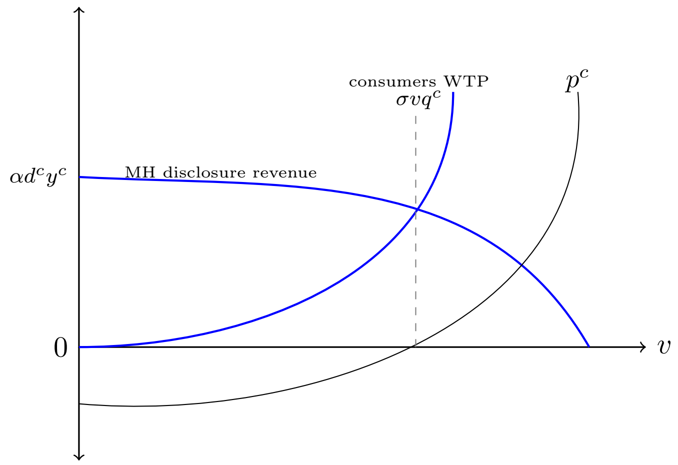

Figure 4: Pricing trade-off with MH

5.2 Monopoly

Let us study the equilibrium when A and B act as a monopoly, for instance, after a merger. We consider a multi-product monopoly, which supplies services A and B at prices p_A and p_B, with disclosure levels d_A and d_B, respectively.

At Stage 4, the level of information y_i is determined in the same way as the previous sub-section: for SH consumers, it is given by (2) while for MH consumers, it is given $y_i^m(d_i) \equiv \arg\max_{y_i} U_{ij}$. As before, MH and SH consumers provide the same level of information when using service i.

At Stage 3, the consumer chooses which service(s) to patronize. Proceeding in a similar way as in the previous sub-section, the demand functions of the monopoly for services A and B are given by equation (8).

At Stage 2, the monopolist determines the access price of user data charged to the data broker. We consider the possibility that the monopolist sells access to MH user data once rather than twice by incurring the cost c_D of data processing. It implies that it would sell access to the data of MH users to the left (right) of 1/2 at a disclosure level d_A (d_B).

We therefore write the following lemma.

Lemma 3. *(i) If $c_D \leq \underline{c}_D \equiv (1-\alpha)D^{mh}$, the data broker obtains access to both MH and SH user data at a price $r^m = v^b = 1$.*

(ii) If $c_D > \underline{c}_D$, the data broker obtains access to SH user data at a price $r^m = 1$, whereas she obtains access MH user data at a price $r^{m'} = \alpha \leq r^m$.

From Lemma 3(i), we observe that if the cost of data processing is sufficiently low ($c_D \leq \underline{c}_D$), the monopolist sells access to MH user data once at a price $r^m = 1$ to the data broker. This cost is all the more lower that the value of MH user data α increases ($\partial\underline{c}_D/\partial\alpha < 0$) and that the number of MH consumers is low ($\partial\underline{c}_D/\partial D^{mh} \geq 0$). Therefore, the monopolist opts for selling access to MH user data once if, for instance, with a high value of MH user data α and a small MH demand D^{mh}, it is cheaper (in terms of cost c_D) than selling access to MH user data twice. An economic intuition for a low cost c_D could be that the data from users of both services A and B have quite similar formats, and it is thus technically possible to split two distinct datasets with a cheap data mining technology.

From Lemma 3(ii), if the cost of data processing is too high ($c_D > \underline{c}_D$), the monopolist does not incur it and sells access to MH user data twice, at a price $r^{m'} = \alpha$. The cost of data processing c_D increases as the value of MH α decreases and the number of MH users D^{mh} increases. Therefore, the monopolist opts for selling access to MH user data twice rather than once if, for instance, MH user data has a low value α and MH demand D^{mh} is high (i.e., c_D is too high). An economic intuition for a high cost c_D may be that data from users of services A and B have specific formats, and there is either a technical issue complicating the handling of the dataset or the datamining technology needed is too expensive.

In what follows, we distinguish between two cases: (i) when the monopolist does not incur the cost of data processing, and (ii) when it incurs it.

5.2.1 The monopolist does not incur a cost of data processing ($c_D > \underline{c}_D$)

The monopoly profit is given by

$$\Pi^m(p_i, p_j) = \sum_{i=A,B} (p_i + \alpha d_i y_i^c) D_i(p_i) + d_i y_i^c (1-\alpha) D_i^{sh}(p_j). \tag{14}$$

The monopoly chooses p_A and p_B to maximize (14) by setting $\partial \Pi^m / \partial p_i = 0$. Solving for the first-order conditions, we obtain the equilibrium price

$$p_i^m(d_i, d_j) = \frac{\sigma v q_i - \alpha d_i y_i^c + (1-\alpha) d_j y_j^c}{2}, \quad i = A, B. \tag{15}$$

As in equation (10), we observe that the monopolist prices service i by trading-off between consumers willingness to pay ($\sigma v q_i$) and expected dislosure revenues per MH user ($\alpha d_i y_i^c$). On top of that, the firm now takes into account potential disclosure revenues that can be raised on users of service j: $p_i^m(d_i, d_j)$ depends (i) negatively on expected disclosure revenues per MH user ($-\alpha d_j y_j^c$) (i.e., those consumers signing up to service i in addition to service j), and (ii) positively on disclosure revenues per SH user of service j ($d_j y_j^c$), and the overall effect is positive ($(1-\alpha) d_j y_j^c$).

An implication is that for all d_i and d_j, $p^c(d_i) \leq p^m(d_i, d_j)$. It is a consequence from the monopolist who internalizes the impact of the pricing of service i on the consumption of service j: if it earns more disclosure revenues from MH users who signs up to service i in addition to service j, it decreases its price for service i; there is a similar impact if it earns higher disclosure revenues from MH users who sign up to service j in addition to service i. However, higher disclosure revenues raised on service j' SH users induce the firm to increase the price for service i.

The monopoly demand now writes,

$$D_i(d_i, d_j) = \frac{\sigma v q_i + \alpha d_i y_i^c - (1-\alpha) d_j y_j^c}{2t}, \quad D_i^{sh}(d_i, d_j) = 1 - D_i.$$

Plugging the equilibrium prices into the profit function (14), we now solve for the optimal disclosure levels of the monopoly at Stage 1. The monopoly can be written as

$$\Pi^m(d_i, d_j) = t\left((D_i(d_i, d_j))^2 + (D_j(d_i, d_j))^2\right) + (1-\alpha)(d_i y_i^c + d_j y_j^c).$$

The monopoly sets its disclosure levels to maximize its profit $\Pi^m(d_i, d_j)$. We obtain the equilibrium disclosure levels d_A^m and d_B^m.

The following proposition summarizes the analysis.

Proposition 4. *(i) In the monopoly equilibrium with a covered market, if* $(\sigma v q^m + (2\alpha - 1)d^m y^m)/2 < t < v q^m + (2\alpha - 1)d^m y^m$*, consumers provide information* y^m*. Optimal prices and disclosure levels are* $p^m = (\sigma v q^m - (2\alpha - 1)d^m y^m)/2$ *and* d^m*.*

(ii) The monopolist's choice to charge positive disclosure levels depend on consumer valuations v*, the value of MH* σ*, MH user data valuation* α*, and the level of product differentiation* t*.*

(a) If $\alpha \leq 1/2$*,* $\partial d^m/\partial v \leq 0$*,* $\partial d^m/\partial \sigma \leq 0$*,* $\partial d^m/\partial t \geq 0$*, and*

$$\begin{cases} \partial d^m/\partial \alpha \leq 0 & \text{if} \quad |A| \geq B \quad \text{in} \quad (12), \\ \partial d^m/\partial \alpha > 0 & \text{otherwise.} \end{cases}$$

(b) If $\alpha > 1/2$ *and* $A \leq 0$*, we find the same results as in (a);*

(c) If $\alpha > 1/2$ *and* $A > 0$*, we obtain a corner solution where* $d^m = 1$*.*

From Proposition 4, we observe different types of equilibria when we open up for the possibility of multi-homing. As in the duopoly, the monopolist's choice to charge disclosure levels depends on the value of MH σ, product differentiation t, the value of MH user data α, and consumer valuations v. We obtain qualitatively similar results and intuitions, as examined in Proposition 3.

We now analyze the variations of the monopolist's price p^m with respect to v, σ, t, and α, and the possibility of consumer subsidization.

Corollary 4. *(i) If $\alpha \leq 1/2$, the variations of the monopolist price p^m with respect to v, σ, t, and α are ambiguous, whereas if $\alpha > 1/2$, $\partial p^m/\partial v \geq 0$, $\partial p^m/\partial \sigma \geq 0$, $\partial p^m/\partial t \leq 0$, and $\partial p^m/\partial \alpha \gtreqless 0$.*

(ii) There is price subsidization if and only if $\alpha > 1/2$ and consumers willingness to pay ($\sigma v q^m$) are lower than expected disclosure revenues ($(2\alpha - 1)d^m y^m$).

From Corollary 4(i), we observe that the variations of the monopoly price p^m can be analyzed by distinguishing between two cases: either the value of MH user data is low ($\alpha \leq 1/2$) or it is high ($\alpha > 1/2$). Rewriting the monopoly price p^m, we have:

$$p^m = \frac{\sigma v q^m + d^m y^m - 2\alpha d^m y^m}{2}.$$

We see that the monopolist trades-off between consumer willingness to pay ($\sigma v q^m$), disclosure revenues per SH user ($1 \times d^m y^m$), and disclosure revenues per MH user ($\alpha(d^m y^m + d^m y^m)$), i.e., consumers purchasing service A in addition to service B, and those purchasing service B in addition to service A. We have seen in equation (15) that this trade-off comes from the monopolist which prices service i by internalizing the impact on the consumption of service j. The firm decreases its price for service i when it earns higher disclosure revenues from MH users, whereas it increases it when it earns higher disclosure revenues from service j's SH users.

If $\alpha \leq 1/2$, disclosure revenues per SH user (of service j) are comparatively higher than disclosure revenues per MH user: consumers willingness to pay ($\sigma v q^m$) and total disclosure revenues ($(2\alpha - 1)d^m y^m$) varies in opposite signs as d^m increases, which is why the varations of p^m are unclear. However, if $\alpha > 1/2$, disclosure revenues per MH user prevails: p^m increases with consumer valuations v and the value of MH σ, while it decreases with product differentiation t.

We now analyze how p^m varies with α and write the following equation:

$$\frac{\partial p^m}{\partial \alpha} = \frac{1}{2}\left(\sigma v \frac{dq^m}{d\alpha} - (2\alpha - 1)\frac{d(d^m y^m)}{d\alpha} - 2d^m y^m\right). \tag{16}$$

If $\partial d^m/\partial \alpha \leq 0$, there are three effects at work. First, p^m increases since a higher α lowers consumers willingness to pay (first term in (16)). Second, a higher α lowers disclosure revenues but the impact on price p^m depends on α: if $\alpha \leq 1/2$, the impact is negative, whereas if $\alpha > 1/2$, it is positive (second term in (16)). Third, a higher α entails that firms suffer from less monetization issues on MH user data, thereby impacting positively disclosure revenues and driving the price p^m down (third term in (16)). The sign of $\partial p^m/\partial \alpha$ is therefore unclear. However, if $\partial d^m/\partial \alpha > 0$ and $\alpha > 1/2$, p^m decreases as α increases ($\partial p^m/\partial \alpha > 0$); this is due the negative effect of α on consumer willingness to pay, the positive effect of α on disclosure revenues, and higher disclosure revenues (due to lower monetization issues).

From Corollary 4(ii), we find that a necessary condition for the monopolist to subsidize consumers is that the value of MH user data is sufficiently high ($\alpha > 1/2$). It then trades-off between consumer willingness to pay ($\sigma v q^m$) and disclosure revenues raised on its users ($(2\alpha - 1)d^m y^m$). In other words, if monetization issues on MH user data are too significant ($\alpha \leq 1/2$), the monopolist never subsidizes consumers.

5.2.2 The monopolist incurs the cost of data processing ($c_D > \bar{c}_D$)

For simplicity, we normalize c_D to zero.

The monopoly profit is given by

$$\Pi^m(p_i, p_j) = \sum_{i=A,B} p_i D_i(p_i) + \frac{1}{2} d_i y_i^m. \tag{17}$$

The monopoly chooses p_A and p_B to maximize (14) by setting $\partial \Pi^m/\partial p_i = 0$. Solving for the first-order conditions, we obtain the equilibrium price

$$p_i^m(d_i) = \frac{\sigma v q_i}{2}, \quad i = A, B.$$

Notice that here, the monopolist always charges a positive price ($p^m(d_i) > 0$) and this price is decreasing in d_i ($\partial p_i^m(d_i)/\partial d_i \leq 0$).

The monopoly demand now writes,

$$D_i(d_i, d_j) = \frac{\sigma v q_i}{2t}, \quad D_i^{sh}(d_i, d_j) = 1 - D_i.$$

Plugging the equilibrium prices into the profit function (14), we now solve for the optimal disclosure levels of the monopoly at Stage 1. The monopoly can be written as

$$\Pi^m(d_i, d_j) = t\left((D_i(d_i))^2 + (D_j(d_j))^2\right) + \frac{1}{2}(d_i y_i^c + d_j y_j^m)$$

The monopoly sets its disclosure levels to maximize its profit $\Pi^m(d_i, d_j)$. We obtain the equilibrium disclosure levels d_A^m and d_B^m.

The following proposition summarizes the analysis.

Proposition 4′. *(i) In the monopoly equilibrium with a covered market, if $\sigma v q^m/2 < t \leq v q^m$, consumers provide information y^m ($i = A, B$). Optimal prices and disclosure levels are $p^m = \sigma v q^m/2$ and d^m.*

(ii) The monopolist's choice to charge positive disclosure levels depends on consumer valuations v, the value of MH σ, and the level of product differentiation t, where $\partial d^m/\partial v \leq 0$, $\partial d^m/\partial \sigma \leq 0$, and $\partial d^m/\partial t \geq 0$.

From Proposition 4', we observe that when the monopolist incurs the cost c_D of data processing, the monopolist's choice to charge disclosure levels depend on consumer valuations v, product differentiation t, the value of MH σ. However, this choice no longer depends on the value of MH user data α because the monopolist now sells access to MH user data once to the data broker. Consumer valuations v and the incremental value of signing up to a second service σ, which drive up consumers willingness to pay, have a similar and intuitive impact on d^m: as v or σ increase, the firm decreases disclosure levels, increases prices, and consumers provide more information. More product differentiation t decreases the number of MH consumers and increases the number of SH consumers; the monopolist is therefore a bottleneck on a higher amount of user data. This is why as t increases, d^m increases, while price (p^m) and the level of information provisions (y^m) decreases.

Since the monopolist sells MH user data once by incurring the cost of data processing c_D, the pricing strategy no longer includes the possibility of consumer subsidization. Indeed, the incentives to subsidize consumers come from monetization issues on MH user data when it is sold twice, and the firm would have to trade-off between consumer willingness to pay and expected disclosure revenues. This trade-off no longer exists when the monopolist incurs c_D, prices charged to consumers are always positive.

6 Conclusion

In this paper, we investigate the business strategies of digital firms when they compete with privacy. We found that privacy is indeed an essential dimension of competition in digital markets. The level of privacy was measured by disclosure levels set by firms on user data, while consumers, who have privacy concerns on the disclosure of their personal information, choose the level of information to provide to the firm they patronized. We examined a framework with single-homing only and a second one where there is a possibility of multi-homing.

With SH only, firms tend to adopt two types of business models by arbitrating between disclosure levels, consumer information provision, and consumer valuations. If consumer valuations for the service are sufficiently high, firms adopt a strict privacy regime and rely exclusively on purchase revenues. However, if consumer valuations are low enough, firms adopt a flexible privacy regime. As consumer valuations increase, firms charge lower disclosure levels and consumers are willing to provide more information. Firms increasingly rely on disclosure revenues as consumer valuations decrease. We show that there exists consumer subsidization if consumer valuations are very low or if competition intensifies. Firms represent bottlenecks on the data from their users and charge the data broker a price that leaves her with zero surplus.

With the possibility of multi-homing, firms may face monetization issues on MH user data, on which they are no longer bottlenecks. Firms' business models are altered compared to SH. Indeed, firms' trade-off with respect to privacy is not only impacted by consumer valuations, but also by the value of purchasing a second service (i.e., the value of MH), the level of product differentiation, and the value of MH user data. Firms' adoption of privacy regimes thus depends on the interaction between these parameters. Three distinct scenarios emerge with the possibility of MH. The first one occurs when monetization issues on MH user data are important. A second scenario happens when monetization issues on MH user data are small and a higher disclosure level induces a sharp decrease in consumer willingness to pay. A third scenario exists where monetization issues are small and a higher disclosure level induces a slight decrease in consumer willingness to pay. In this case, firms engage in a zero-privacy regime by charging maximal disclosure levels. Finally, we show that firms choose to subsidize consumers by trading-off between consumer willingness to pay and expected disclosure revenues per MH user.

Our framework deals with horizontally differentiated firms where a symmetric equilibrium is obtained. A direction for further research would be to examine a double differentiation framework to gain more intuition on the impact of competition between digital players. Moreover, a more full-fledged modeling of the interaction with the data broker side would be insightful.

References

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016) "The economics of privacy," *Journal of Economic Literature*, Vol. 54, pp. 442–92.

Anderson, Simon P, Oystein Foros, and Hans Jarle Kind (2017) "Product functionality, competition, and multipurchasing," *International Economic Review*, Vol. 58, pp. 183–210.

——— (2019) "The importance of consumer multihoming (joint purchases) for market performance: Mergers and entry in media markets," *Journal of Economics & Management Strategy*, Vol. 28, pp. 125–137.

Anderson, Simon P and Joshua S Gans (2011) "Platform siphoning: Ad-avoidance and media content," *American Economic Journal: Microeconomics*, Vol. 3, pp. 1–34.

Argenziano, Rossella and Alessandro Bonatti (2020) "Information Revelation and Privacy Protection."

Armstrong, Mark (2006) "Competition in two-sided markets," *The RAND Journal of Economics*, Vol. 37, pp. 668–691.

Belleflamme, Paul, Wing Man Wynne Lam, and Wouter Vergote (2020) "Competitive imperfect price discrimination and market power," *Marketing Science*, Vol. 39, pp. 996–1015.

Belleflamme, Paul and Martin Peitz (2019) "Platform competition: Who benefits from multihoming?" *International Journal of Industrial Organization*, Vol. 64, pp. 1–26.

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

Calzolari, Giacomo and Alessandro Pavan (2006) "On the optimality of privacy in sequential contracting," *Journal of Economic theory*, Vol. 130, pp. 168–204.

Casadesus-Masanell, Ramon and Andres Hervas-Drane (2015) "Competing with privacy," *Management Science*, Vol. 61, pp. 229–246.

Choi, Jay Pil (2010) "Tying in two-sided markets with multi-homing," *The Journal of Industrial Economics*, Vol. 58, pp. 607–626.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim (2019) "Privacy and personal data collection with information externalities," *Journal of Public Economics*, Vol. 173, pp. 113–124.

De Corniere, Alexandre and Romain De Nijs (2016) "Online advertising and privacy," *The RAND Journal of Economics*, Vol. 47, pp. 48–72.

De Corniere, Alexandre and Greg Taylor (2020) "Data and competition: a general framework with applications to mergers, market structure, and privacy policy."

Hermalin, Benjamin E and Michael L Katz (2006) "Privacy, property rights and efficiency: The economics of privacy as secrecy," *Quantitative Marketing and Economics*, Vol. 4, pp. 209–239.

Hui, Kai-Lung and PL Ivan (2006) "Png. 2006," *The Economics of Privacy*, pp. 471–498.

Ichihashi, Shota (2020a) "Online privacy and information disclosure by consumers," *American Economic Review*, Vol. 110, pp. 569–95.

——— (2020b) "Non-competing data intermediaries,"Technical report, Bank of Canada.

Kim, Jin-Hyuk, Liad Wagman, and Abraham L Wickelgren (2019) "The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing," *Journal of Economics & Management Strategy*, Vol. 28, pp. 373–391.

Kim, Soo Jin (2020) "Privacy, Information Acquisition, and Market Competition," *Working Paper*.

Lefouili, Yassine and Ying Lei Toh (2020) "Privacy and quality," *TSE Working Paper 17 (795)*.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2018) "The value of personal information in online markets with endogenous privacy," *Management Science*.

Noam, Eli (1995a) "Privacy in telecommunications, part I," *New Telecommunications Quarterly*, pp. 52–59.

——— (1995b) "Privacy in telecommunications, part III," *New Telecommunications Quarterly*.

Posner, Richard A (1981) "The economics of privacy," *The American economic review*, Vol. 71, pp. 405–409.

Prat, Andrea and Tommaso M Valletti (2019) "Attention oligopoly," *Working paper*.

Prufer, Jens and Christoph Schottmuller (2017) "Competing with big data," *TILEC Discussion Paper No. 2017-006*.

Rochet, Jean-Charles and Jean Tirole (2003) "Platform competition in two-sided markets," *Journal of the european economic association*, Vol. 1, pp. 990–1029.

Shy, Oz and Rune Stenbacka (2016) "Customer privacy and competition," *Journal of Economics & Management Strategy*, Vol. 25, pp. 539–562.

Spulber, Daniel F (2009) "The map of commerce: Internet search, competition, and the circular flow of information," *Journal of Competition Law and Economics*, Vol. 5, pp. 633–682.

Stigler, George J (1980) "An introduction to privacy in economics and politics," *The Journal of Legal Studies*, Vol. 9, pp. 623–644.

Tag, Joacim (2009) "Paying to remove advertisements," *Information Economics and Policy*, Vol. 21, pp. 245–252.

Taylor, Curtis and Liad Wagman (2014) "Consumer privacy in oligopolistic markets: Winners, losers, and welfare," *International Journal of Industrial Organization*, Vol. 34, pp. 80–84.

Appendix

Appendix A: Proofs

Lemma 1

If $d_i y_i^c(d_i)$ is concave in d_i, it first means that

$$\frac{\partial^2(d_i y_i^c)}{\partial d_i^2}(d_i) = 2\frac{\partial y_i^c}{\partial d_i} + d_i \frac{\partial^2 y_i^c}{\partial d_i^2} \leq 0,$$
$$\text{and} \quad y_i^c(0) > y_i^c(1) + \left.\frac{\partial y_i^c}{\partial d_i}\right|_{d_i=1}.$$

Second, let $F(d_i) \equiv \partial(d_i y_i^c)/\partial d_i$. Assuming continuity and monoticity of $F(d_i)$ on $[0,1]$, by the intermediary values theorem, there exists a d_i such that $F(d_i) = 0$. Consequently, $d_i y_i^c$ is concave in d_i.

Proposition 1

We look for an interior solution and therefore restrict analysis to the case where $\partial(d_i y_i^c)/\partial d_i \geq 0$.

Note that

$$\frac{\partial D_i}{\partial d_i}(d_i) = \frac{1}{6t}\left(v\frac{\partial q_i}{\partial d_i} + \frac{\partial(d_i y_i^c)}{\partial d_i}\right),$$

The first-order conditions of firm i are given by

$$\frac{\partial \Pi_i}{\partial d_i} = 0 \Leftrightarrow 2t\left(2D_i \frac{\partial D_i}{\partial d_i}\right) = 0.$$

We look for an interior solution and check that the second-order conditions are satistied:

$$\frac{\partial^2 \Pi_i}{\partial d_i^2} = 4t\left(\left(\frac{\partial D_i}{\partial d_i}\right)^2 + D_i \frac{\partial^2 D_i}{\partial d_i^2}\right) \leq 0, \quad \text{by} \quad \text{Assumption 1}.$$

We know check the variations of d^c with respect to v.

$$\frac{\partial^2 \Pi_i}{\partial d_i \partial v}_{|\{d_i=d^c, d_j=d^c\}} = 4t \left(D_i \frac{\partial^2 D_i}{\partial d_i \partial v} + \underbrace{\frac{\partial D_i}{\partial v}}_{=0} \frac{\partial D_i}{\partial d_i} \right) \leq 0.$$

From the implicit function theorem, we obtain

$$\frac{\partial d_i}{\partial v}_{|\{d_i=d^c, d_j=d^c\}} = -\frac{\partial^2 \Pi_i / \partial d_i \partial v}{\partial^2 \Pi_i / \partial d_i^2} \leq 0.$$

It means that $d^c > 0$ if and only if consumer valuations are sufficiently low, i.e., $\lim_{v \to v^c} d^c(v) = 0$.

If $d^c > 0$, we have the following variations:

$$\frac{\partial y^c}{\partial v} = \frac{\partial y^c}{\partial d_i} \frac{\partial d^c}{\partial v} \geq 0; \quad \frac{\partial p^c}{\partial v} = -\left(\frac{\partial d^c}{\partial v} \left(y^c + \frac{\partial y^c}{\partial d_i} \right) \right) \geq 0; \quad \frac{\partial \left(d^c y^c \right)}{\partial v} \leq 0.$$

Note that the optimal disclosure level d^c does not depend on t:

$$\frac{\partial^2 \Pi_i}{\partial d_i \partial t}_{|\{d_i=d^c, d_j=d^c\}} = 0.$$

All consumers are served if the marginal consumer derives a positive utility in equilibrium, that is, if $vq(y^c, d^c) - p^c - tx^* \geq 0$, i.e., if $0 < t \leq vq(y^c, d^c) + d^c y^c/2$.

Proposition 2

Since we look for an interior solution, we restrict analysis to the case where $\partial(d_i y_i^m)/\partial d_i \geq 0$.

$$\frac{\partial D_i}{\partial d_i} = \frac{v}{2t} \frac{\partial q_i}{\partial d_i} + \frac{1}{4t} \frac{\partial (d_i y_i^m)}{\partial d_i}; \frac{\partial D_j}{\partial d_i} = -\frac{\partial D_i}{\partial d_i}.$$
$$\frac{\partial p_i}{\partial d_i} = \frac{v}{2} \frac{\partial q_i}{\partial d_i} - \frac{1}{4} \frac{\partial (d_i y_i^m)}{\partial d_i} < 0; \frac{\partial p_j}{\partial d_i} = t \frac{\partial D_i}{\partial d_i}; \frac{\partial^2 D_i}{\partial d_i \partial t} = -\frac{1}{t} \frac{\partial D_i}{\partial d_i}.$$

We determine the first-order conditions for the maximization of d_i:

$$\begin{aligned}
\frac{\partial \Pi^m}{\partial d_i} &= \left(\frac{\partial p_i}{\partial d_i} + \frac{\partial (d_i y_i^m)}{\partial d_i}\right) D_i + (p_i + d_i y_i^m)\frac{\partial D_i}{\partial d_i} + (p_j + d_j y_j)\frac{\partial D_j}{\partial d_i} + \frac{\partial p_j}{\partial d_i} D_j = 0 \\
&= \left(\frac{v}{2}\frac{\partial q_i}{\partial d_i} + \frac{1}{4}\frac{\partial (d_i y_i^m)}{\partial d_i} + \frac{1}{2}\frac{\partial (d_i y_i^m)}{\partial d_i}\right) D_i + (p_i + d_i y_i^m)\frac{\partial D_i}{\partial d_i} + t\frac{\partial D_i}{\partial d_i} D_j - (p_j + d_j y_j^m)\frac{\partial D_i}{\partial d_i} = 0 \\
&= \frac{\partial D_i}{\partial d_i}\left(t(D_i + D_j)\right) + \frac{1}{2}\frac{\partial (d_i y_i^m)}{\partial d_i} = 0.
\end{aligned}$$

Note that with a symmetric equilibrium, $(p_i + d_i y_i^m) - (p_j + d_j y_j^m) = 0$.

We now check that the second-order conditions are satisfied, that is

$$\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i^2} &= \frac{\partial^2 D_i}{\partial d_i^2}\left(t(D_i + D_j)\right) + \frac{1}{2}\frac{\partial^2 (d_i y_i^m)}{\partial d_i^2} + \frac{\partial D_i}{\partial d_i}\left(t\left(\frac{\partial D_i}{\partial d_i} - \frac{\partial D_i}{\partial d_i}\right)\right) \\
&= \frac{\partial^2 D_i}{\partial d_i^2}\left(t(D_i + D_j)\right) + \frac{1}{2}\frac{\partial^2 (d_i y_i^m)}{\partial d_i^2} \leq 0.
\end{aligned}$$

We now check the conditions under which d^c is positive.

$$\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i \partial v}_{|\{d_i = d^m, d_j = d^m\}} &= \frac{\partial^2 D_i}{\partial d_i \partial v}\left(t(D_i + D_j)\right) + \frac{\partial D_i}{\partial d_i}\left(t\left(\frac{\partial D_i}{\partial v} + \frac{\partial D_j}{\partial v}\right)\right) \\
&= \frac{\partial^2 D_i}{\partial d_i \partial v}\left(t(D_i + D_j)\right) \leq 0
\end{aligned}$$

We therefore use the implicit function theorem to find that

$$\frac{\partial d_i}{\partial v}_{|\{d_i = d^m, d_j = d^m\}} = -\frac{\partial^2 \Pi^m / \partial d_i \partial v}{\partial^2 \Pi^m / \partial d_i^2}_{|\{d_i = d^m, d_j = d^m\}} \leq 0.$$

It means that $d_i > 0$ if and only if consumer valuations are sufficiently low, i.e., $\lim_{v \to v^m} d^m(v) = 0$.

If $d^m > 0$, we then have that

$$\frac{\partial p^m}{\partial v} = q(y^m, d^m) + v\left(\frac{\partial q^m}{\partial y}\frac{\partial y^m}{\partial d}\frac{\partial d^m}{\partial v} + \frac{\partial q^m}{\partial d}\frac{\partial d^m}{\partial v}\right) \geq 0 \quad \frac{\partial y^m}{\partial v} \geq 0.$$

Note that the optimal disclosure level d^c does not depend on t:

$$\begin{aligned}\frac{\partial^2 \Pi^m}{\partial d_i \partial t}_{|\{d_i=d^m, d_j=d^m\}} &= \frac{\partial^2 D_i}{\partial d_i \partial t}\left(t(D_i+D_j)\right) + \frac{\partial D_i}{\partial d_i}\left(D_i + D_j + t\left(\frac{\partial D_i}{\partial t} + \frac{\partial D_j}{\partial t}\right)\right) \\ &= \frac{\partial D_i}{\partial d_i}\left(D_i + D_j - \frac{t}{t}(D_i+D_j)\right) = 0.\end{aligned}$$

Since we wish to restrict our attention to parameter values such that market is covered, we check under which conditions the monopoly chooses to cover the market. The monopolist covers the market if $\partial \Pi^u / \partial p_i|_{\{p_i=p_i^m, d_i=d_i^m\}} < 0$, which holds if $0 < t < vq(y^m, d^m) + d^m y^m$.

Lemma 2

The data broker gets access to SH user data from firm i if $r_i \leq v^b$ while she gets access to MH user data from both firms if $r_i \leq \alpha v^b$, if this data is sold twice.

Firm i is a bottleneck on SH user data and therefore charges the data broker $r_i = v^b = 1$. By contrast, if firms A and B compete for selling access to the information of MH users, the equilibrium access price of this data depends on how the data broker values it, that is, $r_i = \alpha$.

However, firms can strategically sell MH user data by selling access once. Firm A would sell the data of MH users to the left of 1/2 only whereas firm B would sell the data of MH users to the right of 1/2 only; by that, firms thus create bottlenecks on MH user data and $r_A = r_B = 1$). We examine if such outcome exists.

		Firm B: Cooperation	Firm B: No cooperation
Firm A	Cooperation	$(\frac{1}{2} - c_D, \frac{1}{2} - c_D)$	(D_A^{sh}, D_B)
	No cooperation	(D_A, D_B^{sh})	$(D_A^{sh} + \alpha D^{mh}, D_B^{sh} + \alpha D^{mh})$

Table 1: Data selling game

Depending on the situation, payoffs in Table 1 characterize the payments realized by the data broker to each firm when selling access to user data. In each box, the payment to the left of the brackets is received by firm A while the payment to the right of the brackets is received by firm B. If firm A cooperates with firm B, firm B is better off not cooperating since $D_B \geq 1/2$. The same applies if firm B cooperates with firm A. If firm A does not cooperate with firm B, firm B is better of not cooperating too ($D_B^{sh} < D_B^{sh} + \alpha D^{mh}$), and the same applies if firm B does not cooperate with firm A. Therefore, there is only one Nash Equilibrium in pure strategies, i.e., a non-cooperative outcome.Firms therefore charge the data broker $r_A = r_B = \alpha$ for access to MH user data.

Proposition 3

Since we look for interior solutions, we restrict analysis to the case where $\partial(d_i y_i^c)/\partial d_i \geq 0$.

The first-order conditions write as follows,

$$\frac{\partial \Pi_i}{\partial d_i} = 0 \Leftrightarrow 2tD_i(d_i)\frac{\partial D_i}{\partial d_i} + (1-\alpha)\frac{\partial(d_i y_i^c)}{\partial d_i} D_i^{sh}(d_j) = 0.$$

We check that the second-order conditions are satisfied:

$$\frac{\partial^2 \Pi_i}{\partial d_i^2} = 2t\left(\left(\frac{\partial D_i}{\partial d_i}\right)^2 + D_i \frac{\partial^2 D_i}{\partial d_i^2}\right) + (1-\alpha)\frac{\partial^2(d_i y_i^c)}{\partial d_i^2} D_i^{sh}(d_j) \leq 0$$

All consumers are served and there is multi-homing if the following set of conditions is satisfied:

$$\begin{aligned}
&U_i^c = vq(y^c, d^c) - p^c - tx^* \geq 0, \\
&U_{ij}^c = (1+\sigma)vq(y^c, d^c) - 2p^c - t \geq 0, \\
&0 < \underbrace{\frac{\sigma v q^c + \alpha d^c y^c}{t} - 1}_{D^m} < 1.
\end{aligned}$$

The above conditions are satisfied if $(\sigma v q^c + \alpha d^c y^c)/2 < t < \sigma v q^c + \alpha d^c y^c$.

We now check the variations of d^c. Let us inspect how d^c varies with consumer valuations v.

$$\frac{\partial D_i}{\partial v} = \frac{\sigma q_i}{2t} > 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial v} = \frac{\sigma}{2t}\frac{\partial q_i}{\partial d_i} \leq 0; \quad \frac{\partial D_i^{sh}}{\partial v} = -\frac{\sigma q_j}{2t} < 0; \quad \frac{\partial D_i}{\partial d_i} = \frac{1}{2t}\left(\sigma v \frac{\partial q_i}{\partial d_i} + \alpha \frac{\partial(d_i y_i^c)}{\partial d_i}\right).$$

At the symmetric equilibrium with $d_i = d_j = d^c$, $\partial D_i^{sh}/\partial v \equiv -\partial D_i/\partial v$.

$$\begin{aligned}
\frac{\partial^2 \Pi_i}{\partial d_i \partial v}_{|\{d_i=d^c,d_j=d^c\}} &= 2t\left(\frac{\partial D_i}{\partial v}\frac{\partial D_i}{\partial d_i} + D_i\frac{\partial^2 D_i}{\partial d_i \partial v}\right) + (1-\alpha)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}\frac{\partial D_i^{sh}}{\partial v}, \\
&= \frac{\partial D_i}{\partial v}\left(2t\frac{\partial D_i}{\partial d_i} - (1-\alpha)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}\right) + 2tD_i\frac{\partial^2 D_i}{\partial d_i \partial v} \\
&= \frac{\partial D_i}{\partial v}\left(\underbrace{\sigma v\frac{\partial q_i}{\partial d_i} + (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}}_{\gtreqless 0}\right) + \sigma D_i\underbrace{\frac{\partial q_i}{\partial d_i}}_{\leq 0}.
\end{aligned}$$

We observe that if $\alpha \leq 1/2$, the above equation is negative whereas the sign is unclear if $\alpha > 1/2$. In fact, we have

$$\frac{\partial^2 \Pi_i}{\partial d_i \partial v}_{|\{d_i=d^c,d_j=d^c\}} \begin{cases} \leq 0 & \text{if} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| \geq (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i} \quad \text{and} \quad \alpha > \frac{1}{2}, \\ \gtreqless 0 & \text{if} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| < (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i} \quad \text{and} \quad \alpha > \frac{1}{2}. \end{cases}$$

$$\begin{aligned}
&\text{However if} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| < (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i} \quad \text{and} \quad \alpha > \frac{1}{2} \\
&\text{Then} \quad \frac{\partial D_i}{\partial d_i} = \frac{1}{2t}\left(\sigma v\frac{\partial q_i}{\partial d_i} + \alpha\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}\right) > 0 \Rightarrow \frac{\partial \Pi_i}{\partial d_i} > 0 \Rightarrow d^c = 1,
\end{aligned}$$

which means that we obtain a corner solution with a maximal disclosure level $d^c = 1$. This scenario occurs under two conditions. First, MH user data has to be sufficiently valuable ($\alpha > 1/2$) so that the expected marginal disclosure revenue per MH user are high; this induces firm i to charge a lower price for the service. Second, the negative effect of disclosure on consumer willingness to pay has to be sufficiently low, i.e., consumers are not too sensitive to service quality degradation (lower privacy). It might happen if consumer valuations v and the value of MH σ are sufficiently low.

From the implicit function theorem, we therefore obtain

$$\left.\frac{\partial d_i}{\partial v}\right|_{\{d_i=d^c,d_j=d^c\}} = -\frac{\partial^2\Pi_i/\partial d_i\partial v}{\partial^2\Pi_i/\partial d_i^2} \leq 0 \quad \text{if} \begin{cases} \alpha \leq \frac{1}{2}, \\ \alpha > \frac{1}{2} \quad \text{and} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| > (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}. \end{cases}$$

If MH user data has a low valuation ($\alpha \leq 1/2$), the disclosure level d^c decreases with consumer valuations v and is equal to zero if v is sufficiently high. However, if MH user data has a high valuation ($\alpha > 1/2$), there are two possible outcomes: (i) if a higher disclosure level affects more negatively consumers willingness to pay than it affects positively firm i's expected disclosure revenue per MH user, d^c decreases with v. But (ii) if it is not the case, we find a corner solution where firm i charges $d^c = 1$.

We obtain similar results when studying how d^c varies with σ.

We now study how d^c varies with t.

$$\frac{\partial D_i}{\partial t} = -\frac{\sigma v q_i + \alpha d_i y_i^c}{2t^2} < 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial t} = -\frac{\sigma v \partial q_i/\partial d_i + \alpha\partial(d_i y_i^c)/\partial d_i}{2t^2}; \quad \frac{\partial D_i^{sh}}{\partial t} = \frac{\sigma v q_j + \alpha d_j y_j}{2t^2}.$$

At the symmetric equilibrium with $d_i = d_j = d^c$, $D_i \equiv D_j$.

$$\begin{aligned}
\frac{\partial^2\Pi_i}{\partial d_i\partial t}_{|\{d_i=d^c,d_j=d^c\}} &= 2D_i\frac{\partial D_i}{\partial d_i} + 2t\left(\frac{\partial D_i}{\partial t}\frac{\partial D_i}{\partial d_i} + D_i\frac{\partial^2 D_i}{\partial d_i\partial t}\right) + (1-\alpha)\frac{\partial(d_i y_i^c)}{\partial d_i}\frac{\partial D_i^{sh}}{\partial t} \\
&= 2D_i\frac{\partial D_i}{\partial d_i} + 2t(-\frac{D_i}{t}\frac{\partial D_i}{\partial d_i} - \frac{D_i}{t}\frac{\partial D_i}{\partial d_i}) + (1-\alpha)\frac{\partial(d_i y_i^c)}{\partial d_i}\frac{D_j}{t} \\
&= -2D_i\frac{\partial D_i}{\partial d_i} + (1-\alpha)\frac{\partial(d_i y_i^c)}{\partial d_i}\frac{D_j}{t} \\
&= D_i\left(\frac{1-\alpha}{t}\frac{\partial(d_i y_i^c)}{\partial d_i} - 2\frac{\partial D_i}{\partial d_i}\right) \\
&= -\frac{D_i}{t}\left(\sigma v\frac{\partial q_i}{\partial d_i} + (2\alpha-1)\frac{\partial(d_i y_i^c)}{\partial d_i}\right).
\end{aligned}$$

Using the implicit function theorem, we find that

$$\frac{\partial d_i}{\partial t}_{|\{d_i=d^c,d_j=d^c\}} \geq 0 \begin{cases} \text{if} \quad \alpha \leq \frac{1}{2}, \\ \text{if} \quad \alpha > \frac{1}{2} \quad \text{and} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| > (2\alpha-1)\frac{\partial(d_i y_i^c(d_i))}{\partial d_i}. \end{cases}$$

We now study how d^c varies with α.

$$\frac{\partial D_i}{\partial \alpha} = \frac{d_i y_i^c}{2t} > 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial \alpha} = \frac{\partial (d_i y_i^c)/\partial d_i}{2t} > 0; \quad \frac{\partial D_i^{sh}}{\partial \alpha} = -\frac{d_j y_j}{2t} < 0.$$

At the symmetric equilibrium with $d_i = d_j = d^c$, let $\partial D_i^{sh}/\partial \alpha \equiv -\partial D_i/\partial \alpha$.

$$\begin{aligned}
\frac{\partial^2 \Pi_i}{\partial d_i \partial \alpha}_{|\{d_i=d^c, d_j=d^c\}} &= 2t\left(\frac{\partial D_i}{\partial \alpha}\frac{\partial D_i}{\partial d_i} + D_i \frac{\partial^2 D_i}{\partial d_i \partial \alpha}\right) + \frac{\partial (d_i y_i^c)}{\partial d_i}\left((1-\alpha)\frac{\partial D_i^{sh}}{\partial \alpha} - D_i^{sh}\right) \\
&= \frac{\partial D_i}{\partial \alpha}\left(2t\frac{\partial D_i}{\partial d_i} - (1-\alpha)\frac{\partial (d_i y_i^c)}{\partial d_i}\right) + 2tD_i \frac{\partial^2 D_i}{\partial d_i \partial \alpha} - \frac{\partial (d_i y_i^c)}{\partial d_i} D_i^{sh} \\
&= \underbrace{\frac{\partial D_i}{\partial \alpha}\left(\sigma v \frac{\partial q_i}{\partial d_i} + (2\alpha - 1)\frac{\partial (d_i y_i^c)}{\partial d_i}\right)}_{A \gtreqless 0} + \underbrace{\frac{\partial (d_i y_i^c)}{\partial d_i}(\underbrace{D_i + D_j - 1}_{D^m})}_{B \geq 0}. \qquad \text{(A.1)}
\end{aligned}$$

Therefore,

$$\left.\frac{\partial d_i}{\partial \alpha}\right|_{\{d_i=d^c, d_j=d^c\}} \begin{cases} \leq 0 & \text{if} \quad |A| > B, \\ > 0 & \text{otherwise.} \end{cases}$$

Lemma 3

Let $r^m = 1 \geq r^{m'} = \alpha$. If the monopolist chooses to incur the cost of data processing c_D, the payment the monopolist receives from the data broker will be $r^m \times (D_A^{sh} + D_B^{sh} + D^{mh}) - c_D = 1 - c_D$. If it does not incur this cost, the payment received by the data broker would be $r^m(D_A^{sh} + D_B^{sh}) + r^{m'} D^{mh} = 1 - (1-\alpha)D^{mh}$.

Comparing both payments, we find that if $c_D \leq \underline{c}_D \equiv (1-\alpha)D^{mh}$, the monopolist sells access to MH user data once at a price $r^m = 1$, whereas if $c_D > \underline{c}_D$, it sells it twice at a price $r^{m'} = \alpha \leq r^m$.

Proposition 4

$$\frac{\partial D_i}{\partial d_i} = \frac{1}{2t}\left(\sigma v \frac{\partial q_i}{\partial d_i} + \alpha \frac{\partial (d_i y_i^m)}{\partial d_i}\right); \quad \frac{\partial D_j}{\partial d_i} = -\frac{(1-\alpha)}{2t}\frac{\partial (d_i y_i^m(d_i))}{\partial d_i}; \quad \frac{\partial^2 D_j}{\partial d_i^2} = -\frac{(1-\alpha)}{2t}\frac{\partial^2 (d_i y_i^m(d_i))}{\partial d_i^2};$$
$$\frac{\partial^2 D_i}{\partial d_i^2} = \frac{1}{2t}\left(\sigma v \frac{\partial^2 q_i}{\partial d_i^2} + \alpha \frac{\partial^2 (d_i y_i^m(d_i))}{\partial d_i^2}\right).$$

The first-order conditions write as follows,

$$\frac{\partial \Pi^m}{\partial d_i} = 0 \Leftrightarrow 2t\left(\frac{\partial D_i}{\partial d_i}D_i + \frac{\partial D_j}{\partial d_i}D_j\right) + (1-\alpha)\frac{\partial (d_i y_i^m)}{\partial d_i} = 0$$

We check that the second-order conditions for an interior solution are satisfied. With a symmetric equilibrium, we have $D_i \equiv D_j$.

$$\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i^2} =& 2t\left(\frac{\partial^2 D_i}{\partial (d_i)^2}D_i + \left(\frac{\partial D_i}{\partial d_i}\right)^2 + \frac{\partial^2 D_j}{\partial (d_i)^2}D_j + \left(\frac{\partial D_j}{\partial d_i}\right)^2\right) + (1-\alpha)\frac{\partial^2 (d_i y_i^m)}{\partial d_i^2} \\
=& 2t\left(D_i\left(\frac{\partial^2 D_i}{\partial (d_i)^2} + \frac{\partial^2 D_j}{\partial (d_i)^2}\right) + \left(\frac{\partial D_i}{\partial d_i}\right)^2 + \left(\frac{\partial D_j}{\partial d_i}\right)^2\right) + (1-\alpha)\frac{\partial^2 (d_i y_i^m)}{\partial d_i^2} \\
=& 2t\left(D_i\left(\frac{1}{2t}\left(\sigma v\frac{\partial^2 q_i}{\partial d_i^2} + (2\alpha-1)\frac{\partial^2 (d_i y_i^m(d_i))}{\partial d_i^2}\right)\right)^2 + \left(\frac{\partial D_i}{\partial d_i}\right)^2 + \left(\frac{\partial D_j}{\partial d_i}\right)^2\right) + (1-\alpha)\frac{\partial^2 (d_i y_i^m)}{\partial d_i^2}.
\end{aligned}$$

In this case, an interior solution for the symmetric equilibrium exists if we make the following assumption.

Assumption 2.

$$D_i\left|\frac{\partial^2 D_i}{\partial (d_i)^2}\right| + D_j\frac{\partial^2 D_j}{\partial (d_i)^2} \geq \left(\frac{\partial D_i}{\partial d_i}\right)^2 + \left(\frac{\partial D_j}{\partial d_i}\right)^2.$$

Note that Assumption 2 should hold if we have the following conditions.

$$\underbrace{D_i\frac{\partial^2 D_i}{\partial (d_i)^2}}_{\leq 0} + \underbrace{D_j\frac{\partial^2 D_j}{\partial (d_i)^2}}_{\geq 0} \leq 0 \quad \text{if} \quad \begin{cases} \alpha > \frac{1}{2}, \\ \alpha \leq \frac{1}{2} \quad \text{and} \quad D_i\left|\frac{\partial^2 D_i}{\partial (d_i)^2}\right| \quad \text{sufficiently high.} \end{cases}$$

It means that an interior solution require more conditions if $\alpha \leq 1/2$ compared to if $\alpha > 1/2$, and may be less likely to arise.

All consumers are served and there is multi-homing if the following set of conditions is satisfied:

$$
\begin{aligned}
&U_i^m = vq(y^m, d^m) - p^m - tx^* \geq 0,\\
&U_{ij}^m = (1+\sigma)vq(y^m, d^m) - 2p^m - t \geq 0,\\
&0 < \underbrace{\frac{\sigma v q^m + (2\alpha - 1)d^m y^m}{t} - 1}_{D^m} < 1.
\end{aligned}
$$

The above conditions are satisfied if $(\sigma v q^m + (2\alpha-1)d^m y^m)/2 < t < vq^m + (2\alpha-1)d^m y^m$.

We now check the variations of which d^m . Let us examine how d^m varies with consumer valuations v.

$$
\frac{\partial D_i}{\partial v} = \frac{\sigma q_i}{2t} > 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial v} = \frac{\sigma}{2t}\frac{\partial q_i}{\partial d_i} \leq 0, \quad \frac{\partial^2 D_j}{\partial d_i \partial v} = 0.
$$

At the symmetric equilibrium with $d_i = d_j = d^m$, let $\partial D_i/\partial v \equiv \partial D_j/\partial v$.

$$
\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i \partial v}_{|\{d_i=d^m, d_j=d^m\}} &= 2t\left(\frac{\partial^2 D_i}{\partial d_i \partial v}D_i + \frac{\partial D_i}{\partial d_i}\frac{\partial D_i}{\partial v} + \frac{\partial^2 D_j}{\partial d_i \partial v}D_j + \frac{\partial D_j}{\partial d_i}\frac{\partial D_j}{\partial v}\right)\\
&= 2t\left(\frac{\partial^2 D_i}{\partial d_i \partial v}D_i + \frac{\partial D_i}{\partial v}\left(\frac{\partial D_i}{\partial d_i} + \frac{\partial D_j}{\partial d_i}\right)\right)\\
&= \sigma\frac{\partial q_i}{\partial d_i}D_i + \frac{\partial D_i}{\partial v}\left(\sigma v\frac{\partial q_i}{\partial d_i} + (2\alpha-1)\frac{\partial(d_i y_i^m)}{\partial d_i}\right).
\end{aligned}
$$

From the implicit function theorem, we therefore obtain

$$
\left.\frac{\partial d_i}{\partial v}\right|_{\{d_i=d^m, d_j=d^m\}} = -\frac{\partial^2 \Pi^m/\partial d_i \partial v}{\partial^2 \Pi^m/\partial d_i^2} \quad \text{if} \begin{cases} \alpha \leq \frac{1}{2}, \\ \alpha > \frac{1}{2} \quad \text{and} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| > (2\alpha-1)\frac{\partial(d_i y_i^m)}{\partial d_i}. \end{cases}
$$

Note that

$$
\text{If} \quad \sigma v\left|\frac{\partial q_i}{\partial d_i}\right| < (2\alpha-1)\frac{\partial(d_i y_i^m)}{\partial d_i} \quad \text{and} \quad \alpha > \frac{1}{2} \quad \Rightarrow \frac{\partial \Pi^m}{\partial d_i} > 0 \quad \Rightarrow d^m = 1.
$$

We obtain a similar result when studying the variations of d^m with respect to σ.

We also study how d^m varies with t.

$$\frac{\partial D_i}{\partial t} = -\frac{\sigma v q_i + \alpha d_i y_i^m - (1-\alpha) d_j y_j^m}{2t^2} < 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial t} = -\frac{\sigma v \partial q_i/\partial d_i + \alpha \partial(d_i y_i^m)/\partial d_i}{2t^2};$$
$$\frac{\partial^2 D_j}{\partial d_i \partial t} = \frac{(1-\alpha)}{2t^2}\frac{\partial(d_i y_i^m)}{\partial d_i} > 0.$$

At the symmetric equilibrium with $d_i = d_j = d^m$, let $D_i \equiv D_j$.

$$\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i \partial t}_{|\{d_i=d^m, d_j=d^m\}} =& 2\left(\frac{\partial D_i}{\partial d_i} D_i + \frac{\partial D_j}{\partial d_i} D_j\right) \\
&+2t\left(\frac{\partial^2 D_i}{\partial d_i \partial t} D_i + \frac{\partial D_i}{\partial d_i}\frac{\partial D_i}{\partial t} + \frac{\partial^2 D_j}{\partial d_i \partial t} D_j + \frac{\partial D_j}{\partial d_i}\frac{\partial D_j}{\partial t}\right) \\
=& 2\frac{\partial D_i}{\partial d_i} D_i + 2\frac{\partial D_j}{\partial d_i} D_j + 2t\left(-\frac{\partial D_i}{\partial d_i}\frac{D_i}{t} - \frac{\partial D_i}{\partial d_i}\frac{D_i}{t} - \frac{\partial D_j}{\partial d_i}\frac{D_j}{t} - \frac{\partial D_j}{\partial d_i}\frac{D_j}{t}\right) \\
=& -2\left(\frac{\partial D_i}{\partial d_i} D_i + \frac{\partial D_j}{\partial d_i} D_j\right) \\
=& -\frac{D_i}{t}\left(\sigma v \frac{\partial q_i}{\partial d_i} + (2\alpha - 1)\frac{\partial d_i y_i^m}{\partial d_i}\right).
\end{aligned}$$

Using the implicit function theorem, we find that

$$\left.\frac{\partial d_i}{\partial t}\right|_{\{d_i=d^m, d_j=d^m\}} \geq 0 \begin{cases} \text{if} \quad \alpha \leq \frac{1}{2}, \\ \text{if} \quad \alpha > \frac{1}{2} \quad \text{and} \quad \sigma v \left|\frac{\partial q_i}{\partial d_i}\right| < (2\alpha - 1)\frac{\partial(d_i y_i^m)}{\partial d_i}. \end{cases}$$

We finally study how d^m varies with α.

$$\frac{\partial D_i}{\partial \alpha} = \frac{\partial D_j}{\partial \alpha} = \frac{d_i y_i^m + d_j y_j^m}{2t} > 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial \alpha} = \frac{\partial(d_i y_i^m)/\partial d_i}{2t} = \frac{\partial^2 D_j}{\partial d_i \partial \alpha} \geq 0.$$

$$\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i \partial \alpha}_{|\{d_i=d^m, d_j=d^m\}} =& 2t\left(\frac{\partial^2 D_i}{\partial d_i \partial \alpha} D_i + \frac{\partial D_i}{\partial d_i}\frac{\partial D_i}{\partial \alpha} + \frac{\partial^2 D_j}{\partial d_i \partial \alpha} D_j + \frac{\partial D_j}{\partial d_i}\frac{\partial D_j}{\partial \alpha}\right) - \frac{\partial(d_i y_i^m)}{\partial d_i} \\
=& 2t\left(\frac{\partial D_i}{\partial \alpha}\left(\frac{\partial D_i}{\partial d_i} + \frac{\partial D_j}{\partial d_i}\right)\right) + 2t\frac{\partial^2 D_i}{\partial d_i \partial \alpha}(D_i + D_j) - \frac{\partial(d_i y_i^m)}{\partial d_i} \\
=& \frac{\partial D_i}{\partial \alpha}\left(\sigma v \frac{\partial q_i}{\partial d_i} + (2\alpha - 1)\frac{\partial(d_i y_i^m)}{\partial d_i}\right) + \frac{\partial(d_i y_i^m)}{\partial d_i}(\underbrace{D_i + D_j - 1}_{D^m}).
\end{aligned} \tag{A.2}$$

Therefore,

$$\frac{\partial d_i}{\partial \alpha}\bigg|_{\{d_i=d^m, d_j=d^m\}} \begin{cases} \leq 0 & \text{if} \quad |A| > B, \\ > 0 & \text{otherwise.} \end{cases}$$

Proposition 4'

$$\frac{\partial D_i}{\partial v} = \frac{\sigma q_i}{2t} \geq 0; \quad \frac{\partial D_i}{\partial d_i} = \frac{\sigma v}{2t}\frac{\partial q_i}{\partial d_i} \leq 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial v} = \frac{\sigma}{2t}\frac{\partial q_i}{\partial d_i} \leq 0.$$

The first-order conditions write as follows,

$$\frac{\partial \Pi^m}{\partial d_i} = 0 \Leftrightarrow 2t\frac{\partial D_i}{\partial d_i}D_i + \frac{1}{2}\frac{\partial(d_i y_i^m)}{\partial d_i} = 0$$

We check that the second-order conditions for an interior solution are satisfied:

$$\frac{\partial^2 \Pi^m}{\partial d_i^2} = 2t\left(\frac{\partial^2 D_i}{\partial (d_i)^2}D_i + \left(\frac{\partial D_i}{\partial d_i}\right)^2\right) + \frac{1}{2}\frac{\partial^2(d_i y_i^m)}{\partial d_i^2} \leq 0.$$

We now check how d^m varies with v.

$$\frac{\partial^2 \Pi^m}{\partial d_i \partial v}_{|\{d_i=d^m, d_j=d^m\}} = 2t\left(\frac{\partial D_i}{\partial v}\frac{\partial D_i}{\partial d_i} + D_i\frac{\partial^2 D_i}{\partial d_i \partial v}\right) \leq 0$$

From the implicit function theorem, we therefore obtain

$$\frac{\partial d_i}{\partial v}_{|\{d_i=d^m, d_j=d^m\}} = -\frac{\partial^2 \Pi^m/\partial d_i \partial v}{\partial^2 \Pi_i/\partial d_i^2} \leq 0.$$

We obtain a similar result when looking at how d^m varies with σ.

We also check how d^m varies with t.

$$\frac{\partial D_i}{\partial t} = -\frac{\sigma v q_i}{2t^2} = -\frac{D_i}{t} < 0; \quad \frac{\partial^2 D_i}{\partial d_i \partial t} = -\frac{\sigma v}{2t^2}\frac{\partial q_i}{\partial d_i} = -\frac{1}{t}\frac{\partial D_i}{\partial d_i} \geq 0.$$

$$
\begin{aligned}
\frac{\partial^2 \Pi^m}{\partial d_i \partial t}_{|\{d_i=d^m, d_j=d^m\}} &= 2\frac{\partial D_i}{\partial d_i} D_i + 2t\left(\frac{\partial D_i}{\partial t}\frac{\partial D_i}{\partial d_i} + D_i \frac{\partial^2 D_i}{\partial d_i \partial t}\right) \\
&= 2\frac{\partial D_i}{\partial d_i}\left(D_i + t\frac{\partial D_i}{\partial t}\right) + 2tD_i \frac{\partial^2 D_i}{\partial d_i \partial t} \\
&= 2tD_i \frac{\partial^2 D_i}{\partial d_i \partial t} \geq 0.
\end{aligned}
$$

Using the implicit function theorem, we find that

$$
\left.\frac{\partial d_i}{\partial t}\right|_{\{d_i=d^m, d_j=d^m\}} \geq 0.
$$

All consumers are served and there is multi-homing if the following set of conditions is satisfied:

$$
\begin{aligned}
&U_i^m = vq(y^m, d^m) - p^m - tx^* \geq 0, \\
&U_{ij}^m = (1+\sigma)vq(y^m, d^m) - 2p^m - t \geq 0, \\
&0 < \underbrace{\frac{\sigma v q^m}{t} - 1}_{D^m} < 1.
\end{aligned}
$$

The above conditions are satisfied if $\sigma v q^m/2 < t < vq^m$.

Appendix B: numerical example

We solve their model in a horizontal differentiation framework using the utility function of Casadesus-Masanell and Hervas-Drane (2015).

The net utility of the consumer, located at $x \in [0,1]$, when purchasing service i, for given price p_i and information disclosure $d_i \in [0,1]$, is

$$
U_i = vy_i(1 - y_i - d_i) - p_i - tx,
$$

where $v > 0$ is the intrinsic benefit of the service, and transportation costs are equal to tx for firm A and to $t(1-x)$ for firm B.

Duopoly

Stage 4. A consumer chooses the level of information provision to the firm she has patronized, say firm i, by maximizing U_i with respect to y_i, which gives

$$y_i^c = \frac{1-d_i}{2}.$$

Stage 3. The consumer who is indifferent between purchasing service A and service B is given by

$$x^* = \frac{1}{2} - \frac{p_A - p_B}{2t} + \frac{v}{8t}\left((1-d_A)^2 - (1-d_B)^2\right),$$

where $D_A(p_A, p_B) = x^*$ and $D_B(p_A, p_B) = 1 - x^*$.

Stage 2. The selling price of consumer information is the monopoly price, that is, $p_i^b = 1$. Firm i chooses the price p_i that maximizes its profit, given the rival's price p_j. It then solves

$$\max_{p_i} \Pi_i(p_A, p_B), \quad \text{where} \quad \Pi_i(p_A, p_B) = (p_i + d_i y_i^c(d_i))D_i(p_A, p_B).$$

Firm i's equilibrium price is $p_i^c(d_A, d_B)$ (the SOCs are satisfied: $\partial^2\Pi_i/\partial p_i^2 = -1/t < 0$).

Stage 1. Firm i chooses the disclosure level d_i that maximizes its profit, given the rival's disclosure level d_j. Plugging the equilibrium prices $p_A^c(d_A, d_B)$ and $p_B^c(d_A, d_B)$ into the profit function, firm i maximizes $\Pi_i(d_A, d_B)$ with respect to d_i. We find that the equilibrium disclosure level (such that the second order conditions are satisfied) are symmetric and given by $d_i^c = (1-v)/(2-v)$.

To ensure that all consumers are served, *i.e.*, $U_i \geq 0$, the following conditions should hold: (i) $0 < v < 1$ and $0 < t \leq 1/6(2-v)$ or (ii) $v \geq 1$ and $0 < t \leq v/6$.

Result 1. *In the duopoly equilibrium, the market is covered and consumers provide information*

$$y^c = \frac{1}{(1+d^c)^2}.$$

(i) If $0 < v < 1$ *and* $0 < t \leq 1/6(2-v)$, *firms' optimal prices and disclosure levels are*

$$p^c = \frac{v-1}{2(v-2)^2} + t, \quad d^c = \frac{1-v}{2-v}.$$

(ii) If $v \geq 1$ *and* $0 < t \leq v/6$, *firms' optimal prices and disclosure levels are*

$$p^c = t, \quad d^c = 0.$$

Monopoly

Stage 4. We obtain the same information provision function at Stage 4, as in the duopoly.

Stage 3. If the monopolist covers the market, the indifferent consumer is defined in the same way as in the duopoly. If the monopoly does not cover the market, the monopoly demand for service i is (i.e., the consumer who is indifferent between purchasing service i and staying out of the market)

$$D_i^u(p_i) = \frac{v(1-d_i)^2}{4t} - \frac{p_i}{t}.$$

Stage 2. From the profit maximization of the covered monopoly, we get $p_i^m(d_A, d_B)$. For the uncovered monopoly, we obtain

$$p_i^u(d_i) = \frac{(1-d_i)(v(1-d_i)-2d_i)}{8}.$$

Stage 1. Solving for the equilibrium disclosure levels of the covered and uncovered monopolies, we find that $d_i^m = d_i^u = (1-v)/(2-v)$.

To determine the range of values for which the monopolist covers the market, we solve $\partial\Pi^u/\partial p_i|_{\{p_i=p_i^m, d_i=d_i^m\}} < 0$. The monopolist covers the market if (i) $0 < v < 1$ and $t < 1/4(2-v)$, or if (ii) $v \geq 1$ and $t < v$.

Result 2. *The monopolist chooses to cover the market in equilibrium and consumers* *buying service i provide information*

$$y^m = \frac{1}{(1+d^m)^2}$$

(i) If $0 < v < 1$ and $0 < t < 1/4(2-v)$, the monopoly prices and disclosure levels for services A and B are

$$p^m = \frac{v}{4(2-v)^2} - \frac{t}{2}, \quad d^m = \frac{1-v}{2-v}.$$

(ii) If $v \geq 1$ and $0 < t < v$, the monopoly prices and disclosure levels for services A and B are

$$p^m = v - \frac{t}{2}, \quad d^m = 0.$$