CESifo WORKING PAPERS

9339
2021

October 2021

Competition and Mergers with Strategic Data Intermediaries

David Bounie, Antoine Dubus, Patrick Waelbroeck

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9339*

Competition and Mergers with Strategic Data Intermediaries

Abstract

We analyze competition between data intermediaries collecting information on consumers, which they sell to firms for price discrimination purposes. We show that competition between data intermediaries benefits consumers by increasing competition between firms, and by reducing the amount of consumer data collected. We argue that merger policy guidelines should investigate the effect of the data strategies of large intermediaries on competition and consumer surplus in related markets.

JEL-Codes: L130, L400, L860.

Keywords: data, mergers, competition, consumer surplus.

David Bounie
Telecom Paris
Institut Polytechnique de Paris
19 Place Marguerite Perey
France – 91120 Palaiseau
david.bounie@telecom-paris.fr

Antoine Dubus
Department of Management, Technology and Economics, ETH Zurich
Leonhardstrasse 21
Switzerland – 8092 Zurich
antoine1dubus@gmail.com

Patrick Waelbroeck
Telecom Paris
Institut Polytechnique de Paris
19 Place Marguerite Perey
France – 91120 Palaiseau
patrick.waelbroeck@telecom-paris.fr

September 6, 2021
We thank Andreea Cosnita-Langlais, Doh-Shin Jeon, Ulrich Laitenberger, Patrick Legros, Winston Maxwell, as well as participants at the Annual Scientific Seminar on Media and the Dig-ital Economy of the Florence School of Regulation, the CESifo area conference on the Economics of Digitization, and the Paris Seminar on Digital Economics for useful remarks and comments. Patrick Waelbroeck thanks for insightful discussions the members of the Chair Values and Poli-cies of Personal Information of Institut Mines Télécom, Paris. Portions of the work on this paper were done while as Antoine Dubus was receiving financial support from the FNRS Grant PDR T.01.47.19. This research has also been conducted within the Chair “Digital Finance” under the aegis of the Risk Foundation, a joint initiative by Groupement des Cartes Bancaires CB, La Banque Postale, la Caisse des Dépôts et Consignations, Telecom Paris and University of Paris 2 Panthéon - Assas.

1 Introduction

Data is the building block of the digital economy (Hagiu and Wright, 2020). Information technologies enable companies to collect, store, and treat large amounts of data that they combine with artificial intelligence algorithms to deliver more efficient services to customers. Greater access to and use of data provide a significant advantage over market competitors (DalleMule and Davenport, 2017), but also create a wide array of impacts and policy challenges, ranging from privacy and consumer protection to open access issues and market dominance by large digital companies (Crémer et al., 2019).

To strengthen their dominant positions, digital companies such as Google, Amazon, Facebook and Microsoft have initiated an important wave of mergers and acquisitions, including WhatsApp, LinkedIn, FitBit, Skype or Nest (De Loecker et al., 2020). The UK Competition and Market Authority reports that "over the last 10 years the 5 largest firms have made over 400 acquisitions globally. None has been blocked and very few have had conditions attached to approval, in the UK or elsewhere, or even been scrutinised by competition authorities (CMA, 2020)." Recently, economists have questioned this lax approach to mergers in the digital economy that can potentially reduce competition and innovation (Furman et al., 2019; Scott Morton et al., 2019; Crémer et al., 2019; Parker et al., 2021).

Another concern is that mergers and acquisitions not only reshape the competitiveness of markets on which big tech companies initially operate – Facebook on social networking for instance–, they have also indirect effects in related product markets (Tirole, 2020). The main reason is that companies such as Google, Amazon, and Facebook act as data intermediaries on a new market for information (Bergemann and Bonatti, 2019). They sell information to firms such as retailers, banks or insurance companies, to improve their business practices through personalized recommendations, products, and prices (Varian, 1989).[1] For instance, banks purchase consumer information from data intermediaries such as Equifax,

[1]According to Accenture, data marketplaces will be the leading place to exchange information among data-brokers and information buyers in the coming years (The dawn of the data marketplace, Accenture, last accessed May 26, 2021).

which they use to compete in the credit market.[2]

Mergers and acquisitions in the market for information have indirect effects on related product markets through two strategic dimensions of information. First, data intermediaries can influence the intensity of competition among firms active in the product market. More information sold to firms means that they will fight more fiercely for consumers that they have identified as belonging to their business segments. There is thus a competitive effect of information that lowers the profits of the firms. A monopolist data intermediary can strategically choose to withhold information from firms to minimize the competitive effect of information (Bounie et al., 2021). Competing data intermediaries may be forced to sell more information to firms, which intensifies competition in the product market and increases consumer surplus. Therefore, mergers that reduce competition in the market for information can soften competition in product markets and harm consumers.

Secondly, collecting data is costly for intermediaries, but provides more information on consumers, allowing firms to increase their profits through better consumer surplus extraction. This rent extraction effect of information increases the willingness to pay of firms, which also increases the incentives of data intermediaries to collect consumer data. Competition in the market for information determines the price of information and the incentives of data intermediaries to collect consumer data. When data intermediaries compete, the price of information is lower than in a monopoly situation, which changes the amount of consumer data that they collect. Mergers between data intermediaries can reduce consumer surplus through this second indirect effect of information.

The novelty of this paper is to analyze how competition between data intermediaries impacts consumer surplus and competition in related product markets through these indirect effects of information, and to provide merger policy recommendations that can be used to revisit recent flagship mergers.

We build a model of competition between data intermediaries collecting and selling information to firms seeking to price-discriminate consumers in a related

[2] Data brokers: regulators try to rein in the 'privacy deathstars', Financial Times, January 8, 2019.

product market. Data intermediaries collect information that partitions consumer demand into segments of different sizes: collecting more information reduces the size of the segments and allows intermediaries to better identify consumers. We introduce two types of markets that leave some market power to each data intermediary, allowing them to collect consumer data.[3] Each intermediary has a local monopoly market and has exclusive access to consumer data on this market. Intermediaries also collect data on a competitive market where consumer information is available to other intermediaries. Data intermediaries compete by selling information to firms on this market. This framework allows us to analyze how mergers that change the incentives of intermediaries to collect data on each market may reduce consumer surplus.

Using this framework, this article achieves four main results regarding how competition in the market for information impacts consumer surplus. First, competition between data intermediaries lowers their ability to internalize the competitive effect of information. We show that when data intermediaries compete, they sell more consumer segments to firms than under monopoly, which intensifies competition in the product market. Nevertheless, even under fierce competition, intermediaries do not sell all the available consumer segments in order to soften competition between information buyers. Hence, we find that this strategic effect of information does not go away when intermediaries compete, confirming results by Bounie et al. (2021).

Secondly, we find that a monopolist data intermediary strategically limits the number of firms that can purchase information. Conversely, competing data intermediaries sell information to all firms, which allows equal access to information. These results contribute to the burgeoning literature that only considers the selling strategies of a monopolist data intermediary who sells information to only one of the information buyers (Montes et al., 2018; Bounie et al., 2021), and to models where it is assumed that a data intermediary sells information to all active

[3] Diamond (1971) has shown that a perfectly competitive market does not allow intermediaries to invest in costly consumer data collection. This paradox has raised conceptual problems to model the market for information (see for instance Grossman and Stiglitz (1980) for financial information). Our framework solves this issue by providing each intermediary with market power.

firms (Bergemann and Bonatti, 2015; Bergemann et al., 2018, 2019). We challenge these assumptions and results by showing that competing data intermediaries optimally sell information to all firms, while a monopolist data intermediary sells information to only one firm.[4]

Thirdly, competition between data intermediaries reduces the profitability of information, and lowers the amount of consumer information that they collect, decreasing the ability of firms in the product market to extract consumer surplus. This result is new and goes beyond the existing academic literature that does not consider the data collection strategies of competing intermediaries. Accounting for this strategic dimension is important for merger analysis since a merged entity will have more incentives to collect data than two separate firms, thus reducing consumer surplus.

Fourthly, we challenge existing merger policy guidelines by showing how mergers can reduce consumer surplus by changing data collection and selling strategies of data intermediaries. A merged entity benefits from cost efficiencies, and can leverage on a larger market size to collect more data. Our results suggest that mergers between data intermediaries are detrimental to consumers as they increase the amount of data collected and lower consumer surplus. Therefore, we argue that our analysis of merger impacts using indirect effects of information in related product markets should be included in the analysis of mergers and acquisitions.

The remainder of this article is organized as follows. In Section 2 we describe the model. We analyze how competition between data intermediaries changes the number of consumer segments they sell to firms in Section 3, and how competition changes the number of consumer segments they collect in Section 4. We analyze the impact of competition between intermediaries on consumer surplus in Section 5. We apply our framework to recent mergers in Section 6. Section 7 concludes.

[4]See Bergemann and Bonatti (2019) for a recent review.). See also Chen et al. (2020) for an analysis of data-driven mergers in the context of platforms.

2 Description of the model

We build a model of competition between data intermediaries that collect and sell customer data for price discrimination purposes. We describe in this section consumer utility from purchasing a product, the data collection and selling strategies of competing data intermediaries, the incentives of firms to purchase consumer data, and finally, the timing of the game.

2.1 Consumers

Consumers are uniformly distributed on a unit line $[0, 1]$, and they can buy one product at a price p_1 from Firm 1 located at 0, or p_2 from Firm 2 located at 1.[5] Since firms can price discriminate when they have information, different consumers may pay different prices. A consumer located at $x \in [0, 1]$ derives a utility V from purchasing the product. He incurs a transportation cost $t > 0$ so that buying from Firm 1 (resp. from Firm 2), has a total cost tx (resp. $t(1 - x)$). Consumers purchase the product for which they have the highest utility.

Consumers are divided into $n + 1$ mutually exclusive markets that we describe in detail in the next section. Each market is characterized by a Hotelling line of unit mass. On each unit line, consumer located at x has a utility function defined by:

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

2.2 Data intermediaries

Data intermediaries collect information that divides each market into consumer segments. More data is costly to collect but allows an intermediary to have a finer partition of the line. Partitions sold by data intermediaries enable firms to

[5]We assume that the market is covered. This assumption is common in the literature. See for instance Bounie et al. (2021) or Montes et al. (2018).

identify consumers and price discriminate them. We describe in this section the nature of competition between data intermediaries, and the way they collect and sell information on different markets.

2.2.1 Nature of competition between data intermediaries

Analyzing competition between data intermediaries raises a major challenge. Collecting data is costly, but reproducing information is almost costless (Shapiro et al., 1998; Varian, 2018). With high fixed costs and a low marginal cost of producing information, a perfectly competitive market for information with a price close to 0 is not sustainable, as firms would not be able to recover their fixed costs of collecting data. This is the well-known Diamond information paradox (Diamond, 1971).

To solve this issue, we assume that data intermediaries sell information on two types of markets: on the first type of markets, a data intermediary owns proprietary and rival information that competing intermediaries cannot access, which grants him a local monopoly power. For instance, Facebook collects data on its users, and other data intermediaries cannot sell this information in the product market. We denote each monopoly market by m_i. There is also a competitive market, where data intermediaries sell information that all intermediaries possess, and that are therefore non-rival (Jones and Tonetti, 2020). Indeed, Facebook also collects information on users who visit other platforms or online services such as the ones offered by Google; Facebook and Google have therefore similar information on these consumers. We will refer to this market as competitive market l.

Our $m-l$ approach addresses the Diamond paradox and explicitly draws the frontier between rival and non-rival data. We discuss in Section 4 how changes in the sizes of m and l increase or decrease consumer surplus.

We consider n competing data intermediaries that collect and sell consumer information to firms (with $n \geq 2$). Each data intermediary can collect information on a mass m_i of consumers who belong to their monopoly market (with $i = 1, .., n$), or on a market l where they compete with other data intermediaries.[6] As a con-

[6] We analyze in Section 3.1 a situation in which data intermediaries only collect and sell information on their monopoly market.

sequence, each intermediary i has monopoly (rival) information on consumers in m_i, and common (non-rival) information on consumers in l. Consumers, therefore, either belong to a monopoly market or a competitive market, so that the total mass of consumers is $m_1 + ... + m_n + l$.[7] By convention, data intermediary 1 has a larger monopoly market than data intermediary 2 and so on: $m_1 \geq m_2 ... \geq m_n$.[8]

2.2.2 Collecting data

A data intermediary i uses an appropriate technology to collect data on M_i consumers.[9] The technology allows the intermediary to distinguish consumers who are exclusively using its services, and therefore belong to market m_i, and consumers who are also using other services and belong to market l ($M_i = m_i + l$). A data intermediary collects data points such as sex, age, or zip-code, which allows it to partition consumer demand into k segments of size $\frac{1}{k}$.

We illustrate the partition collected by a data intermediary in Figure 1. The k segments of size $\frac{1}{k}$ form a partition $\mathcal{P}_{ref}$ that we refer to as the reference partition.

Figure 1: Reference partition $\mathcal{P}_{ref}$

The number of consumer segments k corresponds to the precision of information, and a firm that has information can third-degree price-discriminate consumers by charging different prices on different segments. For instance, when $k = 2$, the partition is coarse, and firms can only distinguish whether consumers belong to $[0, \frac{1}{2}]$ or to $[\frac{1}{2}, 1]$. At the other extreme, when k converges to infinity, the

[7]We assume that $m_i > 0 \;\; \forall i$ and $l > 0$ in the remainder of the article. This framework has as special cases $l = 0$ and $m_i = 0$, that are analyzed in Section 3.

[8]As a special case, we will also allow for symmetric data intermediaries in terms of size of their monopoly markets: $m_1 = m_2 ... = m_n$; we will show that they collect different amounts of information in the only equilibrium of the game.

[9]There are several technologies to collect data on consumers such as cookies and pixels (Bergemann and Bonatti, 2015; Choe et al., 2018). Cookies used for analytics and research provide information for instance on all the websites visited by consumers and allow firms to know whether consumers have visited their own website or the competitors' websites.

data intermediary knows the exact location of each consumer (first-degree price discrimination).

This approach allows us to analyze varying levels of information precision and characterize the data collection strategies of data intermediaries.[10] We will show how competition between data intermediaries has an indirect effect on consumer surplus in the product market, by changing the amount of data collected, changing in turn the ability of firms to price discriminate consumers.

The cost of collecting data is equal to $c(k)$ for a consumer segment of size one. We assume that a data intermediary cannot distinguish consumers who belong to m_i and to l before collecting information. Therefore, it will collect the same amount of information on markets m_i and l.[11] Thus the total data collection cost is $(l + m_i)c(k)$. This cost encompasses various dimensions of the activity of data intermediaries, such as installing trackers or storing and handling data. Collecting more information by increasing the number of segments allows a firm to extract more surplus on consumers, increasing the willingness to pay for information and the price of information.

2.2.3 Selling information

Data intermediaries can sell any combination of segments of consumer demand, contrary to previous frameworks that assume that intermediaries sell all available information (Liu and Serfes, 2004; Montes et al., 2018). Selling strategic information will allow us to capture the second indirect effect of competition in the market for information on the product market. Bounie et al. (2021) have shown that a data intermediary can weaken or strengthen the intensity of competition in the product market by determining the quantity of information available to firms, which has two effects on consumer surplus. On the one hand, an informed firm can price discriminate consumers, thus increasing its profits through this rent extraction effect. On the other hand, information also increases competition in the product market, which reduces the profits of both firms. An optimal parti-

[10] Our model encompasses the limit case where $k \to \infty$, i.e. information is perfect and firms can first-degree price-discriminate consumers.

[11] We drop index i when there is no confusion.

tion thus maximizes consumer surplus extraction while softening the competitive effect of information. We will see how competition between data intermediaries reduces this ability to soften the competitive effect of information, impacting in turn competition in the product market and consumer surplus.

To illustrate how strategic information changes the intensity of competition in the product market, we consider in Figure 2 a situation in which $k = 4$ segments are available. By allowing Firm 1 to distinguish consumers located close to Firm 2 and to charge them prices p_{13} and p_{14}, the data intermediary increases the competitive pressure on Firm 2 that lowers price p_2. Now suppose that the data intermediary only sells the first segment to Firm 1 that charges consumers price p'_{11}: the competitive pressure will be much lower, and Firm 2 will increase its price $p'_2 > p_2$. By keeping a share of consumers unidentified, the data intermediary will keep a low level of competition between firms, allowing Firm 1 to extract more surplus from identified consumers close to its location.[12]

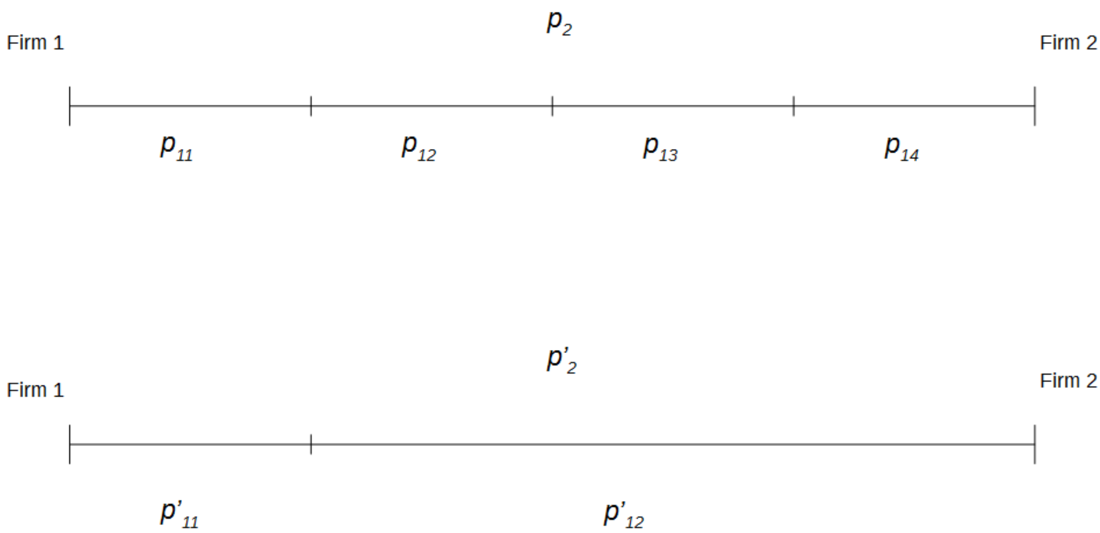

Figure 2: Example of partitions, $k = 4$

On a monopoly market m_i, data intermediary i sells information to one or two firms. We assume that data intermediaries sell information through first-price

[12]This result holds even if the collection cost is equal to zero and under first-degree price discrimination ($k \to \infty$) (see Bounie et al. (2021)).

auctions. This selling mechanism is commonly used in the literature (Montes et al., 2018; Bounie et al., 2021), and allows an intermediary to reach the first-best outcome by maximizing surplus extraction from firms.[13]

On the competitive market, data intermediaries compete à la Bertrand in the sale of information. Since firms only purchase information from one data intermediary, they will choose data intermediary $\overline{DI}$ with the best information precision, so that competition in market l leads to a winner takes all situation. Other data intermediaries will not be active in market l, but the second-best data intermediary $\underline{DI}$ exerts a competitive pressure on the equilibrium price set by $\overline{DI}$.

2.3 Firms

Without information, firms only know that consumers are uniformly distributed on the unit line. Firms can acquire information from the monopolist data intermediary i on the monopoly markets m_i and from one of the competing data intermediaries on the competitive market l.[14] When a firm acquires an information partition $\mathcal{P}^{\theta}$, it knows which interval of this partition a consumer belongs to. Firms simultaneously set prices on each segment of the unit line where they have information. Firm θ sets prices in two stages.[15] First, it sets prices on competitive segments where it shares consumer demand with its competitors. Then, on segments where it is a monopolist, it sets a monopoly price. Each firm knows whether its competitor is informed, and the partition $\mathcal{P}^{-\theta}$.[16]

[13]Auctions are frequently used by major data intermediaries such as Google (First-price Auction, Second-price, and the Header-Bidding, Smartyads, February 2018), and in data marketplaces (Sheehan and Yalif (2001); O'kelley and Pritchard (2009)).

[14]We assume that firms cannot combine information purchased from different data intermediaries.

[15]Sequential pricing decision avoids the nonexistence of Nash equilibrium in pure strategies and is common in the literature supported by managerial practices. For instance, Acquisti and Varian (2005) use sequential pricing to analyze intertemporal price discrimination with incomplete information on consumer demand. Jentzsch et al. (2013) and Belleflamme et al. (2020) also focus on sequential pricing where a higher personalized price is charged to identified consumers after a firm sets a uniform price. Sequential pricing is also common in business practices (see also, Fudenberg and Villas-Boas (2006)). Recently, Amazon has been accused to show higher prices for Amazon Prime subscribers, who pay an annual fee for unlimited shipping services, than for non-subscribers (Lawsuit alleges Amazon charges Prime members for "free" shipping, Consumer affairs, August 29 2017). Thus Amazon first sets a uniform price and then increases prices for high-consumers who are better identified when they join the Prime program.

[16]This assumption is also standard in Braulin and Valletti (2016) and Montes et al. (2018).

We denote by $d_{\theta i}$ the demand of Firm θ on the ith segment. An informed Firm θ maximizes the following profit function with respect to $p_{\theta 1}, .., p_{\theta n}$:

$$\pi_\theta = \sum_{i=1}^{n} d_{\theta i} p_{\theta i}. \tag{2}$$

2.4 Timing

Data intermediaries first collect data and sell partition $\mathcal{P}_m^\theta$ to Firm θ in market m, and partition $\mathcal{P}_l^\theta$ to Firm θ in market l. Then, firms set prices on segments where they compete. Finally, firms set prices on the monopolistic segments. The timing of the game is the following:

- Stage 1: data intermediaries collect data on k consumer segments.
- Stage 2: data intermediaries auction information partition $\mathcal{P}_m^\theta$ and $\mathcal{P}_l^\theta$.
- Stage 3: firms set prices p_1 and p_2 on the competitive segments of each sub market.
- Stage 4: informed firms set prices $p_{\theta i}$ on consumer segments on which they have information.

3 Selling consumer information

We characterize the optimal selling strategies of data intermediaries on their monopoly markets and on the competitive market.

To simplify the exposition, we consider a market of size 1. We will see that the selling strategy does not depend on the size of the market. However, the size of the monopoly market relative to the size of the competitive market will impact the data collection strategy of intermediaries. In Section 4, we explicitly deal with different market sizes.

3.1 Monopoly market

We characterize the price of information in a monopoly market. A data intermediary can sell information to one or two firms.

Consider first the situation where the data intermediary sells information to only one firm, say Firm 1. Let $\pi_1(k, \emptyset)$ and $\pi_2(k, \emptyset)$ be the respective profits of Firm 1 and Firm 2 when they acquire the reference partition $\mathcal{P}_k$ and their competitor has no information. Similarly, let $\pi_1(\emptyset, k)$ and $\pi_2(\emptyset, k)$ be their profits when they are uninformed but face a competitor that has acquired partition $\mathcal{P}_k$. The profits of an uninformed firm are minimized when its competitor has information $\mathcal{P}_k$. Thus, this partition represents the maximal level of threat for a firm that does not purchase information. The resulting price of information is given by the difference between the profits of Firm 1 with information and this maximal threat and is given in Equation 3. A reader uninterested in technical details can skip this part and go directly to Lemma 1

Simultaneous auctions. In order to maximize the price of information, the data intermediary designs two simultaneous auctions, and only the partition with the highest bid will be sold. We are looking for a pure strategy Nash equilibrium. Consider a given partition $\mathcal{P}_1$. We first characterize the price of information and then obtain the optimal partition.

Firm 1 with the highest willingness to pay knows the bid of Firm 2 and has interest to underbid from its true valuation. Thus, a firm can bid just above the willingness to pay of its competitor and win the auction, which reduces the price of information. To avoid underbidding by Firm 1, in auction 1 $\mathcal{P}_1$ is auctioned with a reserve price p_1^m.[17] The reference partition $\mathcal{P}_k$ that includes all k information segments is auctioned in auction 2, in order to exert a maximal threat on Firm 1 and to maximize its willingness to pay for $\mathcal{P}_1$. Participation of both firms is guaranteed as the data intermediary sets no reserve price in auction 2.

Consider the optimal strategies of Firm 1 and Firm 2. Firm 2 will bid $\pi_2(k, \emptyset) - \pi_2(\emptyset, k)$ in auction 2 that corresponds to its willingness to pay for partition $\mathcal{P}_k$, as its worst outside option is to face Firm 1 informed with k. However, Firm 2 will

[17] For instance, Coey et al. (2021) analyze the role of reserve prices in repeated online auctions.

never bid above the reserve price for $\mathcal{P}_1$. Consider now the optimal strategy of Firm 1. Firm 1 can bid for partition $\mathcal{P}_k$, pay a price $\pi_1(k, \emptyset) - \pi_1(\emptyset, k)$, and make profits $\pi_1(\emptyset, k)$. On the other hand, Firm 1 can also participate to the auction with $\mathcal{P}_1$, win the auction by bidding the reserve price p_1^m, and make profits $\pi_1(\mathcal{P}_1, \emptyset) - p_1^m$. The data intermediary will set a reserve price $p_1^{m*} = \pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, k) - \epsilon$, where ϵ is an arbitrary small positive number. Thus, $\pi_1(\mathcal{P}_1, \emptyset) - p_1^{m*} > \pi_1(\emptyset, k)$, and since only one partition is sold, it will be $\mathcal{P}_1$. In equilibrium, Firm 1 bids p_1^{m*} for $\mathcal{P}_1$, and Firm 2 bids $\pi_2(k, \emptyset) - \pi_2(\emptyset, k)$. The partitions are therefore $(\mathcal{P}_1, k)$.

Lemma 1

The monopoly price of information when selling partition $\mathcal{P}_1$ to Firm 1 and auctioning partition $\mathcal{P}_{ref}$ is:

$$p_1^{m*} = max_{\mathcal{P}_1}\{\pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}_{ref})\}. \tag{3}$$

The data intermediary finds a partition that maximizes the price of information given by Lemma 1, by combining elementary segments of the reference partition. For instance, in Figure 2 the data intermediary can combine segments 2, 3 and 4 to sell the partition at the bottom. Even though we allow for any partition of the unit line, some partitions can be easily ruled out. For instance, selling consumer data far away from a firm will only increase the competitive impact of information, while selling coarse segments close to a firm's location is not optimal since more precise information would increase its willingness to pay for information. In Lemma 2, we characterize the features of the optimal partition $\mathcal{P}_1^*$, represented in Figure 3.

Lemma 2

An optimal partition $\mathcal{P}_1^*$ divides the unit line into two intervals:

- *The first interval consists of j_1 segments of size $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$ where consumers are identified.*
- *Consumers in the second interval of size $1 - \frac{j_1}{k}$ are unidentified.*

Proof: see Appendix B.1.1.

Partition $\mathcal{P}_1$ divides the unit line into two intervals. Firm 1 can price discriminate identified consumers, and firms charge a uniform price on the second interval of unidentified consumers. The data intermediary does not sell all consumer segments to Firm 1 to reduce the competitive pressure of information. It is easy to understand that selling all consumer segments is not optimal for a data intermediary: selling more consumer segments increases competition and reduces Firm 1's willingness to pay for information. Partition $\mathcal{P}_1^*$ balances the competition and surplus extraction effects of information.

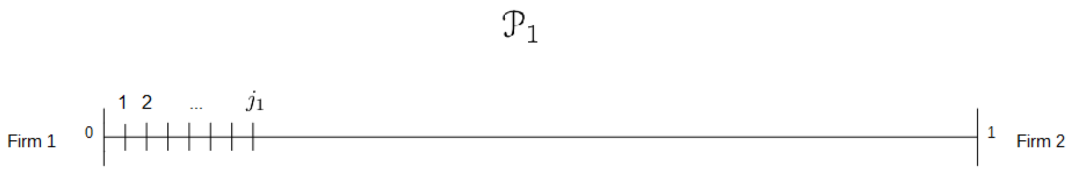

Figure 3: Selling partition $\mathcal{P}_1$ to Firm 1

Since the optimal partition has a structure similar to partition $\mathcal{P}_1^*$, the optimization problem for the data intermediary on its monopoly market boils down to choosing j_1.

Lemma 3

The data intermediary maximizes the price of information paid by Firm 1:

$$\max_{j_1}\{p_1^m(j_1;k)\} = \max_{j_1}\{\pi_1(j_1,\emptyset) - \pi_1(\emptyset,k)\},$$

with

$$j_1^*(k) = \frac{6k-9}{14}.$$

Proof: see Appendix B.1.1.

Lemma 3 shows that the objective functions of the data intermediary and Firm 1 are aligned: they both want j_1^* to maximizes $\pi_1(j_1^*,\emptyset)$.[18] Firm 2 cannot acquire information from other data intermediaries, since we focus on monopoly markets where only one data intermediary has information.

[18]The integer value of j_1^* that maximizes the profits of the data intermediary is chosen by comparing $\pi(\lfloor j_1^*\rfloor)$ and $\pi(\lfloor j_1^*\rfloor+1)$: $max(\pi(\lfloor j_1^*\rfloor),\pi(\lfloor j_1^*\rfloor+1))$.

Suppose now that the data intermediary sells information to both firms. By abuse of notation, let $\mathcal{P}_1$ and $\mathcal{P}_2$ denote now the optimal partitions sold to Firm 1 and Firm 2 respectively. $\pi_1(\mathcal{P}_1, \mathcal{P}_2)$ and $\pi_2(\mathcal{P}_2, \mathcal{P}_1)$ are the respective profits of Firm 1 and Firm 2 with partitions $\mathcal{P}_1$ and $\mathcal{P}_2$. Partitions are potentially different from those found when the data intermediary sells information to Firm 1 only.

Simultaneous auctions, selling to both firms: The data intermediary simultaneously auctions partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ in two separate auctions: Firm 1 (Firm 2) can bid in the two auctions but is only interested in partition $\mathcal{P}_1$ ($\mathcal{P}_2$). Since both firms are guaranteed to obtain their preferred partitions, they will underbid in both auctions from their true valuation. To avoid underbidding, a data intermediary respectively sets reserve prices p_{21}^m and p_{22}^m that correspond to the willingness to pay of Firm 1 for $\mathcal{P}_1$ and of Firm 2 for $\mathcal{P}_2$. Since partition $\mathcal{P}_2$ is optimal for Firm 2, and since Firm 1 has a lower valuation for this partition, Firm 1 will not bid above p_{22}^m in the auction for $\mathcal{P}_2$, and similarly Firm 2 will not bid above p_{21}^m in the auction for $\mathcal{P}_1$. In equilibrium, the data intermediary maximizes the sum of the willingness to pay of each firm for information:

Lemma 4

Partitions $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ maximize the prices of information sold to Firm 1 and Firm 2, $p_{21}^m + p_{22}^m$:

$$p_{21}^m + p_{22}^m = max_{\mathcal{P}_1,\mathcal{P}_2}\{\pi_1(\mathcal{P}_1, \mathcal{P}_2) - \pi_1(\emptyset, \mathcal{P}_2) + \pi_2(\mathcal{P}_2, \mathcal{P}_1) - \pi_2(\emptyset, \mathcal{P}_1)\}. \quad (4)$$

Optimal information partitions $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ that maximize the profit of the data intermediary have the same feature as the one obtained in Lemma 3 when the data intermediary sells information to one firm only. They include all available consumer segments close to a firm up to a cutoff point, and no segments afterward. Let's denote j_{21} and j_{22} the number of segments sold to Firm 1 and to Firm 2. We will show that they are different from the number of segments sold to Firm 1 when the data intermediary sells information in monopoly. Thus, we can derive the profit-maximizing function of the data intermediary using these information structures in equilibrium. Lemma 5 shows that maximizing Eq. 4 with respect to $\mathcal{P}_1$ and $\mathcal{P}_2$ is equivalent to maximizing $p_{21}^m + p_{22}^m$ with respect to j_{21} and j_{22}.

Lemma 5

The profit maximizing function of the data intermediary selling information to both firms in the monopoly market is:

$$p_{21}^m + p_{22}^m = \pi_1(j_{21}, j_{22}) - \pi_1(\emptyset, j_{22}) + \pi_2(j_{22}, j_{21}) - \pi_2(\emptyset, j_{21}).$$

The optimal partitions are given by:

$$j_{21}^*(k) = j_{22}^*(k) = \frac{6k-9}{22}.$$

Proof: see Appendix B.1.2.

Contrary to selling information to Firm 1 only, the data intermediary does not maximize the profits of Firm 1 and Firm 2, but the sum of their willingness to pay for information. Indeed, consider the optimal value of j_{21}, the information partition sold to Firm 1: Firm 1 wants j_{21} that maximizes $\pi_1(j_{21}, j_{22})$, while the data intermediary will set j_{21} that maximizes $p_{21}^m + p_{22}^m$. By choosing j_{21} that takes into account the competitive effect of information on Firm 2, the data intermediary internalizes the negative externality of information on Firm 2. In equilibrium, more consumers are identified by firms and competition is stronger than when the data intermediary sells information to Firm 1 only. To summarize, the last segment sold to both firms will be different from the last segment sold when only Firm 1 purchases information.

The data intermediary compares the profits when selling information to one or to both firms. Proposition 1 shows that the data intermediary will only sell information to Firm 1 on its monopoly market (Firm 2 remains uninformed).

Proposition 1

On monopoly markets m_i, data intermediary i sells information to Firm 1 only.

Proof: see Appendix B.2.

Proposition 1 states that data intermediaries optimally sell information only to Firm 1 on their monopoly markets. Accordingly, Firm 2 does not acquire information and stays uninformed, which allows a monopolist data intermediary to maximize the profit of Firm 1 that is equal to $\pi_1(j_1, \emptyset)$. As there is no competing intermediary from which Firm 1 could acquire information, a monopolist data intermediary can charge a higher price of information by threatening Firm 1 to remain uninformed with profits $\pi_1(\emptyset, k)$.

3.2 Competitive market

We now characterize the information selling strategies in the competitive market. Given our framework, only data intermediary $\overline{DI}$ with the highest information precision $\overline{k}$ will sell information in market l. Nevertheless, data intermediary $\underline{DI}$, which is the second-best intermediary with information collection $\underline{k}$, exerts a competitive pressure and limits the ability of $\overline{DI}$ to extract rent from firms. We denote by $\underline{j_1}$ and $\underline{j_2}$, the information structures proposed to firms by $\underline{DI}$, and by $\underline{p}_1$ and $\underline{p}_2$ the prices of information charged by $\underline{DI}$ to Firm 1 and Firm 2.

Data intermediary $\overline{DI}$ makes an offer to Firm 1 and to Firm 2 that consists of information partitions $\overline{j_1}$ and $\overline{j_2}$ and prices for information $\overline{p}_1$ and $\overline{p}_2$.

3.2.1 Price of information in the competitive market: selling information to one firm

Contrary to the monopoly market, Firm 2 can purchase information from intermediary $\underline{DI}$ even when intermediary $\overline{DI}$ sells information only to Firm 1. The profits of Firm 2 in this case can be written $\pi_2(\underline{j_2}, \overline{j_1})$. Firm 1 can purchase information $\overline{j_1}$ from intermediary $\overline{DI}$ and make profits $\pi_1(\overline{j_1}, \underline{j_2})$. Otherwise Firm 1 can purchase information $\underline{j_1}$ from intermediary $\underline{DI}$ and make profits $\pi_1(\underline{j_1}, \underline{j_2})$. Thus the willingness to pay of Firm 1 for information $\overline{j_1}$ is $\pi_1(\overline{j_1}, \underline{j_2}) - \pi_1(\underline{j_1}, \underline{j_2})$. This corresponds to the price of information charged by data intermediary $\overline{DI}$ when selling information to Firm 1 only.

Lemma 6

The price of information charged by data intermediary $\overline{DI}$ to Firm 1 on the competitive market is:

$$\overline{p}_1 = \pi_1(\overline{j_1}, \underline{j_2}) - \pi_1(\underline{j_1}, \underline{j_2}). \tag{5}$$

3.2.2 Price of information in the competitive market: selling information to both firms

Data intermediary $\overline{DI}$ sells information to Firm 1 and Firm 2. In this case, the price of information corresponds to the willingness to pay of each firm for information. Consider the incentive of Firm 1 to purchase information. Firm 1 has the choice to acquire $\overline{j_1}$ at price $\overline{p}_1$, or to acquire $\underline{j_1}$ at price $\underline{p}_1$ from $\underline{DI}$. Other data intermediaries will have the same strategies as $\underline{DI}$, as competition in market l exhibits a winner-takes-all equilibrium. Because it has the second most precise information, only $\underline{DI}$ will exert a competitive pressure on $\overline{DI}$ by offering firms an alternative way of acquiring information. Thus, the outside option of Firm 1 to acquiring $\overline{j_1}$ is to buy $\underline{j_1}$ from $\underline{DI}$ and make profit $\pi_1(\underline{j_1}, \overline{j_2})$. The willingness to pay of Firm 1 for information is thus $\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})$. The price that Firm 2 is ready to pay for information is defined in a similar way: $\pi_2(\overline{j_2}, \overline{j_1}) - \pi_2(\underline{j_2}, \overline{j_1})$.

Similarly to Lemma 6, we characterize in Lemma 7 the prices of information charged by data intermediary $\overline{DI}$ to Firm 1 and Firm 2.

Lemma 7

The prices of information charged by data intermediary $\overline{DI}$ to Firm 1 and Firm 2 on the competitive market are:

$$\begin{cases} \overline{p}_1(\overline{j_1}, \overline{j_2}) = \pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2}), \\ \qquad \textit{and} \\ \overline{p}_2(\overline{j_2}, \overline{j_1}) = \pi_2(\overline{j_2}, \overline{j_1}) - \pi_2(\underline{j_2}, \overline{j_1}). \end{cases}$$

When data intermediary $\overline{DI}$ sells information to both firms they thus make profits $\pi_1(\overline{j_1}, \overline{j_2})$ and $\pi_2(\overline{j_2}, \overline{j_1})$. The outside option of Firm 1 (w.l.o.g.) is to buy information from intermediary $\underline{DI}$ while facing a competitor informed with $\overline{j}_2$. Data intermediary $\overline{DI}$ can charge a positive price for information as it has the highest k, and firms can thus make higher profits with its information than with information from other intermediaries.

3.2.3 Optimal selling strategy in the competitive market

We can now characterize in Proposition 2 the selling strategy of $\overline{DI}$ in the competitive market.

Proposition 2

The objective function of $\overline{DI}$ when selling information to Firm 1 and Firm 2 is:

$$\max_{\overline{j_1},\overline{j_2}}\{\overline{p}_1(\overline{j_1},\overline{j_2})+\overline{p}_2(\overline{j_2},\overline{j_1})\}. \tag{6}$$

The optimal partitions are:

$$\overline{j_1}^*(\overline{k})=\overline{j_2}^*(\overline{k})=\frac{1}{3}-\frac{1}{9\underline{k}}-\frac{7}{18\overline{k}}.$$

Proof: see Appendix B.3.

Proposition 2 shows that selling the reference partition (with information on all consumer segments) is not optimal even when data intermediaries compete. Indeed, data intermediary $\overline{DI}$ has always incentives to only sell a subset of all available segments to soften the competitive effect of information and to maximize the price of information. Close-by consumers with a high willingness to pay will be identified and the remaining consumers will not be identified.[19]

Therefore, Proposition 2, which analyzes competing data intermediaries, is an important generalization of the result of Bounie et al. (2021) who focus on a monopolist data intermediary. Data intermediaries do not sell all their available data, which echoes the results by (Grossman and Stiglitz, 1980) who found that financial markets are not informationally efficient.

Proposition 3 characterizes the equilibrium on the competitive market l.

[19] A direct comparison with Bertrand competition could lead to the conclusion that competing intermediaries would sell all their available information and each firm would have information $\mathcal{P}_{ref}$. However, this reasoning is incorrect, as data intermediaries have incentives to deviate from such equilibrium. Selling fewer segments than $\mathcal{P}_{ref}$ unambiguously increases profits of a data intermediary.

Proposition 3

In equilibrium on the competitive market l:

- *(a) $\overline{DI}$ sells information to both firms.*
- *(b) Other data intermediaries do not sell information.*
- *(c) Firms identify more consumers than on monopoly markets m_i.*

Proof: see Appendix B.3.

Proposition 3 (a) shows that firms will buy information from data intermediary $\overline{DI}$ that offers the highest precision, as it maximizes their profits. We will show in the next section that $\overline{DI}$ is the data intermediary with the largest monopoly market m_1, i.e. data intermediary 1. Also, we find in Proposition 3 (b) that firms do not purchase information from other data intermediaries. Nevertheless, data intermediary $\underline{DI}$ exerts a competitive pressure on data intermediary $\overline{DI}$ that cannot charge firms the maximal price of information p_{21}^{m*} and p_{22}^{m*}.

Finally, Proposition 3 (c) shows that competition in market l lowers the ability of data intermediary $\overline{DI}$ to fully internalize the business stealing effect of information and that more segments are sold in market l than in monopoly markets m_i. When intermediaries compete, both firms can purchase the same information and compete on a level playing field. Competition between data intermediaries thus increases competition in the product market as more consumers are identified by firms.

4 Collecting consumer data

We analyze in this section the data collection strategies of intermediaries. We rank the number of consumer segments k_i collected by each data intermediary i. The number of consumer segments collected increases the price of information through two effects. First, more data increases rent extraction as firms can better price-discriminate consumers on thinner segments of the demand. Secondly, more precise information lowers the profits of an uninformed firm facing an informed competitor. Since the data intermediary threatens an uninformed firm to sell the reference partition (that includes all available consumer segments) to its competitor, more segments collected by the data intermediary increase the value of this threat.

The two effects of data collection on the price of information vary according to whether data intermediaries are on their monopoly market or on the competitive market l, which in turn changes the number of consumer segments k_i collected by intermediary i. Remember that data intermediaries cannot distinguish whether consumers belong to m or l before collecting their data, and the number of consumer segments collected by each data intermediary is identical on m and l.

A monopolist data intermediary makes the following profit given by the price of information net of the data collection cost, times the size of the monopoly market m:[20]

$$\Pi_m(k) = m[p_m(k) - c(k)].$$

On the competitive market l, $\overline{DI}$ sells information and makes profits equal to the sum of the prices paid by each firm (net of the data collection cost), times the size of the competitive market l:[21]

$$\Pi_l(k) = l[2\overline{p}(k) - c(k)].$$

Additionally, other data intermediaries collect information at cost $lc(k)$, but do not sell information and make zero profits in market l.

We characterize in Proposition 4 the data collection strategies of intermediaries.

Proposition 4

[20] The data collection cost is defined such that profit functions are strictly concave with a unique maximum. Because the price of information is concave with a horizontal asymptote, it is sufficient to assume that $c(.)$ is convex and increasing to ensure the strict concavity and unique maximum of the profit function. In Appendix A.1 we show that concave functions with a low level of concavity can also be used, as we provide an example where $c(k) = ln(k)$.

[21] By assumption, there is a unique optimal $\overline{k}^*$ that maximizes Π_l.

- *(a) The number of consumer segments collected by intermediaries is smaller in the competitive market than in monopoly markets.*

- *(b) The number of consumer segments $\overline{k}$ collected by $\overline{DI}$ increases with the number of consumer segments $\underline{k}$ collected by $\underline{DI}$.*

Proof: See Appendix B.4.

Proposition 4 (a) shows that on the competitive market, $\underline{DI}$ exerts a pressure on $\overline{DI}$, which decreases the profitability of data sold by $\overline{DI}$. The optimal number of segments collected is thus higher on monopoly markets where data intermediaries sell information at the maximal price. Other data intermediaries do not sell information on the competitive market l, and they incur a loss from collecting information on these consumers.

Also, Proposition 4 (b) shows that an important determinant of the data collection strategy of the leading intermediary $\overline{DI}$ is the amount of data collected by its direct competitor $\underline{DI}$ on the competitive market. There is an escape-competition effect: as $\underline{DI}$ collects more consumer segments, $\overline{DI}$ has interest to differentiate and to collect more segments too.[22] As we will see in Proposition 5, the amount of data collected by $\underline{DI}$ is determined by the size of its monopoly market. This effect is new in the literature, as previous research has ignored strategic considerations and instead mainly considered the direct cost as the main determinant for data collection strategies (Varian, 2018). Proposition 4 (b) introduces strategic considerations in the literature on data collection.

We can now characterize the number of consumer segments k_i collected by each data intermediary i according to m_i. Proposition 5 shows that data intermediaries with the highest and the second highest precision, respectively $\overline{DI}$ and $\underline{DI}$, are data intermediary 1 (with market size m_1) and data intermediary 2 (with market size m_2):

Proposition 5

[22] This effect is also similar to models of innovation such as Aghion et al. (2005), and is also present in models of product differentiation with quality improvements.

- *(a) When data intermediaries are asymmetric in terms of market sizes, the larger the size of the monopoly market, the higher the total number of consumer segments collected by a data intermediary:*

$$m_1 > m_2 \geq ... \geq m_n \implies k_1 > k_2 \geq ... \geq k_n.$$

- *(b) When data intermediaries are symmetric in terms of market sizes $m_1 = m_2 = ... = m_n$, an equilibrium has the following property. One data intermediary (1, w.l.o.g.) collects strictly more information than the others who all collect the same number of segments with:*

$$k_1 > k_2 = ... = k_n.$$

Proof: see Appendix B.5.

Proposition 5 (a) highlights a positive relation between market power, captured by the size of the monopoly market, and the data collection strategies of data intermediaries. A data intermediary that is dominant in terms of size of its monopoly market collects more consumer segments than other intermediaries and is the only intermediary that sells information on the competitive market l. We will discuss in Section 6 how merger policies can limit the dominance of data intermediaries by preventing the emergence of large data intermediaries.

Proposition 5 (b) shows that the only possible equilibrium when two intermediaries have monopoly markets of identical sizes $m_1 = m_2$ is such that one of the intermediaries collects more information than the other. It is easy to show that when $m_1 = m_2$, $k_1 > k_2$ is an equilibrium. Indeed, it is not profitable for intermediary 2 to deviate and collect more data than intermediary 1, as collecting more data than intermediary 1 is costly. Conversely, collecting k_1 segments is optimal for intermediary 1 given that the other intermediary collects k_2. Depending on the primitives of the models, an equilibrium does not necessarily exist, however if it exists, it must have the features described in Proposition 5 (b). This has strong implications for competition authorities willing to guarantee competition on a level playing field. It is not sufficient to encourage symmetric competition in terms of market shares, and asymmetry in data collection and data dominance arise even when all data intermediaries have the same monopoly market size.

5 Consumer surplus

In this section, we compare consumer surplus in monopoly markets with consumer surplus in the competitive market. We show that competition between data intermediaries is always beneficial for consumers for two reasons. On the one hand, consumer surplus depends on the number of segments sold to firms, which increases when intermediaries compete. More segments sold increase competition in the product market and benefit consumers. On the other hand, competition between intermediaries also lowers the price of information, which lowers the amount of data k_i that they collect. Competition between intermediaries thus also reduces consumer rent extraction.

To analyze how competition between data intermediaries impacts consumer surplus in the product market, we first consider how consumer surplus changes with the number of consumer segments sold to firms, holding k constant. As we have seen in Proposition 2, competition between data intermediaries increases the number of segments sold on the competitive market. Suppose that Firm 1 has information on j_1 consumer segments, and Firm 2 has information on j_2 consumer segments. If Firm 1 obtains additional information on segment $[\frac{j_1}{k}, \frac{j_1+1}{k}]$, there are two effects on consumer surplus:

1. A rent extraction effect: Firm 1 price discriminates consumers on $[\frac{j_1}{k}, \frac{j_1+1}{k}]$, which reduces their surplus.
2. A competitive effect: Firm 1 lowers its price on $[\frac{j_1+1}{k}, 1]$, which increases the competitive pressure on Firm 2. Firm 2 lowers its price, which has a positive effect on the surplus of consumers over the whole unit line.

Overall, the second effect always dominates the first, and consumer surplus increases when more consumer segments are sold. Indeed, the rent extraction effect only increases profits on one additional segment, while the competitive effect operates on the whole Hotelling line. Proposition 6 shows that consumer surplus, denoted $CS(j_1, j_2, k)$, increases with the number of consumer segments j_1 and j_2 sold to Firm 1 and to Firm 2. The same result holds for j_2 given j_1.

Proposition 6

For a given j_2, consumer surplus always increases with the number of consumer segments j_1 sold to Firm 1:

$$\forall \;\; j_2, k : \quad \frac{\partial CS(j_1, j_2, k)}{\partial j_1} > 0.$$

Proof: See Appendix B.6

We now discuss the effect of a change in the amount of data collected k on consumer surplus. Increasing the value of k reduces the size of the segments and allows firms to better extract consumer surplus. To simplify notations, we denote by $x_1 = \frac{j_1}{k}$ and $x_2 = \frac{j_2}{k}$ the locations of the last consumers identified by Firm 1 and Firm 2. Proposition 7 shows that consumer surplus decreases with k for given x_1 and x_2.

Proposition 7

Consumer surplus always decreases with k:

$$\forall \;\; x_1, \;\; x_2 : \quad \frac{\partial CS(x_1, x_2, k)}{\partial k} < 0.$$

Proof: see Appendix B.7.

We want to compare consumer surplus in a monopoly market with consumer surplus in the competitive market independently of their respective sizes. To do so, we standardize their size to 1, without loss of generality.

Proposition 8

Consumer surplus is higher in the competitive market than in monopoly markets.

Proof: see Appendix B.8.

The proof of proposition 8 proceeds in two steps. First, we compare market l with market m_1 where the numbers of segments collected are identical. More consumers are identified in the competitive market than in market m_1, which increases the intensity of competition between information buyers, decreases product prices, and increases consumer surplus. Thus, consumer surplus is higher on l than on m_1. Secondly, we compare consumer surplus on l with consumer surplus in markets $m_i < m_1$. On the latter markets, fewer consumer segments are collected than on the competitive market since $k_i < k_1$, which lowers the ability of firms to extract consumer surplus. However, more consumers are identified in the competitive market than in markets m_i, and this increases consumer surplus. Overall, the second effect dominates the first, since the competitive effect of information operates on the whole Hotelling line and benefits all consumers, while the rent extraction effect only reduces the surplus of identified consumers.[23]

Proposition 8 is established when both the monopoly markets and the competitive market are of equal size 1. Thus, by changing the relative size of both markets, total consumer surplus can increase or decrease. Proposition 8 suggests that increasing the size of market l will increase consumer surplus, and thus, the competitive market l should be made as large as possible. Increasing the size of the competitive market can be achieved for instance through the right to data portability of the European General Data Protection Regulation. Indeed, data portability allows consumers to bring all personal information owned by a firm to any of its competitors, thus increasing the size of market l in our model. As Crémer et al. (2019) emphasize, access to data has become a critical competitive factor in digital markets, and ensuring access to data to all firms is an efficient way to increase product market competition and consumer surplus. Our results provide a theoretical background for such policies.

[23] This result holds under other distributional assumptions, as long as the density of the distribution around x_1 and x_2 is not too concentrated.

6 Mergers and acquisitions in the market for information

We now analyze the impacts of a merger between two data intermediaries on consumer surplus. A merger between intermediaries will impact consumer surplus by changing the number of consumer segments collected by the merged entity.

Mergers have three distinct effects on the number of consumer segments collected by intermediaries. First, the merged entity benefits from cost efficiencies. Two separate data intermediaries collect data in market l. However, the merged entity only collects information once in market l. This cost efficiency leads the merged entity to collect more consumer data. Secondly, the merged entity can leverage on a larger market size and have more incentives to collect data. Thirdly, larger market sizes may change the equilibrium on market l depending on whether the merged entity has the highest information precision or the second highest.

In the remaining of the section we highlight how these three mechanisms operate in recent flagship mergers, and can overturn standard merger recommendations.[24] The first case analyzed in Section 6.1 deals with a merger between two major companies, such as the acquisition of WhatsApp by Facebook. The second case described in Section 6.2 is the acquisition of a start-up by a dominant intermediary, such as the acquisition of Fitbit by Google. The third case, studied in Section 6.3, illustrates the merger between two intermediate companies, such as the acquisition of DataLogix by Oracle.

6.1 The Facebook/WhatsApp case revisited

We first consider a merger between data intermediaries 1 and 2, which can be used in our model to discuss the acquisition of WhatsApp by Facebook. There are two opposite effects of such mergers on consumer data collection. First, the merged entity benefits from a larger monopoly market $m_1 + m_2$. The marginal gain from collecting consumer data is thus higher after the merger, and this first effect drives up data collection. Secondly, in market l the competitive pressure decreases after the merger as $m_3 < m_2$. Thus the escape-competition effect is

[24] All the proofs of this section are available in Appendix B.9.

weaker: the difference in size between $m_1 + m_2$ and m_3 is larger after the merger, and there is less competitive pressure on the newly merged entity. In turn, this decreases the incentives of the merged entity to collect consumer data.[25]

These two effects go in opposite directions. However, when the size of the merged entity $m_1 + m_2$ is large compared with l, the increase of data collection due to the size effect always dominates the second effect, and data collection increases overall. The acquisition of WhatsApp by Facebook in 2014 illustrates how a data-driven merger can result in more data collection. Facebook was a dominant actor and acquired WhatsApp, a rising competitor at that time.[26] Immediately after the acquisition, Facebook merged the consumer data available on both platforms, which gave more precise information on its users. Recently, WhatsApp changed its privacy policy and no longer allows its users to opt-out of sharing data with Facebook, increasing the amount of consumer data collected further.

6.2 The Google/Fitbit case

In this second case, we consider the acquisition of a small data intermediary (m_i) by data intermediary 1 or 2. This case allows us to discuss startup acquisition by dominant actors, such as Google acquiring Fitbit. The impact of the merger depends on the relative share of the newly merged entity compared to its direct competitor.

There are two opposite effects of such a merger on the amount of data collected by the merged entity. On the one hand, when data intermediary 1 acquires a small data intermediary and the identity of the second best data intermediary in market l remains the same, the escape-competition effect is weaker after the merger, since the newly merged entity now has a relatively larger market size. This first effect reduces the amount of data collected by the merged entity. On the other hand, the merged entity can leverage on its larger market size to collect more consumer data. Overall, this second effect always dominates the first, and more data is collected after the merger. As the amount of data collected increases, consumer surplus

[25] All proofs of this section are available upon request.

[26] See Valletti and Zenger (2019) for a discussion of this acquisition.

decreases through more rent extraction on both the merged monopoly market and the competitive market.[27]

Our model provides a new argument against the acquisition of Fitbit by Google. By acquiring Fitbit, Google acquires new sources of information, leveraging markets where it is already dominant, thus increasing the amount of data collected. Hence, consumer surplus could be reduced in related markets such as healthcare and health insurance market.

6.3 The Oracle/DataLogix case revisited

Two data intermediaries who are not selling in market l ($m_i, m_j \neq m_1$) merge and their information becomes more precise than the second most precise data intermediary: $m_1 > m_i + m_j > m_2$. This case illustrates a merger between a small data intermediary and data intermediary 2, and can be used to understand, for example, the acquisition of DataLogix by Oracle. Following the merger, consumer surplus decreases for two reasons. First, data collection increases on the monopoly market of the merged entity $m_i + m_j$, and consumer surplus decreases on this monopoly market.

Secondly, the intensity of competition exerted by the second best intermediary on m_1 increases. As the competitive pressure faced by data intermediary 1 in market l increases, the escape-competition effect described in Proposition 4 (b) becomes stronger, increasing data collection and decreasing consumer surplus on both the competitive market l and monopoly market m_1.

Hence, our model shows that the acquisition of DataLogix by Oracle may have reduced consumer surplus through improved price discrimination due to better

[27]Suppose now that data intermediary 2 makes the acquisition and that the merged entity becomes larger than data intermediary 1. This example is also valid when two small data intermediaries merge and become larger than data intermediary 1. On the one hand, if the size of the monopoly market of the merged entity is equal to data intermediary 1, they compete à la Bertrand on the competitive market l, and the equilibrium is described in Proposition 5. Data intermediary 1 will collect more consumer data, decreasing consumer surplus on m_1 and l. The merged entity will collect more consumer data than both firms separately, which lowers consumer surplus on the merged monopoly market. On the other hand, if the merged entity has a larger monopoly market than m_1, competition on l after the merger will be fiercer. The merged entity collects then more data than intermediary 1 before the merger. In this case, consumer surplus on m_1 increases after the merger, but consumer surplus decreases on l and on monopoly market $m_2 + m_i$.

consumer segmentation on related markets such as credit and mortgage markets.

7 Conclusion

Our model emphasizes how competition in the market for information has indirect effects in related markets that are relevant for analyzing consumer surplus and merger impact assessments. Competition between intermediaries is beneficial for consumers for three reasons.

First, our model shows that competition changes the selling strategies of data intermediaries: a monopolist data intermediary sells less information than under competition, which reduces the intensity of competition in the product market and harms consumers. Therefore, a merger between data intermediaries will lower consumer surplus through a reduction of the intensity of competition. Most models in the existing literature ignore this effect and assume that data intermediaries sell all available information, which overestimates the competitive effect of information in merger guidelines.

Secondly, the intensity of competition between data intermediaries determines which firm can access information: a monopolist intermediary sells information to one firm while both firms purchase information in the competitive market. Firms compete more fiercely when they both have access to consumer data, which benefits consumers. Thirdly, competition between intermediaries also reduces the amount of data collected, which increases consumer surplus. Policy-makers should therefore promote competition between data intermediaries as it guarantees fair and equal access to information to firms in the product market, and it intensifies competition and increases consumer surplus. Such an outcome can be reached through open data regulations, under which companies have to share consumer data with fair, reasonable, and non-discriminatory terms (Crémer et al., 2019).

Further research could explore the impact of personal data protection on data intermediaries and how data protection agencies could work closer with competition authorities. Recent actions from the FTC call for regulation of the data brokerage industry,[28] and in the US, states such as Vermont or California have

[28] Federal Trade Commission, 2014, Data brokers: A Call for Transparency and Accountability.

recently passed laws to gain control over the practices of data intermediaries.[29] It remains to be shown how recent regulations, such as General Data Protection Regulation in the European Union - which creates new ways to protect consumers through opt-in, right to be forgotten, data minimization, and privacy by design - will change the amount of data collected by large data intermediaries.

References

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

Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, 120(2):701–728, 2005.

Paul Belleflamme, Wing Man Wynne Lam, and Wouter Vergote. Competitive imperfect price discrimination and market power. *Marketing Science*, 39(5):996–1015, 2020.

Dirk Bergemann and Alessandro Bonatti. Selling cookies. *American Economic Journal: Microeconomics*, 7(3):259–94, 2015.

Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11, 2019.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American economic review*, 108(1):1–48, 2018.

Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. 2019.

David Bounie, Antoine Dubus, and Patrick Waelbroeck. Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, 52(2):283–313, 2021. doi: https://doi.org/10.1111/1756-2171.12369. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12369.

Francesco Clavorà Braulin and Tommaso Valletti. Selling customer information to competing firms. *Economics Letters*, 149:10–14, 2016.

Zhijun Chen, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima. Data-driven mergers and personalization. *ISER DP*, (1108), 2020.

Chongwoo Choe, Stephen King, and Noriaki Matsushima. Pricing with cookies: Behavior-based price discrimination and spatial competition. *Management Science*, 64(12):5669–5687, 2018.

CMA. Online platforms and digital advertising. 2020.

[29] Column: Shadowy data intermediaries make the most of their invisibility cloak; Los Angeles Times, November 5, 2019.

Dominic Coey, Bradley J Larsen, Kane Sweeney, and Caio Waisman. Scalable optimal online auctions. *Marketing Science*, 2021.

Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. Competition policy for the digital era. *Report for the European Commission*, 2019.

Leandro DalleMule and Thomas H Davenport. What's your data strategy. *Harvard Business Review*, 95(3):112–121, 2017.

Jan De Loecker, Jan Eeckhout, and Gabriel Unger. The Rise of Market Power and the Macroeconomic Implications*. *The Quarterly Journal of Economics*, 135(2):561–644, 01 2020. ISSN 0033-5533. doi: 10.1093/qje/qjz041. URL https://doi.org/10.1093/qje/qjz041.

Peter A Diamond. A model of price adjustment. *Journal of economic theory*, 3(2): 156–168, 1971.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.

Jason Furman, Diane Coyle, Amelia Fletcher, Derek McAules, and Philip Marsden. Unlocking digital competition: Report of the digital competition expert panel. *Report prepared for the Government of the United Kingdom, March*, 2019.

Sanford J Grossman and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980.

Andrei Hagiu and Julian Wright. When data creates competitive advantage. *Harvard Business Review*, 98(1):94–101, 2020.

Ganesh Iyer, David Soberman, and J Miguel Villas-Boas. The targeting of advertising. *Marketing Science*, 24(3):461–476, 2005.

Nicola Jentzsch, Geza Sapi, and Irina Suleymanova. Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization*, 31(2): 131–144, 2013.

Charles I Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–58, 2020.

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

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

Charles Brian O'kelley and Adam Roger Pritchard. Data marketplace and broker fees, January 8 2009. US Patent App. 11/772,965.

Geoffrey Parker, Georgios Petropoulos, and Marshall W Van Alstyne. Platform mergers and antitrust. *Available at SSRN 3763513*, 2021.

Fiona Scott Morton, P Bouvier, Al Ezrachi, B Jullien, R Katz, G Kimmelman, AD Melamed, and J Morgenstern. Committee for the study of digital platforms: Market structure and antitrust subcommittee-report. *Chicago: Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business*, 2019.

Carl Shapiro, Shapiro Carl, Hal R Varian, et al. *Information rules: a strategic guide to the network economy.* Harvard Business Press, 1998.

Andrew Sheehan and Guy Yalif. Multi-round auction and internet marketplace, December 6 2001. US Patent App. 09/746,022.

Jean Tirole. Competition and the industrial challenge for the digital age. 2020.

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

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

Hal R Varian. Price discrimination. *Handbook of industrial organization*, 1:597–654, 1989.

A Appendix

A.1 Numerical example

We characterize the number of consumer segments collected by a monopolist data intermediary with the following cost function:

$$c(k) = \frac{ln(k)}{10}$$

This cost function is chosen in order to satisfy the existence of a unique positive optimal value of k.

The profit function of the data intermediary in this case is:

$$\Pi(k) = -\frac{19t}{28k} + \frac{11t}{56k^2} + \frac{29t}{56} - \frac{ln(k)}{10}$$

When t=10, this is maximized for $|k| = 67$.

B Mathematical Appendix

B.1 Optimal information structures on the monopoly markets

We show that the data intermediary optimally sells a partition that divides the unit line into two intervals. The first interval identifies the closest consumers to a firm and is partitioned in j segments of size $\frac{1}{k}$. The second interval is of size $1 - \frac{j}{k}$ and leaves the other consumers unidentified. We first establish this claim when the data intermediary sells information to only one firm, and then when it sells information to both firms.[30]

B.1.1 Proof of Lemma 3: the data intermediary sells information to one firm

Suppose that the intermediary sells information to Firm 1 (without loss of generality). The data intermediary can choose any partition in the sigma-field $\mathbb{P}$ generated by the elementary segments of size $\frac{1}{k}$. There are three types of segments to consider:

- Segments A, where Firm 1 serves all consumers but Firm 2 exerts a competitive pressure.
- Segments B, where Firms 1 and 2 compete; both have a positive demand.
- Segments C, where Firm 1 has no demand and makes zero profit.

We proceed in three steps. In step 1 we analyze type A segments. We show that it is optimal to sell a partition where type A segments are of size $\frac{1}{k}$. In step 2, we show that all segments of type A are located closest to Firm 1. In step 3 we analyze segments of type B and we show that it is always more profitable to sell a union of such segments. Therefore, there is only one segment of type B, located furthest away from Firm 1, and of size $1 - \frac{j}{k}$ (with j an integer, $j \leq k$). Finally, we can discard segments of type C because information on consumers on these segments does not increase profits.

[30] All along the proofs, we refer to Liu and Serfes (2004) who prove the continuity and concavity of the profit functions with third-degree price discrimination.

Step 1: We analyze segments of type A where Firm 1 is in constrained monopoly, and show that reducing the size of segments to $\frac{1}{k}$ is optimal.

Consider any segment $I = [\frac{i}{k}, \frac{i+l}{k}]$ of type A with l, i integers verifying $i + l \leq k$ and $l \geq 2$, such that Firm 1 is in constrained monopoly on this segment. We show that dividing this segment into two sub-segments increases the profits of Firm 1. Figure 4 shows on the left panel a partition with segment I of type A, and on the right, a finer partition including segments I_1 and I_2, also of type A. In Figure 4 and in all similar figures, the blue curves represent the demand for Firm 1 (demand for Firm 2 is not represented and corresponds to the complementary demand on the segments). To illustrate, for segments of type A, the blue curve covers the whole segment. For segments of type B, the blue curve only covers part of the segment. We compare profits in both situations and show that the finer segmentation is more profitable for Firm 1. We write $\pi_1^A(\mathcal{P})$ and $\pi_1^{AA}(\mathcal{P}')$ the profits of Firm 1 on I with partitions $\mathcal{P}$ and on I_1 and I_2 with partition $\mathcal{P}'$.

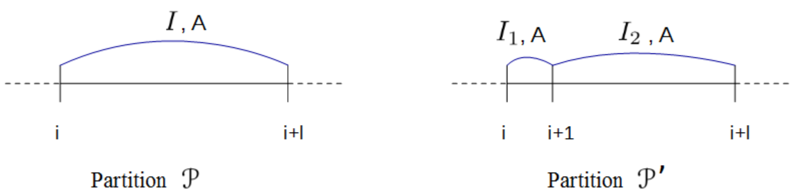

Figure 4: Step 1: segments of type A

To prove this claim, we establish that the profit of Firm 1 is lower with a coarser sub-partition $\mathcal{P}$ with $I = [\frac{i}{k}, \frac{i+l}{k}]$, than with a finer sub-partition $\mathcal{P}'$ obtained by replacing I with two segments: $I_1 = [\frac{i}{k}, \frac{i+1}{k}]$ and $I_2 = [\frac{i+1}{k}, \frac{i+l}{k}]$ (other segments are unchanged).

First, profits with the coarser partition is: $\pi_1^A(\mathcal{P}) = p_{1i} d_1 = p_{1i} \frac{l}{k}$. The demand is $\frac{l}{k}$ as Firm 1 serves all consumers; p_{1i} is such that the indifferent consumer x is located at $\frac{i+l}{k}$:

$$V-tx-p_{1i} = V-t(1-x)-p_2 \implies x = \frac{p_2 - p_{1i} + t}{2t} = \frac{i+l}{k} \implies p_{1i} = p_2 + t - 2t\frac{i+l}{k},$$

with p_2 the price charged by (uninformed) Firm 2. This price is only affected by strategic interactions on the segments where firms compete, and therefore does not depend on the pricing strategy of Firm 1 on type A segments.

We write the profit function for any p_2, replacing p_{1i} and d_1 by their equilibrium values obtained in the previous equations:

$$\pi_1^A(\mathcal{P}) = \frac{l}{k}(t + p_2 - \frac{2(l+i)t}{k}).$$

Secondly, using a similar argument, we show that the profit on $I_1 \cup I_2$ with partition $\mathcal{P}'$ is:

$$\pi_1^{AA}(\mathcal{P}') = \frac{1}{k}(t + p_2 - \frac{2(1+i)t}{k}) + \frac{l-1}{k}(t + p_2 - \frac{2(l+i)t}{k}).$$

Comparing $\mathcal{P}$ and $\mathcal{P}'$ shows that the profit of Firm 1 using the finer partition increases by $\frac{2t}{k^2}(l-1)$, which establishes the claim.

By repeating the previous argument, it is easy to show that the data intermediary will sell a partition of size $\frac{l}{k}$ with l segments of equal size $\frac{1}{k}$.

Step 2: We show that all segments of type A are closest to Firm 1 (located at* 0 *on the unit line by convention).

Going from left to right on the Hotelling line, we look for the first time a type B interval, $J = [\frac{i}{k}; \frac{i+l}{k}]$ of length $\frac{l}{k}$, is followed by an interval $I_1 = [\frac{i+l}{k}, \frac{i+l+1}{k}]$ of type A, shown to be of size $\frac{1}{k}$ in step 1 (right panel of Figure 5). We now show that profits are higher when the data intermediary switches segments I_1 and J. The resulting sub-partition is now $I_1' = [\frac{i}{k}; \frac{i+1}{k}]$ followed by $J' = [\frac{i+1}{k}, \frac{i+l+1}{k}]$ (right panel of Figure 5).

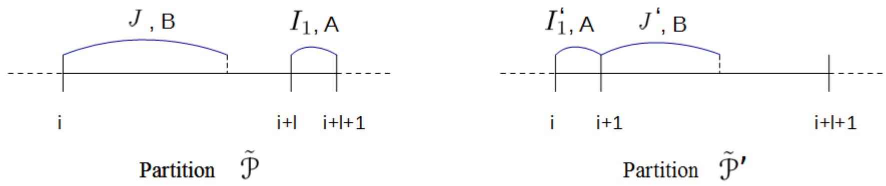

Figure 5: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 5 and correspond respectively to the partitions $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. The curved line represents the demand of Firm 1, which does not cover type B segments. In partition $\tilde{\mathcal{P}}$, a segment of type B of size $\frac{l}{k}$, J, is followed by a segment of type A of size $\frac{1}{k}$, I_1. We show that segments of type A are always located closest to Firm 1 by proving that it is always optimal to change partition starting with segments of type B with a partition starting with segments of type A like in partition $\tilde{\mathcal{P}}'$. To show this claim, we compare the profits of the informed firm with $J \cup I_1$ under partition $\tilde{\mathcal{P}}$ and with $I_1' \cup J'$ under partition $\tilde{\mathcal{P}}'$, and we show that the latter is always higher than the former. The other segments of the partition remain unchanged.

To compare the profits of the informed firm under both partitions, we first characterize type B segments. On segments of type B, both firms must have a positive demand. Eq. 7 gives the conditions for the demands addressed to Firm 1 and to Firm 2 to be positive on such segments.

$$\begin{gathered}\forall \;\; i, l \in \mathbb{N} \;\; s.t. \;\; 0 \leq i \leq k-1 \;\; \text{and} \;\; 1 \leq l \leq k-i-1, \\ \frac{i}{k} \leq \frac{\tilde{p}_2 + t}{2t} \qquad \text{and} \qquad \frac{\tilde{p}_2 + t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}.\end{gathered} \tag{7}$$

Condition $\frac{i}{k} \leq \frac{\tilde{p}_2+t}{2t}$ guarantees that Firm 1 serves consumers on segment J, and $\frac{\tilde{p}_2+t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}$ guarantees that Firm 2 serves positive demand on segment J.

In particular, we use the relation that Eq. 7 characterizes between price $\tilde{p}_2$ and segments endpoint $\frac{i}{k}$ and $\frac{i+l}{k}$ to compare the profits of Firm 1 with $\tilde{\mathcal{P}}'$ and with $\tilde{\mathcal{P}}$.

To facilitate the computation of demands on segments of type A, we introduce intermediary notations that characterize the location of these segments (u_i). Segments of type A are of size $\frac{1}{k}$ and are located at $\frac{u_i-1}{k}$, and segments of type B, are located at $\frac{s_i}{k}$ and are of size $\frac{l_i}{k}$.[31] There are $h \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$, where prices are noted $\tilde{p}^A_{1i}$. On each of these segments, the demand is $\frac{1}{k}$. There are $n \in \mathbb{N}$ segments of type B, where prices are noted $\tilde{p}^B_{1i}$. We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

$$d_{1i} = x - \frac{s_i}{k} = \frac{\tilde{p}_2 - \tilde{p}^B_{1i} + t}{2t} - \frac{s_i}{k}.$$

We can rewrite profits of Firm 1 as the sum of two terms. The first term represents the profits on segments of type A. The second term represents the profits on segments of type B.

$$\pi_1(\tilde{\mathcal{P}}) = \sum_{i=1}^{h} \tilde{p}^A_{1i} \frac{1}{k} + \sum_{i=1}^{n} \tilde{p}^B_{1i} [\frac{\tilde{p}_2 - \tilde{p}^B_{1i} + t}{2t} - \frac{s_i}{k}].$$

Profits of Firm 2 are generated on segments of type B only, where the demand for Firm 2 is:

$$d_{2i} = \frac{s_i + l_i}{k} - x = \frac{\tilde{p}^B_{1i} - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}.$$

Profits of Firm 2 can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^{n} \tilde{p}_2 [\frac{\tilde{p}^B_{1i} - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}]. \tag{8}$$

Firm 1 maximizes profits $\pi_1(\tilde{\mathcal{P}})$ with respect to $\tilde{p}^A_{1i}$ and $\tilde{p}^B_{1i}$, and Firm 2 maximizes $\pi_2(\tilde{\mathcal{P}})$ with respect to $\tilde{p}_2$, both profits are strictly concave.

Equilibrium prices are:

$$\begin{aligned} \tilde{p}^A_{1i} &= t + \tilde{p}_2 - 2\frac{u_i t}{k} \\ \tilde{p}^B_{1i} &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n}[\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]] - \frac{s_i t}{k} \\ \tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]. \end{aligned} \tag{9}$$

[31]With u_i and s_i integers below k.

We can now compare profits with $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. When we move segments of type B from the left of segments of type A to the right of segment of type A, it is important to check that Firm 1 is still competing with Firm 2 on each segment of type B, and that Firm 1 is still in constrained monopoly on segments of type A. The second condition is met by the fact that price $\tilde{p}_2$ is higher in $\tilde{\mathcal{P}}'$ than in $\tilde{\mathcal{P}}$. The first condition is guaranteed by Eq. 7: $\frac{\tilde{p}_2+t}{2t} - \frac{l_i}{k} \leq \frac{s_i+l_i}{k}$ for all segments of type B located at $[\frac{s_i}{k}, \frac{s_i+l_i}{k}]$. Let $\tilde{s}_i$ denote the m segments ($m \in [0, n-1]$) of type B with partition $\tilde{\mathcal{P}}$ located at $[\frac{\tilde{s}_i}{k}, \frac{\tilde{s}_i+\tilde{l}_i}{k}]$ that do not meet these conditions, and therefore are type A segments with partition $\tilde{\mathcal{P}}'$.

Noting $\tilde{p}_2'$ and $\tilde{p}_{1i}^{B'}$ the prices with $\tilde{\mathcal{P}}'$, we have:

$$\begin{aligned}
\tilde{p}_2' &= \frac{4t}{3(n-m)}[-\frac{n}{4} + \sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}] + \frac{m}{4} + \frac{1}{2k} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] \\
&= \tilde{p}_2 + \frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}],
\end{aligned}$$

for segments of type B where inequalities in Eq. 7 hold:

$$\tilde{p}_{1i}^{B'} = \tilde{p}_{1i} + \frac{1}{2}\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}],$$

for segments of type B where inequalities in Eq. 7 do not hold:

$$\tilde{p}_{1i}^{B'} = \tilde{p}_{1i} + \frac{1}{2}\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] - \frac{t}{k}.$$

We now compare the profits of Firm 1 with sub-partition $\tilde{\mathcal{P}}$ ($J \cup I_1$) and with sub-partition $\tilde{\mathcal{P}}'$ ($I_1' \cup J'$). We proceed in two steps. First we show that the profits of Firm 1 on $[\frac{i}{k}, \frac{i+l+1}{k}]$ are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$. Secondly we show that the profits of Firm 1 on type B segments are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$.

First we show that the profits of Firm 1 increase on $[\frac{i}{k}, \frac{i+l+1}{k}]$, that is, we show that $\Delta\pi_1 = \pi_1(\tilde{\mathcal{P}}') - \pi_1(\tilde{\mathcal{P}}) \geq 0$:

$$\begin{aligned}
\Delta\pi_1 =& \pi_1(\tilde{\mathcal{P}}') - \pi_1(\tilde{\mathcal{P}}) \\
=& \frac{1}{k}[\tilde{p}_2' - 2\frac{it}{k} - \tilde{p}_2 + 2\frac{i+l}{k}t] \\
&+ \tilde{p}_{1i}^{B'}[\frac{\tilde{p}_2' - \tilde{p}_{1i}^{B'} + t}{2t} - \frac{i+1}{k}] - \tilde{p}_{1i}^{B}[\frac{\tilde{p}_2 - \tilde{p}_{1i}^{B} + t}{2t} - \frac{i}{k}].
\end{aligned}$$

By definition, $\tilde{s}_i$ verifies the inequalities in Eq. 7, thus $\frac{\tilde{s}_i}{k} \leq \frac{\tilde{p}_2+t}{2t}$, which allows us to establish that $\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m} \frac{\tilde{s}_i}{2k}] \geq \frac{2t}{3nk}$. It is then immediate to show that:

$$\Delta\pi_1 \geq \frac{t}{k}[1 - \frac{1}{3n}][\frac{2}{k}\frac{3nl+1}{3n-1} - \frac{\tilde{p}_2}{2t} - \frac{1}{2} - \frac{1}{6nk} + \frac{i}{k} + \frac{1}{2k}].$$

Also, by assumption, firms compete on $J = [\frac{i}{k}, \frac{i+l}{k}]$ with $\tilde{\mathcal{P}}$, which implies that inequalities in Eq. 7 hold, and in particular, $\frac{\tilde{p}_2+t}{4t} - \frac{i}{2k} \leq \frac{l}{k}$.

Thus:

$$\Delta\pi_1 \geq \frac{t}{k}[1 - \frac{1}{3n}][\frac{2}{k}\frac{3nl+1}{3n-1} - \frac{2l}{k} - \frac{1}{6nk} + \frac{1}{2k}] \geq 0.$$

Profits on segment $[\frac{i}{k}, \frac{i+l+1}{k}]$ are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$.

Second we consider the profits of Firm 1 on the rest of the unit line. We write the reaction function of Firm 1 to an increase in the equilibrium price of Firm 2 ($\tilde{p}_2' \geq \tilde{p}_2$).

For segments of type A:

$$\frac{\partial}{\partial \tilde{p}_2}\pi_{1i}^A = \frac{\partial}{\partial \tilde{p}_2}(\frac{1}{k}[t + \tilde{p}_2 - 2\frac{u_i t}{k}]) = \frac{1}{k},$$

which means that a higher $\tilde{p}_2$ increases the profits.

For segments of type B:

$$\frac{\partial}{\partial \tilde{p}_2}\pi_{1i}^B = \frac{\partial}{\partial \tilde{p}_2}(p_{1i}[\frac{\tilde{p}_2 - \tilde{p}_{1i}^B + t}{2t} - \frac{s_i}{k}]) = \frac{\partial}{\partial \tilde{p}_2}(\frac{1}{2t}[\frac{\tilde{p}_2+t}{2} - \frac{s_i t}{k}]^2) = \frac{1}{2t}[\frac{\tilde{p}_2+t}{2} - \frac{s_i t}{k}],$$

which is greater than 0 as $\frac{\tilde{p}_2+t}{2} - \frac{s_i t}{k}$ is the expression of the demand on this segment, which is positive under Eq. 7.

Thus for any segment, the profits of Firm 1 increase with $\tilde{\mathcal{P}}'$ compared to $\tilde{\mathcal{P}}$.

Intermediary result 1: *By iteration, we conclude that type A segments are always at the left of type B segments.*

Step 3: We now analyze segments of type B where firms compete. Starting from any partition with at least two segments of type B, we show that it is always more profitable to sell a coarser partition.

As there are only two possible types of segments (A and B) and that we have shown that segments of type A are the closest to the firms, segment B is therefore further away from the firm. We prove the claim of step 3 by showing that if Firm 1 has a partition of two segments where it competes with Firm 2, a coarser partition softens competition between firms and yields a higher profit for Firm 1. We compute the profits of the firm on all the segments where firms compete, and compare the two situations described below with partition $\hat{\mathcal{P}}$ and partition $\hat{\mathcal{P}}'$.

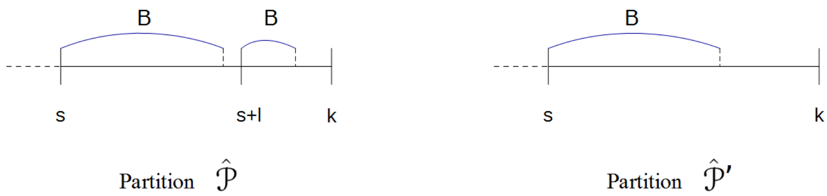

Figure 6: Step 3: demands of Firm 1 on segments of type B

Figure 6 depicts partition $\hat{\mathcal{P}}$ on the left panel, and partition $\hat{\mathcal{P}}'$ on the right panel. Partition $\hat{\mathcal{P}}$ divides the interval $[\frac{i}{k}, 1]$ in two segments $[\frac{i}{k}, \frac{i+l}{k}]$ and $[\frac{i+l}{k}, 1]$, whereas $\hat{\mathcal{P}}'$ only includes segment $[\frac{i}{k}, 1]$. We compare the profits of the firm on the segments where firms compete and we show that $\hat{\mathcal{P}}'$ induces higher profits for Firm 1. There are three types of segments to consider:

1. segments of type A that with partition $\hat{\mathcal{P}}$ that remain of type A with partition $\hat{\mathcal{P}}'$.

2. segments of type B with partition $\hat{\mathcal{P}}$ that are of type A with partition $\hat{\mathcal{P}}'$.

3. segments of type B with partition $\hat{\mathcal{P}}$ that remain of type B with partition $\hat{\mathcal{P}}'$.

1. Profits always increase on segments that are of type A with partitions $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}'$. Indeed, we show that $\hat{p}_2'$ with partition $\hat{\mathcal{P}}'$ is higher than $\hat{p}_2$ with partition $\hat{\mathcal{P}}$, and thus the profits of Firm 1 on type A segments increase.

2. There are $0 \leq m \leq n$ segments of type B in partition $\hat{\mathcal{P}}$ that are no longer of type B in partition $\hat{\mathcal{P}}$ (and are therefore of type A).

3. There are $n+1-m$ segments of type B with partition $\hat{\mathcal{P}}$ that remain of type B with partition $\hat{\mathcal{P}}'$. We compute prices and profits on these $n+1+m$ segments.

We proved in step 2 that prices can be written as:

$$\begin{aligned}
\hat{p}_2 &= -\frac{t}{3} + \frac{4t}{3(n+1)}\sum_{i=1}^{n+1}[\frac{s_i}{2k} + \frac{l_i}{k}],\\
\hat{p}^B_{1i} &= \frac{\hat{p}_2 + t}{2} - \frac{s_i t}{k}\\
&= \frac{t}{3} + \frac{2t}{3(n+1)}\sum_{i=1}^{n+1}[\frac{s_i}{2k} + \frac{l_i}{k}] - \frac{s_i t}{k}.
\end{aligned}$$

Let $\hat{p}^B_{1s}$ and $\hat{p}^B_{1s+l}$ be the prices on the last two segments when the partition is $\hat{\mathcal{P}}$.

$$\begin{aligned}
\hat{p}^B_{1s} &= \frac{\hat{p}_2 + t}{2} - \frac{st}{k},\\
\hat{p}^B_{1s+l} &= \frac{\hat{p}_2 + t}{2} - \frac{s+l}{k}t,
\end{aligned}$$

$\hat{p}'_2$ is the price set by Firm 2 with partition $\hat{\mathcal{P}}'$, and $\hat{p}^{B'}_{1s}$ is the price set by Firm 1 on the last segment of partition $\hat{\mathcal{P}}'$.

Inequalities in Eq. 7 might not hold as price $\hat{p}_2$ varies depending on the partition acquired by Firm 1. As $\hat{p}_2$ is greater with coarser partitions, some segments that are of type B with partition $\hat{\mathcal{P}}$ are then of type A with partition $\hat{\mathcal{P}}'$. We note $\tilde{s}_i$ the m segments for which it is the case. We then have:

$$\begin{aligned}
\hat{p}'_2 &= \frac{4t}{3(n-m)}[-\frac{n-m}{4} + \sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}] - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}]\\
&= \frac{4t}{3(n-m)}[-\frac{n+1}{4} + \sum_{i=1}^{n+1}[\frac{s_i}{2k} + \frac{l_i}{k}] + \frac{m+1}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k} - \frac{s+l}{2k}]\\
&= \hat{p}_2 + \frac{4t}{3(n-m)}[\frac{3(m+1)\hat{p}_2}{4t} + \frac{m+1}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k} - \frac{s+l}{2k}]\\
&\geq \hat{p}_2 + \frac{4t}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}],\\
\hat{p}^{B'}_{1s} &= \frac{\hat{p}_2 + t}{2} - \frac{st}{k},
\end{aligned}$$

$$\pi_1(\hat{\mathcal{P}}) = \sum_{i=1, s_i \neq \tilde{s}_i}^{n} p_{1i}[\frac{\hat{p}_2 + t}{4t} - \frac{s_i}{2k}] + \sum_{i=1}^{m} \hat{p}_{1i}^B[\frac{\hat{p}_2 + t}{4t} - \frac{\tilde{s}_i}{2k}] + \hat{p}_{1s+l}^B[\frac{\hat{p}_2 + t}{4t} - \frac{s+l}{2k}]$$

$$\pi_1(\hat{\mathcal{P}}') = \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B'}[\frac{\hat{p}_2' + t}{4t} - \frac{s_i}{2k}] + \sum_{i=1}^{m} \frac{\tilde{l}_i}{k}[\hat{p}_2' + t - 2t\frac{\tilde{s}_i + \tilde{l}_i}{k}].$$

We compare the profits of Firm 1 in both cases in order to show that $\hat{\mathcal{P}}'$ induces higher profits:

$$\begin{aligned}
\Delta\pi_1 &= \pi_1(\hat{\mathcal{P}}') - \pi_1(\hat{\mathcal{P}}) \\
&= \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B'}[\frac{\hat{p}_2' + t}{4t} - \frac{s_i}{2k}] - \sum_{i=1, s_i \neq \tilde{s}_i}^{n} \hat{p}_{1i}^{B}[\frac{\hat{p}_2 + t}{4t} - \frac{s_i}{2k}] \\
&+ \sum_{i=1}^{m} \frac{\tilde{l}_i}{k}[\hat{p}_2' + t - 2t\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \sum_{i=1}^{m} \hat{p}_{1i}^B[\frac{\hat{p}_2 + t}{4t} - \frac{\tilde{s}_i}{2k}] - \hat{p}_{1s+l}^B[\frac{\hat{p}_2 + t}{4t} - \frac{s+l}{2k}] \\
&= \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2' + t}{2t} - \frac{s_i}{k}]^2 - \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}]^2 \\
&+ \frac{t}{2} \sum_{i=1}^{m} \frac{\tilde{l}_i}{k}[2\frac{\hat{p}_2' + t}{t} - 4\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \frac{t}{2} \sum_{i=1}^{m} [\frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{2k}]^2 - \frac{t}{2}[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}]^2.
\end{aligned}$$

We consider the terms separately. First,

$$\begin{aligned}
&\frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2' + t}{2t} - \frac{s_i}{k}]^2 - \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}]^2 \\
=&\frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^{n} [[\frac{2}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]^2 \\
&+ [\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k}][\frac{4}{3(n-m)}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}]]] \\
\geq&\frac{t}{2}[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}].
\end{aligned} \tag{10}$$

Second, on segments of type B with partition $\hat{\mathcal{P}}$ that are of type A with partition $\hat{\mathcal{P}}'$:

$$\frac{t}{2} \sum_{i=1}^{m} \frac{\tilde{l}_i}{k}[2\frac{\hat{p}_2' + t}{t} - 4\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \frac{t}{2} \sum_{i=1}^{m} [\frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{2k}]^2.$$

On these m segments, inequalities in Eq. 7 hold for price $\hat{p}_2'$ and do not hold for price $\hat{p}_2$. We can rank prices according to $\tilde{s}_i$ and $\tilde{l}_i$:

$$\frac{\tilde{s}_i + \tilde{l}_i}{k} \geq \frac{\hat{p}_2 + t}{2t} - \frac{\tilde{l}_i}{k} \quad \text{and} \quad \frac{\hat{p}_2' + t}{2t} - \frac{\tilde{l}_i}{k} \geq \frac{\tilde{s}_i + \tilde{l}_i}{k}.$$

thus:

$$2\frac{\tilde{l}_i}{k} \geq \frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{k} \quad \text{and} \quad \frac{\hat{p}_2' + t}{2t} - 2\frac{\tilde{l}_i}{k} \geq \frac{\tilde{s}_i}{k}.$$

We replace $\tilde{s}_i$ by its upper bound value and then $\tilde{l}_i$ by its lower bound value. We can rewrite Eq. 10 for all permissible values of $\hat{p}_2'$:

$$\frac{t}{2}\sum_{i=1}^{m}\frac{\tilde{l}_i}{k}[2\frac{\hat{p}_2' + t}{t} - 4\frac{\tilde{s}_i + \tilde{l}_i}{k}] - \frac{t}{2}\sum_{i=1}^{m}[\frac{\hat{p}_2 + t}{2t} - \frac{\tilde{s}_i}{2k}]^2 \geq 0.$$

Getting back to the difference in profits, we obtain:

$$\begin{aligned}\Delta\pi_1 &\geq \frac{t}{2}[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}] - \frac{t}{2}[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}]^2 \\ &\geq \frac{t}{2}[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k}][\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} - \frac{1}{6}].\end{aligned} \tag{11}$$

The first bracket of Equation 11 is positive given Eq. 7. The second bracket is positive if $\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} \geq \frac{1}{6}$. A sufficient condition for this result to hold is $\hat{p}_2 \geq \frac{t}{3}$. We prove that this inequality is always satisfied by showing that the reference partition minimizes the price and profit of Firm 2, and that in this case, $\hat{p}_2 \geq \frac{t}{2}$.[32] And as this price is greater than $\frac{t}{3}$, the second bracket of Equation 11 is positive. This proves that $\Delta\pi_1 \geq 0$.

The price and profit of an uninformed firm are minimized when its competitor acquires $\mathcal{P}_{ref}$.

To prove this claim we consider Firm 1 informed and Firm 2 uninformed. We consider prices and demand on a segment of length $\frac{l}{k}$, $[\frac{s}{k}, \frac{s+l}{k}]$, and we show that partitioning this segment into two subsegments $[\frac{s}{k}, \frac{s+1}{k}]$ and $[\frac{s+1}{k}, \frac{s+l}{k}]$ reduces the price set by Firm 2 as well as it demand on $[\frac{s}{k}, \frac{s+l}{k}]$, which overall lowers its

[32]As shown in Liu and Serfes (2004).

profits. By iterating this argument, we can conclude that the reference partition $\mathcal{P}_{ref}$ minimizes the profit of the uninformed firm.

We have seen that we can write the equilibrium price set by Firm 2 with partition $\mathcal{P}$:

$$p_2 = -\frac{t}{3} + \frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]$$

This term is proportional to the average of $\frac{s_i}{2k} + \frac{l_i}{k}$'s. We show that this value is smaller with finer as partitions.

We rule out the case where Firm 1 is a monopolist on $[\frac{s}{k}, \frac{s+l}{k}]$, as prices and profit of Firm 2 do not change with finer subsegments in this case.

Consider the case where Firm 1 and Firm 2 compete on $[\frac{s}{k}, \frac{s+l}{k}]$. There are two cases to consider when partitioning this segment into two subsegments $[\frac{s}{k}, \frac{s+1}{k}]$ and $[\frac{s+1}{k}, \frac{s+l}{k}]$.

First, Firm 1 is a monopolist on $[\frac{s}{k}, \frac{s+1}{k}]$, and firms compete on $[\frac{s+1}{k}, \frac{s+l}{k}]$. The price set by Firm 2 with this second partition decreases as on segment $[\frac{s+1}{k}, \frac{s+l}{k}]$ we have $\frac{s}{2k} + \frac{l}{k} > \frac{s+1}{2k} + \frac{l-1}{k}$. It is clear that demand for Firm 2 also decreases as Firm 1 sets a price on $[\frac{s+1}{k}, \frac{s+l}{k}]$ instead of $[\frac{s}{k}, \frac{s+l}{k}]$. In reaction the aggregate profit of Firm 2 over the unit line decreases.

Secondly, Firm 1 and Firm 2 compete on $[\frac{s}{k}, \frac{s+1}{k}]$ and on $[\frac{s+1}{k}, \frac{s+l}{k}]$.

In order to show that the price set by Firm 2 decreases with this new partition, we compare the terms in the right hand side of the expression of price p_2: $\frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]$. This term is the average of $\frac{s_i}{2k} + \frac{l_i}{k}$ on the unit line. To prove that the price set by Firm 2 decreases, we need to show that this average is lower with the second partition than with the first one.

Consider a typical element $\frac{s}{2k} + \frac{l}{k}$ of $[\frac{s}{k}, \frac{s+l}{k}]$. Similarly, consider a typical element $\frac{1}{2}[\frac{s}{2k} + \frac{s+1}{2k} + \frac{l-1}{k} + \frac{1}{k}]$ of the finer partition $[\frac{s}{k}, \frac{s+1}{k}] \cup [\frac{s+1}{k}, \frac{s+l}{k}]$.

The first term is larger than the second as

$$\frac{s}{2k} + \frac{l}{k} > \frac{1}{2}[\frac{s}{2k} + \frac{s+1}{2k} + \frac{l-1}{k} + \frac{1}{k}].$$

It is clear that demand for Firm 2 also decreases as Firm 1 can better target consumers and compete more fiercely with finer segments. In reaction the aggregate profit of Firm 2 over the unit line are smaller with the finer partition than with the coarser one. This establishes the result.

To summarize it is always more profitable for the data intermediary to sell a partition with one segment of type B than to sell a partition with several segments of type B.

Conclusion

These three steps prove that the optimal partition includes two intervals, as illustrated in Figure 3. The first interval is composed of j segments of size $\frac{1}{k}$ located at $[0, \frac{j}{k}]$, and the second interval is composed of unidentified consumers, and is located at $[\frac{j}{k}, 1]$. ■

B.1.2 Proof of Lemma 5: the data intermediary sells information to both firms.

The data intermediary can maximize surplus extraction from firm by using the following selling mechanism. The data intermediary simultaneously auctions partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ in two separate auctions, since these partitions can be potentially different. Firm 1 (Firm 2) can bid in the two auctions but is only interested in Partition $\mathcal{P}_1$ ($\mathcal{P}_2$). Since both firms are guaranteed to obtain their preferred partition, they will underbid in both auctions from their true valuation. To avoid underbidding, the data intermediary respectively sets reserve prices w_1 and w_2 that correspond to the willingness to pay of Firm 1 for $\mathcal{P}_1$ and of Firm 2 for $\mathcal{P}_2$. Since partition $\mathcal{P}_2$ is optimal for Firm 2, Firm 1 will not bid above w_2 in the auction for $\mathcal{P}_2$ and similarly Firm 2 will not bid above w_1 in the auction for $\mathcal{P}_1$. Thus, the subgame perfect equilibrium is characterized by the following strategies: Firm 1 bids the reserve price w_1 for $\mathcal{P}_1$ and Firm 2 bids the reserve price w_2 for $\mathcal{P}_2$. The data intermediary thus maximizes the sum of the prices of information for Firm 1 and Firm 2. We now derive the optimal information structure.

Part a: optimal information structure when the data intermediary sells information to both firms

We prove that the partition described in Lemma 5 is optimal when information is sold to both firms. For each firm, the partition divides the unit line into two intervals. The first interval identifies the closest consumers to a firm and is partitioned in j segments of size $\frac{1}{k}$. The second interval is of size $1-\frac{j}{k}$ and leaves unidentified the other consumers.

Three types of segments are defined as before:

- Segments A, where Firm θ is in constrained monopoly;
- Segments B, where Firms 1 and 2 compete;
- Segments C, where Firm θ gets no demand.

We assume that the unit line is composed of one interval where firms compete, located at the middle of the line. Information structures that are ruled out by this assumption are those that allow firms to poach consumers located far away from their locations. Selling consumer segments far away from the location of a firm has two conflicting effects on the profits of the data intermediary. On the one hand, partitions ruled out by this assumption lower the valuation of the firms for information because they intensify competition in the market. On the other hand, these partitions also worsen the outside options of firms by lowering their profits if they remain uninformed, which increases their valuation for information. Hence the two effects go in opposite directions. Showing that the first effect dominates the second is not tractable without this assumption, given the high cardinality of the possible combinations of consumers segments. Additionally, there is no evidence of firm strategies targeting consumers who do not belong to their core market. On the contrary, the marketing literature has emphasized the benefits of targeting ads to consumer segments with the strongest preferences (Iyer et al., 2005). As we will show, the optimal partition under this assumption is similar to the optimal partition when the data intermediary sells information to one firm.

Inequalities in Eq. 7 characterize segments $[\frac{s_i}{k}, \frac{s_{i+1}}{k}]$ where both firms have positive demand:

$$\frac{s_i}{k} \leq \frac{p_2+t}{2t} \quad \text{and} \quad \frac{p_2+t}{2t} \leq \frac{2s_{i+1}-s_i}{k}.$$

The first part of Eq. 7 guarantees that there is positive demand for Firm 1, whereas the second part guarantees positive demand for Firm 2. Inequalities in Eq. 7 are expressed as a function of p_2 without loss of generality. We use Eq. 7 to characterize type A and type B segments, in order to compute the profits of the firms.

The profits of the data intermediary when it sells information to both firms is the difference between the profits of the firms when they are informed and their outside option, when they do not have information, but their competitor is informed:

$$\Pi_2 = (\pi_1^{I,I}(\mathcal{P}_1, \mathcal{P}_2) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2)) + (\pi_2^{I,I}(\mathcal{P}_1, \mathcal{P}_2) - \pi_2^{NI,I}(\emptyset, \mathcal{P}_1)).$$

Firm θ buys a partition composed of segments of type A and one segment of type B. To show that a partition in which type A segments are of size $\frac{1}{k}$ is optimal, we prove that 1) such a partition maximizes $\pi_{\mathcal{P},\theta}^{I,I}$ and 2) such a partition does not change $\pi_{\mathcal{P},\theta}^{NI,I}$.

1) *A partition which maximizes $\pi_{\mathcal{P},\theta}^{I,I}$ is necessarily composed of type A segments of size $\frac{1}{k}$.*

The proof of this claim is similar to step 1 of the proof in Appendix B.1 the price of the competing firm $-\theta$ does not change when Firm θ gets more precise information on type A segments, and the profits of Firm θ increase as it can target more precisely consumers with this information.

2) *Changing from a partition with type A segments of arbitrary size to a partition where type A segments are of size $\frac{1}{k}$ does not change $\pi_{\mathcal{P},\theta}^{NI,I}$.*

It is immediate to show that the profit of the uninformed firm does not depend on the fineness of type A segments. As a result, Π_2 is maximized when segments of type A are of size $\frac{1}{k}$.

We conclude that the optimal partition is composed of two intervals, sold to each firm. For Firm 1, the first interval is partitioned in j_1 segments of size $\frac{1}{k}$, and is located at $[0, \frac{j_1}{k}]$. Consumers are unidentified on the second interval of size $1 - \frac{j_1}{k}$ located at $[\frac{j_1}{k}, 1]$. For Firm 2, the first interval is partitioned in j_2 segments of size $\frac{1}{k}$, and is located at $[1 - \frac{j_2}{k}, 1]$. Consumers are unidentified on the second interval of size $1 - \frac{j_2}{k}$ located at $[0, 1 - \frac{j_2}{k}]$. ■

Part b: the data intermediary sells symmetric information to both firms

We show now that selling symmetric information is optimal for the data intermediary, that is, in equilibrium $j_1 = j_2$.

We compute prices and profits in equilibrium when both firms are informed with the optimal partition found above.

Firm 1 is a monopolist on the j_1 segments of size $\frac{1}{k}$ in $[0, \frac{j_1}{k}]$ and Firm 2 has information on $[1-\frac{j_2}{k}, 1]$. On $[\frac{j_1}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1. Similarly, on $[0, 1-\frac{j_2}{k}]$ Firm 2 sets a unique price p_2 and gets demand d_2.

We write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

Firm $\theta = 1, 2$ sets a price $p_{\theta i}$ for each segment of size $\frac{1}{k}$, and a unique price p_θ on the rest of the unit line. The demand for Firm θ on type A segments is $d_{\theta i} = \frac{1}{k}$. The corresponding prices are computed using the indifferent consumer located on the right extremity of the segment, $\frac{i}{k}$. For Firm 1:

$$\begin{aligned}
& V - t\frac{i}{k} - p_{1i} = V - t(1-\frac{i}{k}) - p_2 \\
\implies & \frac{i}{k} = \frac{p_2 - p_{1i} + t}{2t} \\
\implies & p_{1i} = p_2 + t - 2t\frac{i}{k}.
\end{aligned}$$

p_2 is the price set by Firm 2 on interval $[0, \frac{j_2}{k}]$ where it cannot identify consumers. Prices set by Firm 2 on segments in interval $[\frac{j_2}{k}, 1]$ are:

$$p_{2i} = p_1 + t - 2t\frac{i}{k}.$$

Let denote d_1 the demand for Firm 1 (resp. d_2 the demand for Firm 2) where firms compete. d_1 is found in a similar way as when information is sold to one firm, which gives us $d_1 = \frac{p_2-p_1+t}{2t} - \frac{j_1}{k}$ (resp. $d_2 = 1 - \frac{j_2}{k} - \frac{p_2-p_1+t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\pi_1 = \sum_{i=1}^{j_1} d_{1i}p_{1i} + d_1 p_1 = \sum_{i=1}^{j_1} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k})p_1,$$

$$\pi_2 = \sum_{i=1}^{j_2} d_{2i}p_{2i} + d_2 p_2 = \sum_{i=1}^{j_1} \frac{1}{k}(p_1 + t - 2t\frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{j_2}{k})p_2.$$

Step 3: prices, demands and profits in equilibrium.

We now compute the optimal prices and demands, using first order conditions on π_θ with respect to p_θ. Prices in equilibrium are:

$$p_1 = t[1 - \frac{2}{3}\frac{j_2}{k} - \frac{4}{3}\frac{j_1}{k}],$$

$$p_2 = t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}].$$

Replacing these values in the above demands and prices gives:

$$p_{1i} = 2t - \frac{4}{3}\frac{j_2 t}{k} - \frac{2}{3}\frac{j_1 t}{k} - 2\frac{it}{k},$$

$$p_{2i} = 2t - \frac{4}{3}\frac{j_1 t}{k} - \frac{2}{3}\frac{j_2 t}{k} - 2\frac{it}{k}.$$

and

$$d_1 = \frac{1}{2} - \frac{2}{3}\frac{j_1}{k} - \frac{1}{3}\frac{j_2}{k},$$

$$d_2 = \frac{4}{3}\frac{j_2}{k} - \frac{1}{2} - \frac{1}{3}\frac{j_1}{k}.$$

Profits are:

$$\pi_1^* = \sum_{i=1}^{j_1} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_1}{k} - \frac{1}{3}\frac{j_2}{k})t[1 - \frac{2}{3}\frac{j_2}{k} - \frac{4}{3}\frac{j_1}{k}]$$

$$= \frac{t}{2} - \frac{7}{9}\frac{j_1^2 t}{k^2} + \frac{2}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{2}{3}\frac{j_2 t}{k} - \frac{j_1 t}{k^2}.$$

$$\pi_2^* = \sum_{i=1}^{j_2} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_2}{k} - \frac{1}{3}\frac{j_1}{k})t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}]$$

$$= \frac{t}{2} - \frac{7}{9}\frac{j_2^2 t}{k^2} + \frac{2}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{2}{3}\frac{j_1 t}{k} - \frac{j_2 t}{k^2}.$$

The data intermediary maximizes the following profit function:

$$
\begin{aligned}
\Pi_2(j_1, j_2) &= (\pi_1^{I,I}(j_1, j_2) - \pi_1^{NI,I}(\emptyset, j_2)) + (\pi_2^{I,I}(j_1, j_2) - \pi_2^{NI,I}(\emptyset, j_1)) \\
&= -\frac{7}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{j_2 t}{k^2} - \frac{7}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{j_1 t}{k^2}.
\end{aligned}
$$

At this stage, straightforward FOCs with respect to j_1 and j_2 confirm that, in equilibrium, $j_1 = j_2$. The fact that the solution is a maximum is directly found using the determinant of the Hessian matrix. ■

B.2 Proof of Proposition 1

We compare profits when the data intermediary sells information to both firms and to Firm 1 only in the monopoly market, and we prove that the data intermediary sells information to Firm 1 only in equilibrium. We first compute the price of information in both cases. Then we find the optimal information structures, that is, the optimal number of segments sold to Firm 1 or to both firms. We show in particular that when the data intermediary sells information to both firms, it sells symmetric information structures, that is the same number of consumer segments to Firm 1 and to Firm 2. Then we derive prices of information in equilibrium and we show that profits are higher when selling information to Firm 1 only.

B.2.1 Profits of Firm 1 on the monopoly market.

We compute prices and profits in equilibrium when information is sold to one firm. Without loss of generality we consider the situation where Firm 1 is informed only. Firm 1 owns the optimal partition on $[0, \frac{j_1}{k}]$ that includes j_1 segments of size $\frac{1}{k}$, and has no information on consumers on $[\frac{j_1}{k}, 1]$.

We write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

On each segment of size $\frac{1}{k}$, Firm 1 sets a price p_{1i}, $i = 1, .., j_1$, and consumer demand is: $d_{1i} = \frac{1}{k}$. Let's p_2 denote the unique price set by Firm 2. Prices on each segment are determined by the indifferent consumer of each segment located at its right extremity, $\frac{i}{k}$:

$V - t\frac{i}{k} - p_{1i} = V - t(1-\frac{i}{k}) - p_2 \implies \frac{i}{k} = \frac{p_2 - p_{1i} + t}{2t} \implies p_{1i} = p_2 + t - 2t\frac{i}{k}$.

On the rest of the unit line, Firm 1 sets a price p_1 and competes with Firm 2. Firm 2 sets a unique price p_2 for all consumers on the segment $[0,1]$. We note d_1 the demand for Firm 1 on this segment, which is determined by the indifferent consumer:

$V - tx - p_1 = V - t(1-x) - p_2 \implies x = \frac{p_2 - p_1 + t}{2t}$ and $d_1 = x - \frac{j_1}{k} = \frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k}$.

Firm 2 sets p_2 and the demand, d_2, is found similarly to d_1, and $d_2 = 1 - \frac{p_2 - p_1 + t}{2t} = \frac{p_1 - p_2 + t}{2t}$.

Step 2: profits.

The profits of both firms can be written as follows:

$$\pi_1 = \sum_{i=1}^{j_1} d_{1i} p_{1i} + d_1 p_1 = \sum_{i=1}^{j_1} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k})p_1,$$
$$\pi_2 = d_2 p_2 = \frac{p_1 - p_2 + t}{2t} p_2.$$

Step 3: prices, demands and profits in equilibrium.

We solve prices and profits in equilibrium. First order conditions on π_θ with respect to p_θ give us $p_1 = t[1 - \frac{4}{3}\frac{j_1}{k}]$ and $p_2 = t[1 - \frac{2}{3}\frac{j_1}{k}]$. By replacing these values in profits and demands we find that: $p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k}]$, $d_1 = \frac{1}{2} - \frac{2}{3}\frac{j_1}{k}$ and $d_2 = \frac{1}{2} - \frac{1}{3}\frac{j_1}{k}$.

Profits are:[33]

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j_1} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j_1}{k})^2 \\ &= \frac{t}{2} + \frac{2j_1 t}{3k} - \frac{7t}{9}\frac{j_1^2}{k^2} - \frac{tj_1}{k^2} \\ \pi_2^* &= \frac{t}{2} + \frac{2t}{9}\frac{j_1^2}{k^2} - \frac{2}{3}\frac{j_1 t}{k}.\end{aligned} \tag{12}$$

[33]For $p_{1i} \geq 0 \implies \frac{j_1}{k} \leq \frac{3}{4}$. Profits are equal whatever $\frac{j_1}{k} \geq \frac{3}{4}$.

B.2.2 The data intermediary sells symmetric information structures to Firm 1 and Firm 2.

We show now that selling symmetric information is optimal for the data intermediary, that is, in equilibrium $j_1 = j_2$.

We compute prices and profits in equilibrium when both firms are informed with the optimal partition found above.

Firm 1 is a monopolist on the j_1 segments of size $\frac{1}{k}$ in $[0, \frac{j_1}{k}]$ and Firm 2 has information on $[1-\frac{j_2}{k}, 1]$. On $[\frac{j_1}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1, similarly on $[0, 1-\frac{j_2}{k}]$ Firm 2 sets a unique price p_2 and gets demand d_2.

We write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

Firm $\theta = 1, 2$ sets a price $p_{\theta i}$ for each segment of size $\frac{1}{k}$, and a unique price p_θ on the rest of the unit line. The demand for Firm θ on segments of size $\frac{1}{k}$ is $d_{\theta i} = \frac{1}{k}$. The corresponding prices are computed using the indifferent consumer located on the right extremity of the segment, $\frac{i}{k}$. For Firm 1:

$$\begin{aligned} & V - t\frac{i}{k} - p_{1i} = V - t(1-\frac{i}{k}) - p_2 \\ \Longrightarrow & \frac{i}{k} = \frac{p_2 - p_{1i} + t}{2t} \\ \Longrightarrow & p_{1i} = p_2 + t - 2t\frac{i}{k}. \end{aligned}$$

p_2 is the price set by Firm 2 on interval $[0, \frac{j_2}{k}]$ where it cannot identify consumers. Prices set by Firm 2 on segments in interval $[\frac{j_2}{k}, 1]$ are:

$$p_{2i} = p_1 + t - 2t\frac{i}{k}.$$

Let denote d_1 the demand for Firm 1 (resp. d_2 the demand for Firm 2) where firms compete. d_1 is found in a similar way as when information is sold to one firm, which gives us $d_1 = \frac{p_2-p_1+t}{2t} - \frac{j_1}{k}$ (resp. $d_2 = 1 - \frac{j_2}{k} - \frac{p_2-p_1+t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\pi_1 = \sum_{i=1}^{j_1} d_{1i}p_{1i} + d_1 p_1 = \sum_{i=1}^{j_1} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j_1}{k})p_1,$$
$$\pi_2 = \sum_{i=1}^{j_2} d_{2i}p_{2i} + d_2 p_2 = \sum_{i=1}^{j_1} \frac{1}{k}(p_1 + t - 2t\frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{j_2}{k})p_2.$$

Step 3: prices, demands and profits in equilibrium.

We now compute the optimal prices and demands, using first order conditions on π_θ with respect to p_θ. Prices in equilibrium are:

$$p_1 = t[1 - \frac{2}{3}\frac{j_2}{k} - \frac{4}{3}\frac{j_1}{k}],$$
$$p_2 = t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}].$$

Replacing these values in the above demands and prices gives:

$$p_{1i} = 2t - \frac{4}{3}\frac{j_2 t}{k} - \frac{2}{3}\frac{j_1 t}{k} - 2\frac{it}{k},$$
$$p_{2i} = 2t - \frac{4}{3}\frac{j_1 t}{k} - \frac{2}{3}\frac{j_2 t}{k} - 2\frac{it}{k}.$$

and

$$d_1 = \frac{1}{2} - \frac{2}{3}\frac{j_1}{k} - \frac{1}{3}\frac{j_2}{k},$$
$$d_2 = \frac{4}{3}\frac{j_2}{k} - \frac{1}{2} - \frac{1}{3}\frac{j_1}{k}.$$

Profits are:

$$\pi_1^* = \sum_{i=1}^{j_1} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_1}{k} - \frac{1}{3}\frac{j_2}{k})t[1 - \frac{2}{3}\frac{j_2}{k} - \frac{4}{3}\frac{j_1}{k}]$$
$$= \frac{t}{2} - \frac{7}{9}\frac{j_1^2 t}{k^2} + \frac{2}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{2}{3}\frac{j_2 t}{k} - \frac{j_1 t}{k^2}.$$

$$\pi_2^* = \sum_{i=1}^{j_2} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j_2}{k} - \frac{1}{3}\frac{j_1}{k})t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}]$$
$$= \frac{t}{2} - \frac{7}{9}\frac{j_2^2 t}{k^2} + \frac{2}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{2}{3}\frac{j_1 t}{k} - \frac{j_2 t}{k^2}.$$

The data intermediary maximizes the following profit function:

$$\begin{aligned}\Pi_2(j_1,j_2) &= (\pi_1^{I,I}(j_1,j_2) - \pi_1^{NI,I}(\emptyset,j_2)) + (\pi_2^{I,I}(j_1,j_2) - \pi_2^{NI,I}(\emptyset,j_1))\\ &= -\frac{7}{9}\frac{j_2^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_2 t}{k} - \frac{j_2 t}{k^2} - \frac{7}{9}\frac{j_1^2 t}{k^2} - \frac{4}{9}\frac{j_1 j_2 t}{k^2} + \frac{2}{3}\frac{j_1 t}{k} - \frac{j_1 t}{k^2}.\end{aligned}$$

At this stage, straightforward FOCs with respect to j_1 and j_2 confirm that, in equilibrium, $j_1 = j_2$. The fact that the solution is a maximum is directly found using the determinant of the Hessian matrix. ■

B.2.3 Firms' profits on the competitive market.

Firm 1 is a monopolist on the j_1 segments of size $\frac{1}{k}$ in $[0, \frac{j_1}{k}]$ and Firm 2 has symmetric information, composed of j_1 segments of size $\frac{1}{k}$ on $[1-\frac{j_1}{k}, 1]$. On $[\frac{j_1}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1, similarly on $[0, 1-\frac{j_1}{k}]$ Firm 2 sets a unique price p_2 and gets demand d_2.

We do not go through the computation of prices and demand which have already been described, and we directly give prices and profits in equilibrium.

Prices in equilibrium are $p_1 = p_2 = t[1-2\frac{j_1}{k}]$, $p_{\theta i} = 2t[1-\frac{j_1}{k}-\frac{i}{k}]$ and $d_\theta = \frac{1}{2}-\frac{j_1}{k}$.

Profits are:[34]

$$\begin{aligned}\pi_\theta^* &= \sum_{i=1}^{j_1} \frac{2t}{k}[1-\frac{i}{k}-\frac{j_1}{k}] + \frac{1}{2}(1-2\frac{j_1}{k})^2 t\\ &= \frac{t}{2} - \frac{j_1^2}{k^2}t - \frac{j_1 t}{k^2}.\end{aligned}$$

■

B.2.4 The data intermediary sells information to Firm 1 only on the monopoly market

In this section we compute the optimal numbers of segment sold by taking j_1 and j_2 continuous. The optimal number of segments sold will then be the integer closest to the optimum found in the continuous case.

[34]For $\frac{j_1}{k} < \frac{1}{2}$. Profits are equal as soon as $\frac{j_1}{k} > \frac{1}{2}$.

We determine the optimal size j_1^* when the data intermediary only sells information to Firm 1, by maximizing profits with respect to j_1. When the data intermediary sells information to both firms, we determine the optimal number j_2^* in a similar way. We then compare the maximized profits of the data intermediary to find whether it sells information to one or to both firms in equilibrium.

1) Optimal partition when the data intermediary sells information to one firm.

The profits of the data intermediary when it sells to one firm are:[35]

$$\begin{aligned}\Pi_1(j) = w_1(j) &= \pi^{I,NI}(j,\emptyset) - \pi^{NI,I}(\emptyset,\mathcal{P}_{ref})\\ &= \frac{3t}{8} + \frac{2jt}{3k} - \frac{t}{4k} - \frac{7j^2t}{9k^2} - \frac{jt}{k^2} - \frac{t}{8k^2}.\end{aligned}$$

FOC on j leads to the following maximizing value: $j^* = \frac{6k-9}{14}$ and:

$$\Pi_1^* = \frac{29t}{56} - \frac{19t}{28k} + \frac{11t}{56k^2}.$$

2) Optimal partition when the data intermediary sells information to both firms.

We maximize the profit function with respect to the j segments sold to Firm 1 and Firm 2. The profits of the data intermediary when both firms are informed are:

$$\Pi_2(j) = 2w_2 = 2[\frac{2jt}{3k} - \frac{11j^2t}{9k^2} - \frac{jt}{k^2}].$$

FOC on j leads to $j^* = \frac{6k-9}{22}$ and:

$$\Pi_2^* = \frac{2t}{11} - \frac{6t}{11k} + \frac{9t}{22k^2}.$$

3) DI's selling strategy in equilibrium.

[35] The expression of $\pi^{NI,I}(\emptyset,\mathcal{P}_{ref})$ is provided in Liu and Serfes (2004).

We compare the profits of the data intermediary when it sells information to one firm or to both firms. The difference between the two profits is:

$$\Pi_1^* - \Pi_2^* = \frac{(207k^2 - 82k - 131)t}{616k^2}.$$

which is positive for any $k \geq 2$.

■

B.3 Proofs of Propositions 3 and 2

Proofs of Proposition 3 (a) and (b).

We consider $\overline{k} \geq \underline{k}$. We show that in equilibrium data intermediary $\overline{DI}$ with the highest information precision $\overline{k}$ sells information to both firms, while $\underline{DI}$ with information $\underline{k}$ does not sell information. Firms can either buy no information, or information from a competing data intermediary. The information acquisition game can be described by the Nash table, for clarity we focus on information acquisition from data intermediaries $\overline{DI}$ and $\underline{DI}$ with the two best information precision:

		Firm 1		
		NI	$\overline{DI}$	$\underline{DI}$
Firm 2	NI	$(\pi_1(\emptyset,\emptyset),\pi_2(\emptyset,\emptyset))$	$(\pi_1(\overline{j_1},\emptyset),\pi_2(\emptyset,\overline{j_1}))$	$(\pi_1(\underline{j}_1,\emptyset),\pi_2(\emptyset,\underline{j}_1))$
	$\overline{DI}$	$(\pi_1(\emptyset,\overline{j_2}),\pi_2(\overline{j_2},\emptyset))$	$(\pi_1(\overline{j_1},\overline{j_2}),\pi_2(\overline{j_2},\overline{j_1}))$	$(\pi_1(\underline{j}_1,\overline{j_2}),\pi_2(\overline{j_2},\underline{j}_1))$
	$\underline{DI}$	$(\pi_1(\emptyset,\underline{j}_2),\pi_2(\underline{j}_2,\emptyset))$	$(\pi_1(\overline{j_1},\underline{j}_2),\pi_2(\underline{j}_2,\overline{j_1}))$	$(\pi_1(\underline{j}_1,\underline{j}_2),\pi_2(\underline{j}_2,\underline{j}_1))$

Data intermediaries propose information to both firms and, contrary to monopoly markets, there is no exclusivity contract in equilibrium. Assume the opposite. Then, only one firm acquires information from one data intermediary, and the other firms cannot purchase information from the same intermediary. Then the other data intermediary can make positive profits by selling information to the uninformed firm.

In this situation, each data intermediary has interest to propose information to both firms. Each firm acquires information from the data intermediary that maximizes its profit. Necessarily, firms acquire information from data intermediary $\overline{DI}$ since any partition that data intermediary $\underline{DI}$ can propose, data intermediary $\overline{DI}$ can propose too.

Proofs of Propositions 2 and 3 (c).

We now prove that the number of consumers identified by firms is larger on the monopoly market than on competitive markets. This number is defined on the monopoly market by $\frac{j_1^m(k)}{k} = \frac{6k-9}{14k}$, and on the competitive market by $\frac{\overline{j_1}(k)+\overline{j_2}(k)}{k}$, for any k, k'. We first need to compute the optimal number of segments sold on the competitive market to characterize the share of identified consumers in this case.

Optimal information structure on the competitive market.

We compute the number of segments proposed by data intermediaries $\overline{DI}$ and $\underline{DI}$ to Firm 1 and Firm 2 in equilibrium. On the competitive market, data intermediary $\overline{DI}$ with the highest information precision $\overline{k}$ maximizes its profits by maximizing with respect to $\overline{j_1}$ and $\overline{j_2}$ the sum:

$$p_l(\overline{j_1}, \overline{j_2}) + p_l(\overline{j_2}, \overline{j_1}) = \pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2}) + \pi_2(\overline{j_2}, \overline{j_1}) - \pi_2(\underline{j_2}, \overline{j_1})$$

$$= \frac{(7\underline{k}\overline{k}(\underline{j_2})^2 + (4\underline{k}\overline{j_1} - 6\underline{k} + 9)\overline{k}\underline{j_2} + 7\underline{k}\overline{k}(\underline{j_1})^2 + (4\underline{k}\overline{j_2} - 6\underline{k} + 9)\overline{k}\underline{j_1})t}{9\underline{k}\overline{k}}$$
$$+ \frac{((-7\underline{k}(\overline{j_2})^2 + (6\underline{k} - 8\underline{k}\overline{j_1})\overline{j_2} - 7\underline{k}(\overline{j_1})^2 + 6\underline{k}\overline{j_1})\overline{k} - 9\underline{k}\overline{j_2} - 9\underline{k}\overline{j_1})t}{9\underline{k}\overline{k}} \tag{13}$$

Data intermediary $\underline{DI}$ with the second-best information precision competes à la Bertrand with $\overline{DI}$. It exerts the maximal competitive pressure by proposing respectively to Firm 1 and Firm 2 information partitions $\underline{j_1}$ and $\underline{j_2}$ that maximize their profits $\pi_1(\underline{j_1}, \overline{j_2})$ and $\pi_2(\underline{j_2}, \overline{j_1})$.

By replacing variables $\underline{j_1}$ and $\overline{j_2}$ into π_1 (and respectively for π_2), we obtain the following expressions:

$$\pi_1(\underline{j_1}, \overline{j_2}) = \frac{t}{2} - \frac{7}{9}\frac{(\underline{j_1})^2 t}{\underline{k}^2} + \frac{2}{9}\frac{(\overline{j_2})^2 t}{\overline{k}^2} - \frac{4}{9}\frac{\underline{j_1}\overline{j_2} t}{\underline{k}\overline{k}} + \frac{2}{3}\frac{\underline{j_1} t}{\underline{k}} - \frac{2}{3}\frac{\overline{j_2} t}{\overline{k}} - \frac{\underline{j_1} t}{\underline{k}^2}$$

$$\pi_2(\underline{j_2}, \overline{j_1}) = \frac{t}{2} - \frac{7}{9}\frac{(\underline{j_1})^2 t}{\underline{k}^2} + \frac{2}{9}\frac{(\overline{j_1})^2 t}{\overline{k}^2} - \frac{4}{9}\frac{\underline{j_2}\overline{j_1} t}{\underline{k}\overline{k}} + \frac{2}{3}\frac{\underline{j_2} t}{\underline{k}} - \frac{2}{3}\frac{\overline{j_1} t}{\overline{k}} - \frac{\underline{j_2} t}{\underline{k}^2}$$

Data intermediary $\underline{DI}$ maximizes simultaneously these two profit functions with respect to $\underline{j_1}$ and $\underline{j_2}$. Simultaneously, data intermediary $\overline{DI}$ maximizes $p_l(\overline{j_1}, \overline{j_2}) + p_l(\overline{j_2}, \overline{j_1})$ with respect to $\overline{j_1}$ and $\overline{j_2}$.

Thus equilibrium variables $\overline{j_1}, \overline{j_2}, \underline{j_1}, \underline{j_2}$ are chosen as simultaneous best response. FOCs on $\overline{j_1}$, $\overline{j_2}$, $\underline{j_1}$ and $\underline{j_2}$ give respectively in equilibrium:

$$\overline{j_1}^* = \overline{j_2}^* = \frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}$$

$$\underline{j_1}^* = \underline{j_2}^* = \frac{1}{3} - \frac{11}{18\underline{k}} + \frac{1}{9\overline{k}}$$

We show that more consumers are identified in the competitive market than in monopoly markets by comparing $\frac{\overline{j_1}(k)+\overline{j_2}(k)}{k} = 2[\frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}]$ with $\frac{j_1^m(k')}{k'} = \frac{6k'-9}{14k'}$:

$$2[\frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}] - \frac{6k'-9}{14k'} = \frac{((30\underline{k}-28)k' + 81\underline{k})\overline{k} - 98\underline{k}k'}{126\underline{k}k'\overline{k}}$$

which is clearly positive for $k', \underline{k}, \overline{k} \geq 2$. ■

B.4 Proof of Proposition 4

B.4.1 Proof of Proposition 4 (a)

We show that the optimal number of segments collected by an intermediary is larger in a monopoly market than in the competitive market.

We first write the price of information on the competitive market. We substitute the values of $\overline{j_1}^*$, $\overline{j_2}^*$ and $\underline{j_1}^*$ in $\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})$ in the profit functions of Firm 1 and Firm 2. The price of information is

$$p_l(\overline{k}, \underline{k}) = [\pi_1(\overline{j_1}, \overline{j_2}) - \pi_1(\underline{j_1}, \overline{j_2})]$$
$$= \frac{((12\underline{k} - 11)\overline{k}^2 + (4\underline{k} - 12\underline{k}^2)\overline{k} + 7\underline{k}^2)t}{36\underline{k}^2\overline{k}^2} \tag{14}$$

which increases in $\overline{k}$.

When selling information to Firm 1 on the monopoly market, a monopolist data intermediary i has revenue

$$p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}$$

with marginal revenue equal to:

$$\frac{\partial p_m(k_i)}{\partial k_i} = \frac{3t}{7k_i^2} - \frac{9t}{14k_i^3}$$

We prove that the optimal number of segments collected is larger for $p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}$ than on the competitive market where the total revenue of $\overline{DI}$ is $2p_l(k_l)$.

For $k_1 > k_2$, we have

$$p_l(k_1) = \frac{((12k_2 - 11)k_1^2 + (4k_2 - 12k_2^2)k_1 + 7k_2^2)t}{36k_1^2k_2^2}$$

$$2\frac{\partial p_l(k_1)}{\partial k_1} = \frac{((6k_1 - 2)k_2 - 7k_1)t}{9k_1^3k_2}$$

$$p_m(k_i) = \frac{t}{7} - \frac{3t}{7k_i} + \frac{9t}{28k_i^2}.$$

and

$$\frac{\partial p_m(k_i)}{\partial k_i} = \frac{3t}{7k_i^2} - \frac{9t}{14k_i^3} \geq \frac{\partial p_l(k_1)}{\partial k_1} = \frac{((6k_1 - 2)k_2 - 7k_1)t}{9k_1^3k_2}$$

Consider k^* such that

$$2\frac{\partial p_l(k_1)}{\partial k_1}|_{k_1=k^*} = \frac{((6k^* - 2)k_2 - 7k^*)t}{9k^{*3}k_2} = \frac{\partial c(k_1)}{\partial k_1}|_{k_1=k^*}$$

Since

$$\frac{\partial^2 p_m(k_\gamma)}{\partial k_\gamma^2} = \frac{27t}{14k^4} - \frac{6t}{7k^3}) \leq 0$$

for $k \in [2, \infty[$

revenues are concave, and necessarily, $\tilde{k}^*$ such that

$$\frac{\partial p_{m_1}(k_1)}{\partial k_1} = \frac{\partial c(k_1)}{\partial k_1}$$

verifies $\tilde{k}^* \leq k^*$.

The optimal amount of consumer data collection is higher in monopoly markets than in the competitive market l. ■

B.4.2 Proof of Proposition 4 (b)

We prove that the number of consumer segments $\overline{k}$ collected by $\overline{DI}$ increases with the number of consumer segments $\underline{k}$ collected by $\underline{DI}$. To do so, we show that the concavity of the price of information $p_l(k_1, k_2)$ on the competitive market l increases with k_2.

Consider the second degree derivative of p_l with respect to k_1 and k_2:

$$\frac{\partial^2 p_l(k_1, k_2)}{\partial k_1 \partial k_2} = \frac{1}{9k_1^2 k_2^2} \geq 0$$

Thus, the larger the k_2, the larger the value of the first degree derivative of p_l with respect to k_1, and the higher the marginal gain from collecting data. An increase of k_2 will thus increase the value of k_1^* in equilibrium.

B.5 Proof of Proposition 5

We first prove that the number of consumer segments collected by data intermediaries increases with m_i. We first show that collecting information with the highest precision is an equilibrium for DI_1 and not for other data intermediaries. We then show that k_1^* decreases with k_2. Finally, we show that for $i > 1$, k_i^* decreases with m_i.

B.5.1 Proof of Proposition 5 (a)

Let's $k_1 = argmax\{m_1 p_m(k) - (m_1 + l)c(k)\}$. Such a maximum necessarily exists, otherwise $m_1 p_m(k) + 2lp_l(k) - (m_1 + l)c(k)$ has no maximum either, and data intermediary 1 always collects an infinite amount of information. This is a degenerate scenario that we rule out from our analysis as we focus on data collection strategies of data intermediaries. Note that as $m_1 > m_i$, necessarily $k_i = argmax_k\{m_i p_m(k) - (m_i + l)c(k)\}$ verifies $k_i < k_1$. We assume that $m_1, .., m_n$ and c are such that $m_i p_m(k_i) + 2lp_l(k_i, k_1) - (m_i + l)c(k_i)$ has a unique maximum $\forall i$.

As $k_1 = argmax\{m_1 p_m(k) - (m_1 + l)c(k)\}$ and $m_1 > m_2$, necessarily $k_2 = argmax_k\{m_2 p_m(k) + 2lp_l(k, k_1) - (m_2 + l)c(k)\}$ verifies $k_2 < k_1$. In particular $\max_k\{m_2 p_m(k) - (m_2 + l)c(k)\} \geq \max_k\{m_2 p_m(k) + 2lp_l(k, k_1) - (m_2 + l)c(k)\}$, and deviation to $k_2 > k_1$ is never profitable for firm 2.

B.5.2 Proof of Proposition 5 (b)

Consider data intermediaries that are symmetric in terms of market size $m_1 = m_2$ (the reasoning generalizes easily to any number of intermediaries). In this case there is no symmetric pure strategy Nash equilibrium in consumer data collection. Indeed, consider a symmetric equilibrium under which both intermediaries collect the same amount of consumer data k^*. Necessarily we have $mp'_m(k^*) - (m + l)c'(k^*) = 0$. However, a data intermediary has interest do deviate from this situation by increasing consumer data collection and make $\pi(k_i) = mp_m(k_i) + 2lp_l(k_i, k^*) - (m + l)c(k_i)$, whose first degree derivative is strictly positive at $k^* + \epsilon$ with ϵ very small.

We now prove that asymmetric equilibrium can exist, in which one intermediary collects $\hat{k}^*$ that maximizes $\pi(k_i)$ and the other collects k^* that maximizes the profits on its monopoly market m $mp_m(k^*) - (m + l)c(k^*)$, with $\hat{k}^* > k^*$.[36]

Clearly the data intermediary with the highest precision has no interest to deviate since its profits are maximized at $\hat{k}^*$. The other intermediary has interest to

[36]By concavity of $\pi(k_i)$ with respect to k_i over $[k_i^*, \infty[$, there exists an optimal $\hat{k}^*$ that maximizes $\pi(k_i)$, and there exists k^* that maximizes $mp_m(k^*) - (m + l)c(k^*)$.

deviate and to collect $\tilde{k}^* > \hat{k}^*$, that is, to collect more data than the intermediary with the highest precision, if the profits of doing so is greater than its monopoly profits:

$$mp_m(\tilde{k}^*) + 2lp_l(\tilde{k}^*, \hat{k}^*) - (m+l)c(\tilde{k}^*) \geq mp_m(k^*) - (m+l)c(k^*). \tag{15}$$

If Equation 15 is satisfied, deviation to $\tilde{k}^*$ is profitable for intermediary j, and intermediary i has interest to collect k^* data. In this case, $\tilde{k}^*$ is not an equilibrium, and no equilibrium exist.

Thus, there exists an asymmetric equilibrium if Equation 15 is not satisfied. In this case, there are two equilibrium of the game. In each equilibrium, one data intermediary collects k^*, and the other collects $\hat{k}^*$.

Thus asymmetry arises naturally even when data intermediaries are symmetric in terms of market size and of data collection costs. This has strong implications for competition in digital markets. Even when data intermediaries are symmetric, asymmetry arises in equilibrium and one of the company dominates the other.

B.6 Proof of Proposition 6

We show that consumer surplus always increases with the number of consumer segments sold. Consumer surplus when Firm 1 has j_1 consumer segments and Firm 2 has j_2 consumer segments is defined as follows:

$$
\begin{aligned}
CS(j_1, j_2, k) = & \sum_{i=1}^{j_1}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
& + \int_{\frac{j_1}{k}}^{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}} V - t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] - tx\mathrm{d}x + \int_{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}}^{1-\frac{j_2}{k}} V - t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}] - tx\mathrm{d}x \\
& + \sum_{i=1}^{j_2}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
& = \sum_{i=1}^{j_1} \frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}]) - \frac{j_1 t}{2k^2} \\
& + \sum_{i=1}^{j_2} \frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}]) - \frac{j_2 t}{2k^2} \\
& + V[1 - \frac{j_2}{k} - \frac{j_1}{k}] - [\frac{1}{2} - \frac{2j_1}{3k} - \frac{j_2}{3k}]t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] \\
& - [\frac{1}{2} - \frac{2j_2}{3k} - \frac{j_1}{3k}]t[1 - \frac{4}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] - t[\frac{1}{4} - \frac{1}{9}\frac{j_1 j_2}{k^2} - \frac{7}{18}\frac{j_2^2}{k^2} - \frac{7}{18}\frac{j_1^2}{k^2}] \\
& = \frac{j_1}{k}[V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + \frac{j_1(j_1+1)t}{k^2} - \frac{j_1 t}{2k^2} \\
& + \frac{j_2}{k}[V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + \frac{j_2(j_2+1)t}{k^2} - \frac{j_2 t}{2k^2} \\
& + V[1 - \frac{j_2}{k} - \frac{j_1}{k}] + t[-\frac{5}{4} + \frac{1}{3}\frac{j_1}{k} + \frac{1}{3}\frac{j_2}{k} + \frac{5}{6}\frac{j_1^2}{k^2} + \frac{5}{6}\frac{j_2^2}{k^2} - 2\frac{j_1 j_2}{k^2}] \\
& = V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{k^2} + \frac{17}{18}\frac{j_2^2}{k^2} + \frac{j_1 j_2}{k^2}] + \frac{1}{2}\frac{j_1 t}{k^2} + \frac{1}{2}\frac{j_2 t}{k^2}
\end{aligned} \tag{16}
$$

The first degree derivative with respect to j_1 is

$$\frac{\partial CS}{\partial j_1} = \frac{17 j_1}{9k} + \frac{j_2}{k} + \frac{1}{2k}$$

which is larger than zero, for $\frac{j_1}{k} \geq -\frac{18j_2+9}{34k}$, that is, it is always above zero. ■

B.7 Proof of Proposition 7

We consider the first degree derivative of CS with respect to k, for given $\frac{j_1}{k}$, $\frac{j_2}{k}$: $\frac{\partial CS}{\partial k} = -\frac{j_1 t}{k^3} - \frac{j_2 t}{k^3}$.

This is clearly always negative, and consumer surplus always decreases with information precision. ■

B.8 Proof of Proposition 8

We compare consumer surplus on the competitive market l and on monopoly markets m_i.

First in markets m_1 and l the same number of consumer segments are collected by intermediary 1, and more consumer segments are sold in market l than in market m_1. It is thus immediate that consumer surplus is higher in market l than in market m_1.

In markets $m_i < m_1$, fewer consumer segments are collected which drives up consumer surplus compared to market m_1. We show that nevertheless, consumer surplus is higher in market l because a larger share of consumers are identified, and thus that the increase of surplus due to consumer identification overcome the reduction of surplus due to more segments collected.

In market l,

$$\begin{aligned} CS_l(\overline{j_1}^*(\overline{k}) &= \frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}, \overline{j_2}^*(\overline{k}) \\ &= \frac{1}{3} - \frac{1}{9\underline{k}} - \frac{7}{18\overline{k}}, \overline{k}) \\ &\geq V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{\overline{k}^2} + \frac{17}{18}\frac{j_2^2}{\overline{k}^2} + \frac{j_1 j_2}{\overline{k}^2}] \\ &\geq V + t[-\frac{5}{4} + \frac{17}{18}\frac{9}{121} + \frac{17}{18}\frac{9}{121} + \frac{9}{121}] \end{aligned} \tag{17}$$

In market m_i:

$$CS_{m_i}(j_1, k_i) = V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{k_i^2}] + \frac{1}{2}\frac{j_1 t}{k_i^2} \leq V + t[-\frac{5}{4} + \frac{17}{18}\frac{9}{121}] + \frac{1}{2}\frac{3t}{121^2}$$

It is straightforward that $CS_l > CS_{m_i} \forall k_i$, and surplus is higher on the competitive market than on monopoly markets.

■

B.9 Proofs of Section 6: the impacts of mergers and acquisitions on data collection

B.9.1 Merger between data intermediaries 1 and 2:

All things equal, the incentives of the merged intermediary to collect data increase compared to intermediaries 1 and 2 before the merger as the size of the monopoly market of the merged entity is $m_1 + m_2 > m_1 > m_2$. Thus the marginal gains from collecting data are higher, and the merger increases data collection by the merged entity.

A second effect that decreases data collection comes from lower competition in market l as the second-best intermediary after the merger is 3. Thus the profits of the merged entity on the competitive market lose concavity as $k_3 < k_2$, and the incentives of the merged entity to collect data.

Both effects go in opposite directions, and data collection will increase only if the second one is smaller than the first one, that is, if the merged entity is sufficiently large compared with data intermediary 3.

B.9.2 Data intermediary 2 acquires a small data intermediary i:

We focus on the case where the merged entity remains smaller than intermediary 1. The size of the merged entity increases, and thus it has more incentives to collect data than intermediaries 2 and i. As the competitive pressure on intermediary 1 increases, so does the concavity of profits in market l, which increases its incentive to collect data.

B.9.3 Data intermediary 1 acquires a small data intermediary i:

All things equal, the incentives of the merged intermediary to collect data are higher than those of intermediary 1 before the merger as the size of the monopoly market of the merged entity is $m_1 + m_i > m_1 > m_i$. Thus the marginal gains from collecting data are higher, and the merger increases data collection by the merged entity.

B.9.4 Two small intermediaries i, j merge:

In this case where $m_i + m_j < m_2$, the only effect is that data collection increases on $m_i + m_j$.