



HAL
open science

Competition and Mergers with Strategic Data Intermediaries

David Bounie, Antoine Dubus, Patrick Waelbroeck

► **To cite this version:**

David Bounie, Antoine Dubus, Patrick Waelbroeck. Competition and Mergers with Strategic Data Intermediaries. 2022. hal-03336520v2

HAL Id: hal-03336520

<https://hal.science/hal-03336520v2>

Preprint submitted on 1 Mar 2022 (v2), last revised 28 Apr 2023 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Competition and Mergers with Strategic Data Intermediaries*

David Bounie,[†] Antoine Dubus[‡] and Patrick Waelbroeck[§]

February 28, 2022

Abstract

We analyze mergers between strategic data intermediaries collecting consumer information that they sell to firms competing in a product market. We show that a merger: (a) reduces the intensity of competition in the product market through a change in the selling strategies of merging intermediaries; (b) increases data collection, reducing consumer surplus through a better rent extraction. We argue that the role of Big Tech companies acting as strategic data intermediaries in the market for information should be included in antitrust analysis.

*We thank Andreea Cosnita-Langlais, Doh-Shin Jeon, Ulrich Laitenberger, Patrick Legros, Winston Maxwell and Flavio Pino as well as participants at the DG Competition Economic Seminar, the Annual Scientific Seminar on Media and the Digital 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 Policies 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 Cartes Bancaires CB, La Banque Postale, la Caisse des Dépôts, Telecom Paris and University of Paris 2 Panthéon - Assas.

[†]i3, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; david.bounie@telecom-paris.fr.

[‡]Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 Switzerland – 8092 Zürich, Switzerland; antoine1dubus@gmail.com.

[§]i3, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; patrick.waelbroeck@telecom-paris.fr.

1 Introduction

In recent years, Big Tech companies such as Alphabet, 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). For the year 2017 alone, Alphabet, Amazon, Apple, Facebook and Microsoft spent a total of \$31.6bn on acquisitions of start-ups.¹ 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).”

Economists have recently 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). The debate centers first on whether more Big Tech acquisitions should be reviewed by antitrust authorities and, secondly, on whether additional or different theories of harm should be developed (Motta and Peitz, 2021). The most common arguments against Big Tech mergers are the elimination of potential competition (Tirole, 2020), or the acquisition of strategic assets such as talents and intellectual property rights (Gans and Stern, 2003).

Surprisingly, however, their dominant position as strategic data intermediaries has been overlooked by antitrust authorities. Major Big Tech companies such as Google, Amazon, and Facebook act indeed as data intermediaries on a new market for information (Bergemann and Bonatti, 2019). They extensively collect information on consumers, which they provide to firms that compete in product markets. In this article, we therefore analyze how a merger between strategic data intermediaries may impact competition in related product markets.

We identify two indirect effects of data-driven mergers on competition and consumers surplus in product markets, which we argue, should be included in antitrust analysis of Big Tech mergers.

¹American tech giants are making life tough for startups; *The Economist*, June 2 2018.

The first indirect effect results from the incentives of data intermediaries to collect information on consumers, which change after a merger. Intermediaries have interest to collect information that allows firms to increase their profits through a better extraction of consumer surplus. This rent extraction effect of information increases the willingness to pay of firms, and a merger that increases data collection will lower consumer surplus through this first indirect effect of information in product markets.

The second indirect effect is related to the ability of data intermediaries to sell information to firms strategically. They 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, a merger that reduces competition in the market for information can also soften competition in product markets through this second indirect effect of information.

We include these indirect effects in a model of competition between data intermediaries collecting and selling information to firms seeking to price-discriminate consumers in a 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. Each intermediary has a local monopoly market and has exclusive access to consumer data in this market. Intermediaries also collect data on a competitive market where consumer information is available to other intermediaries. This framework allows us to analyze how mergers that change the incentives of intermediaries to collect and sell data in each market may reduce consumer surplus.

This article achieves four main results regarding how a merger in the market for information impacts competition and consumer surplus in related product markets. The first two results are related to the data collection strategies of intermediaries. First, mergers between data intermediaries increase the profitability of information, which also heightens the incentives of the merged intermediary to collect data. More data collected reduces consumers surplus in related product markets.

Secondly, we show that the equilibrium in the competitive market has the structure of a monopoly facing a competitive fringe. In this equilibrium, there exists an escape-competition effect: the incentives of the dominant intermediary to collect consumer information increase with the intensity of the competitive pressure,² which will both reduce consumer surplus in the competitive market and in its local monopoly market.

Overall, the first two results related to data collection show that intermediaries collect more consumer data after a merger. Accounting for strategic data collection in merger analysis is essential to fully assess the effects of a merger on consumer surplus in related product markets. This result is new and goes beyond existing academic literature that does not consider the data collection strategies of competing intermediaries.

The last two results are related to the selling strategies of data intermediaries. We establish that a merger between data intermediaries decreases the amount of information sold in the product market when they have a better ability to internalize the competitive effect of information after a merger.

First, we show that when data intermediaries compete, they sell more consumer segments to firms than under monopoly. Hence, mergers in the market for information can reduce competition in related product markets through this second indirect effect: intermediaries sell information on fewer consumers after a merger and firms compete less fiercely in the product market. Nevertheless, we show that even under fierce competition, intermediaries do not sell all available

²Aghion et al. (2005) find a similar escape-competition effect when two firms compete neck-to-neck, i.e. when they have the same technology. On the contrary, here the escape-competition effect applies to the leading technological firm.

consumer segments, but keep low-valuation consumers unidentified to soften competition between information buyers. This effect is robust to symmetric Bertrand competition between information sellers, and highlights a fundamental inefficiency of markets for information, as a large share of collected data is not used by intermediaries and firms. This is a striking finding that challenges recent results in the literature on markets for information, which assumes that intermediaries sell all available information to prospective buyers in a frictionless environment (Bergemann and Bonatti, 2015; Bergemann et al., 2018; Montes et al., 2019; Bergemann and Bonatti, 2019; Ichihashi, 2021).

Secondly, we find that data intermediaries may engage into exclusionary practices after a merger, as 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 and a higher intensity of competition in the product market. Hence mergers also harm competition in product markets by providing a data advantage to one of the firms over its competitor, which reduces consumer surplus. This result contributes to the burgeoning literature on markets for information that considers a monopolist data intermediary selling information to only one of the information buyers (Montes et al., 2019; Bounie et al., 2021), and to models where it is assumed that a data intermediary sells information to all active 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.³

The remainder of this article is organized as follows. In Section 2 we describe the model. We characterize the data strategies of intermediaries in the competitive market in Section 3 and in their monopoly markets in Section 4. We analyze mergers between intermediaries in Section 5, and we show how they impact competition and consumer surplus in product markets. Section 6 concludes.

³See 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 and mergers between data intermediaries that collect and sell customer data for price-discrimination purposes. Our analysis applies to mergers involving companies such as Facebook and Google that industry observers and academic scholars consider as data intermediaries (Venkatadri et al., 2018).⁴

For the purpose of their online advertising activities, it is indeed well known that Big Tech companies collect data to improve the value of their services. However, they also directly or indirectly sell consumer data to firms or third-parties through position auctions and real-time bidding. This has led Google users to sue the company in the U.S..⁵

In this section, we first characterize the nature of competition between data intermediaries. We then describe 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 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 other intermediaries cannot access, which grants it a local monopoly power. For instance, Facebook collects data on its

⁴See also Facebook and Google Are the New Data Brokers; Digital Life Initiative, January 6 2021.

⁵Google faces \$5 billion lawsuit in U.S. for tracking 'private' internet use; Reuters, June 3 2020.

users, which other data intermediaries cannot sell 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.⁶ We will refer to this market as competitive market l in which all intermediaries compete. Our $m - l$ approach addresses the Diamond paradox and explicitly draws the frontier between rival and non-rival data.

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 its monopoly market (with $i = 1, \dots, n$), and on a market of mass l where all intermediaries compete.⁷ As a consequence, 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 $\mu = m_1 + \dots + m_n + l$.⁸ By convention, data intermediary 1 has a larger monopoly market than data intermediary 2 and so on: $m_1 \geq m_2 \dots \geq m_n$.⁹

Our $m - l$ approach is an important methodological contribution to the literature. This flexible framework allows us to characterize different degrees of competition depending on the respective sizes of the monopoly markets and the competitive market. To the best of our knowledge, the literature has not yet analyzed the data collection and selling strategies of intermediaries in a competitive framework, precisely because of the conceptual challenge raised by the Diamond

⁶Ichihashi (2021) considers a different framework in which consumers sell their (non-rival) personal information to intermediaries. In our framework, firms collect consumer data without compensating consumers, as for the vast majority of business models based on the collection of consumer information, and the non-rivalry of data implies that firms will compete à la Bertrand for the sale of information.

⁷We analyze in Section 4 a situation in which data intermediaries only collect and sell information on their monopoly market.

⁸We 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 Sections 3 and 4.

⁹As a special case, we will also allow for symmetric data intermediaries in terms of size of their monopoly markets: $m_1 = m_2 \dots = m_n$; we will show that they collect different amounts of information in the only equilibrium of the game.

information paradox. We will show that the properties of the equilibrium in the competitive market differ radically from those of the monopoly markets, which impacts significantly consumer surplus in product markets where intermediaries compete.

2.2 Consumers

Consumers are divided into $n + 1$ mutually exclusive markets of different sizes such that $\mu = m_1 + \dots + m_n + l$. Each market is characterized by a Hotelling line $[0, 1]$ on which consumers are uniformly distributed. On each line, a consumer can buy one product at a price p_1 from Firm 1 located at 0, or p_2 from Firm 2 located at 1.¹⁰

A consumer located at $x \in [0, 1]$ derives a utility V from purchasing the product. They incur 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. Hence on each unit line, consumers located at x have a utility function defined by:¹¹

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

2.3 Data Intermediaries

Data intermediaries collect information that divides each market into consumer segments.¹² More data is costly to collect but allows an intermediary to have a finer partition of a unit line. Partitions sold by data intermediaries enable firms

¹⁰We assume that all markets are covered. This assumption is common in the literature. See for instance [Thisse and Vives \(1988\)](#), [Liu and Serfes \(2004\)](#), [Stole \(2007\)](#), [Ulph and Vulkan \(2000\)](#), [Montes et al. \(2019\)](#), and [Bounie et al. \(2021\)](#).

¹¹Where p_1 and p_2 can be potentially different on different unit lines, and for distinct locations of a given unit line.

¹²Other modeling choices are possible, such as the sale of a signal analyzed by [Admati and Pfleiderer \(1988\)](#); [Bergemann et al. \(2018\)](#) among others, in the case of a monopolist information seller. Our focus on information that partitions a Hotelling line is motivated by the tractability of the model, yet which provides us with a rich set of strategy space for data intermediaries.

to identify consumers and price discriminate them. We describe in this section the data collection and selling strategies of intermediaries on different markets.

2.3.1 Collecting Data

A data intermediary i uses an appropriate technology to collect data on $m_i + l$ consumers.¹³ 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 . A data intermediary collects data points such as gender, age, or zip-code, which allows it to partition consumer demand into k segments of size $\frac{1}{k}$.¹⁴

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}^k that we refer to as the reference partition.



Figure 1: Reference partition \mathcal{P}^k

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 $[\frac{1}{2}, 1]$.

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

¹³Several technologies allow firms to collect data on consumers such as cookies and pixels (Bergemann and Bonatti, 2015; Choe et al., 2018). Cookies used for analytics can provide information on the websites visited by consumers, and allow a firm to know whether consumers have visited the websites of its competitors.

¹⁴We drop index i when there is no confusion.

The cost of collecting data is equal to $c(k)$ for a mass one of consumers, and satisfies standard convexity conditions.¹⁵ 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 . 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 in turn its willingness to pay for information and the price of information.

2.3.2 Selling Information

Data intermediaries can sell any combination of segments of the consumer demand. We denote by $\mathcal{P}_1(DI_i)$ and $\mathcal{P}_2(DI_i)$ the partitions offered to Firm 1 and Firm 2 respectively by intermediary i on a given market. These partitions can be potentially different for each firm, and can also differ between monopoly market m_i and the competitive market l . Selling strategic information will allow us to capture the second indirect effect of competition between intermediaries on competition and consumer surplus in the product market.

Bounie et al. (2021) have shown that a monopolist 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 partition thus maximizes consumer surplus extraction while softening the competitive effect of information. We will see how a merger between data intermediaries increases their 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

¹⁵We assume that $c(0) = 0$ and $c(x), c''(x) > 0$.

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 also 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, while still allowing Firm 1 to extract more surplus from identified consumers close to its location.

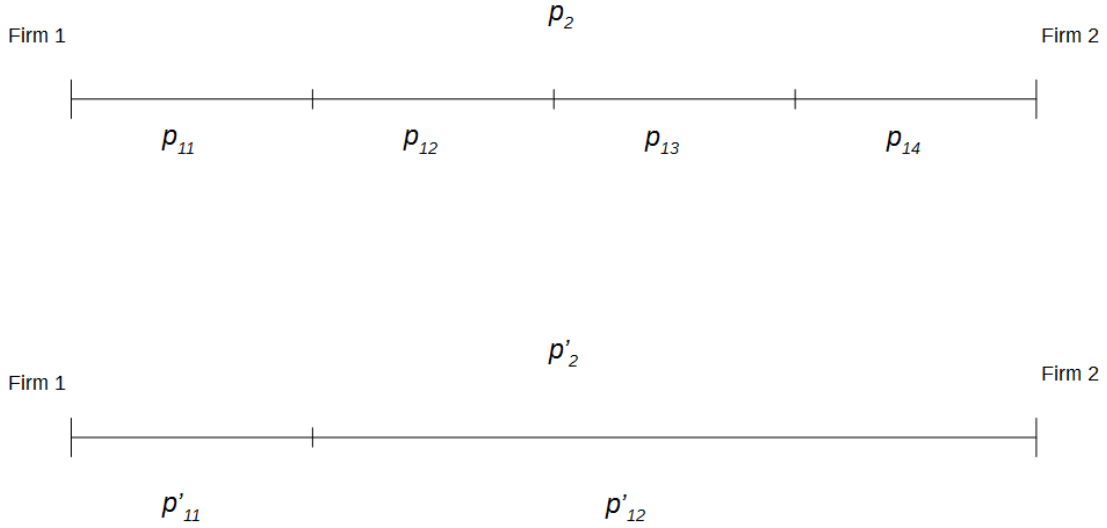


Figure 2: Example of partitions, $k = 4$

In the competitive market, data intermediaries compete à la Bertrand in the sale of information. Since firms only purchase information from one data intermediary,¹⁶ they choose intermediary \overline{DI} with the highest information precision, so that competition in market l leads to a winner takes all situation. Other data intermediaries collect data in market l , but the second-best data intermediary \underline{DI} exerts a competitive pressure on the equilibrium prices set by \overline{DI} .

In monopoly market m_i , data intermediary i sells information to one or two

¹⁶We assume that firms cannot combine information purchased from different data intermediaries.

firms. We assume that data intermediaries sell information through first-price auctions. This selling mechanism is commonly used in the literature (Montes et al., 2019; Bounie et al., 2021), and allows an intermediary to reach the first-best outcome by maximizing surplus extraction from firms.¹⁷

2.4 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 in the competitive market l . When a firm acquires an information partition, it knows to which interval of this partition a consumer belongs. Firms simultaneously set prices on each segment of the unit line where they have information. Firm θ sets prices in two stages.¹⁸ 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. On a given Hotelling line, each firm knows whether its competitor is informed, and the partition $\mathcal{P}_{-\theta}$.¹⁹

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

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

¹⁷Auctions 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)).

¹⁸Sequential 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.

¹⁹This assumption is also standard in Braulin and Valletti (2016) and Montes et al. (2019).

2.5 Timing

A data intermediary i first collects data and sells potentially different partitions to firms in market m_i and 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 intermediary i collects data on k_i consumer segments in markets m_i and l .
- Stage 2: data intermediary i offers information partitions for sale in its monopoly market and in the competitive market.
- Stage 3: firms set prices p_1 and p_2 in the competitive segments of each market.
- Stage 4: firms set prices p_{θ_i} on consumer segments on which they have information.

We analyze in the next sections the data collection and selling strategies of intermediaries in the competitive market, and then in their respective monopoly markets. Analyzing first competition and then monopoly allows us to understand the different impacts of a merger on the data strategies of intermediaries, which we analyze in Section 5.

3 Competitive Market

We characterize the selling strategies of intermediaries in the competitive market, as well as their incentives to collect consumer data.²⁰

Intermediaries compete à la Bertrand in prices and information structures in the competitive market, and firms purchase information from only one buyer. We denote by \overline{DI} the intermediary with the highest information precision \bar{k} , and by \underline{DI} the second-best intermediary with information precision \underline{k} . We characterize the equilibrium in Proposition 1, which we derive in the remaining of the section.

²⁰To simplify the exposition, when analyzing the selling strategies of intermediaries we consider a market of size 1. We will see that the selling strategy does not depend on the size of the market.

Proposition 1

An equilibrium in competitive market l has the following properties:

- \overline{DI} sells information to both firms; other data intermediaries do not sell information.
- The optimal partitions sold to Firm 1 and Firm 2 divide the unit line into two intervals:
 - For Firm 1, 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.
 - For Firm 2, the second interval consists of j_2 segments of size $\frac{1}{k}$ on $[1 - \frac{j_2}{k}, 1]$ where consumers are identified; consumers in the second interval of size $1 - \frac{j_2}{k}$ are unidentified.
- The optimal numbers of segments sold to each firms are:

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

- The incentives of \overline{DI} to collect information increase with the intensity of the competitive pressure exerted by \underline{DI} .

Proposition 1 shows that data intermediary \overline{DI} internalizes the competitive effect of information by selling to each firm information partitions on their high-valuation consumers, keeping low-valuation consumers unidentified. Such selling strategy softens the competitive effect of information, and reduces consumer surplus compared to a case where firms have information on all consumers.

3.1 Selling Consumer Information

We determine the profits of the firms, the price of information, the optimal selling strategy and the information structure offered by intermediaries in the competitive market.

Profits of the firms in the competitive market.

We denote by $\pi_1(\mathcal{P}_1(DI_i), \mathcal{P}_2(DI_j))$ and $\pi_2(\mathcal{P}_2(DI_j), \mathcal{P}_1(DI_i))$ the profits of Firm 1 and Firm 2 when they respectively acquire partitions $\mathcal{P}_1(DI_i)$ from data intermediary i and $\mathcal{P}_2(DI_j)$ from data intermediary j . Both firms can purchase information from the same intermediary ($i = j$).

Intermediaries compete à la Bertrand in prices and information structures in the competitive market, and firms purchase information from only one buyer. Hence, each intermediary will charge the highest price for its information, determined by the difference of profits of each firm with their partitions and with the partition that maximizes their profits among those proposed by the remaining intermediaries.

As firms compete a la Bertrand, only intermediary \overline{DI} with the highest information precision \bar{k} will sell information in market l . Other data intermediaries make zero profit in this market. Nevertheless, data intermediary \underline{DI} exerts a competitive pressure and limits the ability of \overline{DI} to extract rent from firms.

Prices of information in the competitive market.

Data intermediary \overline{DI} sells information to Firm 1 and Firm 2, at prices corresponding to the willingness to pay of each firm for information. Consider the incentive of Firm 1 to purchase information. Firm 1 can acquire $\mathcal{P}_1(\overline{DI})$ at price \bar{p}_1 from \overline{DI} , or $\mathcal{P}_1(\underline{DI})$ at price \underline{p}_1 from \underline{DI} and make profit $\pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI}))$. The willingness to pay of Firm 1 for information is thus $\pi_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) - \pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI}))$. The price that Firm 2 is ready to pay for information is defined in a similar way: $\pi_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) - \pi_2(\mathcal{P}_2(\underline{DI}), \mathcal{P}_1(\overline{DI}))$. Lemma 1 summarizes this discussion.

Lemma 1

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

$$\left\{ \begin{array}{l} \bar{p}_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) = \pi_1(\mathcal{P}_1(\overline{DI}), \mathcal{P}_2(\overline{DI})) - \pi_1(\mathcal{P}_1(\underline{DI}), \mathcal{P}_2(\overline{DI})), \\ \text{and} \\ \bar{p}_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) = \pi_2(\mathcal{P}_2(\overline{DI}), \mathcal{P}_1(\overline{DI})) - \pi_2(\mathcal{P}_2(\underline{DI}), \mathcal{P}_1(\overline{DI})). \end{array} \right.$$

Optimal information structure

Data intermediary \overline{DI} chooses the partitions that maximize the profits of the firms by combining segments of the reference partition $\mathcal{P}^{\bar{k}}$. For instance, in Figure 2 the data intermediary can combine segments 2, 3 and 4 to sell to Firm 1 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 segments far away from a firm will only increase the competitive effect 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 partitions $\mathcal{P}_1^*(\overline{DI})$ and $\mathcal{P}_2^*(\overline{DI})$, represented in Figure 3.

Lemma 2

An optimal partition $\mathcal{P}_1^(\overline{DI})$ for Firm 1 (and $\mathcal{P}_2^*(\overline{DI})$ for Firm 2) divides the unit line into two intervals:*

- *The first interval consists of \overline{j}_1 (\overline{j}_2) segments of size $\frac{1}{k}$ on $[0, \frac{\overline{j}_1}{k}]$ ($[\frac{1 - \overline{j}_2}{k}, 1]$) where consumers are identified.*
- *Consumers in the second interval of size $1 - \frac{\overline{j}_1}{k}$ ($1 - \frac{\overline{j}_2}{k}$) are unidentified.*

Proof: see Appendix A.1.

The optimal partitions divide the unit line into two intervals. Firms can price discriminate identified consumers, and charge a uniform price on the second interval of unidentified consumers. Data intermediary \overline{DI} does not sell all consumer segments 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 segments increases competition and reduces the willingness to pay of firms for information. Partitions $\mathcal{P}_1^*(\overline{DI})$ and $\mathcal{P}_2^*(\overline{DI})$ balance the competition and surplus extraction effects of information. Similarly, \underline{DI} will offer to Firm 1 and Firm 2 partitions composed of \underline{j}_1 and \underline{j}_2 segments closest to their locations.

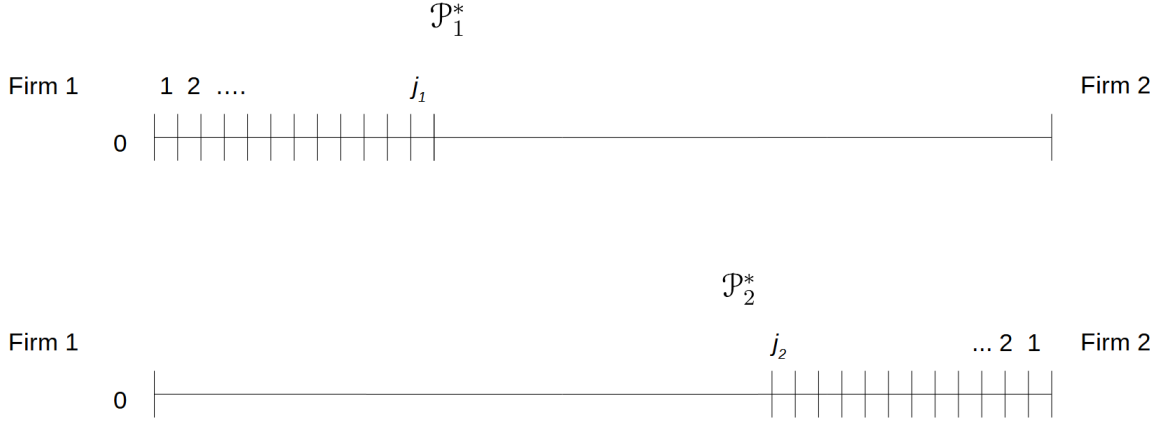


Figure 3: Selling partition \mathcal{P}_1^* to Firm 1

Hence, the optimization problem for data intermediary \overline{DI} in the competitive market boils down to choosing \overline{j}_1 and \overline{j}_2 under the constraint exerted by intermediary \underline{DI} .

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:

$$\max_{\overline{j}_1, \overline{j}_2} \{\overline{p}_1(\overline{j}_1, \overline{j}_2) + \overline{p}_2(\overline{j}_2, \overline{j}_1)\}. \quad (3)$$

The optimal partitions are:

$$\overline{j}_1^*(\overline{k}) = \overline{j}_2^*(\overline{k}) = \frac{\overline{k}}{3} - \frac{\overline{k}}{9\underline{k}} - \frac{7}{18}.$$

Proof: see Appendix A.2.

Proposition 2 shows that selling the reference partition is not optimal. Instead, data intermediary \overline{DI} has incentives to only sell a subset of all available segments

to soften the competitive effect of information and to maximize the willingness of firms to pay for information. Close-by consumers with a high willingness to pay are identified and the remaining consumers remain unidentified.

A direct comparison with Bertrand competition could lead to the conclusion that in the competitive market, intermediary \overline{DI} would sell all its available information and each firm would have information $\mathcal{P}^{\bar{k}}$. However, this reasoning is incorrect, as data intermediaries have incentives to deviate from such equilibrium. Selling fewer segments than $\mathcal{P}^{\bar{k}}$ unambiguously increases the profits of a data intermediary. By showing that data intermediaries do not sell all their available data under competition, Proposition 2 makes therefore an important contribution to the literature.

3.2 Collecting Consumer Data

We analyze in this section the incentives of intermediary \overline{DI} to collect consumer data in the competitive market. Clearly other intermediaries make zero profit in this market, and overall, they have incentives to collect data only through the profits they make on their respective monopoly markets, and which we analyze in Section 4.2.

Remember that data intermediaries cannot distinguish to which market consumers belong before collecting their data, and the number of consumer segments collected by each data intermediary i is identical on m_i and l . For this reason, we analyze in this section the incentives of intermediaries to collect data in the competitive market.

In 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

$$\Pi_l(k) = l[\bar{p}_1(k) + \bar{p}_2(k) - c(k)].$$

²¹The concavity of the cost function ensure that the profit functions are strictly concave with a unique maximum at \bar{k}^* .

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

Collecting more segments increases the prices of information charged by \overline{DI} , as finer segments allow firms to better price-discriminate consumers on whom they have information. Moreover, the prices of information also depend on the competitive pressure exerted by \underline{DI} , which depends itself on \underline{k} . An important determinant of the data collection strategy of the leading intermediary \overline{DI} is therefore the amount of data collected by its direct competitor \underline{DI} in the competitive market. There is an escape-competition effect in data collection: as the competitive pressure exerted by \underline{DI} on \overline{DI} increases – in the sense that \underline{k} becomes closer to \overline{k} – the incentives of \overline{DI} to differentiate by collecting more segments increase as well.²² Proposition 3 summarizes this discussion.

Proposition 3

Escape-competition effect in data collection: the incentives of \overline{DI} to collect consumer segments in the competitive market increase with \underline{k} .

Proof: See Appendix A.3.

Thus, a merger that increases the competitive pressure exerted on data intermediary \overline{DI} in market l , will also increase its incentives to collect consumer data. We apply this result in Section 5, to assess the impacts of mergers between intermediaries on consumers in the product market, and we show that accounting for this effect can overturn standard merger analysis.

4 Monopoly Markets

We now characterize the selling strategies of data intermediaries in their monopoly markets, as well as their incentives to collect consumer data.

Proposition 4 characterizes the equilibrium, which we prove and discuss in the remaining of this section (we drop subscripts i , m_i and DI_i from the notations in this section):

²²This effect is also reminiscent of models of innovation in a competitive environment (Aghion et al., 2005).

Proposition 4

In monopoly market m_i , the selling strategies of data intermediary i are characterized by the following properties:

- Intermediary i sells information to Firm 1 only and Firm 2 remains uninformed.
- The optimal partition \mathcal{P}_1^* sold to Firm 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.
- Intermediary i sells $j_1^*(k) = \frac{6k-9}{14}$ segments to Firm 1.
- Firms identify fewer consumers than in the competitive market.
- The incentives to collect data are higher than in the competitive market.

Proposition 4 shows that a monopolist data intermediary can fully internalize the competitive effect of information by selling information to only one firm, keeping its competitor uninformed, and by selling fewer segments than in the competitive market. This selling strategy softens the competitive effect of information, and reduces consumer surplus compared to a case where both firms have information.

4.1 Selling Consumer Information

We determine the price of information, the optimal selling strategy and the information structure sold by an intermediary in a monopoly market when selling information to Firm 1 only. This allows us to fully characterize the equilibrium in monopoly markets as we then show that the profits of an intermediary in the monopoly market are always higher when selling information to one firm than when selling information to both firms.

Price of information in a monopoly market.

The data intermediary sells information to only one firm, say Firm 1.²³ Let $\pi_1(\mathcal{P}^k, \emptyset)$ and $\pi_2(\mathcal{P}^k, \emptyset)$ 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 $\pi_1(\emptyset, \mathcal{P}^k)$ and $\pi_2(\emptyset, \mathcal{P}^k)$ 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 4.

Lemma 3

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

$$p_1^{m*} = \max_{\mathcal{P}_1} \{ \pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) \}. \quad (4)$$

Optimal information structure.

The partition that maximizes the price of information given by Lemma 3 is similar to the partitions sold in the competitive market: consumers located close to Firm 1 are identified, and far-away consumers are kept unidentified to soften the competitive effect of information. In Lemma 4, we characterize the features of this optimal partition \mathcal{P}_1^* , represented in Figure 4.

Lemma 4

In a monopoly market, the 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.*

²³We characterize the equilibrium when selling information to both firms in the Appendix and we show that it yields lower profits than when selling to Firm 1 only.

- Consumers in the second interval of size $1 - \frac{j_1}{k}$ are unidentified.

Proof: see Appendix A.4.

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 effect of information.



Figure 4: Selling partition \mathcal{P}_1^* to Firm 1

Optimal selling strategy in monopoly markets.

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 5

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

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

Fewer consumers are identified than in the competitive market:

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

Proof: see Appendix A.5.

Lemma 5 shows that the objective functions of the data intermediary and Firm 1 are aligned: they both want j_1^* to maximize $\pi_1(j_1^*, \emptyset)$.²⁴ Firm 2 cannot acquire information from other data intermediaries, since we focus on monopoly markets where only one data intermediary has information.

Selling information to one or to both firms.

The data intermediary compares its profits when selling information to one or to both firms. Lemma 6 shows that the data intermediary will only sell information to Firm 1 on its monopoly market (Firm 2 remains uninformed).

Lemma 6

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

Proof: see Appendix A.6.

Lemma 6 states that data intermediaries optimally sell information only to Firm 1 in 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, 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 high price of information by threatening Firm 1 to remain uninformed with profits $\pi_1(\emptyset, \mathcal{P}^k)$.

4.2 Consumer Data Collection: Monopoly Markets vs Competitive Market

We analyze in this section the data collection strategies of intermediaries in their monopoly markets, and we compare their incentives to collect data with those in the competitive market. We then rank the number of consumer segments k_i collected by each data intermediary i in equilibrium.

A data intermediary makes the following profit on its monopoly market given by the price of information net of the data collection cost, times the size of the monopoly market m :

²⁴The integer value of j_1^* that maximizes the profits of the data intermediary is chosen by comparing $\pi(|j_1^*|)$ and $\pi(|j_1^*| + 1)$: $\max(\pi(|j_1^*|), \pi(|j_1^*| + 1))$.

$$\Pi_m(k) = m[p_{1m}(k) - c(k)].$$

As only intermediary \overline{DI} makes positive profits in the competitive market, all other intermediaries have higher incentives to collect data on their monopoly markets. Hence we only need to show that the incentives of intermediary \overline{DI} to collect consumer data are higher in its monopoly market than in the competitive market.

Data collection has two effects on the price of information and on the profits of a data intermediary in its monopoly market, which change the incentives of \overline{DI} to collect data compared with the competitive market. First, similarly to the competitive market, more data increases rent extraction as firms can better price-discriminate consumers on thinner segments of the demand. This effect is stronger when intermediaries sell information to firms on a larger share of the consumer demand: the marginal gains from having finer segments are higher when more consumers are identified. Secondly, more precise information lowers the profits of an uninformed firm facing an informed competitor. Since in a monopoly market, a data intermediary threatens a prospective buyer to sell the reference partition (that includes all available consumer segments) to its competitor, more segments collected increase the value of this threat.

Competition has three impacts on the incentives of \overline{DI} to collect data in market l compared to its monopoly market. First, we have shown in Proposition 4 that \overline{DI} sells consumer segments on a larger share of consumers in the competitive market than in its monopoly market, and the marginal gains of an increase in \bar{k} are greater in market l according to this first effect. Secondly, \overline{DI} has an additional incentive to collect data in the competitive market, due to the escape-competition effect in data collection described in Proposition 3. Thirdly, \overline{DI} cannot exert a threat on firms in market l , because of the competitive pressure exerted by \underline{DI} . Hence an increase in \bar{k} will have no impact on the outside option of a firm in market l , and this third effect lowers the incentives of \overline{DI} to collect data in the competitive market. Overall, we show in Proposition 5 that the last effect dominates the first two effects and \overline{DI} has higher incentives to collect data in its monopoly market

than in the competitive market.

Proposition 5

Data intermediaries have higher incentives to collect data in their monopoly markets than in the competitive market.

Proof: See Appendix A.7.

Proposition 5 shows that the marginal gains from data collection are higher in monopoly markets than in the competitive market. This is a central result of this article that has important implications for competition policy, which we analyze in Section 5. While Proposition 5 is intuitive, we have shown that the incentives of \overline{DI} to collect data in its monopoly market and in the competitive market result from several opposite effects of different natures. Other intermediaries do not sell information in the competitive market l , and they incur a loss from collecting information on these consumers.

We can now characterize the number of consumer segments k_i collected by each data intermediary i according to m_i . Proposition 6 shows that data intermediaries with the highest and the second highest precisions, 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 6

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

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

- (b) *When data intermediaries have identical market sizes $m_1 = m_2 = \dots = m_n$, 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:*

$$k_1 > k_2 = \dots = k_n.$$

Proof: see Appendix A.8.

Proposition 6 (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 in the competitive market l .

Proposition 6 (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 6 (b).

5 Merger Analysis with Strategic Data Intermediaries

We analyze how a merger between data intermediaries changes their data strategies, and impacts in turn competition and consumer surplus in product markets.²⁵ There are two cases to consider depending on the number of competing intermediaries. First, consider a competitive market with only 2 data intermediaries. When they merge, the merged entity becomes a monopolist as there are no other intermediaries in the competitive market, and the competitive pressure on this market disappears after the merger. We analyze this first case in Section 5.1. Secondly, when some of the intermediaries do not participate to the merger, they still

²⁵Alternatively, De Corniere and Taylor (2020) consider a merger between a monopolistic firm on a market and a firm operating in a competitive related market.

compete with the merged intermediary in market l . We analyze this second case in Section 5.2 by considering the merger of two intermediaries in an oligopolistic setting. Finally, we generalize our competitive framework in Section 5.3 and we show how mergers can have ripple effects on consumers surplus in related markets.

While mergers may yield benefits for consumers and society, stemming for instance from a greater ability of the merged firm to invest in innovation, in product quality, or from increased network externalities. We show that changes in the data collection and selling strategies of intermediaries resulting from a merger usually reduce consumer surplus. Hence, accounting for changes in the data strategies of intermediaries requires mergers and acquisitions to yield higher welfare improvements compared with current practices.

5.1 Merger Between Duopolist Intermediaries

In this section, we consider a merger in the competitive market when only two intermediaries compete, and we compare consumer surplus before and after the merger. A merger will eliminate competition in market l , where the merged entity will collect and sell information as a monopolist.

In our model that focuses on the competitive effects of information, we show that such mergers always reduce consumer surplus in product markets for two reasons. On the one hand, consumer surplus depends on the number of segments sold to firms, which is smaller after the merger. Fewer segments sold reduces the intensity of competition in the product market and harms consumers. On the other hand, intermediaries have higher incentives to collect data after the merger, which intensifies consumer rent extraction.

Selling information and consumer surplus

To analyze how a merger between duopolist intermediaries impacts consumer surplus in product markets, 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, more consumer segments are sold in the competitive market than in monopoly markets. 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. In turn, Firm 2 also lowers its price, which has a positive effect on the surplus of consumers over the whole 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. Lemma 7 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 .

Lemma 7

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

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

Proof: See Appendix A.9

We have seen in Proposition 4 that fewer segments are sold in monopoly markets than in the competitive market. Hence, a merger between duopolist intermediaries will have a negative impact on consumer surplus in market l through a change of their information selling strategy. Fewer segments are sold in l after the merger, and to only one firm, which reduces the intensity of competition and decreases consumer surplus.

Data collection and consumer surplus

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. Lemma 8 shows that consumer surplus decreases with k for given x_1 and x_2 .

Lemma 8

Consumer surplus always decreases with k :

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

Proof: see Appendix A.10.

A merger between duopolist intermediaries will reduce consumer surplus through a change in the data collection strategies of the intermediaries. More segments are collected after the merger, which increases rent extraction from consumers.

The impact of a merger on consumer surplus.

We can now consider a merger between intermediaries 1 and 2 with monopoly markets m_1, m_2 , and competing in market l . Proposition 7 compares consumer surplus before and after the merger.

Proposition 7

A merger between duopolist intermediaries always reduces the intensity of competition in the competitive market l and lowers consumer surplus in markets m_1, m_2 and l .

5.2 Mergers in an Oligopolistic Market for Information

We now consider a merger between two intermediaries in market l where there are $n > 2$ active data intermediaries.²⁶ For instance, Kesler et al. (2020) describe

²⁶All proofs of this section are available upon request.

competition between data-driven mobile applications, and highlight a high intensity of competition in this market. In such market, even after a merger between several of these applications, the market would remain highly competitive. Contrary to a merger between duopolists, there are still $n - 2$ competitors after the merger, and market l remains competitive. Hence, a merger changes only at the margin the selling strategy of the merged firm. However, a merger still has an impact on consumer surplus in the product markets through changes in the data collection strategy of the merged entity, and potentially also of other firms through the escape-competition effect in data collection.

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 , while 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.

A crucial element will be the respective sizes of the merged firms before and after the merger, and whether the merger changes the competitive pressure between the two largest intermediaries. We will analyze three relevant cases in the following sections.

5.2.1 Start-up Merger

We consider a merger between two small intermediaries m_i, m_j , such that the intensity of competition in the competitive market remains unchanged: $m_i + m_j < m_2$. This first case allows us to identify how the cost efficiency resulting from a merger can increase the incentives of the merged intermediary to collect data, thus reducing consumer surplus in product markets through better consumer rent extraction.

We denote by Π_{ij} the profits of the merged intermediary, and by k_{ij} the number of segments that it collects. The profits of the intermediaries before and after the

merger are as follows:

$$\left\{ \begin{array}{l} \Pi_i(k_i) = m_i p_{1m_i}(k_i) - (m_i + l)c(k), \\ \Pi_j(k_j) = m_j p_{1m_j}(k_j) - (m_j + l)c(k), \\ \Pi_{ij}(k_{ij}) = (m_i + m_j) p_{1m_i}(k_{ij}) - (m_i + m_j + l)c(k). \end{array} \right.$$

The size of the monopoly market of the merged intermediary is relatively greater than the respective sizes of intermediaries i and j before the merger: $\frac{m_i+m_j}{m_i+m_j+l} > \frac{m_i}{m_i+l}, \frac{m_j}{m_j+l}$, and the incentives of the intermediaries to collect data increase after the merger due to this cost efficiency. Hence, the merged entity collects more data than the separate intermediaries: $k_{ij} > k_j, k_i$. A merger between start-ups thus always reduces consumer surplus as the merged intermediary has higher incentives to collect data after the merger, which intensifies consumer rent extraction in monopoly markets m_i and m_j .

5.2.2 Start-up Acquisition

A large data intermediary can purchase a smaller one and change the competitive balance on market l . We consider two cases that occur depending on whether intermediary 1 or 2 makes the acquisition, and on the resulting size of the merged intermediary.

Overall, start-up acquisitions can change the competitive pressure in market l , and thus the intensity of the escape-competition effect in data collection. This will impact in turn the data collection strategies of intermediaries that are not involved in the merger.

Start-up acquisition by intermediary 1.

When the largest intermediary acquires a smaller one of market size m_i , two opposite effects may drive up or down its incentives to collect consumer data. On the one hand, its incentives to collect information increase following the cost efficiency described in the previous section. Data intermediary 2 does not change its data collection strategy, and the competitive pressure in market l remains

the same as before the acquisition. On the other hand, the escape-competition effect is now weaker as the merged entity is larger than before the acquisition: $m_1 + m_i > m_1$.

Overall the first effect always dominates the second, and data intermediary 1 collects more consumer segments after the acquisition than intermediary 1 before the acquisition. Hence such an acquisition reduces consumer surplus in markets m_1 , m_i , and l through an increased ability of firms to extract rent. Consumers in market m_i incur the highest increase in data collection and the greatest loss of surplus from the merger.

Start-up acquisition by intermediary 2.

When intermediary 2 acquires a smaller intermediary of market size m_i , there are two cases to consider. First, if $m_2 + m_i \leq m_1$, intermediary 2 collects more segments due to the efficiency gains, and intermediary 1 collects more segments due to the increased escape-competition effect. This reduces surplus in markets m_1, m_2, m_i and l .

Secondly, if $m_2 + m_i > m_1$, intermediary 1 collects more segments after the acquisition (\tilde{k}_1) than intermediary 2 collected before the acquisition: $\tilde{k}_1 > k_2$. Moreover, intermediary 2 also collects more segments after the acquisition than intermediary 1 before the acquisition following two effects. On the one hand, the relative size of the merged entity is greater than intermediary 1: $m_2 + m_i > m_1$. On the other hand, the competitive pressure after the acquisition is exerted by intermediary 1 with $\tilde{k}_1 > k_2$. Overall, aggregate consumer surplus decreases after the acquisition. There are however winners and losers of such acquisition. Consumer surplus decreases in markets m_2 , m_i , and l where more segments are collected and sold after the acquisition. However, as $k_1 > \tilde{k}_1$, consumers on market m_1 are better identified before the acquisition, and their surplus increases after the acquisition.

5.2.3 Big Tech Merger

Finally the two largest intermediaries may choose to merge. The competitive pressure on the leading intermediary is lower after the merger, which lowers the escape-competition effect and decreases the incentives of the merged intermediary to collect data. Indeed, before the merger, intermediary 2 exerts a competitive pressure on intermediary 1; after the merger, the competitive pressure is exerted by intermediary 3 with $m_3 < m_2$, and the escape-competition effect is weaker than before the merger.

A merger between the two largest intermediaries will thus have opposite effects on the incentives of the merged entity to collect consumer data. The respective sizes of the monopoly markets and the competitive market determine whether the cost efficiency dominates the reduced escape-competition effect. When the competitive market is sufficiently large compared with markets m_1 and m_2 , and if the intensity of competition drops in market l after the merger, the merged intermediary has lower incentives to collect data than intermediary 1, which benefits consumers in markets m_1, m_2 and l . Otherwise, the merger increases data collection and reduces consumer surplus in these markets.

5.3 Mergers with Several Competitive Markets

We have shown in previous sections that a merger can impact the data strategies of the competitors of the merged intermediary, impacting in turn consumer surplus in the monopoly markets of intermediaries that do not participate to the merger. As we focus on Bertrand competition for the sale of information, this effect impacts the data strategies of the two largest intermediaries, and changes consumer surplus in their monopoly markets. We generalize our $m - l$ competitive framework in this section to show that a merger between intermediaries can also impact consumer surplus in competitive markets where none of the merging intermediaries sell information, thus generating ripple effects beyond the relevant market. This result supports a broad definition of relevant markets by competition authorities when assessing the impacts for consumers of a merger between data intermediaries.

We formalize this statement in proposition 8, and we prove it in the remaining of the section using an illustrative example.

Proposition 8

Mergers in the market for information can lower consumer surplus in competitive markets where none of the intermediaries involved in the merger or their competitors operate.

General competitive framework.

Our $m-l$ competitive framework can be generalized to account for several markets where different intermediaries compete. For instance, data on consumer interests or online purchasing habits can be provided by certain intermediaries, while health data or credit scoring data are provided by other intermediaries that are not necessarily active in other markets.

There are potentially $2^n - n$ markets where the n intermediaries may be competing, with variable sizes $l_j \geq 0$ ($j = 1, \dots, 2^n - n$). We denote by \mathcal{M}_i the set of competitive markets where intermediary i operates. Its total profits Π_i can be written as the sum of profits on its monopoly market and on all markets in \mathcal{M}_i :

$$\Pi_i(k_i) = m_i[p_{1m_i}(k_i) - c(k_i)] + \sum_{l_h \in \mathcal{M}_i} l_h[p_{1h}(k_i) + p_{2h}(k_i) - c(k_i)]. \quad (5)$$

Prices p_{1h} and p_{2h} on the different competitive markets can be null or positive, and vary depending on the value of k_i on each market compared with the precision of information proposed by other competing intermediaries.

We use this general formulation to show that a merger between two intermediaries can have indirect effects on competitive markets where other intermediaries operate.

Example of indirect effects in related competitive markets.

Consider three intermediaries with monopoly markets $m_1 > m_2 > m_3$, and two competitive markets l_1 and l_2 . We focus on the case where intermediaries 1 and

2 compete in market l_1 and intermediaries 2 and 3 compete in market l_2 . Market sizes are chosen so that intermediary 1 sells information at a positive price in market l_1 and intermediary 2 sells information at a positive price in market l_2 .

Using this set-up, we consider a merger between intermediaries 2 and 3, and we show how it impacts intermediary 1 through the escape-competition effect in data collection in market l_1 . We assume $m_1 > m_2 + m_3 + l_2$, so that intermediary 1 collects more data than the merged intermediary after the merger.

The profits of each intermediary can be written:

$$\begin{cases} \Pi_1(k_1) = m_1 p_{1m_1}(k_1) + l_1(p_1(k_1) + p_2(k_1)) - (m_1 + l_1)c(k_1), \\ \Pi_2(k_2) = m_2 p_{1m_2}(k_2) + l_2(p_1(k_2) + p_2(k_2)) - (m_2 + l_1 + l_2)c(k_2), \\ \Pi_3(k_3) = m_3 p_{1m_3}(k_3) - (m_3 + l_2)c(k_3). \end{cases}$$

After a merger between intermediaries 2 and 3, the merged intermediary collects k_{23} consumer segments, with a monopoly market of size $m_2 + m_3 + l_2$, and competes with intermediary 1 in market l_1 . The efficiency gain allows the merged intermediary to collect more segments than the separate entities, which reduces consumer surplus on markets l_2, m_2 and m_3 . Moreover, even though the merger does not directly impact competition in market l_1 – the number of firms in this market remains the same before and after the merger –, the merged intermediary exerts a stronger competitive pressure than intermediary 2 before the merger. This increases in turn the incentives of intermediary 1 to collect data in markets l_1 and m_1 through the escape-competition effect.

As intermediary 1 collects more data after the merger, the escape-competition effect also takes place in any competitive market where intermediary 1 has the second highest information precision and so on. Hence, even markets where none of the merging intermediaries and none of their direct competitors operate can be impacted by the merger. In these markets, more data is collected after the merger following the escape-competition effect, and consumer surplus always decreases after the merger. This simple example highlights the need for competition authorities to use a broad definition of the different product markets that can be impacted by a merger between data intermediaries.

6 Conclusion

Our model of competition between strategic data intermediaries emphasizes how mergers in the market for information have indirect effects in related product markets that are relevant for antitrust and consumer surplus analysis. Our analysis allows us to reach two main recommendations for policymakers.

First, we characterize negative effects of mergers on consumer surplus in related product markets, resulting from changes in the data collection and selling strategies of the newly merged entity. Competition authorities should thus account for how a merger changes the data strategies of intermediaries, as this will change the intensity of competition in products markets. Moreover, a merger can have ripple effects on markets in which none of the merging entities sell information. Hence, competition authorities should take into account all markets where intermediaries operate in their merger impact assessments, as mergers are likely to reduce consumer surplus through these indirect effects of information.

Secondly, our model highlights the need to maintain a competitive market for information. In particular, consumer surplus is the lowest when a monopolist data intermediary sells to firms rival information – that no other intermediary can sell. Any policy that reduces the amount of rival data will increase competition between data intermediaries, and increase in turn consumer surplus in product markets. This result is in line with [Cr mer et al. \(2019\)](#) who call for open data regulations and the creation of data-pools, under which data could be accessed by any company. In the same line, the right to data portability enacted in the European General Data Protection Regulation, which allows consumers to access and move their data from one intermediary to the other is likely to increase competition in the market for information and increase consumer surplus. Moreover, competition between data intermediaries also determines which firm can access information: a monopolist intermediary sells information to one firm while both firms purchase information when data intermediaries compete. Having access to more consumer data increases competition between firms, which in turn benefits consumers. Competition between data intermediaries guarantees fair and equal access to information to firms, which can also be reached for instance

by implementing fair, reasonable, and non-discriminatory clauses in markets for information (Crémer et al., 2019).

Finally, further research could analyze the impact of personal data protection on competition between data intermediaries and on related product markets, and how data protection agencies could work closer with competition authorities to protect consumer privacy and consumer surplus. Recent actions from the FTC call for regulation of the data brokerage industry,²⁷ and in the US, states such as Vermont or California have recently passed laws to gain control over the practices of data intermediaries.²⁸ 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.
- Anat R Admati and Paul Pfleiderer. Selling and trading on information in financial markets. *The American Economic Review*, 78(2):96–103, 1988.
- 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.

²⁷Federal Trade Commission, 2014, Data brokers: A Call for Transparency and Accountability.

²⁸Column: Shadowy data intermediaries make the most of their invisibility cloak; Los Angeles Times, November 5, 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.
- 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.
- Alexandre De Corniere and Greg Taylor. Data and competition: a general framework with applications to mergers, market structure, and privacy policy. 2020.
- 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.
- Joshua S Gans and Scott Stern. The product market and the market for “ideas”: commercialization strategies for technology entrepreneurs. *Research policy*, 32(2): 333–350, 2003.
- Shota Ichihashi. Competing data intermediaries. *The RAND Journal of Economics*, 52 (3):515–537, 2021.
- 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.
- Reinhold Kesler, Michael Kummer, and Patrick Schulte. Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings*, volume 2020, page 20978. Academy of Management Briarcliff Manor, NY 10510, 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*, 65(3): 1342–1362, 2019.
- Massimo Motta and Martin Peitz. Big tech mergers. *Information Economics and Policy*, 54:100868, 2021.
- 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.
- Lars A Stole. Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299, 2007.
- Jacques-Francois Thisse and Xavier Vives. On the strategic choice of spatial price policy. *The American Economic Review*, pages 122–137, 1988.
- Jean Tirole. Competition and the industrial challenge for the digital age. 2020.
- David Ulph and Nir Vulkan. *Electronic commerce and competitive first-degree price discrimination*. University of Bristol, Department of Economics, 2000.

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

Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P Gum-madi, Patrick Loiseau, and Oana Goga. Privacy risks with facebook’s pii-based targeting: Auditing a data broker’s advertising interface. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 89–107. IEEE, 2018.

A Appendix

A.1 Proof of Lemma 2: Optimal information structures in the competitive market

We characterize the partitions that maximize the profits of the firms in the competitive market. We show that a 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 characterize the optimal information structure for one of the firms (say Firm 1), and the result directly applies to its competitor.

Suppose that the intermediary sells information to Firm 1 (without loss of generality). The data intermediary can choose any partitions 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 θ 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 θ 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 a firm. 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 firms, 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.

Throughout the analysis we consider the sale of information to Firm 1, and the result directly generalizes to Firm 2. Moreover, we solve the different steps when Firm 2 charges a homogeneous price on the subsegments of the line considered.²⁹ This simplification allows us to focus on segments that allow firms to extract surplus from high-valuation consumers while softening the competitive effect of information. It also avoids us to compute profits with all possible partitions, since there is a high cardinality of the possible combinations of consumers segments, leaving a general resolution of the problem intractable without this assumption.

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 5 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 5 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}' .

²⁹Bounie et al. (2021) make a similar assumption for the tractability of the model.

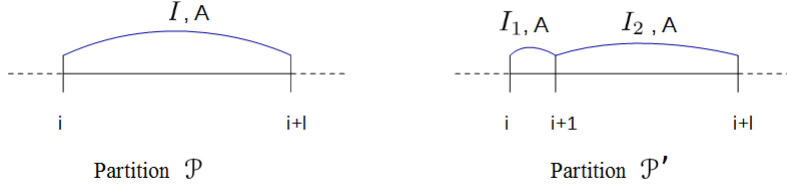


Figure 5: Step 1: segments of type A

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 6). 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 6).

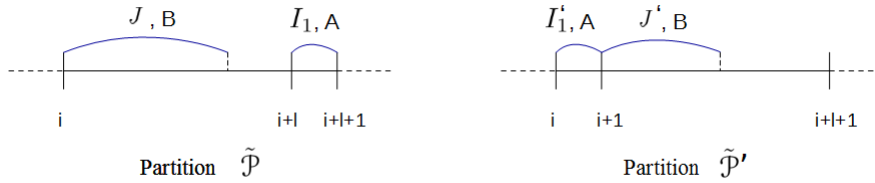


Figure 6: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 6 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. 6 gives the conditions for the demands addressed to Firm 1 and to Firm 2 to be positive on such segments.

$$\forall i, l \in \mathbb{N} \text{ s.t. } 0 \leq i \leq k-1 \text{ and } 1 \leq l \leq k-i-1, \quad (6)$$

$$\frac{i}{k} \leq \frac{\tilde{p}_2 + t}{2t} \quad \text{and} \quad \frac{\tilde{p}_2 + t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}.$$

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. 6 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 $\frac{l_i}{k}$.³⁰ There are $h \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$, where prices are noted \tilde{p}_{1i}^A . 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}_{1i}^B . We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

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

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.

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

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

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

³⁰With u_i and s_i integers below k .

Profits of Firm 2 can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^n \tilde{p}_2 \left[\frac{\tilde{p}_{1i}^B - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k} \right]. \quad (7)$$

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

Equilibrium prices are:

$$\begin{aligned} \tilde{p}_{1i}^A &= t + \tilde{p}_2 - 2 \frac{u_i t}{k} \\ \tilde{p}_{1i}^B &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n} \left[\sum_{i=1}^n \left[\frac{s_i}{2k} + \frac{l_i}{k} \right] \right] - \frac{s_i t}{k} \\ \tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n} \sum_{i=1}^n \left[\frac{s_i}{2k} + \frac{l_i}{k} \right]. \end{aligned} \quad (8)$$

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. 6: $\frac{\tilde{p}_2 + t}{2t} - \frac{l_i}{k} \leq \frac{s_i + l_i}{k}$ 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} the prices with $\tilde{\mathcal{P}}'$, we have:

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

for segments of type B where inequalities in Eq. 6 hold:

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

for segments of type B where inequalities in Eq. 6 do not hold:

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

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 $\Delta\pi_1 = \pi_1(\tilde{\mathcal{P}}') - \pi_1(\tilde{\mathcal{P}}) \geq 0$:

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

By definition, \tilde{s}_i verifies the inequalities in Eq. 6, thus $\frac{\tilde{s}_i}{k} \leq \frac{\tilde{p}_2+t}{2t}$, which allows us to establish that $\frac{4t}{3(n-m)} \left[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^m \frac{\tilde{s}_i}{2k} \right] \geq \frac{2t}{3nk}$. It is then immediate to show that:

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

Also, by assumption, firms compete on $J = [\frac{i}{k}, \frac{i+l}{k}]$ with $\tilde{\mathcal{P}}$, which implies that inequalities in Eq. 6 hold, and in particular, $\frac{\tilde{p}_2+t}{4t} - \frac{i}{2k} \leq \frac{l}{k}$.

Thus:

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

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:

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

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

For segments of type B:

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

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. 6.

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 $\hat{\mathcal{P}}'$.

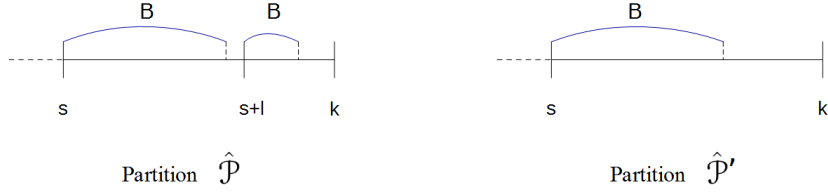


Figure 7: Step 3: demands of Firm 1 on segments of type B

Figure 7 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:

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

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

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

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

Inequalities in Eq. 6 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:

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

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

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

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

We consider the terms separately. First,

$$\begin{aligned}&\frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^n \left[\frac{\hat{p}'_2 + t}{2t} - \frac{s_i}{k} \right]^2 - \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^n \left[\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k} \right]^2 \\ &= \frac{t}{2} \sum_{i=1, s_i \neq \tilde{s}_i}^n \left[\frac{2}{3(n-m)} \left[\frac{3}{4t} \hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k} \right] \right]^2 \\ &\quad + \left[\frac{\hat{p}_2 + t}{2t} - \frac{s_i}{k} \right] \left[\frac{4}{3(n-m)} \left[\frac{3}{4t} \hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k} \right] \right] \\ &\geq \frac{t}{2} \left[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k} \right] \frac{4}{3} \left[\frac{3}{4t} \hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k} \right].\end{aligned} \tag{9}$$

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

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

On these m segments, inequalities in Eq. 6 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. 9 for all permissible values of \hat{p}'_2 :

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

Getting back to the difference in profits, we obtain:

$$\begin{aligned} \Delta\pi_1 &\geq \frac{t}{2} \left[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k} \right] \frac{4}{3} \left[\frac{3}{4t} \hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k} \right] - \frac{t}{2} \left[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k} \right]^2 \\ &\geq \frac{t}{2} \left[\frac{\hat{p}_2 + t}{2t} - \frac{s+l}{k} \right] \left[\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} - \frac{1}{6} \right]. \end{aligned} \quad (10)$$

The first bracket of Equation 10 is positive given Eq. 6. 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, $\hat{p}_2 \geq \frac{t}{2}$.³¹ And as this price is greater than $\frac{t}{3}$, the second bracket of Equation 10 is positive. This proves that $\Delta\pi_1 \geq 0$.

Conclusion

These three steps prove that the optimal partition for Firm 1 (and similarly for Firm 2) includes two intervals, as illustrated in Figure 4. 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 $[\frac{j}{k}, 1]$. ■

³¹As shown in Liu and Serfes (2004).

A.2 Proofs of Proposition 2

We characterize the equilibrium prices and numbers of segments sold in 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.

We compute prices and profits in equilibrium when both firms are informed with the optimal partitions found above. Firm 1 is a monopolist on the \overline{j}_1 segments of size $\frac{1}{k}$ in $[0, \frac{\overline{j}_1}{k}]$ and Firm 2 has information on $[1 - \frac{\overline{j}_2}{k}, 1]$. On $[\frac{\overline{j}_1}{k}, 1]$ Firm 1 sets a unique price p_1 and gets demand d_1 . Similarly, on $[0, 1 - \frac{\overline{j}_2}{k}]$ 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 $[0, \frac{\overline{j}_2}{k}]$ where it cannot identify consumers. Prices set by Firm 2 on segments in interval $[\frac{\overline{j}_2}{k}, 1]$ 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: $d_1 = \frac{p_2 - p_1 + t}{2t} - \frac{\overline{j}_1}{k}$ (resp. $d_2 = 1 - \frac{\overline{j}_2}{k} - \frac{p_2 - p_1 + t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\begin{aligned}\pi_1 &= \sum_{i=1}^{\bar{j}_1} d_{1i} p_{1i} + d_1 p_1 = \sum_{i=1}^{\bar{j}_1} \frac{1}{k} (p_2 + t - 2t \frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{\bar{j}_1}{k}) p_1, \\ \pi_2 &= \sum_{i=1}^{\bar{j}_2} d_{2i} p_{2i} + d_2 p_2 = \sum_{i=1}^{\bar{j}_1} \frac{1}{k} (p_1 + t - 2t \frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{\bar{j}_2}{k}) p_2.\end{aligned}$$

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:

$$\begin{aligned}p_1 &= t[1 - \frac{2}{3} \frac{\bar{j}_2}{k} - \frac{4}{3} \frac{\bar{j}_1}{k}], \\ p_2 &= t[1 - \frac{2}{3} \frac{\bar{j}_1}{k} - \frac{4}{3} \frac{\bar{j}_2}{k}].\end{aligned}$$

Replacing these values in the above demands and prices gives:

$$\begin{aligned}p_{1i} &= 2t - \frac{4}{3} \frac{\bar{j}_2 t}{k} - \frac{2}{3} \frac{\bar{j}_1 t}{k} - 2 \frac{it}{k}, \\ p_{2i} &= 2t - \frac{4}{3} \frac{\bar{j}_1 t}{k} - \frac{2}{3} \frac{\bar{j}_2 t}{k} - 2 \frac{it}{k}.\end{aligned}$$

and

$$\begin{aligned}d_1 &= \frac{1}{2} - \frac{2}{3} \frac{\bar{j}_1}{k} - \frac{1}{3} \frac{\bar{j}_2}{k}, \\ d_2 &= \frac{4}{3} \frac{\bar{j}_2}{k} - \frac{1}{2} - \frac{1}{3} \frac{\bar{j}_1}{k}.\end{aligned}$$

Profits are:

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{\bar{j}_1} \frac{2t}{k} [1 - \frac{i}{k} - \frac{1}{3} \frac{\bar{j}_1}{k} - \frac{2}{3} \frac{\bar{j}_2}{k}] + (\frac{1}{2} - \frac{2}{3} \frac{\bar{j}_1}{k} - \frac{1}{3} \frac{\bar{j}_2}{k}) t [1 - \frac{2}{3} \frac{\bar{j}_2}{k} - \frac{4}{3} \frac{\bar{j}_1}{k}] \\ &= \frac{t}{2} - \frac{7}{9} \frac{\bar{j}_1^2 t}{k^2} + \frac{2}{9} \frac{\bar{j}_2^2 t}{k^2} - \frac{4}{9} \frac{\bar{j}_1 \bar{j}_2 t}{k^2} + \frac{2}{3} \frac{\bar{j}_1 t}{k} - \frac{2}{3} \frac{\bar{j}_2 t}{k} - \frac{\bar{j}_1 t}{k^2}.\end{aligned}$$

$$\begin{aligned}
\pi_2^* &= \sum_{i=1}^{\bar{j}_2} \frac{2t}{\bar{k}} \left[1 - \frac{i}{\bar{k}} - \frac{1}{3} \frac{\bar{j}_2}{\bar{k}} - \frac{2}{3} \frac{\bar{j}_1}{\bar{k}} \right] + \left(\frac{1}{2} - \frac{2}{3} \frac{\bar{j}_2}{\bar{k}} - \frac{1}{3} \frac{\bar{j}_1}{\bar{k}} \right) t \left[1 - \frac{2}{3} \frac{\bar{j}_1}{\bar{k}} - \frac{4}{3} \frac{\bar{j}_2}{\bar{k}} \right] \\
&= \frac{t}{2} - \frac{7}{9} \frac{\bar{j}_2^2 t}{\bar{k}^2} + \frac{2}{9} \frac{\bar{j}_1^2 t}{\bar{k}^2} - \frac{4}{9} \frac{\bar{j}_1 \bar{j}_2 t}{\bar{k}^2} + \frac{2}{3} \frac{\bar{j}_2 t}{\bar{k}} - \frac{2}{3} \frac{\bar{j}_1 t}{\bar{k}} - \frac{\bar{j}_2 t}{\bar{k}^2}.
\end{aligned}$$

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 $\pi_1(\underline{j}_1, \bar{j}_2)$ and $\pi_2(\underline{j}_2, \bar{j}_1)$.

By replacing variables \underline{j}_1 and \bar{j}_2 into π_1 (and respectively for π_2), we obtain the following expressions:

$$\begin{aligned}
\pi_1(\underline{j}_1, \bar{j}_2) &= \frac{t}{2} - \frac{7}{9} \frac{(\underline{j}_1)^2 t}{\bar{k}^2} + \frac{2}{9} \frac{(\bar{j}_2)^2 t}{\bar{k}^2} - \frac{4}{9} \frac{\underline{j}_1 \bar{j}_2 t}{\bar{k} \bar{k}} + \frac{2}{3} \frac{\underline{j}_1 t}{\bar{k}} - \frac{2}{3} \frac{\bar{j}_2 t}{\bar{k}} - \frac{\underline{j}_1 t}{\bar{k}^2} \\
\pi_2(\underline{j}_2, \bar{j}_1) &= \frac{t}{2} - \frac{7}{9} \frac{(\underline{j}_1)^2 t}{\bar{k}^2} + \frac{2}{9} \frac{(\bar{j}_1)^2 t}{\bar{k}^2} - \frac{4}{9} \frac{\underline{j}_2 \bar{j}_1 t}{\bar{k} \bar{k}} + \frac{2}{3} \frac{\underline{j}_2 t}{\bar{k}} - \frac{2}{3} \frac{\bar{j}_1 t}{\bar{k}} - \frac{\underline{j}_2 t}{\bar{k}^2}
\end{aligned}$$

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} with the highest information precision \bar{k} maximizes its profits by maximizing with respect to \bar{j}_1 and \bar{j}_2 the sum:

$$\begin{aligned}
p_l(\bar{j}_1, \bar{j}_2) + p_l(\underline{j}_2, \bar{j}_1) &= \pi_1(\bar{j}_1, \bar{j}_2) - \pi_1(\underline{j}_1, \bar{j}_2) + \pi_2(\bar{j}_2, \bar{j}_1) - \pi_2(\underline{j}_2, \bar{j}_1) \\
&= \frac{(7\bar{k}\bar{k}(\bar{j}_2)^2 + (4\bar{k}\bar{j}_1 - 6\bar{k} + 9)\bar{k}\bar{j}_2 + 7\bar{k}\bar{k}(\bar{j}_1)^2 + (4\bar{k}\bar{j}_2 - 6\bar{k} + 9)\bar{k}\bar{j}_1)t}{9\bar{k}\bar{k}} \\
&\quad + \frac{((-7\bar{k}(\bar{j}_2)^2 + (6\bar{k} - 8\bar{k}\bar{j}_1)\bar{j}_2 - 7\bar{k}(\bar{j}_1)^2 + 6\bar{k}\bar{j}_1)\bar{k} - 9\bar{k}\bar{j}_2 - 9\bar{k}\bar{j}_1)t}{9\bar{k}\bar{k}}
\end{aligned} \tag{11}$$

Thus equilibrium variables $\bar{j}_1, \bar{j}_2, \underline{j}_1, \underline{j}_2$ are chosen as simultaneous best responses. FOCs on $\bar{j}_1, \bar{j}_2, \underline{j}_1$ and \underline{j}_2 give respectively in equilibrium:

$$\bar{j}_1^* = \bar{j}_2^* = \frac{\bar{k}}{3} - \frac{\bar{k}}{9\underline{k}} - \frac{7}{18}$$

$$\underline{j}_1^* = \underline{j}_2^* = \frac{\underline{k}}{3} - \frac{11}{18} + \frac{\underline{k}}{9\bar{k}}$$

A.3 Proof of Proposition 3

We characterize the escape-competition effect in data collection. We first write the price of information in the competitive market. We substitute the values of \bar{j}_1^* , \bar{j}_2^* and \underline{j}_1^* in $\pi_1(\bar{j}_1, \bar{j}_2) - \pi_1(\underline{j}_1, \bar{j}_2)$ in the profit functions of Firm 1 and Firm 2. The price of information is identical for both firms and can be written as

$$\begin{aligned} p_l(\bar{k}, \underline{k}) &= [\pi_1(\bar{j}_1, \bar{j}_2) - \pi_1(\underline{j}_1, \bar{j}_2)] \\ &= \frac{((12\underline{k} - 11)\bar{k}^2 + (4\underline{k} - 12\underline{k}^2)\bar{k} + 7\underline{k}^2)t}{36\underline{k}^2\bar{k}^2} \end{aligned} \quad (12)$$

We prove that the number of consumer segments \bar{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 $p_l(\bar{k}, \underline{k})$ in the competitive market l increases with \underline{k} .

Consider the second degree derivative of p_l with respect to \bar{k} and \underline{k} :

$$\frac{\partial^2 p_l(\bar{k}, \underline{k})}{\partial \bar{k} \partial \underline{k}} = \frac{1}{9\bar{k}^2 \underline{k}^2} \geq 0$$

Thus, the larger the \underline{k} , the larger the value of the first degree derivative of p_l with respect to \bar{k} , and the higher the marginal gain from collecting data. An increase of \underline{k} will thus increase the value of \bar{k}^* in equilibrium.

A.4 Proof of Lemma 4: Optimal information structure in a monopoly market when the data intermediary sells information to one firm

We begin with describing the mechanism used by an intermediary to sell information in its monopoly market.

Selling mechanism.

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 .³² 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 $\pi_2(\mathcal{P}^k, \emptyset) - \pi_2(\emptyset, \mathcal{P}^k)$ 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 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 $\pi_1(\mathcal{P}^k, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k)$, and make profits $\pi_1(\emptyset, \mathcal{P}^k)$. 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 $p_1^{m*} = \pi_1(\mathcal{P}_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) - \epsilon$, where ϵ is an arbitrary small positive number. Thus, $\pi_1(\mathcal{P}_1, \emptyset) - p_1^{m*} > \pi_1(\emptyset, \mathcal{P}^k)$, 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 $\pi_2(\mathcal{P}^k, \emptyset) - \pi_2(\emptyset, \mathcal{P}^k)$. The partitions are therefore $(\mathcal{P}_1, \mathcal{P}^k)$.

The proof of the optimal partition is identical to the proof of Lemma 2, in the special case where one of the firms is not informed. Hence we only need to prove that the profit of an uninformed firm are minimized when its competitor acquires \mathcal{P}^k .

³²For instance, Coey et al. (2021) analyze the role of reserve prices in repeated online auctions.

The profit of an uninformed firm are minimized when its competitor acquires \mathcal{P}^k .

To prove this claim we consider Firm 1 that is informed and Firm 2 that is 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 its demand on $[\frac{s}{k}, \frac{s+l}{k}]$, which overall lowers its profits. By iterating this argument, we can conclude that the reference partition \mathcal{P}^k 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 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 profits of Firm 2 over the unit line are smaller with the finer partition than with the coarser one. This establishes the result.

A.5 Proof of Lemma 5.

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} , $i = 1, \dots, j_1$, 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:

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

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:

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

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

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j_1} \frac{2t}{k} [1 - \frac{i}{k} - \frac{1}{3} \frac{j_1}{k}] + \frac{t}{2} (1 - \frac{4}{3} \frac{j_1}{k})^2 \\ &= \frac{t}{2} + \frac{2j_1 t}{3k} - \frac{7t j_1^2}{9k^2} - \frac{t j_1}{k^2} \\ \pi_2^* &= \frac{t}{2} + \frac{2t j_1^2}{9k^2} - \frac{2j_1 t}{3k}.\end{aligned}\tag{13}$$

We can now determine the optimal size j_1^* when the data intermediary only sells information to Firm 1, by maximizing profits with respect to j_1 . The profits of the data intermediary when it sells to one firm are:³⁴

$$\begin{aligned}\Pi_1(j_1) &= \pi_1(j_1, \emptyset) - \pi_1(\emptyset, \mathcal{P}^k) \\ &= \frac{3t}{8} + \frac{2j_1 t}{3k} - \frac{t}{4k} - \frac{7j_1^2 t}{9k^2} - \frac{j_1 t}{k^2} - \frac{t}{8k^2}.\end{aligned}$$

FOC with respect to j_1 leads to the following maximizing value: $j_1^* = \frac{6k-9}{14}$

We show that more consumers are identified in the competitive market than in monopoly markets by comparing $\frac{\bar{j}_1(k) + \bar{j}_2(k)}{k} = 2[\frac{1}{3} - \frac{1}{9k} - \frac{7}{18k}]$ with $\frac{j_1^m(k')}{k'} = \frac{6k'-9}{14k'}$:

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

³⁴The expression of $\pi(\emptyset, \mathcal{P}^k)$ is provided in Liu and Serfes (2004).

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

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

A.6 Proof of Lemma 6

Suppose now that the data intermediary sells information to both firms. We first describe the selling mechanism. Then we characterize the optimal information structure in a lemma that we prove below, and that we will use to derive the profits of the intermediary in equilibrium, and to show that it is more profitable to sell information to only one firm.

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 in a monopoly market:

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:

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

We now derive the optimal information structure.

Optimal information structure: selling information to both firms in a monopoly market

Lemma 9

The optimal partitions include all available consumer segments:

$$j_{21}^*(k) = j_{22}^*(k) = k.$$

There are two classes of partitions to consider in order to find the optimal partitions sold to firms. A first class \mathcal{C} is composed of segments closest to a firm's location. The other class is composed of the remaining partitions. There are two local maximum, one in each class, which we will characterize and then compare profits in both cases.

We first characterize the optimal information structure in class \mathcal{C} .

Part a: optimal information structure when the data intermediary sells information to both firms in class \mathcal{C}

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. 6 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. 6 guarantees that there is positive demand for Firm 1, whereas the second part guarantees positive demand for Firm 2. Inequalities in Eq. 6 are expressed as a function of p_2 without loss of generality. We use Eq. 6 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:

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

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 the profits of a firm when both of them are informed and 2) such a partition does not change the profits of an uninformed firm facing an informed competitor.

1) *A partition which maximizes the profits of a firm when both of them are informed 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 the profits of an uninformed firm facing an informed competitor.*

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 $[0, 1 - \frac{j_2}{k}]$.

In the following section we will compute interior solutions with $j_1, j_2 \in [0, \frac{k}{2}]$, and we will compare profits with the interior solution with profits with the corner solution where all information is sold to both firms. ■

Part b: the data intermediary sells symmetric information to both firms in class \mathcal{C}

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

Prices and profits in equilibrium are provided in Appendix A.2:

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

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

The data intermediary maximizes the following profit function:

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

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.

The profit of the data intermediary when both firms are informed with partitions $j_1 = j_2 = j \in [0, \frac{k}{2}]$ is:

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

FOC with respect to j leads to $j_2^* = \frac{6k-9}{22}$ and:

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

The remaining partitions out of class \mathcal{C} necessarily include all segments closest to a firm's location and nothing after $\frac{1}{2}$. We show that the price of information is maximized in this class when all information is sold to both firms.

Different partitions in this class include different numbers of segments closest to a firm's competitor. Because selling only part of the segments does not yield the highest feasible threat, and that the profits of the firms when both of them are informed are identical for all partitions in this class, we can conclude that the profits are maximized when all information is sold to both firms.

We can write the profit of the data intermediary when all information is sold by replacing j_1, j_2 by $\frac{k}{2}$ to obtain firms' profits when both firms are informed ($\pi_\theta(k, k) = \frac{t}{4} - \frac{t}{2k}$), and by considering the profits of an uninformed firm facing a

competitor informed with all segments, given in Liu and Serfes (2004) ($\pi_\theta(\emptyset, \mathcal{P}^k) = \frac{t}{8} + \frac{t}{4k} + \frac{t}{8k^2}$).

$$\Pi_2^{all} = \frac{t}{4} - \frac{3t}{2k} - \frac{t}{4k^2}.$$

Profits are higher with the corner solution where all information is sold than with partitions belonging to class \mathcal{C} , and the data intermediary sells all information to both firms. ■

Profit comparison.

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.

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

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

FOC with respect to 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) *Profits when the data intermediary sells information to both firms.*

The profits of the data intermediary when both firms are informed are:

$$\Pi_2^{all} = \frac{t}{4} - \frac{3t}{2k} - \frac{t}{4k^2}.$$

3) *DI's selling strategy in equilibrium.*

³⁵The expression of $\pi_\theta(\emptyset, \mathcal{P}^k)$ 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:

$$\Pi_1^* - \Pi_2^* \geq \frac{15t}{56k^2}.$$

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

A.7 Proof of Proposition 5

We show that the incentives for an intermediary to collect segments are greater in its monopoly market than in the competitive market. This result is straightforward for intermediaries that do not sell information and make zero profits in the competitive market, and we focus our proof on intermediary \overline{DI} .

We first write the price of information in 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

$$\begin{aligned} p_l(\overline{k}, \underline{k}) &= [\pi_1(\overline{j}_1, \overline{j}_2) - \pi_1(\underline{j}_1, \overline{j}_2)] \\ &= \frac{((12\underline{k} - 11)\overline{k}^2 + (4\underline{k} - 12\underline{k}^2)\overline{k} + 7\underline{k}^2)t}{36\underline{k}^2\overline{k}^2} \end{aligned} \quad (14)$$

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 in the competitive market where the total revenue of \overline{DI} is $2p_l(\overline{k})$.

For $\overline{k} > \underline{k}$, we have

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

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

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

and

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

Consider \bar{k}^* such that

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

Since

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

for $\bar{k} \in [2, \infty[$

revenues are concave, and necessarily, k such that

$$\frac{\partial p_m(k)}{\partial k} = \frac{\partial c(k)}{\partial k}$$

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

The incentives to collect consumer data are higher in monopoly markets than in the competitive market l . ■

A.8 Proof of Proposition 6

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^* is higher than k_2 . Finally, we show that for $i > 1$, k_i^* decreases with m_i .

A.8.1 Proof of Proposition 6 (a)

Let's $k_1 = \operatorname{argmax}\{m_1 p_m(k) - (m_1 + l)c(k)\}$. 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 the largest amount of segments. 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 $k_i = \operatorname{argmax}_k\{m_i p_m(k) - (m_i + l)c(k)\}$ verifies $k_i < k_1$. We assume that m_1, \dots, m_n 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 $k_1 = \operatorname{argmax}\{m_1 p_m(k) - (m_1 + l)c(k)\}$ and $m_1 > m_2$, necessarily $k_2 = \operatorname{argmax}_k\{m_2 p_m(k) + 2lp_l(\mathcal{P}^k, \emptyset)k_1 - (m_2 + l)c(k)\}$ verifies $k_2 < k_1$. In particular $\max_k\{m_2 p_m(k) - (m_2 + l)c(k)\} \geq \max_k\{m_2 p_m(k) + 2lp_l(\mathcal{P}^k, \emptyset)k_1 - (m_2 + l)c(k)\}$, and deviation to $k_2 > k_1$ is never profitable for firm 2.

A.8.2 Proof of Proposition 6 (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 to 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 deviate and to collect $\tilde{k}^* > \hat{k}^*$, that is, to collect more data than the intermediary

³⁶By 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 $mp_m(k^*) - (m + l)c(k^*)$.

with the highest precision, if the profits of doing so is greater than its monopoly profits:

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

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.

A.9 Proof of Lemma 7

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} \left[\int_0^{\frac{1}{k}} V - 2t \left[1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} - \frac{i}{k} \right] - t x dx \right] \\
&+ \int_{\frac{j_1}{k}}^{\frac{1}{2} + \frac{j_1}{3k} - \frac{j_2}{3k}} V - t \left[1 - \frac{4}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] - t x dx + \int_{\frac{1}{2} + \frac{j_1}{3k} - \frac{j_2}{3k}}^{1 - \frac{j_2}{k}} V - t \left[1 - \frac{2}{3} \frac{j_1}{k} - \frac{4}{3} \frac{j_2}{k} \right] - t x dx \\
&+ \sum_{i=1}^{j_2} \left[\int_0^{\frac{1}{k}} V - 2t \left[1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} - \frac{i}{k} \right] - t x dx \right] \\
&= \sum_{i=1}^{j_1} \frac{1}{k} \left(V - 2t \left[1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} - \frac{i}{k} \right] \right) - \frac{j_1 t}{2k^2} \\
&+ \sum_{i=1}^{j_2} \frac{1}{k} \left(V - 2t \left[1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} - \frac{i}{k} \right] \right) - \frac{j_2 t}{2k^2} \\
&+ V \left[1 - \frac{j_2}{k} - \frac{j_1}{k} \right] - \left[\frac{1}{2} - \frac{2j_1}{3k} - \frac{j_2}{3k} \right] t \left[1 - \frac{4}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] \\
&- \left[\frac{1}{2} - \frac{2j_2}{3k} - \frac{j_1}{3k} \right] t \left[1 - \frac{4}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} \right] - t \left[\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} \right] \\
&= \frac{j_1}{k} \left[V - 2t \left[1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] \right] + \frac{j_1(j_1 + 1)t}{k^2} - \frac{j_1 t}{2k^2} \\
&+ \frac{j_2}{k} \left[V - 2t \left[1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} \right] \right] + \frac{j_2(j_2 + 1)t}{k^2} - \frac{j_2 t}{2k^2} \\
&+ V \left[1 - \frac{j_2}{k} - \frac{j_1}{k} \right] + t \left[-\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} \right] \\
&= V + t \left[-\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} \right] + \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

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

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

A.10 Proof of Lemma 8

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. ■