Collecting and Selling Consumer Information: Selling Mechanisms Matter

David Bounie, Antoine Dubus, Patrick Waelbroeck

Collecting and Selling Consumer Information: Selling Mechanisms Matter*

David Bounie,† Antoine Dubus‡ and Patrick Waelbroeck§

November 4, 2022

Abstract

We study the strategies of a data intermediary collecting and selling information to competing firms under different selling mechanisms. We characterize the amount of data collected and sold as well as the price of information with posted prices, sequential bargaining, first-price and second-price auctions. We generalize pair-wise comparisons to establish the economic properties of classes of mechanisms.

Keywords: Selling mechanisms; Data intermediaries; Data collection; Selling strategies; Price of information.

*We would like to thank Alessandro Bonatti, Andreea Cosnita-Langlais, Axel Gautier, Victor Ginsburgh, Doh Shin Jeon, Brad Larsen, Patrick Legros, Flavio Pino, and Wilfried Sand-Zantman, as well as participants at the 46th EARIE conference, the CESifo Area Conference on the Economics of Digitization, and the 14th Digital Economics Conference in Toulouse 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 have been done while Antoine Dubus was receiving financial support from the FNRS Grant PDR T.01.47.19. This research has also been conducted within the Research Chair "Digital Finance" under the aegis of the Risk Foundation, a joint initiative by Groupement des Cartes Bancaires CB, la Banque Postale, CDC, Telecom Paris and University of Paris 2 Panthéon-Assas.

†i3, CNRS, Télécom Paris, Institut Polytechnique de Paris; david.bounie@telecom-paris.fr.

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

§i3, CNRS, Télécom Paris, Institut Polytechnique de Paris; patrick.waelbroeck@telecom-paris.fr.

1 Introduction

The use of large quantities of consumer data has rapidly become an essential factor for most industries to remain competitive in today's global markets. Companies now devote a significant portion of their budgets to acquire information on potential customers. Data intermediaries are emerging as market leaders in this new rush for information (see for instance Elliott et al. (2021)). They collect consumer data to build segmented profiles of similar groups of people using machine-learning algorithms, and they sell consumer segments to companies seeking to improve their data strategies through personalized ads, products, and prices (Varian, 1989). Because data can provide a company with a significant advantage over its competitors, data intermediaries play a central role in shaping competition in product markets by determining how much data they collect from consumers and then sell to firms.

A recent economic literature has emerged to study the optimal information selling strategies of data intermediaries. For example, Bergemann et al. (2018) analyze a monopolist intermediary selling data to a single buyer using posted prices. The authors investigate how much information the data seller should provide and how to price the access to this data when the seller has no information on the initial private valuation of the data buyer. Then, Braulin and Valletti (2016) study vertically differentiated products, for which consumers have hidden valuations; the data broker can sell information on these valuations to companies using first-price auctions. Following this, Montes et al. (2019) analyze the strategy of a monopolist data intermediary selling information to one or to two firms willing to price-discriminate consumers, a framework also used by Thisse and Vives (1988), and more recently by Nageeb Ali et al. (2022). The data intermediary uses first-price auctions to sell information, and the authors show that it is optimal to sell all the available information to only one firm. Finally, Bounie et al. (2021) investigate the case of a monopolist data intermediary using first-price auctions to sell segments of the consumer demand to companies competing in the product market. They find that it is optimal for the intermediary to sell information to only one firm, but not to sell all the available consumer segments.

Two important questions arise from previous articles: do optimal selling strategies depend on the selling mechanism used by the data intermediary? Does the use of posted prices or auctions have an impact on the amount of information sold to firms, and possibly on the number of firms informed in the markets?

Economists have for long acknowledged the central role of the selling mechanism on market outcomes. Wang (1993) studies under which circumstances first-price auctions dominate posted prices, and in a companion paper Wang (1995) compares posted prices and bargaining.[1] The variety of selling mechanisms studied in the theoretical literature is echoed in empirical studies. With the rapid development of online platforms for consumer auctions such as eBay in 2001, researchers have devoted a great deal of time combing through auctions and posted prices (Bajari and Hortaçsu, 2004; Zeithammer and Pengxuan, 2006; Hammond, 2010, 2013; Einav et al., 2018; Jindal and Newberry, 2018). In this article, we show that the selling mechanism not only influences the information selling strategy of a data intermediary but also the amount of data that the intermediary collects, a dimension that has not been explored in the recent economic literature that focuses on the sale of information by an intermediary.

Consider first the effect of the selling mechanism on the amount of information sold to firms. A monopolist data intermediary can sell consumer data using posted prices or first-price auctions. Data divides consumer demand into segments. The data intermediary charges a price of information and chooses which consumer segments to sell to a firm. Consider posted prices: the intermediary can personalize its offer to the need of Firm 1, and Firm 2 is not interested in an offer that is not tailored to its need. Therefore, there are no negative consequences for Firm 1 for not purchasing information as the outside option of Firm 1 when purchasing information does not depend on the information sold to Firm 2. Consider now first-price auctions. If Firm 2 wins the auction, Firm 1 faces a negative externality; uninformed Firm 1 competes with informed Firm 2. In order to reach the highest bid of Firm 1, the intermediary can threaten Firm 1 to sell all consumer segments to Firm 2 in case Firm 1 loses the auction. Therefore, the value of the threat of the outside option of Firm 1 with auctions is stronger than with posted prices. We show that it is optimal for the intermediary to

[1] Wang (1995) also notes that "three major selling methods, bargaining, posted-price selling, and auctions, have now become part of our everyday life" and that "new and used cars, for example, are typically sold through bargaining."

sell information to one firm with auctions and to two firms with posted prices. Using auctions, the data intermediary can leverage on the negative externality related to the threat of being uninformed, which increases the willingness to pay of a prospective buyer. This threat is weaker with posted prices as the data intermediary cannot threaten a firm to sell information to its competitor. Thus, the selling mechanism influences the number of firms that are informed on the product market and consumer surplus. Therefore, a central result of this paper illustrates how consumer surplus changes with different selling mechanisms.

Now consider how much personal data the intermediary collects. In the aforementioned studies, the intermediary chooses the quantity of information to sell to firms, but has no strategic choice to make about the quantity of data to collect from consumers. The optimal sale of information by the intermediary is then supposed to be independent of its data collection strategy. However, collecting data is expensive for technical reasons (e.g. setting up trackers and processing information) and regulations are also complex (laws on the protection of personal data). This paper addresses this challenging context by showing that the optimal amount of data collected by an intermediary depends on the selling mechanism, which will in turn impact consumer surplus.

As an illustration, consider the sale of information to one firm, say Firm 1, using first-price auctions and posted prices. The amount of data collected depends on the price of information, which is determined by the outside option of Firm 1 that varies with the selling mechanism. With first-price auctions, the data intermediary can maximize the value of the threat of the outside option, and maximize Firm 1's willingness to pay. On the contrary, with posted prices, both firms are uninformed when one rejects the offer of the data intermediary, resulting in a lower willingness to pay for information. Therefore, the number of segments collected by the intermediary naturally changes with the selling mechanism as the price of information (and the outside option) changes with the selling mechanism.

We propose in this article a simple framework that links the selling mechanism to the data collection and selling strategies of a data intermediary. We analyze how selling mechanisms are central to understanding consumer surplus and competition in product markets. In Section 2, we consider a model of competition à la Hotelling on the product market. Consumers purchase one product from two competing firms that are located at the extremities of the Hotelling line. A monopolist data intermediary costly collects information on consumers and sells data that segment consumer demand. The more data is collected on consumers, the higher the precision of information, which allows firms to locate consumers more precisely. A firm that purchases consumer segments, i.e. an information partition, can set a price on each segment and is thus considered as an informed firm. On the contrary, a firm that does not purchase consumer segments, i.e. that is uninformed, cannot distinguish consumers, and sets a single price on the Hotelling line. The data intermediary then chooses the optimal information partition by selling more or less segments to firms.

To understand the underlying mechanisms of the model, we first study the simple case of the sale of information to only one firm, called Firm 1 (w.l.o.g). This simplification streamlines the data collection and selling strategies of data intermediaries. We then analyze in Section 5 under which conditions the intermediary sells information to both competing firms.

The strategies of the firms and the data intermediary critically depend on the way information is sold, i.e. the selling mechanism, which influences the price of information (defined as the profits of Firm 1 with information minus its profits without information), and the incentive of the intermediary to collect data. In Section 3, we define what is a selling mechanism, and we determine the optimal information partition sold to firms as well as the equilibrium price of information for any selling mechanism. We study the properties of classes of mechanisms, and we apply the main results to four selling mechanisms commonly studied in the economic literature: posted prices, sequential bargaining, first-price auctions, and second-price auctions.

First, we show that for selling mechanisms for which the data intermediary sells a partition that maximizes the profits of Firm 1, namely first-price auctions, sequential bargaining and posted prices, the optimal partition divides the unit line into two intervals: the first interval consists of segments where consumers are identified, and consumers in the second interval are unidentified. Moreover, we show that first-price auctions, sequential bargaining and posted prices belong to a class of selling mechanisms that share similar properties and that we call "independent offers"; the number of segments sold to Firm 1 (denoted by j_1) is independent of the information proposed to Firm 2 (denoted by j_2) if Firm 1 does not acquire information.

Secondly, we also find that there are mechanisms for which the data intermediary does not maximize the profit of Firm 1. For instance, in the case of second-price auctions, the intermediary may choose to sell a partition that yields a higher price of information by worsening the outside option of Firm 1. The number of segments sold to Firm 1 clearly depends on the information proposed to Firm 2. Overall, we find that we can classify selling mechanisms into three classes according to whether the profits of Firm 1 are independent from the number of consumer segments proposed to Firm 2 (independent offers), or depend positively (increasing function) or negatively (decreasing function) on the information proposed to Firm 2 if Firm 1 does not acquire information.

In Section 4, we exploit the properties of these three classes of selling mechanisms to study the optimal number of consumer segments sold and collected by the data intermediary in order to determine how consumer surplus is impacted by the selling mechanism. First, we find that the number of segments sold to Firm 1 in equilibrium, for a given precision of information, varies with the classes of selling mechanisms. The number of segments is higher for selling mechanisms where the number of consumer segments sold to Firm 1 depends positively on the number of segments proposed to Firm 2. In particular, we find that the number of consumer segments sold to Firm 1 is higher with second-price auctions than with selling mechanisms that belong to independent offers, i.e. posted prices first-price auctions and sequential bargaining. These results confirm the crucial role of selling mechanisms on the amount of data sold in equilibrium.

Secondly, the selling mechanism will also change the impact of the rent-extraction and the outside-option effects on the incentives of the intermediary to collect data. Consider first the rent-extraction effect. More data collected allows a firm to better extract surplus from identified consumers. This gain in profits increases with the number of consumers on whom a firm has information as a marginal increase in the precision of information is greater when more consumers are identified. Now consider the outside-option effect. The willingness to pay of a firm for information depends on its profits if it does not purchase information and faces an informed competitor. These profits decrease when the competitor acquires more precise information. Hence, collecting data allows the intermediary to exert a threat on a prospective buyer through its outside option. The net impact of the rent-extraction and outside-option effects varies with different selling mechanisms, which will impact the incentives of the intermediary to collect data.

For instance, the number of segments collected is the lowest with posted prices, as the marginal gain of data collection does not impact the outside option of Firm 1. The marginal effect of data collection on the outside option of Firm 1 is higher with sequential bargaining than with first-price auctions where the outside option is already the harshest. The marginal impact of data collection on the outside option of Firm 1 with these two mechanisms is stronger than with second-price auctions. Therefore, the incentives to collect data are weaker with second-price auctions. We provide results for the three classes of mechanisms, including independent offers that encompass posted prices, sequential bargaining and first-price auctions.

In Section 5, we extend the model to analyze a setting in which the data intermediary can sell information to two firms. We find that it is optimal for the intermediary to sell information to only one firm with first-price and second-price auctions, and to two firms with posted prices and sequential bargaining. We therefore conclude that the selling mechanism has an impact on the number of firms that are informed on a market, and thus on the competitiveness of markets.

Finally, we discuss in Section 6 how the strategies of the data intermediary impact consumer surplus. For instance, we find that consumer surplus is always higher with second-price auctions than with first-price auctions. Moreover, as firms can better extract consumer surplus when data collection increases, we can show that consumer surplus decreases with the number of segments collected by the intermediary.

Our article contributes to the economic literature on three main points. First, several articles have analyzed the optimal selling strategy of a monopolist data intermediary using first-price auctions (Montes et al., 2019; Bounie et al., 2021). In this article, we show that the results are not robust to a change of selling mechanisms such as posted prices, bargaining or second-price auctions (Wang, 1993, 1995; Jindal and Newberry, 2018; Larsen and Zhang, 2018). Therefore, this article contributes to the literature by showing that the selling mechanism changes the optimal information selling strategy of the data intermediary, and that there are selling mechanisms that yield higher consumer surplus than others. In particular, consumer surplus is maximized when the data intermediary sells all information to both firms.

Secondly, we go a step further than the existing literature by considering the data collection strategy of the intermediary. In previous articles, the intermediary chooses the quantity of information to sell to firms, but makes no strategic choice regarding the quantity of data to collect from consumers (Bergemann et al., 2022). Our paper shows that the optimal amount of data collected by an intermediary depends on the selling mechanism that will in turn determine the optimal information selling strategy to firms. This has important implications for competition policies and contributes to recent debates on the regulation of privacy and competition (Furman et al., 2019; Scott Morton et al., 2019; Tirole, 2020).

Finally, our results also contribute to the literature that compares selling mechanisms pair-wise when a firm sells a homogeneous product to consumers (Einav et al., 2018). First, we have developed a framework that facilitates the comparison of selling mechanisms by determining whether they belong to a specific class. By doing so, we can compare not only selling mechanisms pair-wise, but also classes of mechanisms. Secondly, the seller designs an information partition that will change with the selling mechanism. In other words, the object under study is not a homogeneous product but a heterogeneous product the quality of which can change with the selling mechanism. Finally, we consider a market where a data intermediary sells information to competing firms, thus fully taking into account the competition effect of information and the outside option effect of the selling mechanism.

2 Model

The purpose of this article is to provide a tractable general model of information intermediation, in which a data intermediary strategically collects and sells consumer data to firms competing in a product market. We describe in this section the nature of competition, the different agents, as well as our approach to modeling information.

2.1 Nature of Competition in the Product Market

We consider a model of competition à la Hotelling on the product market. This model has been used to analyze the impact of information on market competition starting from Thisse and Vives (1988) and Liu and Serfes (2004), and has been recently used by Nageeb Ali et al. (2022) to analyze consumers' decisions to reveal personal information. Information will be modeled as partitions of the unit line, an approach largely used in finance, statistics, games with common knowledge, and also in the economics of information (Laffont, 1989).

Consumers are assumed to be uniformly distributed on a unit line $[0, 1]$. They purchase one product from two competing firms that are located at the two extremities of the line, 0 and 1. A monopolist data intermediary collects and sells data that segment consumer demand on the Hotelling line. A firm that acquires an information partition can set a price on each consumer segment and will be considered to be informed. On the contrary, a firm that does not purchase consumer segments remains uninformed and cannot distinguish consumers. This firm sets a single price on the entire line.

2.2 Consumers

Consumers buy one product at a price p_1 from Firm 1 located at 0, or at a price p_2 from Firm 2 located at 1. Consumers located at $x \in [0, 1]$ receive a utility V from purchasing the product, but incur a cost $t > 0$ of consuming a product that does not perfectly fit their taste x. Therefore, buying from Firm 1 (resp. from Firm 2) incurs a cost tx (resp. $t(1 - x)$). Consumers choose the product that gives the highest level of utility:

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

2.3 Data Intermediary

We consider a data intermediary collecting consumer information that allows firms to distinguish consumer segments on the unit line. The data intermediary then chooses the optimal information partitions to sell to firms that price discriminate consumers.[2]

The data intermediary collects $k \in \mathbb{N}_+^*$ consumer segments at a cost $c : \mathbb{N}_+^* \to \mathbb{R}_+$.[3] The cost of collecting information encompasses various dimensions of the activity of the data intermediary such as installing trackers and online cookies allowing it to observe the behavior of web users (Bergemann and Bonatti, 2015), or storing and handling data to eventually build detailed consumer scores (Bonatti and Cisternas, 2020) (see Varian (2018) for a detailed discussion on the structure of the costs associated with data collection). The data collection cost $c(.)$ captures the sum of the costs associated with these activities.

Collecting data is costly for the intermediary but provides more information on consumers. Following Liu and Serfes (2004) and Bounie et al. (2021), data divides the unit line into k segments of equal size, which allow firms to locate consumers more precisely. For instance, when $k = 2$, information is coarse, and firms can only distinguish whether consumers belong to $[0, \frac{1}{2}]$ or to $[\frac{1}{2}, 1]$. At the other extreme, when k converges to infinity, the data intermediary knows the exact location of each consumer. Thus, $\frac{1}{k}$ can be interpreted as the precision of the information collected by the data intermediary. The k segments of size $\frac{1}{k}$ form a partition $\mathcal{P}^k$, illustrated in Figure 1, that we refer to as the reference partition.

[2] We assume that the market is covered, and that information is used for price discrimination. This has been extensively studied since Thisse and Vives (1988), Liu and Serfes (2004), Stole (2007), and Ulph et al. (2007). More recently, price discrimination in models of horizontal differentiation has been studied by Belleflamme et al. (2020), Elliott et al. (2021) and Nageeb Ali et al. (2022).

[3] We assume that $c(k)$ satisfies standard convexity conditions: $c(0) = 0$ and $c(x), c''(x) > 0$.

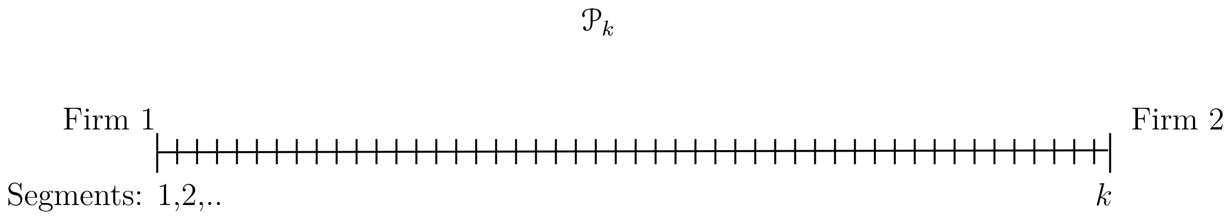

Figure 1: Reference partition $\mathcal{P}^k$

The elementary segments of size $\frac{1}{k}$ of partition $\mathcal{P}^k$ generate a sigma-field $\mathbb{P}_k$ containing the 2^{k-1} possible partitions of the unit line, among which the intermediary can select the partition to sell. The choice of the partition corresponds to the selling strategy of the intermediary, which we analyze in detail in Section 3. We denote by $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{P}_k$ the partitions proposed respectively to Firm 1 and Firm 2.

Before describing firms in the product market, it is worth noting that we do not allow the intermediary to use disjoint intervals from the set of possible partitions. Including such intervals in the analysis would greatly increase the set of possible combinations of consumer segments. To better understand the implications of disjoint intervals on our results, consider disjoint intervals of the type $[0, \frac{1}{k}] \cup [1 - \frac{1}{k}, 1]$. In this case, firms do not know whether they are facing close-by consumers or consumers located at the other extremity of the Hotelling line. These partitions could sustain collusive monopoly prices and thus increase the willingness to pay for information of both firms. However, disjoint intervals also have a competitive effect on firms. Firms can decide to fight for consumers on the other side of the unit line, which lowers the prices and profits of firms. We can show that for arbitrary values of the primitives of the model (t and V) the competitive effect is stronger than the collusive effect and firms have interest to undercut prices, leading to lower profits than without disjoint intervals.[4]

Finally, to simplify the exposition, we assume that the data intermediary sells information to only one firm, say Firm 1, keeping Firm 2 uninformed. In Section 5, we

[4] The proof is available upon request. There are two other reasons why we do not consider disjoint intervals. First, equilibria with disjoint intervals do not converge to first-degree price discrimination when information is perfect. Indeed, when t, for each identified consumer, firms have an uncertainty on their location, which creates a paradoxical situation: firms have perfect precision of information but still cannot locate consumers. Secondly, the marketing literature has emphasized the importance of targeting advertising to consumers who have the strongest preferences for the products of a firm.

give the conditions under which the intermediary sells information to one firm or to both firms. The intuitions that we develop in the next sections remain valid when the intermediary sells information to two firms.

2.4 Firms

Firm 1 can purchase information to price discriminate identified consumers. If Firm 1 remains uninformed, it only knows that consumers are uniformly distributed on the unit line and sets a unique price.

In order to compute the profits of the firms, we need to determine the demands and prices for both firms in each consumer segment. Firm 2 has no information and sets a uniform price on the whole interval $[0, 1]$. Firm 1 has partition $\mathcal{P}_1$ and sets different prices in each segment of the partition. There are two types of segments to analyze: segments on which both firms have a strictly positive demand, and segments on which Firm 1 is a monopolist. We assume that Firm 1 sets prices in two stages.[5] First, Firm 2 sets a homogeneous price p_1 on the whole unit line, and Firm 1 simultaneously sets prices on segments where it shares consumer demand with Firm 2. Then, on segments where it is a monopolist, Firm 1 sets monopoly prices, constrained by p_2. Finally, consumers observe prices and choose which product to purchase.

For any partition $\mathcal{P}_1$ composed of n segments, we denote by $d_{\theta i}$ the demand of Firm $\theta = \{1, 2\}$ on the ith segment. Firm 1 is informed and maximizes the following profit function with respect to the vector of prices $\mathbf{p}_1 := (p_{11}, .., p_{1i}, .., p_{1n})$:

$$\pi_1(\mathbf{p}_1) = \sum_{i=1}^{n} d_{1i} p_{1i}$$

[5] Sequential pricing decision avoids the nonexistence of Nash equilibrium in pure strategies, and allows an informed firm to charge consumers a higher price. This practice is common in the literature and is supported by managerial evidence. For instance, Acquisti and Varian (2005) use sequential pricing to analyze inter-temporal 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-value consumers who are better identified when they join the Prime program.

Firm 2 is uninformed, charges a homogeneous price p_2 on the whole unit line, and maximizes $\pi_2 = \sum_{i=1}^{n} d_{2i} p_2$ with respect to p_2.

2.5 Timing

The timing of the game is the following:

- Stage 1: the data intermediary collects data on k consumer segments at cost $c(k)$.
- Stage 2: the data intermediary sells information partition $\mathcal{P}_1$ to Firm 1.
- Stage 3: firms set prices p_{1i} and p_2 on the competitive segments.
- Stage 4: Firm 1 sets prices on the monopoly segments.

3 Selling Mechanisms and Optimal Information Partitions

The strategies of the firms and of the data intermediary critically depend on the way information is sold, i.e. the selling mechanism, which influences the price of information – defined as the profits of Firm 1 with information minus its profits without information – and the incentive of the intermediary to collect data.

In this section, we first define what is a selling mechanism as well as the price of information for any mechanism. Secondly, we study the properties of classes of mechanisms, and we apply the main results to four selling mechanisms commonly studied in the economic literature: posted prices (pp), sequential bargaining (seq), first-price auctions (a) and second-price auctions (a_2). With posted prices, the data intermediary proposes an information partition to Firm 1 that accepts or rejects the offer. If Firm 1 declines the offer, both firms remain uninformed. The second mechanism, sequential bargaining, allows the data intermediary to propose information to Firm 2 if Firm 1 declines the offer, and so on until one of the firms acquires information. The last two selling mechanisms are first and second-price auctions with a negative externality: a firm that loses the auction faces an informed competitor, similarly to sequential bargaining. A full characterization of these mechanisms is given in Appendix A.

3.1 Selling Mechanisms: Definition

The objective of the intermediary is to choose information partitions that maximize the willingness to pay of firms. When the data intermediary sells information to one firm only, the intermediary chooses an optimal partition that corresponds to the willingness to pay of Firm 1, i.e. the price of information paid by Firm 1.

We introduce notations that simplify the exposition of the model. Let $\mathcal{P}_1 \in \mathbb{P}_k$ denote the partition sold to Firm 1 if it purchases information, and $\mathcal{P}_2 \in \mathbb{P}_k$ the partition sold to Firm 2 in case Firm 1 does not purchase information. We also denote by $\pi_1(\mathcal{P}_1)$ the profit of Firm 1 with partition $\mathcal{P}_1$ (Firm 2 is uninformed), and by $\bar{\pi}_1(\mathcal{P}_2)$ the profit of Firm 1 when it is uninformed and faces Firm 2 that has partition $\mathcal{P}_2$.

Let $p_1, p_2 : \mathbb{P}_k \times \mathbb{P}_k \to \mathbb{R}$ be the prices proposed respectively to Firm 1 and Firm 2. A selling mechanism $s_i(k)$ is a quadruplet $\{\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)}, p_1, p_2\}$ for a given $k \in \mathbb{N}_+^*$, where $i = 1, .., 2^{2k-2}$ indexes all potential mechanisms and has the cardinality of $\mathbb{P}_k \times \mathbb{P}_k$. Using these notations, when the data intermediary has collected k segments, the price charged to Firm 1 for any selling mechanism can be written as:

$$p_1(\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)}) = \pi_1(\mathcal{P}_1^{(i)}) - \bar{\pi}_1(\mathcal{P}_2^{(i)}). \tag{1}$$

The data intermediary chooses partitions $\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)} \in \mathbb{P}_k$ to maximize the price of information $p_1(\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)})$. It is clear from this expression that prices are fully determined by the choice of partitions $\mathcal{P}_1^{(i)}$ and $\mathcal{P}_2^{(i)}$, and therefore, we can simplify the notations by denoting a selling mechanism as a couple $s_i(k) = \{\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)}\}$. Let $\mathcal{S}_k = \{s_i(k)\}_{i=1}^{2^{2k-2}}$ be the set of all selling mechanisms generated by $\mathbb{P}_k \times \mathbb{P}_k$ for a given k, and $\mathcal{S} = \{\mathcal{S}_k\}$ for $k \in \mathbb{N}_+^*$ be the set of all possible selling mechanisms. In the remainder of the article, we drop reference to (i) when there is no confusion.

3.2 Optimal Information Partitions

In this section, we characterize the optimal information partitions. First, we study the case where the data intermediary sells partition $\mathcal{P}_1$ that maximizes the profits of Firm 1 with $\pi_1(\mathcal{P}_1)$. Let $\mathcal{M} \subset \mathcal{S}$ denote this class of selling mechanisms. We can show that for any selling mechanisms belonging to $\mathcal{M}$, the optimal partition has the following features. Partition $\mathcal{P}_1$ divides the unit line into two intervals: the first interval consists of j_1 segments (with $j_1 \in [\![0, k]\!]$) of size $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$. We refer to this interval as the share of identified consumers.[6] The data intermediary does not sell information on consumers in the second interval of size $1 - \frac{j_1}{k}$, and firms charge a uniform price on this second interval. We refer to this interval as the share of unidentified consumers.

Theorem 1. *(Optimal partition for selling mechanisms in $\mathcal{M}$)*

For any selling mechanisms in $\mathcal{M}$, the optimal partition divides the unit line into two intervals:

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

Proof: see Appendix B.

The optimal partition described in Theorem 1 balances the rent extraction and the competitive effects of information. Indeed, when choosing partition $\mathcal{P}_1$, there are two opposite effects on the willingness to pay of Firm 1 for information. On the one hand, more information allows Firm 1 to extract more surplus from consumers. This rent extraction effect increases the price of information. On the other hand, selling more consumer segments increases competition because Firm 1 has information on consumers that are closer to Firm 2, and thus can lower prices for these consumers (Thisse and Vives, 1988). This competition effect lowers the profit of firms, which decreases the price of information.

Partition $\mathcal{P}_1$ is described in Figure 2.[7]

[6]Thus $\frac{j_1}{k} \in [0, 1]$.

[7]Similarly, we focus our analysis on information partitions that are optimal for Firm 2, such that they identify all consumer segments closest to its location up to a cutoff point.

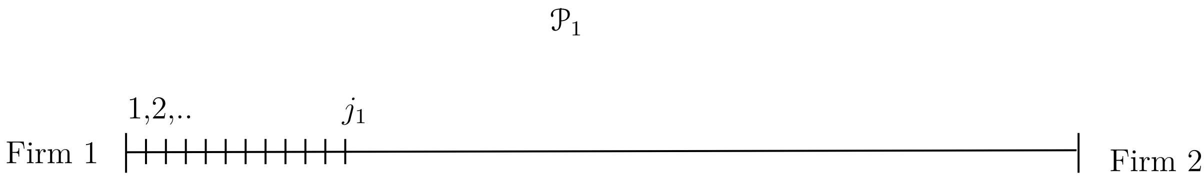

Figure 2: Selling partition $\mathcal{P}_1$ to Firm 1

A large set of selling mechanisms belongs to $\mathcal{M}$ such as various forms of Nash bargaining and infinite sequential bargaining with discount factors, as well as posted prices, sequential bargaining and first-price auctions. We use as an illustration three standard selling mechanisms in Example 1.

Example 1 (Optimal partitions for first-price auctions, sequential bargaining and posted prices)**.**

First-price auctions, sequential bargaining and posted prices belong to $\mathcal{M}$*, and the optimal partitions are given by Theorem 1.*[8] *With posted prices, Firm 2 remains uninformed regardless of the decision of Firm 1 to purchase information, and partition* $\mathcal{P}_1$ *is chosen to maximize the profits of Firm 1. With first-price auctions, the intermediary auctions simultaneously two partitions: partition* $\mathcal{P}_1$ *is chosen to maximize the profits of Firm 1, while partition* $\mathcal{P}_2$ *minimizes its profits if it does not purchase information. With sequential bargaining, at each stage of the process, the firm that declines the offer has no information, even though the competitor can acquire information at the following stage. Hence, in the unique subgame perfect equilibrium, the intermediary offers at each stage a partition that maximizes the profits of a prospective buyer, i.e. the profits of Firm 1.*

However, there are also selling mechanisms that do not belong to $\mathcal{M}$, and for which the partition does not maximize the profits of Firm 1. Proving that the partition described in Theorem 1 is optimal for mechanisms that are not in $\mathcal{M}$ is a complex task given the high dimensionality of the optimization problem, since the data intermediary can

[8]We characterize the equilibrium prices and partitions with these three mechanisms in Appendix A.

potentially recombine consumer segments in any arbitrary fashion.[9] For the tractability of the model, we assume for the remainder of the article that the intermediary can only sell information partitions satisfying Assumption 1.

Assumption 1.

Feasible partitions divide the unit line into the two intervals given by Theorem 1.

We have shown in Theorem 1 that partitions restricted by Assumption 1 are optimal for mechanisms in $\mathcal{M}$, which include among other first-price auctions, posted prices and sequential bargaining. We conjecture that the optimal partitions for mechanisms not included in $\mathcal{M}$ should have the same structure as partitions given by Theorem 1. Indeed, information partitions that are ruled out by Theorem 1 are possibly partitions that allow firms to poach consumers located far away from their locations. In the context of advertising, Iyer et al. (2005) have shown that firms optimally target consumers with the highest willingness to pay for their products, which is similar to Assumption 1.

Example 2 illustrates how we can characterize the optimal partitions for second-price auctions in which the winning bidder pays the price of the second-highest bid. We show that second-price auctions do not belong to $\mathcal{M}$ and how using Assumption 1 will allow us to provide a tractable analysis.

Example 2 (Optimal partitions for second-price auctions)**.**

With second-price auctions, the data intermediary auctions partitions $\mathcal{P}_1^{a_2}$ *and* $\mathcal{P}_2^{a_2}$*, and Firm 1 (the highest bidder) pays the price corresponding to the bid of Firm 2 (the lowest bidder) for partition* $\mathcal{P}_2^{a_2}$*. Consequently, the willingness to pay of Firm 1 and Firm 2 for information are respectively* $\pi_1(\mathcal{P}_1^{a_2}) - \bar{\pi}_1(\mathcal{P}_2^{a_2})$ *and* $\pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$*, and the profit of the intermediary corresponds to the minimum of these bids. Hence, the objective function of the intermediary is to maximize* $\pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$ *under the constraint that* $\pi_1(\mathcal{P}_1^{a_2}) - \bar{\pi}_1(\mathcal{P}_2^{a_2}) \geq \pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$*. In equilibrium, the constraint is binding, which is achieved when* $\mathcal{P}_1^{a_2}$ *and* $\mathcal{P}_2^{a_2}$ *are*

[9] For mechanisms that do not belong to $\mathcal{M}$, we have to compare how different partitions impact not only the profits of Firm 1, but also its outside option. The resulting set of cases to compare increases exponentially.

symmetric with respect to $\frac{1}{2}$, which we denote: $\mathcal{P}_2^{a_2} = sym(\mathcal{P}_1^{a_2})$.[10] *Hence, $\mathcal{P}_1^{a_2}$ is chosen not only to allow Firm 1 to reach a high profit, but as it also impacts partition $\mathcal{P}_2^{a_2}$ sold to Firm 2 in case Firm 1 does not purchase information, $\mathcal{P}_1^{a_2}$ is also chosen to increase the threat on Firm 1 if it remains uninformed, and the optimal partition balances these two effects.*

Under Assumption 1, the data intermediary auctions partitions $j_1^{a_2}$ and $j_2^{a_2}$, the equilibrium constraint $\mathcal{P}_2^{a_2} = sym(\mathcal{P}_1^{a_2})$ is equivalent to $j_1^{a_2} = j_2^{a_2}$ and thus we can use the same general notations than for mechanisms in $\mathcal{M}$ in the rest of the article.

For the remainder of the article, we assume that the data intermediary can sell any partition given by Theorem 1. Let $\mathcal{S}' \subset \mathcal{S}$ be the set of all such selling mechanisms. We can characterize any element of $\mathcal{S}'$ by $s'(k) = \{j_1, j_2\}$ with $j_1, j_2 \in [\![0, k]\!]$.

The optimization problem for the data intermediary boils down to choosing the values of $j_1(k)$ and $j_2(k)$ corresponding to the number of consumer segments sold to Firm 1 and Firm 2. We drop subscript k when there is no confusion.

3.3 Price of Information in Equilibrium

In the remaining of the analysis, we simplify notations by dropping references to the null partition sold to Firm 2, so that $\pi_1(j_1, k)$ denotes the profit of Firm 1 when it has information on the $j_1 \in [\![0, k]\!]$ consumer segments closest to its location and Firm 2 is uninformed. We denote by $\bar{\pi}_1(j_2, k)$ the profit of Firm 1 when it is uninformed and faces Firm 2 that has information on the $j_2 \in [\![0, k]\!]$ consumer segments closest to its location.

Using these notations, the price of information can be written as:

$$p_1(j_1, j_2, k) = \pi_1(j_1, k) - \bar{\pi}_1(j_2, k). \tag{2}$$

Using Equation 2, we can characterize in Example 3 the equilibrium prices for the four selling mechanisms illustrated in this article. Proofs are given in Appendix A.

[10] Application $sym : \mathbb{P}_k \to \mathbb{P}_k$ is such that $\mathcal{P}_1^{a_2}$ and $sym(\mathcal{P}_1^{a_2})$ are symmetric with respect to $\frac{1}{2}$.

Example 3 (Equilibrium prices for posted prices, sequential bargaining, first-price and second-price auctions).

Let π be the profits in the standard Hotelling model without information, and $\mathcal{P}^k$ be the reference partition containing all available consumer segments. The equilibrium prices for the four mechanisms are as follows:[11]

$$\textit{Posted prices}: \quad p_1^{pp}(j_1^{pp}, \emptyset, k) = \pi_1(j_1^{pp}, k) - \pi.$$

$$\textit{Sequential bargaining}: \quad p_1^{seq}(j_1^{seq}, j_2^{seq}, k) = \pi_1(j_1^{seq}, k) - \bar{\pi}_1(j_2^{seq}, k).$$

$$\textit{First-price auctions}: \quad p_1^{a}(j_1^{a}, \mathcal{P}^k, k) = \pi_1(j_1^{a}, k) - \bar{\pi}_1(\mathcal{P}^k, k).$$

$$\textit{Second-price auctions}: \quad p_1^{a_2}(j_1^{a_2}, j_1^{a_2}, k) = \pi_1(j_1^{a_2}, k) - \bar{\pi}_1(j_1^{a_2}, k).$$

3.4 Classes of Selling Mechanisms

We have shown in Section 3.2 that the optimization problem for the data intermediary boils down to choosing a single value $j_1(k)$ corresponding to the number of consumer segments sold to Firm 1. In this section, we characterize three classes of selling mechanisms based on the relationship between the number of segments j_1 sold to Firm 1 and the number of segments j_2 proposed to Firm 2 if Firm 1 does not acquire information.

We first note that, given Assumption 1, the number of consumer segments j_2 proposed to Firm 2 by the data intermediary for a given k – the outside option of Firm 1 – can be written as a function of j_1: $j_2(.) : [\![0, k]\!] \to [\![0, k]\!]$. In order to find the optimal integer value of j_2, we will consider $j_2(.)$ as a continuous function, differentiable with respect to j_1, and throughout the resolution we will use $j_2(.) : [0, k] \to [0, k]$ to characterize the equilibrium strategies of the intermediary. Moreover, for the sake of clarity

[11] Note that for second-price auctions $j_2^{a_2}$ and $j_1^{a_2}$ are equal in equilibrium.

we focus our analysis on mechanisms for which the second-order conditions with respect to j_1 and k are satisfied: $\frac{\partial^2 p_1(j_1,j_2,k)}{\partial j_1^2}|_{j_2^*} < 0$ and $\frac{\partial^2 p_1(j_1,j_2,k)-c(k)}{\partial k^2}|_{j_1^*,j_2^*} < 0$. Nevertheless, our results directly apply to any mechanisms with corner solutions and with multiple interior equilibria.

We work in the following sections with the set of mechanisms $\mathcal{S}_k^c$ such that for a given k, j_1, j_2 are continuous and belong to $[0, k]$, and all the above conditions are satisfied. Let $\mathcal{S}^c = \{\mathcal{S}_k^c\}$ denote the set of all such mechanisms, for $k \in \mathbb{N}_+^*$. Note that $\mathcal{S}'$, the restricted – and discrete – set of mechanisms given by Assumption 1, is included in $\mathcal{S}^c$ so that we will be able to find the optimal integers in $\mathcal{S}'$ by using the best outcome among the two integers closest to the optimum values of j_1 and j_2. We define for $j_1, j_2 \in [0, k]$ the set $\mathcal{M}_k^c \subset \mathcal{S}_k^c$ including all mechanisms with $j_2(.)$ a continuous function of j_1 for which the intermediary sells an information partition that maximizes the profit of Firm 1. The set of such mechanisms for any value of k is defined by $\mathcal{M}^c = \{\mathcal{M}_k^c\}$ for $k \in \mathbb{N}_+^*$, and we have in particular $\mathcal{M}^c \subset \mathcal{S}^c$.

Using this specification, we focus on the following inequality: $\frac{\partial j_2(j_1)}{\partial j_1} \lesseqgtr 0$. We define accordingly three classes of selling mechanisms.

Independent offers $\mathcal{I}$. The first class of mechanisms includes all offers for which $j_2(.)$ does not vary with j_1. We call $\mathcal{I}_k$ the class of independent mechanisms for a given value of k, as j_1 and j_2 are chosen independently in equilibrium. We also define the set of independent mechanisms for any value of k by $\mathcal{I} = \{\mathcal{I}_k\}$ for $k \in \mathbb{N}_+^*$. Mechanisms in $\mathcal{I}$ are formally characterized by $\frac{\partial j_2(j_1)}{\partial j_1} = 0,\ \forall\ j_1$. For mechanisms in $\mathcal{I}$, the intermediary chooses j_1 that maximizes the profit of Firm 1, and therefore $\mathcal{I} \subset \mathcal{M}^c$. This implies that all partitions satisfying Assumption 1 are optimal for mechanisms in $\mathcal{I}$. We show in Example 4 that posted prices, sequential bargaining and first-price auctions belong to independent offers.

Example 4 (Posted prices, sequential bargaining and first-price auctions belong to independent offers $\mathcal{I}$)**.**

With posted prices, Firm 2 remains uninformed regardless of the decision of Firm 1 to purchase information, and the outside option of Firm 1 is independent of *the information partition proposed by the data intermediary. With first-price auctions, when Firm 1 does not acquire information, Firm 2 has information on all consumer segments. Thus, the outside option of Firm 1 that is impacted by the partition proposed to Firm 2 is independent of the partition proposed to Firm 1. With sequential bargaining, at each stage of the process, the firm that declines the offer has no information, even though the competitor can acquire information at the following stage. Here again, the outside option of Firm 1 is independent of the information partition proposed by the data intermediary to Firm 1.*

Dependent offers: increasing functions $\mathcal{S}^{c+}$. Mechanisms in the second class are characterized by a uniformly positive relation between j_1 and j_2: $\frac{\partial j_2(j_1)}{\partial j_1} > 0, \ \forall \ j_1$. We denote by $\mathcal{S}_k^{c+}$ the set of such mechanisms for a given k and by $\mathcal{S}^{c+}$ the total set of such mechanisms. We show in Example 5 that second-price auctions belong to $\mathcal{S}^{c+}$.

Example 5 (Second-price auctions belong to $\mathcal{S}^{c+}$)**.**

For second-price auctions, we have seen in Example 2 that $j_2(j_1) = j_1$, and the price of information is equal to $p_1^{a_2} = \pi_1(j_1^{a_2}, k) - \bar{\pi}_1(j_1^{a_2}, k)$. Hence, $\frac{\partial j_2^{a_2}(j_1^{a_2})}{\partial j_1^{a_2}} = 1$, and the number of segments sold $j_1^{a_2}$ does not maximize the profit of Firm 1 anymore, since the intermediary also takes into account the impact of $j_1^{a_2}$ on the outside option of Firm 1 $\bar{\pi}_1(j_1^{a_2}, k)$.

Dependent offers: decreasing functions $\mathcal{S}^{c-}$. The third class of mechanisms, which we denote for a given k by $\mathcal{S}_k^{c-}$ and overall by $\mathcal{S}^{c-} = \{\mathcal{S}_k^{c-}\}$ for $k \in \mathbb{N}_+^*$, is characterized by a uniformly negative relation between j_1 and j_2: $\frac{\partial j_2(j_1)}{\partial j_1} < 0, \ \forall \ j_1$. We illustrate in Example 6 a selling mechanism that belongs to $\mathcal{S}^{c-}$.

Example 6 (Cap J on j_1 and j_2)**.**

A total cap J on the number of consumer segments sold on the market (for a given value of k) introduces a negative relation between j_1 and j_2: $j_1 + j_2 = J$ and $\frac{\partial j_2(j_1)}{\partial j_1} = -1$.

Relation between the classes of selling mechanisms. We can characterize the relations between the three different classes of mechanisms described previously. The set of independent offers $\mathcal{I}$ is included in $\mathcal{M}^c$: $\mathcal{I} \subset \mathcal{M}^c$. However, there are mechanisms within $\mathcal{M}^c$, for which the intermediary sells an information partition that maximizes the profits of Firm 1, but where j_2 depends on j_1, hence that do not belong to $\mathcal{I}$ (see Appendix C for an example). Classes $\mathcal{S}^{c-}$, $\mathcal{S}^{c+}$ and $\mathcal{M}^c$ are mutually exclusive by definition: $\mathcal{S}^{c+} \cap \mathcal{S}^{c-} = \emptyset$, $\mathcal{S}^{c+} \cap \mathcal{M}^c = \emptyset$ and $\mathcal{S}^{c-} \cap \mathcal{M}^c = \emptyset$.[12]

There are many other possible relations between j_1 and j_2, such as $j_2(.)$ being a non-monotonous or non-differentiable function. As mentioned above, such functions may yield multiple interior equilibria or corner solutions, which are ruled out of the analysis for simplicity. Nevertheless, the insights that we gain from the analysis of classes $\mathcal{I}$, $\mathcal{S}^{c-}$ and $\mathcal{S}^{c+}$ generally apply to any mechanism in $\mathcal{S}^c$, and by extension in $\mathcal{S}'$.

In the next section, we characterize the data collection and selling strategies of the intermediary for the three classes of selling mechanisms, and we analyze their impact on market competition and consumer surplus.

4 Characterization of the Equilibrium

We solve the game by backward induction for any selling mechanisms in $\mathcal{S}^c$. In Section 4.1, the intermediary sells information to Firm 1 only and we characterize the optimal selling strategy. We characterize in section 4.2 the data collection strategy of the intermediary.

4.1 Number of Consumer Segments Sold in Equilibrium

We characterize the number of consumer segments sold to Firm 1 for a given precision k. The number of segments sold to Firm 1 will impact the intensity of competition in the product market, as well as consumer surplus.

The price of information $p_1(j_1, j_2, k)$ corresponds to the willingness to pay of Firm 1, which can be written as the difference between its profits with information j_1 and

[12] We summarize the notations and the properties of classes in Appendix D.

its profits when it remains uninformed and Firm 2 acquires partition j_2, as defined in Eq. 2. The optimal number of consumer segments sold to Firm 1 satisfies the following first-order condition with respect to j_1:

$$\frac{\partial p_1(j_1, j_2, k)}{\partial j_1} = \underbrace{\frac{\partial \pi_1(j_1, k)}{\partial j_1}}_{\substack{\text{First-order effect} \\ \text{on profits}}} \underbrace{- \frac{\partial \bar{\pi}_1(j_2, k)}{\partial j_2} \frac{\partial j_2(j_1)}{\partial j_1}}_{\substack{\text{Outside option} \\ \text{effect}}} = 0. \tag{3}$$

We can characterize the equilibrium partitions by studying the signs of the two terms of Eq. 3. The sign of the first-order effect on profits depends on the equilibrium value of j_1 compared to the number of segments $\hat{j}_1$ that maximizes the profits of Firm 1. The sign of the outside-option effect depends on the impact of j_2 on the profits of Firm 1 when it is uninformed, $\frac{\partial \bar{\pi}_1(j_2,k)}{\partial j_2}$, and on the relation between j_1 and j_2 through $\frac{\partial j_2(j_1)}{\partial j_1}$, which depends on the selling mechanism.

We characterize in Lemma 1 the variations of $\frac{\partial \pi_1(j_1,k)}{\partial j_1}$ and $\frac{\partial \bar{\pi}_1(j_2,k)}{\partial j_2}$. We will then discuss the impact of $\frac{\partial j_2(j_1)}{\partial j_1}$ on market outcomes for mechanisms in $\mathcal{I}, \mathcal{S}^+$ and $\mathcal{S}^{c-}$.

Lemma 1.

The profits of Firm 1 with and without information have the following properties:

- *(a) The profits of Firm 1 with information are quasi-concave with respect to j_1 and have a unique maximum at:*

$$\hat{j}_1(k) = \frac{6k - 9}{14}.$$

- *(b) The profits of Firm 1 when it is uninformed always decrease with j_2:*

$$\frac{\partial \bar{\pi}_1(j_2, k)}{\partial j_2} \leq 0.$$

Proof: see Appendix E.

Lemma 1 (a) characterizes the partition that maximizes the profits of Firm 1. From Eq. 3, it is clear that the optimal partition for the intermediary may differ from $\hat{j}_1(k)$, according to the outside option effect. Indeed, Lemma 1 (b) shows that the profits of Firm 1 always decrease when more segments are sold to Firm 2. Hence, a crucial element of the selling strategy of the intermediary is the value of $\frac{\partial j_2(j_1)}{\partial j_1}$, which depends on the selling mechanism. There are three cases to analyze: $\frac{\partial j_2(j_1)}{\partial j_1} = 0$, $\frac{\partial j_2(j_1)}{\partial j_1} < 0$ and $\frac{\partial j_2(j_1)}{\partial j_1} > 0$. These three cases correspond respectively to mechanisms belonging to classes $\mathcal{I}$, $\mathcal{S}^{c-}$ and $\mathcal{S}^{c+}$, which we now discuss.

Independent offers. Consider a selling mechanism in $\mathcal{I}$ with $\frac{\partial j_2(j_1)}{\partial j_1} = 0$, for which offers are independent. Proposition 1 immediately follows from Eq. 3 and Lemma 1 (a):

Proposition 1.

For all mechanisms in $\mathcal{I}$, the data intermediary sells a partition that maximizes the profits of Firm 1:

$$\frac{\partial j_2(j_1)}{\partial j_1} = 0 \implies j_1^*(k) = \hat{j}_1(k) = \frac{6k-9}{14}.$$

Proposition 1 characterizes an important property of mechanisms belonging to $\mathcal{I}$, for which the information sold to Firm 1 (j_1) is independent of the information proposed to Firm 2 (j_2) if Firm 1 does not acquire information. For all mechanisms in $\mathcal{I}$, the same number of consumer segments $\hat{j}_1$ is sold to Firm 1, and for a given precision k, they will yield an identical market outcome in terms of surplus and industry profits.

As we have emphasized in Section 3.4, a large set of selling mechanisms satisfy this property such as various forms of Nash bargaining and infinite sequential bargaining with discount factors, but also first-price auctions, sequential bargaining, and posted prices. Hence, we establish the uniqueness of the optimal number of segments sold j_1 for any selling mechanism in $\mathcal{I}$ – where j_1 and j_2 are independent. We illustrate in Example 7 the optimal selling strategy with first-price auctions, sequential bargaining, and posted prices.

Example 7 (Optimal selling strategy with posted prices, sequential bargaining, and first-price auctions)**.**

First-price auctions, sequential bargaining and posted prices belong to $\mathcal{I}$, and the *data intermediary sells in equilibrium the same number of consumer segments to Firm 1.*

$$j_1^{pp*}(k) = j_1^{seq*}(k) = j_1^{a*}(k) = \hat{j}_1(k). \quad (4)$$

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

The fact that the data intermediary chooses the same number of segments with independent information partitions is far from being trivial from an economic point of view. Indeed, the outside options in posted prices, sequential bargaining, and first-price auctions reflect different levels of threats. For example, with posted prices, there is no threat to Firm 1 if it does not purchase information. On the contrary, if Firm 1 declines the offer with first-price auctions, the data intermediary sells to Firm 2 the partition that minimizes the profits of Firm 1. Thus, the strength of the threat of the outside option greatly varies between the different selling mechanisms. Different mechanisms in $\mathcal{I}$ will lead to different market outcomes as the intermediary will have different incentives to collect data through changes of the outside option. We analyze data collection strategies in the next section.

Increasing and decreasing functions. The result established in Proposition 1 does not necessarily hold for mechanisms that do not belong to $\mathcal{I}$. We characterize in Proposition 2 the equilibrium numbers of segments sold by the intermediary with mechanisms in $\mathcal{S}^{c+}$ and $\mathcal{S}^{c-}$ for which j_1 and j_2 are not independent.[13]

Proposition 2.

The optimal amount of information sold to Firm 1 satisfies:

- *(a) For mechanisms in $\mathcal{S}^{c+}$: $\frac{\partial j_2(j_1)}{\partial j_1} > 0$, and $j_1^*(k) > \hat{j}_1(k)$.*
- *(b) For mechanisms in $\mathcal{S}^{c-}$: $\frac{\partial j_2(j_1)}{\partial j_1} < 0$, and $j_1^*(k) < \hat{j}_1(k)$.*

Proposition 2 characterizes the amount of information sold to Firm 1 for mechanisms in $\mathcal{S}^{c+}$ and $\mathcal{S}^{c-}$. Proposition 1 and Proposition 2 are crucial to understand the impacts of

[13]The proof is straightforward from Eq. 3 and Lemma 1 (a).

selling mechanisms on consumer surplus. As we will see in Section 6, surplus increases with the number of consumer segments sold to Firm 1, and Proposition 2 allows us to compare surplus with different selling mechanisms by only considering the relations between j_1 and j_2.

We characterize in Example 8 the equilibrium number of segments sold to Firm 1 when the intermediary uses second-price auctions, a selling mechanism included in $\mathcal{S}^{c+}$ and for which we have $\frac{\partial j_2(j_1)}{\partial j_1} = 1$. A similar effect takes place for mechanisms in $\mathcal{S}^{c-}$ for which $\frac{\partial j_2(j_1)}{\partial j_1} < 0$.

Example 8 (Optimal selling strategy with second-price auctions)**.**
The number of consumer segments sold in equilibrium with second-price auctions satisfies:[14]

$$j_1^{a_2*}(k) = \frac{4k-3}{6} > \hat{j}_1(k). \tag{5}$$

With second-price auctions, the number of segments chosen by the data intermediary does not maximize the profits of Firm 1 anymore, as the data intermediary internalizes the outside-option effect of j_1 through the variations of $j_2(j_1)$. Hence, our results emphasize the crucial role of selling mechanisms on the outside option, and thus on the amount of data sold in equilibrium.

4.2 Consumer Data Collection in Equilibrium

We now analyze how selling mechanisms impact the number of consumer segments collected k and the profits of the data intermediary.

The intermediary maximizes its profits by collecting k consumer segments. The profits of Firm 1 without information $\bar{\pi}_1$ only depend on k through $\frac{j_2^*(k)}{k}$, and we can write $\bar{\pi}_1(j_2^*, k) = \bar{\pi}_1(j_2^*(k))$. Thus, the data intermediary maximizes the following profits with respect to k:[15]

[14] The equilibrium value of $j_1^{a_2*}(k)$ is derived in Appendix F.
[15] We focus on data collection costs for which each profit is concave and reaches a unique maximum on $\mathbb{R}^+$.

$$p_1(j_1^*(k), j_2^*(k), k) - c(k) = \pi_1(j_1^*(k), k) - \bar{\pi}_1(j_2^*(k)) - c(k).$$

We characterize in Proposition 3 the variations of the price of information with the number of segments collected k:

Proposition 3.

The price of information in equilibrium satisfies:

- *(a) For all mechanisms that do not belong to $\mathcal{I}$:*

$$\frac{\partial j_2(j_1)}{\partial j_1} \neq 0, \quad and \quad \frac{\partial p_1(j_1^*(k), j_2^*(k), k)}{\partial k} = \frac{j_1^*(k)}{k}\frac{t}{k^2}.$$

- *(b) For all mechanisms in $\mathcal{I}$:*

$$j_1^*(k) = \hat{j}_1(k) = \frac{6k-9}{14}, \quad and \quad \frac{\partial p_1(\hat{j}_1(k), j_2^*(k), k)}{\partial k} = \frac{\hat{j}_1(k)}{k}\frac{t}{k^2} - \frac{\partial \bar{\pi}_1(j_2^*(k))}{\partial k}.$$

The proof follows from a direct application of the envelope theorem. The term $\frac{j_1(k)}{k}\frac{t}{k^2}$ corresponds to the derivative of $\pi_1(j_1(k), k)$ with respect to k. When $\frac{\partial j_2(j_1)}{\partial j_1} \neq 0$, the profits of uninformed Firm 1 facing Firm 2 informed with j_2 segments only vary with j_2 and not directly with k. In this case, the variations of j_2 are internalized in the equilibrium best response of j_1^* and $\frac{\partial}{\partial k}p_1(j_1^*(k), j_2^*(k), k)) = \frac{j_1^*(k)}{k}\frac{t}{k^2}$. This term captures a first-order rent extraction effect: more data collected reduces the size of the segments and allows Firm 1 to extract more surplus from identified consumers. The magnitude of this effect is proportional to the number of consumers identified $\frac{j_1(k)}{k}$. Hence, when $\frac{\partial j_2(j_1)}{\partial j_1} \neq 0$, rent extraction is the only effect that drives data collection by the intermediary, and the number of segments collected is fully determined by the equilibrium value of j_1. This is a new and general result in the literature.

When partitions j_1 and j_2 are independent, the amount of data collected impacts the willingness to pay of Firm 1 through two dimensions: the rent extraction effect identified above and a change of profits in the outside option characterized by $-\frac{\partial \bar{\pi}_1(j_2^*(k))}{\partial k}$. We have shown in Proposition 1 that the same number of segments $\hat{j}_1(k) = \frac{6k-9}{14}$ is sold to Firm 1 for all independent mechanisms, and therefore, the strength of the rent-extraction effect is also the same.

However, different selling mechanisms may have different outside options, and the incentives of the intermediary to collect consumer data vary according to this second effect. We characterize in Lemma 2 how the amount of data collected k impacts the profits of Firm 1 in its outside option $\bar{\pi}_1$ through the equilibrium number of segments sold to Firm 2, j_2^*.

Lemma 2.

For all mechanisms in $\mathfrak{I}$, the amount of data collected by the intermediary has the following impact on the outside option of Firm 1:

$$\frac{\partial}{\partial k}\left(\frac{j_2^*(k)}{k}\right) \geq 0 \implies \frac{\partial \bar{\pi}_1(j_2^*, k)}{\partial k} \leq 0.$$

Proof: Straightforward from Lemma 1 (b).

Consider mechanisms that satisfy Lemma 2. The profits of Firm 1 decrease when it remains uninformed since it faces Firm 2 with information on more consumers. For such mechanisms, there are two positive effects of collecting more segments: the rent extraction effect described in Proposition 3, and the effect on the outside option described in Lemma 2. These two effects go in the same direction, and Proposition 4 characterizes the conditions for the price of information to increase with k.

Proposition 4.

The price of information always increases with k:

- *For all mechanisms that do not belong to $\mathfrak{I}$.*
- *For all mechanisms in $\mathfrak{I}$ such that $\frac{\partial}{\partial k}\left(\frac{j_2^*(k)}{k}\right) \geq 0$.*

To summarize, when partitions j_1 and j_2 are independent, the incentives of the intermediary to collect data are determined by the effect of data collection on the outside option of Firm 1. When partitions j_1 and j_2 are not independent, data collection is determined by the optimal number of consumers identified by Firm 1. Note that $\frac{\partial}{\partial k}\left(\frac{j_2^*(k)}{k}\right) \geq 0$ is a sufficient but not necessary condition for the price of information to increase with k.[16]

[16]There are mechanisms that do not satisfy Lemma 2, and for which a higher k decreases the value of $\frac{j_2^*(k)}{k}$ which increases the profits of Firm 1 if it remains uninformed. In this case, a higher information precision k can increase or decrease the price of information depending on its impact on $\frac{j_2^*(k)}{k}$.

We discuss in Example 9 the optimal data collection strategy of the intermediary with posted prices, sequential bargaining, first and second-price auctions.

Example 9 (Optimal data collection strategies with posted prices, sequential bargaining and first and second-price auctions)**.**

Posted prices, sequential bargaining and first-price auctions belong to the class of independent offers and lead to the same number of segments sold according to Proposition 1. Hence, the marginal impact of data collection on rent extraction is identical, and the incentives of the intermediary to collect data vary only through changes in the outside option of Firm 1, as characterized by Lemma 2.

The number of segments collected is minimized with posted prices, as the marginal gain of data collection does not impact the outside option of Firm 1. The marginal effect of data collection on the outside option of Firm 1 is higher with sequential bargaining than with first-price auctions where the outside option is already the harshest. With these last two mechanisms, the marginal impact of data collection on the outside option of Firm 1 is strong enough to yield higher incentives to collect data than with second-price auctions, despite the stronger rent-extraction effect in this last mechanism.

We can rank profits and the amounts of data collected with the four mechanisms as follows:[17]

$$\begin{aligned} &\Pi_a > \Pi_{a_2} > \Pi_{seq} > \Pi_{pp}, \\ &k_{seq} > k_a > k_{a_2} > k_{pp}. \end{aligned} \tag{6}$$

Up to now, we have characterized the strategy of the intermediary selling information to Firm 1 only. We analyze in the next section the incentives of the intermediary to sell information to Firm 1 or to both firms.

[17] See Appendix G for a proof.

5 Comparing Profits when One or Two Firms are Informed

Selling information to both firms allows the intermediary to increase the total number of segments sold, which may be profitable despite the resulting increasing intensity of competition between firms and their lower willingness to pay for information. Hence, the sale of information to both firms may dominate profits when information is sold to only one firm.

We first characterize the upper bound on profits when selling information to both firms for any mechanisms. We next show that, for the four mechanisms of our focus, profits when selling information to both firms are equal to this upper bound. We then characterize a general class of mechanisms for which the profits of the intermediary when selling information to only one firm are greater than the upper bound of profits when selling information to both firms. We show that first and second-price auctions belong to this class, while profits when selling information to only Firm 1 with posted prices and sequential bargaining are lower than this upper bound, and therefore, the intermediary sells information to both firms with these two mechanisms.

Upper bound of profits when selling information to both firms. We can characterize a general class of mechanisms for which the intermediary always sells information to only one firm. To define this class, we first characterize the upper bound of the profits of the intermediary when selling information to both firms. We denote by $\bar{\Pi}_{both}$ this highest feasible profit, and by $\pi_1(j_1, j_2, k), \pi_2(j_2, j_1, k)$ the profits of Firm 1 and Firm 2 when both firms are informed respectively with partitions j_1, j_2. The resulting prices charged to Firm 1 and Firm 2 are denoted $\rho_1(j_1, j_2, k), \rho_2(j_2, j_1, k)$ to avoid any confusion with prices p_1, p_2 used in previous analysis where information is sold to Firm 1 only. As before, $\bar{\pi}_1(j_2, k)$ and $\bar{\pi}_2(j_1, k)$ correspond to the profits of Firm 1 and Firm 2 when they are uninformed but face a competitor informed with j_2 and j_1.

The highest feasible profit when selling information to both firms $\bar{\Pi}_{both}$ is obtained by maximizing $\rho_1(j_1, j_2, k) + \rho_2(j_2, j_1, k)$ with respect to j_1 and j_2:[18]

[18]The equilibrium value of $\bar{\Pi}_{both}(k)$ is computed in Appendix K.

$$\begin{aligned}\bar{\Pi}_{both}(k) &= \max_{j_1,j_2}\{\rho_1(j_1,j_2,k)+\rho_2(j_2,j_1,k)\} \\ &= \max_{j_1,j_2}\{\pi_1(j_1,j_2,k)-\bar{\pi}_1(j_2,k)+\pi_2(j_2,j_1,k)-\bar{\pi}_2(j_1,k)\}.\end{aligned} \tag{7}$$

We compute in Example 10 the profits of the intermediary selling information to both firms with posted prices, sequential bargaining, first and second-price auctions.

Example 10 (Profits with posted prices, sequential bargaining, first and second-price auctions when the data intermediary sells information to both firms)**.**

The following equalities hold:[19]

$$\Pi^{seq}_{both} = \Pi^{a}_{both} = \Pi^{a_2}_{both} = \Pi^{pp}_{both} = \bar{\Pi}_{both}. \tag{8}$$

The proof of this claim first establishes that the optimal partitions with the four selling mechanisms are identical – and contain all available segments – and then that the outside option for each firm is the same regardless of the selling mechanism. Hence, profits are identical with the four selling mechanisms.

Class of mechanisms for which information is sold to one firm. We can characterize a class of mechanisms for which the intermediary always sells information to Firm 1 only, by computing the values of j_1 and j_2 such that:

$$p_1(j_1,j_2,k) = \pi_1(j_1,k) - \bar{\pi}_1(j_2,k) \geq \bar{\Pi}_{both}(k). \tag{9}$$

We denote by $\mathcal{A}$ the set of mechanisms that satisfy Eq. 9. For any mechanisms in $\mathcal{A}$, the intermediary sells information in exclusivity to only one firm. There are some simple cases in which it is easy to determine whether the data intermediary sells information to one firm. For instance, for all mechanisms in $\mathcal{M}^c$ we have $j_1 = \hat{j}_1$, and the value of j_2 fully determines whether the intermediary sells information to one or to both firms. The

[19]Proof: see Appendix L.

profit of uninformed Firm 1 decreases when j_2 increases, so that there exists $\bar{j}_2$ above which the intermediary sells information to one firm for all mechanisms in $\mathcal{M}^c$.

We determine in Example 11 whether the intermediary sells information to one or to two firms with posted prices, sequential bargaining, first and second-price auctions.

Example 11 (Selling information to one or to two firms with posted prices, sequential bargaining, first and second-price auctions)**.**

(a) With first and second price auctions, the price of information when selling information to one firm satisfies Eq. 9. These mechanisms belong to $\mathcal{A}$, and the intermediary sells information to Firm 1 only.

(b) With posted prices and sequential bargaining, the intermediary sells information to both firms.

First and second-price auctions belong to $\mathcal{A}$, as the data intermediary can exert a strong threat on the outside option of Firm 1. On the contrary, with posted prices, both firms are uninformed when a firm rejects the offer of the data intermediary, resulting in a lower willingness to pay for information. We prove this claim by comparing the prices of information provided in Appendices G and K, which allows to show that Eq. 9 is not satisfied and these mechanisms do not belong to $\mathcal{A}$. As we have shown in Example 10 that the profits of the intermediary when selling information to both firms are equal to $\bar{\Pi}_{both}$ with these two mechanisms, selling information to only one firm necessarily leads to lower profits.

6 Consumer Surplus

We finally discuss in this section how the strategies of the data intermediary impact consumer surplus. We first characterize how consumer surplus changes when firms can identify more consumers (j_1 and j_2 increase), and when more segments are collected by the intermediary (k increases). We then provide conditions on k under which mechanisms in $\mathcal{M}^c, \mathcal{I}$ or $\mathcal{S}^{c+}$ yield the highest surplus. Finally, we provide conditions on the number of consumers identified to compare surplus between different mechanisms.

Consider first the selling strategies. For a given number of segments collected k, we consider two selling mechanisms for which different amounts of consumer segments are sold to firm: (j_1, j_2) and (j_1', j_2'). Increasing the number of segments sold has two effects on consumer surplus. On the one hand, newly identified consumers can be charged a higher price, which decreases consumer surplus (rent extraction effect). On the other hand, all consumers benefit from the increased competitive pressure (competition effect), which increases consumer surplus. The competitive effect benefits all consumers while the rent extraction effect reduces only the surplus of identified consumers. Hence, we can show that the competition effect always dominates the rent extraction effect, regardless of the size of the segment of newly identified consumers.

Secondly, we consider the data collection strategies of the intermediary. If two selling mechanisms allow the intermediary to identify the same number of consumers $x_1, x_2 \in [0,1]$ so that $x_1 = \frac{j_1^*}{k} = \frac{j_1^{*'}}{k'}$ and $x_2 = \frac{j_2^*}{k} = \frac{j_2^{*'}}{k'}$, consumer surplus decreases with the number of consumer segments collected k. In this case, the competitive pressure does not increase when more data are collected, as the location of the last consumer identified remains the same. However, consumer surplus decreases when the intermediary collects more segments. This discussion is summarized in Proposition 5.

Proposition 5.

Consumer surplus varies with the data collection and selling strategies:

$$(a)\quad \forall \quad k,\ j_2,\ j_1 > j_1' : \quad CS(j_1, j_2, k) > CS(j_1', j_2, k),$$

$$(b)\quad \forall \quad k > k',\ x_1,\ x_2 : \quad CS(x_1, x_2, k) < CS(x_1, x_2, k').$$

Proof: see Appendix H.

We cannot unambiguously rank consumer surplus for any mechanisms as the effects characterized in Proposition 5 (a) and (b) may go in opposite directions. Consider mechanisms in $\mathcal{S}^{c+}$. On the one hand, the share of consumers identified by Firm 1 is greater for mechanisms in $\mathcal{S}^{c+}$ than for mechanisms in $\mathcal{I}$ (Proposition 2). Hence, according to Proposition 5, mechanisms in $\mathcal{S}^{c+}$ will yield a higher consumer surplus than mechanisms in $\mathcal{I}$ for a given number of segments collected k, due to the competitive effect of information.

On the other hand, Proposition 3 has shown that the number of segments collected increases with the share of identified consumers. Hence, the intermediary may collect more or fewer data using mechanisms in $\mathcal{S}^{c+}$ – under which a larger share of consumers are identified – than with mechanisms in $\mathcal{I}$, and these different amounts of data collected can have conflicting impacts on consumer surplus. For mechanisms in $\mathcal{S}^{c+}$ for which more data are collected than with mechanisms in $\mathcal{I}$, the competitive effect of information is stronger, but so is the rent-extraction effect, and we cannot conclude in general.

Finally, the two effects of information on consumer surplus also go in opposite directions for mechanisms in $\mathcal{S}^{c-}$ compared with mechanisms in $\mathcal{S}^{c+}$ and $\mathcal{I}$, for which more consumers are always identified, but also more segments are collected.

Surplus comparison: data collection. There are however cases for which we can unambiguously rank consumer surplus. First, consumer surplus is higher for mechanisms in $\mathcal{S}^{c+}$ for which fewer data are collected than with mechanisms in $\mathcal{I}$. Secondly, for all mechanisms in $\mathcal{M}^c$, the same number of consumers are identified and consumer surplus varies only through changes in data collection strategies. More data collected will yield a lower surplus through a stronger rent extraction effect. Overall, there is a negative relation between consumer surplus and data collection for mechanisms in $\mathcal{M}^c$. Proposition 6 summarizes these two comparison rules.

Proposition 6.

Consider two mechanisms s, s' resulting respectively in k and k' segments collected:

- *If $s \in \mathcal{I}$, $s' \in \mathcal{S}^{c+}$, and $k \geq k'$: surplus is always lower with s than with s'.*
- *If $s, s' \in \mathcal{M}^c$, surplus is lower with s than with s' if and only if $k \geq k'$.*

The proof is a direct application of Proposition 5.

For any pair of mechanisms satisfying the conditions of Proposition 6, we can determine which mechanism yields the highest surplus considering only classes of mechanisms and the data collection strategy of the intermediary. We compare in Example 12 consumer surplus with first and second-price auctions.

Example 12 (Consumer surplus with first and second-price auctions)**.**

We can use the data collection ranking of Example 9 and the result of Proposition 6 to compare consumer surplus with first and second-price auction, belonging respectively to $\mathcal{I}$ *and* $\mathcal{S}^{c+}$*:*

$$CS_{a_2} > CS_a. \tag{10}$$

More data are collected, and fewer consumers are identified with first than with second-price auctions. Hence the competition and rent-extraction effects go in the same direction, leading to the above ranking of surplus.

We now compare surplus with different mechanisms by only considering the number of segments sold by the intermediary, regardless of the number of segments collected.

Surplus comparison: consumer identification. We can establish an additional comparison rule based on the competitive effect of information. We provide in Appendix I a sufficient condition to determine for each pair of mechanisms whether one unambiguously yields a higher surplus than the other.

Proposition 7.

For any mechanism $s_1 = (x_1, x_2, k) \in \mathcal{S}^c$*, there exists a set of mechanisms* $\{s'_1 = (x'_1, x'_2, k')\} \subset \mathcal{S}^c$ *such that surplus is lower with* s'_1 *than with* s_1 *for any value of* k, k'*.*

Proof: see Appendix I.

Proposition 7 is based on the following rationale. If the share of consumers identified with mechanism s_1 is sufficiently greater than the share of consumers identified with s'_1, the resulting competition effect will be strong enough to dominate the rent-extraction effect. For instance, when $x_2 = x'_2 = 0$, for any x_1, there exists $\underline{x_1}$ such that consumer surplus is greater with s_1 than with s'_1 if $x'_1 \leq \underline{x_1}$. We compare in Example 13 consumer surplus with posted prices, sequential bargaining, first and second-price auctions.

Example 13 (Consumer surplus with posted prices, sequential bargaining, first and second-price auctions)**.**

When selling information to both firms with posted prices and sequential bargaining, and for any amounts of data collected, the intermediary sells data on all consumers and surplus is higher with these two mechanisms than with first and second-price auctions.

$$CS_{both} > CS_{a_2} > CS_a. \tag{11}$$

Consider posted prices. The intermediary always benefits from selling more information to firms as an uninformed firm faces a worse outside option when its competitor has more information. The price of information when selling information to both firms has two components: the profits of a firm with information minus the profits when it is uninformed and its competitor has information. We can show that the first component is always dominated by the second so that the intermediary reaches the maximum value for $j_1 = j_2 = 1$. The competitive effect of information is the strongest in this case. On the contrary, with first and second-price auctions the intermediary sells to Firm 1 information on respectively $\hat{j}_1$ and $j_1^* = \frac{4k-3}{6}$ consumers. The resulting competitive effect of information is therefore much weaker than with posted prices and sequential bargaining, as well as consumer surplus.

7 Conclusion

Our analysis has shown that the selling mechanism influences the price of information through its two main components: the profit of a firm with information and its outside option (profit without information, facing an informed competitor). Both components are determined by the data collection and selling strategies of the intermediary.

Our results have important implications for both competition policy and data protection regulations. First, the selling mechanism influences the competitiveness of markets through two main channels. On the one hand, we have argued that the intermediary does not sell all the available information in order to reduce the competitive effect of information. Selling mechanisms that increase (decrease) the number of consumer segments sold in equilibrium will increase (decrease) the intensity of competition in the product market.

On the other hand, the selling mechanism also impacts the intensity of competition by changing the number of firms informed in the market, potentially leading to differentiated access to data. While the data intermediary prefers to sell information to only one firm for a large class of mechanisms, including auctions, this is not the case with sequential bargaining and posted prices. Consumer surplus is higher when both firms are informed than when information is sold to only one firm. Regulators can restore a level playing field by enforcing non-discriminatory clauses or price caps. Such regulatory tools are already used for essential patents in patent pools by requiring a fair, reasonable, and non-discriminatory licensing clause (Lerner and Tirole, 2004; Layne-Farrar et al., 2007; Tirole, 2020). Our results contribute to the ongoing debate on competition policy in a digital era, which is on the horizon of acknowledging the strategic role of information on competition. As Crémer et al. (2019) highlight, data create a high barrier to entry on a market, which encourages the emergence of dominant firms. The strategic role of data has led the FTC and the European Commission to increase their anti-competitive scrutiny of big-tech-company and data-broker practices.[20]

Secondly, the existing literature has only focused on the selling strategies of the intermediary. However, understanding data collection strategies is essential to study the impact of a selling mechanism on information available on markets. As a matter of fact, we have shown that for a class of selling mechanisms that we have referred to as "independent offers," that includes first-price auctions, sequential bargaining and posted prices, the intermediary sells the same number of consumer segments, even though the price of information might differ with different selling mechanism. Endogenizing data collection changes the price of information and thus the number of consumer segments sold in equilibrium.

We find that the amount of consumer data collected is the highest for mechanisms for which information is sold to both firms. Further research could investigate the relationship between privacy and competition. Our results open a new research pathway linking data collection strategies and personal data protection on the one hand, and

[20] Congress, Enforcement Agencies Target Tech; Google, Facebook and Apple could face US antitrust probes as regulators divide up tech territory; If you want to know what a US tech crackdown may look like, check out what Europe did; CNBC, last accessed October 26, 2022.

competition policy on the other.

References

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

Bajari, P. and A. Hortaçsu (2004, June). Economic insights from internet auctions. *Journal of Economic Literature 42*(2), 457–486.

Belleflamme, P., W. M. W. Lam, and W. Vergote (2020). Competitive imperfect price discrimination and market power. *Marketing Science 39*(5), 996–1015.

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

Bergemann, D., A. Bonatti, and T. Gan (2022). The economics of social data. *The RAND Journal of Economics*.

Bergemann, D., A. Bonatti, and A. Smolin (2018). The design and price of information. *American economic review 108*(1), 1–48.

Bonatti, A. and G. Cisternas (2020). Consumer scores and price discrimination. *The Review of Economic Studies 87*(2), 750–791.

Bounie, D., A. Dubus, and P. Waelbroeck (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics 52*(2), 283–313.

Braulin, F. C. and T. Valletti (2016). Selling customer information to competing firms. *Economics Letters 149*, 10–14.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review 110*(10), 3267–97.

Crémer, J., Y.-A. de Montjoye, and H. Schweitzer (2019). Competition policy for the digital era. *Report for the European Commission*.

Einav, L., C. Farronato, J. Levin, and N. Sundaresan (2018). Auctions versus Posted Prices in Online Markets. *Journal of Political Economy 126*(1), 178–215.

Elliott, M., A. Galeotti, A. Koh, and W. Li (2021). Market segmentation through information. *Available at SSRN 3432315*.

Fudenberg, D. and J. M. Villas-Boas (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems 1*, 377–436.

Furman, J., D. Coyle, A. Fletcher, D. McAules, and P. Marsden (2019). Unlocking digital competition: Report of the digital competition expert panel. *Report prepared for the Government of the United Kingdom, March*.

Hammond, R. G. (2010). Comparing revenue from auctions and posted prices. *International Journal of Industrial Organization 28*(1), 1–9.

Hammond, R. G. (2013). A structural model of competing sellers: Auctions and posted prices. *European Economic Review 60*, 52–68.

Iyer, G., D. Soberman, and J. M. Villas-Boas (2005). The targeting of advertising. *Marketing Science 24*(3), 461–476.

Jentzsch, N., G. Sapi, and I. Suleymanova (2013). Targeted pricing and customer data sharing among rivals. *International Journal of Industrial Organization 31*(2), 131–144.

Jindal, P. and P. Newberry (2018). To bargain or not to bargain: The role of fixed costs in price negotiations. *Journal of Marketing Research 55*(6), 832–851.

Laffont, J.-J. (1989). *Economie de l'incertain et de l'information.* MIT press.

Larsen, B. and A. L. Zhang (2018). A mechanism design approach to identification and estimation. Technical report, National Bureau of Economic Research.

Layne-Farrar, A., A. J. Padilla, and R. Schmalensee (2007). Pricing patents for licensing in standard-setting organizations: Making sense of frand commitments. *Antitrust Law Journal 74*(3), 671–706.

Lerner, J. and J. Tirole (2004). Efficient patent pools. *American Economic Review 94*(3), 691–711.

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

Montes, R., W. Sand-Zantman, and T. Valletti (2019). The value of personal information in online markets with endogenous privacy. *Management Science 65*(3), 1342–1362.

Nageeb Ali, S., G. Lewis, S. Vasserman, et al. (2022). Voluntary disclosure and personalized pricing. *forthcoming at The Review of Economic Studies*.

Scott Morton, F., P. Bouvier, A. Ezrachi, B. Jullien, R. Katz, G. Kimmelman, A. Melamed, and J. Morgenstern (2019). 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*.

Stole, L. A. (2007). Price discrimination and competition. *Handbook of industrial organization 3*, 2221–2299.

Thisse, J.-F. and X. Vives (1988). On the strategic choice of spatial price policy. *The American Economic Review*, 122–137.

Tirole, J. (2020). Competition and the industrial challenge for the digital age.

Ulph, D., N. Vulkan, et al. (2007). Electronic commerce, price discrimination and mass customisation. Technical report, Technical Report, Said Business School, Oxford 2007. http://vulkan. worc. ox ....

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

Varian, H. R. (1989). Price discrimination. *Handbook of industrial organization 1*, 597–654.

Wang, R. (1993). Auctions versus posted-price selling. *The American Economic Review*, 838–851.

Wang, R. (1995). Bargaining versus posted-price selling. *European Economic Review 39*(9), 1747–1764.

Zeithammer, R. and L. Pengxuan (2006). When is auctioning preferred to posting a fixed selling price? Technical report, Manuscript, Univ. Chicago.

A Proof of Examples 1 and 2

A.1 Posted Prices

We focus our analysis on pure strategy Nash equilibrium where the data intermediary posts an information partition $\mathcal{P}_1^{pp}$, and a price of information p_1^{pp}. Firm 1 can either purchase and make profits $\pi_1(\mathcal{P}_1^{pp}) - p_1^{pp}$, or remain uninformed and make the standard Hotelling profits without information that we denote by π. The intermediary can tailor partition $\mathcal{P}_1^{pp}$ for Firm 1, so that Firm 2 has no interest in acquiring it at the posted price. Thus, the willingness to pay of Firm 1 for information is $\pi_1(\mathcal{P}_1^{pp}) - \pi$. The price of information is found by equalizing the profits of Firm 1 with and without information, which yields:

$$p_1^{pp}(\mathcal{P}_1^{pp}) = \pi_1(\mathcal{P}_1^{pp}) - \pi. \tag{12}$$

A.2 Sequential Bargaining

A data intermediary that uses a sequential bargaining mechanism proposes information to each firm sequentially, in a potentially infinite bargaining game. In the mechanism that we consider, there is no discount factor and the game stops when one firm acquires information. At each stage, the data intermediary proposes information $\mathcal{P}_\theta^{seq}$ to Firm θ ($\theta = 1, 2$), and no information to Firm $-\theta$.

Firm 1 can acquire information $\mathcal{P}_1^{seq}$ and make profits $\pi_1(\mathcal{P}_1^{seq}) - p_1^{seq}$, or decline the offer, and the data intermediary proposes information $\mathcal{P}_2^{seq}$ to Firm 2. If Firm 2 acquires information, the profits of Firm 1 are $\tilde{\pi}_1(\mathcal{P}_2^{seq})$. If Firm 2 declines the offer, the two previous stages are repeated.

To compute the price of information with the sequential bargaining mechanism, we characterize the equilibrium of this game when a transaction takes place. Suppose without loss of generality that Firm 1 purchases information. The data intermediary will propose a price $p_1^{seq}(\mathcal{P}_1^{seq})$ that will be accepted by Firm 1 in equilibrium (minus $\epsilon > 0$). This price is the difference between the profit of Firm 1 when it accepts and when it declines the offer. If Firm 1 accepts the offer, it makes profits $\pi_1(\mathcal{P}_1^{seq})$. If Firm 1 declines the offer, the data intermediary will propose a partition to Firm 2. This partition and its price will be chosen such that Firm 2 will accept the offer, and thus constitute a credible threat to Firm 1.

We propose a candidate equilibrium function. We consider partitions $\mathcal{P}_1^{seq} = sym(\mathcal{P}_2^{seq})$ where sym is the function that, for any partition, gives the symmetric partition with respect to $\frac{1}{2}$. These partitions are each chosen independently from the other, and maximize respectively the profit of Firm 1 and Firm 2. We show that $p_1^{seq} = \pi_1(\mathcal{P}_1^{seq}) - \bar{\pi}_1(\mathcal{P}_2^{seq})$ is an equilibrium. As only the data intermediary has a non binary choice, uniqueness will result naturally.

We write V_1 the value function of Firm 1 in stage 1 to determine its willingness to pay:

$$\begin{cases} V_1 + \pi_1(\mathcal{P}_1^{seq}) - p_1^{seq} & \text{if Firm 1 accepts the offer,} \\ \bar{\pi}_1(\mathcal{P}_2^{seq}) & \text{if Firm 1 declines the offer and Firm 2 accepts the offer,} \\ V_1 & \text{if Firm 2 declines the offer.} \end{cases}$$

Thus, the overall value of Firm 1 is:

$$V_1 + \pi_1(\mathcal{P}_1^{seq}) - p_1^{seq} - \bar{\pi}_1(\mathcal{P}_2^{seq}) - V_1 = \pi_1(\mathcal{P}_1^{seq}) - p_1^{seq} - \bar{\pi}_1(\mathcal{P}_2^{seq})$$

The price of information in the unique equilibrium with sequential bargaining can therefore be written:

$$p_1^{seq}(\mathcal{P}_1^{seq}) = \pi_1(\mathcal{P}_1^{seq}) - \bar{\pi}_1(\mathcal{P}_2^{seq}). \tag{13}$$

The data intermediary has no interest in deviating from this price, as lowering p_1^{seq} would decrease its profits, and increasing p_1^{seq} would have Firm 1 rejecting the offer. Thus $p_1^{seq} = \pi_1(\mathcal{P}_1^{seq}) - \bar{\pi}_1(\mathcal{P}_2^{seq})$ is the unique equilibrium of this game.

Moreover, the data intermediary has no interest in deviating from partitions $\mathcal{P}_1^{seq} = \mathcal{P}_2^{seq}$. Indeed, consider $\mathcal{P}_1 \neq \mathcal{P}_1^{seq}$. Necessarily, $\pi_1(\mathcal{P}_1) \leq \pi_1(\mathcal{P}_1^{seq})$ as $\mathcal{P}_1^{seq}$ is profit maximizing for Firm 1. This lowers the price of information sold to Firm 1, and thus decreases the profit of the data intermediary. Similarly, consider $\mathcal{P}_2 \neq \mathcal{P}_2^{seq}$. For the same reason, proposing such partition is not optimal for the data intermediary when making an offer to Firm 2. Thus it cannot constitute a credible threat on Firm 1 when deciding to acquire information or not as it is not subgame perfect. Thus the partitions used to derive the price of information under sequential bargaining are $\mathcal{P}_1^{seq}$ and $\mathcal{P}_2^{seq}$, and are symmetric.

A.3 First-Price Auctions

Designing first-price auctions in our model is challenging for the following reason. Firms and the data intermediary know the willingness to pay of all bidders. Therefore, firms have incentives to underbid from their true valuation. Indeed, the firm with the highest willingness to pay knows the bid of the other firm. Thus, a firm can bid just above the willingness to pay of its competitor and win the auction. This reduces the price of information achieved through first-price auctions. We solve this problem by setting a reserve price that corresponds to the true valuation of the highest bidder.[21] We describe such strategy in the following paragraph, which a reader uninterested in technical details can skip.

Let $\pi_1(\mathcal{P}^k), \pi_2(\mathcal{P}^k)$ and $\bar{\pi}_1(\mathcal{P}^k), \bar{\pi}_2(\mathcal{P}^k)$ be the respective profits of Firm 1 and Firm 2 when they acquire the reference partition $\mathcal{P}^k$, and when they are uninformed but face a competitor that has acquired partition $\mathcal{P}^k$. 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 Eq. 14.

Simultaneous auctions. In order to maximize the price of information, the data intermediary designs two simultaneous auctions with a reserve price, and only the partition with the highest bid will be sold. The reserve price will be such that Firm 1 does not underbid. We are looking for a pure strategy Nash equilibrium. In auction 1, $\mathcal{P}_1^a$ is auctioned with a reserve price p_1^a to avoid underbidding. 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 $\mathcal{P}_1^a$. Participation of both firms is ensured 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 $\pi_2(\mathcal{P}^k) - \bar{\pi}_2(\mathcal{P}^k)$ 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 $\mathcal{P}_1^a$. Consider now the optimal strategy of Firm 1. Firm 1 can bid for partition $\mathcal{P}^k$, pay a price $\pi_1(\mathcal{P}^k) - \bar{\pi}_1(\mathcal{P}^k)$, and make profits $\bar{\pi}_1(\mathcal{P}^k)$. On the other hand, Firm 1 can also participate to the auction with $\mathcal{P}_1^a$, win the auction by bidding the reserve price p_a, and make profits $\pi_1(\mathcal{P}_1^a) - p_a$. The data intermediary will set a reserve price $p_a = \pi_1(\mathcal{P}_1^a) - \bar{\pi}_1(\mathcal{P}^k) - \epsilon$, where ϵ is an arbitrary small positive number. Thus, $\pi_1(\mathcal{P}_1^a) - p_a > \bar{\pi}_1(\mathcal{P}^k)$, and since only one partition is sold, it will be $\mathcal{P}_1^a$. In equilibrium, Firm 1 bids p_a for $\mathcal{P}_1^a$, and Firm 2 bids $\pi_2(\mathcal{P}^k) - \bar{\pi}_2(\mathcal{P}^k)$.

The price paid by Firm 1 for information is:

$$p_1^a(\mathcal{P}_1^a) = \pi_1(\mathcal{P}_1^a) - \bar{\pi}_1(\mathcal{P}^k). \tag{14}$$

We have just described how to implement simultaneous auctions allowing to reach

[21] Analyzing auctions is important as underbidding is more and more likely to occur in markets for information where bidders acquire valuable information on other bidders through repeated interactions, big data, and artificial intelligence. For instance, Calvano et al. (2020) show that algorithmic pricing by competing firms leads to collusive outcomes even without information transmission.

the first-best price for the data intermediary.[22] Any selling mechanism that allows the data intermediary to reach the first-best price would result in the same equilibrium, and will share the features of the equilibrium partitions found for first-price auctions.[23]

A.4 Second-Price Auctions

With second-price auctions, the data intermediary auctions partitions $\mathcal{P}_1^{a_2}$ and $\mathcal{P}_2^{a_2}$, and Firm 1 (the highest bidder) pays the price corresponding to the bid of Firm 2 (the lowest bidder) for partition $\mathcal{P}_2^{a_2}$.

The willingness to pay of Firm 1 and Firm 2 for information are respectively $\pi_1(\mathcal{P}_1^{a_2}) - \bar{\pi}_1(\mathcal{P}_2^{a_2})$ and $\pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$, and the profit of the intermediary corresponds to the minimum of these bids. Hence, the objective function of the intermediary is to maximize $\pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$ under the constraint that $\pi_1(\mathcal{P}_1^{a_2}) - \bar{\pi}_1(\mathcal{P}_2^{a_2}) \geq \pi_2(\mathcal{P}_2^{a_2}) - \bar{\pi}_2(\mathcal{P}_1^{a_2})$. It is clear that in equilibrium, the constraint is binding. This situation is achieved when $\mathcal{P}_1^{a_2}$ and $\mathcal{P}_2^{a_2}$ are symmetric with respect to $\frac{1}{2}$, which we denote: $\mathcal{P}_2^{a_2} = sym(\mathcal{P}_1^{a_2})$.[24] As a direct consequence of symmetry, we have that $\pi_1(\mathcal{P}_1^{a_2}) = \pi_1(\mathcal{P}_2^{a_2})$ and $\bar{\pi}_1(\mathcal{P}_2^{a_2}) = \bar{\pi}_2(\mathcal{P}_1^{a_2})$.

Hence, the price paid by Firm 1 for information can be written:

$$p_1^{a_2}(\mathcal{P}_1^{a_2}) = \pi_1(\mathcal{P}_1^{a_2}) - \bar{\pi}_1(sym(\mathcal{P}_1^{a_2})). \tag{15}$$

B Proof of Theorem 1

We prove that the partitions described in Theorem 1 are optimal for Firm 1. The proof is a generalization of the proof used in Bounie et al. (2021) since first-price auctions belong to $\mathcal{M}$ and maximize the profits of Firm 1. The data intermediary can choose any partition in the sigma-field $\mathbb{P}_k$ generated by the elementary segments of size $\frac{1}{k}$, to sell to Firm 1 (without loss of generality). There are three types of segments to consider:

- Segments A, where Firm 1 is in constrained monopoly;
- Segments B, where Firms 1 and 2 compete.
- Segments C, where Firms 1 makes zero profit.

We find the partition that maximizes the profits of Firm 1, we will see that it maximizes the profit of the data intermediary. We drop superscript l when there is no confusion. We proceed in three steps. In step 1 we analyze type A segments. We show that it is optimal to sell a partition where type A segments are of size $\frac{1}{k}$. In step 2, we show that all segments of type A are located closest to Firm 1. In step 3 we analyze

[22] The price is maximized as, on the one hand, the profit of Firm 1 with information is the highest possible. On the other hand, the partition sold to Firm 2 if Firm 1 remains uninformed minimizes the profit of Firm 1.

[23] For instance, sequential bargaining with commitment to sell the reference partition to a competitor would lead to the same result.

[24] Application $sym : \mathbb{P}_k \to \mathbb{P}_k$ is such that $\mathcal{P}_1^{a_2}$ and $sym(\mathcal{P}_1^{a_2})$ are symmetric with respect to $\frac{1}{2}$.

segments of type B and we show that it is always more profitable to sell a union of such segments. Therefore, there is only one segment of type B, located furthest away from Firm 1, and of size $1-\frac{j}{k}$ (with j an integer, $j \leq k$). Finally, we can discard segments of type C because information on consumers on these segments does not increase profits.

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 3 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. We compare profits in both situations and show that the finer segmentation is more profitable for Firm 1. We write $\pi_1^A(\mathcal{P})$ and $\pi_1^{AA}(\mathcal{P}')$ the profits of Firm 1 on I with partitions $\mathcal{P}$ and on I_1 and I_2 with partition $\mathcal{P}'$.

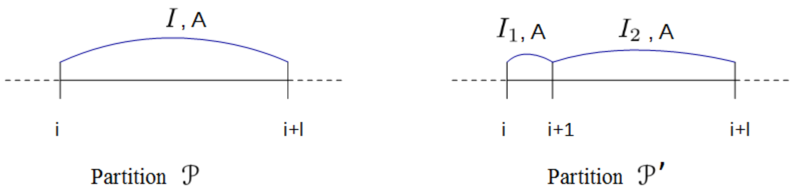

Figure 3: Step 1: segments of type A

To prove this claim, we establish that the profit of Firm 1 is higher with a finer partition $\mathcal{P}'$ with two segments : $I_1 = [\frac{i}{k}, \frac{i+1}{k}]$ and $I_2 = [\frac{i+1}{k}, \frac{i+l}{k}]$ than with a coarser partition $\mathcal{P}$ with I.

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 gets all consumers by assumption; 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:

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

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, look for the first time where 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. Consider a reordering of the overall interval $J \cup I_1 = [\frac{i}{k}, \frac{i+l+1}{k}]$ in two intervals $I_1' = [\frac{i}{k}; \frac{i+1}{k}]$ and $J' = [\frac{i+1}{k}, \frac{i+l+1}{k}]$. We show in this step that such a transformation increases the profits of Firm 1.

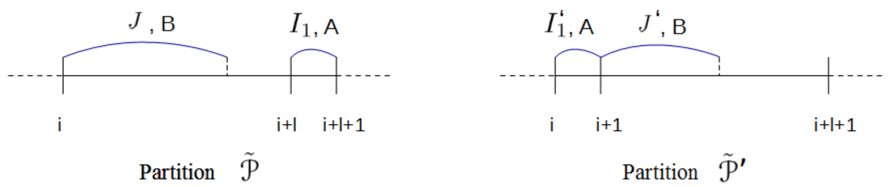

Figure 4: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 4 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, I_1 under partition $\tilde{\mathcal{P}}$ and with I_1', 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 partition, we first characterize type B segments. Segment J of type B is non null (has a size greater than $\frac{1}{k}$), if the following restrictions imposed by the structure of the model, are met: respectively positive demand and the existence of competition on segments of type B. In order to characterize type A and type B segments, it is useful to consider the following inequality:

$$\forall \;\; i, l \in \mathbb{N} \;\; s.t. \;\; 0 \leq i \leq k-1 \;\; \text{and} \;\; 1 \leq l \leq k-i-1,$$
$$\frac{i}{k} \leq \frac{\tilde{p}_2 + t}{2t} \quad \text{and} \quad \frac{\tilde{p}_2 + t}{2t} - \frac{l}{k} \leq \frac{i+l}{k}. \tag{16}$$

In particular, we use the relation that Eq. 16 draws 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 $\tilde{\mathcal{P}}$.

Without loss of generality, we rewrite the notation of type A and B segments. 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 $\frac{l_i}{k}$.[25] There are $h \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$, where prices are noted $\tilde{p}^A_{1i}$. On each of these segments, the demand is $\frac{1}{k}$. There are $n \in \mathbb{N}$ segments of type B, where prices are noted $\tilde{p}^B_{1i}$. We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

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

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

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

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

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

Profits of Firm 2 can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^{n} \tilde{p}_2[\frac{\tilde{p}^B_{1i} - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k}]. \tag{17}$$

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

Equilibrium prices are:

$$\begin{aligned}
\tilde{p}^A_{1i} &= t + \tilde{p}_2 - 2\frac{u_i t}{k} \\
\tilde{p}^B_{1i} &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n}[\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}]] - \frac{s_i t}{k} \\
\tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n}\sum_{i=1}^{n}[\frac{s_i}{2k} + \frac{l_i}{k}].
\end{aligned} \tag{18}$$

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. 16: $\frac{\tilde{p}_2+t}{2t} - \frac{l_i}{k} \leq \frac{s_i+l_i}{k}$ for some segments located at s_i of size l_i. By abuse of notation, let s_i denote the segment located at $[\frac{s_i}{k}, \frac{s_i+l_i}{k}]$, which corresponds to segments of type

[25] With u_i and s_i integers below k.

B that satisfy these condition. Let $\tilde{s}_i$ denote the m segments ($m \in [0, n-1]$) of type B with partition $\tilde{\mathcal{P}}$ located at $[\frac{\tilde{s}_i}{k}, \frac{\tilde{s}_i+\tilde{l}_i}{k}]$ that do not meet these conditions, and therefore are type A segments with partition $\tilde{\mathcal{P}}'$.

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

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

for segments of type B where inequalities in Eq. 16 hold:

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

for segments of type B where inequalities in Eq. 16 do not hold:

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

Let us compare the profits between $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. To compare profits that result by reordering J, I_1 into I_1', J', that is, by moving the segment located at $\frac{i+l}{k}$ to $\frac{i}{k}$ (A to B), 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 $\tilde{\mathcal{P}}$, and that $\tilde{p}_2$ increases as well; and secondly we show that the profits of Firm 1 on type B segments are higher with $\tilde{\mathcal{P}}'$ than with $\tilde{\mathcal{P}}$.

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

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

By definition, $\tilde{s}_i$ verifies the inequalities in Eq. 16, thus $\frac{\tilde{s}_i}{k} \leq \frac{\tilde{p}_2+t}{2t}$, which allows us to establish that $\frac{4t}{3(n-m)}[\frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^{m}\frac{\tilde{s}_i}{2k}] \geq \frac{2t}{3nk}$. It is then immediate to show that:

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

Also, by assumption, firms compete on $J = [\frac{i}{k}, \frac{i+l}{k}]$ with $\tilde{\mathcal{P}}$, which implies that inequalities in Eq. 16 hold, and in particular, $\frac{\tilde{p}_2+t}{4t} - \frac{i}{2k} \leq \frac{l}{k}$.

Thus:

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

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

Second we consider the profits of Firm 1 on the rest of the unit line. We write the reaction functions for the profits on each type of segments, knowing that $\tilde{p}_2' \geq \tilde{p}_2$.

For segments of type A:

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

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

For segments of type B:

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

which is greater than 0 as $\frac{\tilde{p}_2+t}{2} - \frac{s_i t}{k}$ is the expression of the demand on this segment, which is positive under Eq. 16.

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 produces a higher profits. 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 $\hat{\mathcal{P}}'$.

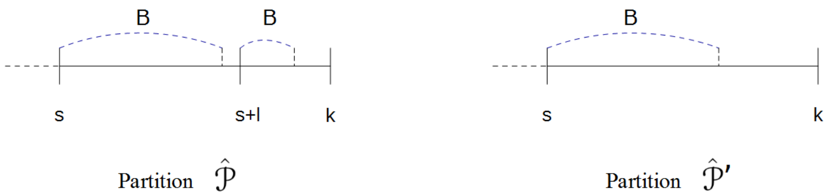

Figure 5: Step 3: demands of Firm 1 on segments of type B (dashed line)

Figure 5 depicts partition $\hat{\mathcal{P}}$ on the left panel, and partition $\hat{\mathcal{P}}'$ on the right panel (on each segment the dashed line represents the demand for Firm 1). 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 will 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 m segments which were type B in partition $\hat{\mathcal{P}}$ are no longer necessarily of type B in partition $\hat{\mathcal{P}}$ (and are therefore of type A).

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

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

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

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

Inequalities in Eq. 16 might not hold as price $\hat{p}_2$ varies depending on the partition acquired by Firm 1. This implies that segments which are of type B with partition $\hat{\mathcal{P}}$ are then of type A with partition $\hat{\mathcal{P}}'$. This is due to the fact that the coarser the partition, the higher $\hat{p}_2$. We note $\tilde{s}_i$ the m segments where it is the case. We then have:

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

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

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

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

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

We consider the terms separately. First,

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

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

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

On these m segments, inequalities in Eq. 16 hold for price $\hat{p}_2'$ but not for $\hat{p}_2$. Thus 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:

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

By replacing $\tilde{s}_i$ by its upper bound value and then $\tilde{l}_i$ by its lower bound value we obtain:

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

Getting back to the profits difference, we obtain:

$$\begin{aligned} \Delta\pi_1 &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}]\frac{4}{3}[\frac{3}{4t}\hat{p}_2 + \frac{m\hat{p}_2}{2t} + \frac{1}{4} - \frac{s+l}{2k}] - \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}]^2 \\ &\geq \frac{t}{2}[\frac{\hat{p}_2+t}{2t} - \frac{s+l}{k}][\frac{\hat{p}_2}{2t} + \frac{s+l}{3k} - \frac{1}{6}]. \end{aligned} \tag{19}$$

The first bracket of Eq. 19 is positive given Eq. 16. 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, $\hat{p}_2 \geq \frac{t}{2}$.[26] And as this price is greater than $\frac{1}{6}$, the second bracket of Eq. 19 is positive. This proves that $\Delta\pi_1 \geq 0$.

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

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

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

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

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.

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

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 after this change decreases, 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.

The element of the sum for segment $[\frac{s}{k}, \frac{s+l}{k}]$ is $\frac{s}{2k} + \frac{l}{k}$. For segments $[\frac{s}{k}, \frac{s+1}{k}]$ and $[\frac{s+1}{k}, \frac{s+l}{k}]$ the term inside the sum is equal to $\frac{1}{2}[\frac{s}{2k} + \frac{s+1}{2k} + \frac{l-1}{k} + \frac{1}{k}]$.

Thus the first term is larger than the second as

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

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

This result allows us to establish that it is always more profitable for the data intermediary to sell a partition with one segment of type B than to sell a partition with several segments of type B.

Conclusion

These three steps prove that the optimal partition includes two intervals, as illustrated in Figure 2. 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 $[\frac{j}{k}, 1]$. □

C Example of selling mechanisms in $\mathcal{M}^c$ and that do not belong to $\mathcal{I}$ (Section 3.4)

There exists selling mechanisms in $\mathcal{M}^c$ where partitions are not independent, and that lead to the same optimal value of $j_1^*(k)$. Consider a selling mechanism in which $j_1^*(k) = \frac{6k-9}{14}$. We will prove that it does not necessarily imply that partitions are independent. The price of information can be written:

$$p(j_1, j_2) = \pi_1(j_1) - \bar{\pi}_1(j_2).$$

Consider j_1 and j_2 such that there exists a function f: $j_2 = f(j_1)$. (for the sake of simplicity we restrict our discussion to continuous and differentiable functions).

We can write the price of information:

$$p(j_1) = \pi_1(j_1) - \bar{\pi}_1(f(j_1)).$$

Thus, solving for the optimal value of j_1 we have:

$$\frac{\partial p(j_1)}{\partial j_1} = \frac{\partial \pi_1(j_1)}{\partial j_1} - \frac{\partial \bar{\pi}_1(f(j_1))}{\partial f(j_1)} \frac{\partial f(j_1)}{\partial j_1} = 0.$$

As this selling mechanism verifies $j_1^*(k) = \frac{6k-9}{14}$, we have:

$$\left.\frac{\partial \pi_1(j_1)}{\partial j_1}\right|_{j_1=\frac{6k-9}{14}} = \left.\frac{\partial \bar{\pi}_1(f(j_1))}{\partial f(j_1)}\right|_{j_1=\frac{6k-9}{14}} \left.\frac{\partial f(j_1)}{\partial j_1}\right|_{j_1=\frac{6k-9}{14}} = 0.$$

Thus, either

$$\left.\frac{\partial \bar{\pi}_1(f(j_1))}{\partial f(j_1)}\right|_{j_1=\frac{6k-9}{14}} = 0$$

or

$$\left.\frac{\partial f(j_1)}{\partial j_1}\right|_{j_1=\frac{6k-9}{14}} = 0.$$

Necessarily, $\left.\frac{\partial \bar{\pi}_1(f(j_1))}{\partial f(j_1)}\right|_{j_1=\frac{6k-9}{14}} < 0$ since the profits of an uninformed firm always decrease with the amount of information purchased by the competitor.

Thus $\left.\frac{\partial f(j_1)}{\partial j_1}\right|_{j_1=\frac{6k-9}{14}} = 0$.

For instance, the data intermediary can commit to selling $j_2(j_1) = f(j_1) = -\frac{j_1^2}{2} + j_1\frac{6k-9}{14}$, and the number of segments sold in equilibrium is $j_1^*(k) = \frac{6k-9}{14}$. □

D Summary of Classes of Mechanisms

Name	Notation: discrete	Notation: continuous	Characterization
All mechanisms	$\mathcal{S} = \{\mathcal{S}_k\}$		$\mathcal{S}_k = \{s_i(k)\}_{i=1}^{2^{2k-2}}$
Restricted set	$\mathcal{S}'$	$\mathcal{S}^c = \{\mathcal{S}_k^c\}$	Satisfy Assumption 1
Maximize Firm 1's profits	$\mathcal{M} = \{\mathcal{M}_k\}$	$\mathcal{M}^c$	$\mathcal{M}_k = \{s_i(k) = \{\mathcal{P}_1^{(i)}, \mathcal{P}_2^{(i)}\},$ $s.t.\ \mathcal{P}_1^{(i)} = argmax\{\pi_1(\mathcal{P}_1)\}\}$
Independent offers		$\mathcal{I} = \{\mathcal{I}_k\}$	$\frac{\partial j_2(j_1)}{\partial j_1} = 0,\ \forall\ j_1$
Dependent offers: Increasing functions		$\mathcal{S}^{c+} = \{\mathcal{S}_k^{c+}\}$	$\frac{\partial j_2(j_1)}{\partial j_1} > 0,\ \forall\ j_1$
Dependent offers: Decreasing functions		$\mathcal{S}^{c-} = \{\mathcal{S}_k^{c-}\}$	$\frac{\partial j_2(j_1)}{\partial j_1} < 0,\ \forall\ j_1$

Table 1: Classes of Selling mechanisms: notations and characterization.

E Proof of Lemma 1

We compute prices and profits in equilibrium when Firm 1 owns the optimal partition on $[0, \frac{j}{k}]$, that includes j segments of size $\frac{1}{k}$, and no information on consumers on $[\frac{j}{k}, 1]$. We define prices and demand functions in step 1. In step 2, we give the expressions for the profits of the firms. Finally we find equilibrium prices and profits in step 3.

Step 1: prices and demands.

Segments of identified consumers are of size $\frac{1}{k}$, and the last one is located at $\frac{j-1}{k}$. Firm 1 sets a price p_{1i} for each segment $i = 1, .., j$ and where it is in constrained monopoly: $d_{1i} = \frac{1}{k}$. Prices on each segment are determined by the indifferent consumer of each segment located at its right extremity, $\frac{i}{k}$:[27]

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

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

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

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

Step 2: profits.

The profits of both firms can be written as follows:

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

Step 3: prices, demands and profits in equilibrium.

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

Profits are:[28]

$$\begin{aligned} \pi_1^* &= \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j}{k})^2 \\ &= \frac{t}{2} + \frac{2jt}{3k} - \frac{7t}{9}\frac{j^2}{k^2} - \frac{tj}{k^2} \\ \pi_2^* &= \frac{t}{2} + \frac{2t}{9}\frac{j^2}{k^2} - \frac{2}{3}\frac{jt}{k}. \end{aligned} \tag{20}$$

Thus, first-order conditions on π_1 gives us

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

F Proof of Example 8

We characterize the equilibrium under second-price auctions.

[27] Assume it is not the case. Then, either p_{1i} is higher and the indifferent consumer is at the left of $\frac{i}{k}$, which is in contradiction with the fact that we deal with type A segments, or p_{1i} is lower and as the demand remain constant, the profits are not maximized.

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

The willingness to pay of firms when the data intermediary auctions information $j_1^{a_2}$ to Firm 1 and $j_2^{a_2}$ to Firm 2 are:

$$\begin{cases} \pi_1(j_1^{a_2}) - \bar{\pi}_1(j_2^{a_2}), \\ \pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2}) \end{cases}$$

We show that in equilibrium $j_1^{a_2} = j_2^{a_2}$.
Assume $\pi_1(j_1^{a_2}) - \bar{\pi}_1(j_2^{a_2}) > \pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2})$ (the other case is solved similarly).

- If $j_1^{a_2} > j_2^{a_2}$: $\pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2})$ increases when $j_2^{a_2}$ increases.
- If $j_1^{a_2} < j_2^{a_2}$: $\pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2})$ increases when $j_1^{a_2}$ increases

Thus the data intermediary chooses $j_1^{a_2} = j_2^{a_2}$.
This implies that

$$p_1^{a_2} = -\frac{t}{3k}\left((3\frac{j_1^{a_2 2}}{k^2} - 4\frac{j_1^{a_2}}{k})k + 3\frac{j_1^{a_2}}{k}\right)$$

Maximizing $p_1^{a_2}$ with respect to $j_1^{a_2}$ and using the FOC give:

$$j_1^{alt*} = \frac{4k-3}{6},$$

$$p_1^{a_2*} = \frac{4t}{9} - \frac{2t}{3k} + \frac{t}{4k^2}$$

and

$$\frac{\partial p_1^{a_2*}}{\partial k} = \frac{(4k-3)t}{6k^3}.$$

□

G Proof of Example 9

We compare the profits of the data intermediary in the different selling mechanisms. The profits of the firms depending on the information partition are the following:

- Profits without information are those in the standard Hotelling competition model:

$$\pi = \frac{t}{2}.$$

- Profit of Firm 1 with j segments of information is:

$$\pi_1^* = \frac{t}{2} + \frac{2jt}{3k} - \frac{7t}{9}\frac{j^2}{k^2} - \frac{tj}{k^2}$$

- Plugging the optimal number of consumer segments $j_1^*(k) = \frac{6k-9}{14}$, we obtain:

$$\pi_1(j_1^*) = \frac{(18k^2 - 12k + 9)t}{28k^2}.$$

- Similarly, the profit of uninformed Firm 1 when facing Firm 2 informed with j segments of information is:

$$\pi_1(j_1) = \frac{t}{2} + \frac{2t}{9}\frac{j_1^2}{k^2} - \frac{2}{3}\frac{j_1 t}{k}$$

- When plugging the optimal number of consumer segments $j_1^*(k) = \frac{6k-9}{14}$ we obtain:

$$\pi_1(j_1^*) = \frac{(25k^2 + 30k + 9)t}{98k^2}.$$

- The profit of an uninformed firm facing a competitor informed with k information segments is provided in Liu and Serfes (2004):

$$\bar{\pi}_1(\mathcal{P}^k) = \frac{(k^2 + 2k + 1)t}{8k^2}.$$

- Profits with second price auctions are provided in Appendix F and are equal to

$$p_1^{a_2*} = \frac{4t}{9} - \frac{2t}{3k} + \frac{t}{4k^2}$$

Profits of the data intermediary under the four selling mechanisms are found directly from these values:

$$p_1^{a*} = \pi_1(j_1^*) - \bar{\pi}_1(\mathcal{P}^k) = \frac{(29k^2 - 38k + 11)t}{56k^2}$$

$$p_1^{pp*} = \pi_1(j_1^*) - \pi = \frac{(4k^2 - 12k + 9)t}{28k^2}$$

$$p_1^{seq*} = \pi_1(j_1^*) - \bar{\pi}_1(j_1^*) = \frac{(76k^2 - 144k + 45)t}{196k^2}$$

$$p_1^{a_2*} = \frac{4t}{9} - \frac{2t}{3k} + \frac{t}{4k^2}$$

We compare the first derivative of the profits of the data intermediary in the different mechanisms in order to compare the optimal amount of data collected in equilibrium.

$$\frac{\partial p_1^{a*}}{\partial k} = \frac{(19k - 11)t}{28k^3},$$

$$\frac{\partial p_1^{pp*}}{\partial k} = \frac{(6k - 9)t}{14k^3},$$

$$\frac{\partial p_1^{seq*}}{\partial k} = \frac{(72k-45)t}{98k^3}.$$

$$\frac{\partial p_1^{a_2*}}{\partial k} = \frac{(4k-3)t}{6k^3}.$$

Comparing the derivatives gives us:

$$\frac{\partial p_1^{seq*}}{\partial k} > \frac{\partial p_1^{a*}}{\partial k} > \frac{\partial p_1^{a_2*}}{\partial k} > \frac{\partial p_1^{pp*}}{\partial k}.$$

From the convexity of the cost function, it is straightforward that:

$$k_{seq} > k_a > k_{a_2} > k_{pp}$$

H Proof of Proposition 5

Consumer surplus when Firm 1 has j_1 consumer segments and Firm 2 has j_2 consumer segments is defined as follows:

$$\begin{aligned}
CS(j_1,j_2,k) &= \sum_{i=1}^{j_1}[\int_0^{\frac{1}{k}} V - 2t[1-\frac{1}{3}\frac{j_1}{k}-\frac{2}{3}\frac{j_2}{k}-\frac{i}{k}] - tx\mathrm{d}x] \\
&+ \int_{\frac{j_1}{k}}^{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}} V - t[1-\frac{4}{3}\frac{j_1}{k}-\frac{2}{3}\frac{j_2}{k}] - tx\mathrm{d}x + \int_{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}}^{1-\frac{j_2}{k}} V - t[1-\frac{2}{3}\frac{j_1}{k}-\frac{4}{3}\frac{j_2}{k}] - tx\mathrm{d}x \\
&+ \sum_{i=1}^{j_2}[\int_0^{\frac{1}{k}} V - 2t[1-\frac{1}{3}\frac{j_2}{k}-\frac{2}{3}\frac{j_1}{k}-\frac{i}{k}] - tx\mathrm{d}x] \\
&= \sum_{i=1}^{j_1}\frac{1}{k}(V-2t[1-\frac{1}{3}\frac{j_1}{k}-\frac{2}{3}\frac{j_2}{k}-\frac{i}{k}]) - \frac{j_1 t}{2k^2} \\
&+ \sum_{i=1}^{j_2}\frac{1}{k}(V-2t[1-\frac{1}{3}\frac{j_2}{k}-\frac{2}{3}\frac{j_1}{k}-\frac{i}{k}]) - \frac{j_2 t}{2k^2} \\
&+ V[1-\frac{j_2}{k}-\frac{j_1}{k}] - [\frac{1}{2}-\frac{2j_1}{3k}-\frac{j_2}{3k}]t[1-\frac{4}{3}\frac{j_1}{k}-\frac{2}{3}\frac{j_2}{k}] \\
&- [\frac{1}{2}-\frac{2j_2}{3k}-\frac{j_1}{3k}]t[1-\frac{4}{3}\frac{j_2}{k}-\frac{2}{3}\frac{j_1}{k}] - t[\frac{1}{4}-\frac{1}{9}\frac{j_1 j_2}{k^2}-\frac{7}{18}\frac{j_2^2}{k^2}-\frac{7}{18}\frac{j_1^2}{k^2}] \\
&= \frac{j_1}{k}[V-2t[1-\frac{1}{3}\frac{j_1}{k}-\frac{2}{3}\frac{j_2}{k}] + \frac{j_1(j_1+1)t}{k^2} - \frac{j_1 t}{2k^2} \\
&+ \frac{j_2}{k}[V-2t[1-\frac{1}{3}\frac{j_2}{k}-\frac{2}{3}\frac{j_1}{k}] + \frac{j_2(j_2+1)t}{k^2} - \frac{j_2 t}{2k^2} \\
&+ V[1-\frac{j_2}{k}-\frac{j_1}{k}] + t[-\frac{5}{4}+\frac{1}{3}\frac{j_1}{k}+\frac{1}{3}\frac{j_2}{k}+\frac{5}{6}\frac{j_1^2}{k^2}+\frac{5}{6}\frac{j_2^2}{k^2}-2\frac{j_1 j_2}{k^2}] \\
&= V + t[-\frac{5}{4}+\frac{17}{18}\frac{j_1^2}{k^2}+\frac{17}{18}\frac{j_2^2}{k^2}+\frac{j_1 j_2}{k^2}] + \frac{1}{2}\frac{j_1 t}{k^2} + \frac{1}{2}\frac{j_2 t}{k^2}
\end{aligned} \tag{21}$$

When only Firm 1 is informed, $j_2 = 0$, and the expressions reduces to;

$$CS(j_1, k) = V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{k^2}] + \frac{1}{2}\frac{j_1 t}{k^2}.$$

Clearly this function decreases with k and increases with j_1, which establishes the result.

□

I Proof of Proposition 7

Surplus is always greater with $x_1 > 0$ and/or $x_2 > 0$ than with $x_1 = x_2 = 0$. Hence for any $s = (x_1, x_2) \neq (0, 0)$, by continuity there exists a neighborhood of $(0, 0)$ such that regardless of k, surplus is lower for any mechanism in this neighborhood than with s. This establishes the claim.

Consider now $x_1, x_1' \in [0, 1]$, and $k, \tilde{k} \in \mathbb{R}_+$.

$$\begin{aligned} CS(x_1, k) &= V + t[-\frac{5}{4} + \frac{17}{18}x_1^2] + \frac{1}{2}\frac{x_1 t}{k} \\ CS(x_1', \tilde{k}) &= V + t[-\frac{5}{4} + \frac{17}{18}x_1'^2] + \frac{1}{2}\frac{x_1' t}{\tilde{k}} \end{aligned} \tag{22}$$

$$x_1 > \sqrt{x_1'^2 + \frac{9}{34}x_1'} \implies CS(x_1, k) > CS(x_1', \tilde{k})$$

Moreover, the profit of Firm 1 do not vary for $x_1 = \frac{j_1}{k} > \frac{3}{4}$ according to the constraint of positivity of prices. Hence, we also require that the greatest value of x_1 is lower than $\frac{3}{4}$, which implies that $x_1' < 0,629$. □

J Proof of Example 13

The fact that consumer surplus decreases with k is established in Appendix H, and this also establishes the claim for posted prices, first-price auctions and sequential bargaining.

To show that surplus is the greatest with second-price auctions, we plug in the equilibrium value of $\hat{j}_1(k)$ and of $j_1^{a_2*}$ in the expression of CS for respectively k and $\tilde{k}$ segments collected:

$$\begin{aligned} CS(\hat{j}_1(k), k) &= -\frac{(170k^2 - 144k - 9)t - 56Vk^2}{56k^2} \\ CS(j_1^{a_2*}(k), k) &= V - \frac{5t}{4} - \frac{112\tilde{k}^2 + 48\tilde{k} - 99}{648\tilde{k}^2}t \end{aligned} \tag{23}$$

A direct comparison of surplus in both cases allows to show that surplus is always greater with second-price auctions than with the three other mechanisms for $k, \tilde{k} \geq 2$. □

K Equilibrium value of the upper bound: Eq. 7

We characterize the equilibrium profits, information partitions and surplus when the data intermediary sells information to Firm 1 and to Firm 2. We first derive the interior solution with $j, j' \in [0, \frac{k}{2}]$, which we will compare with the corner solution, where all information is sold to both firms. We compute in step 1 prices and demands, and in step 2 we give the profits. We solve for equilibrium prices and profits in equilibrium in step 3. Finally we show that selling all information is optimal for the data intermediary.

Step 1: prices and demands.

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

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

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

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

Let denote d_1 the demand for Firm 1 (resp. d_2 the demand for Firm 2) where firms compete. d_1 is found in a similar way as when information is sold to one firm, which gives us $d_1 = \frac{p_2 - p_1 + t}{2t} - \frac{j}{k}$ (resp. $d_2 = 1 - \frac{j'}{k} - \frac{p_2 - p_1 + t}{2t}$).

Step 2: profits of the firms.

The profits of the firms are:

$$\begin{aligned} \pi_1 &= \sum_{i=1}^{j} d_{1i} p_{1i} + d_1 p_1 = \sum_{i=1}^{j} \frac{1}{k}(p_2 + t - 2t\frac{i}{k}) + (\frac{p_2 - p_1 + t}{2t} - \frac{j}{k})p_1, \\ \pi_2 &= \sum_{i=1}^{j'} d_{2i} p_{2i} + d_2 p_2 = \sum_{i=1}^{j} \frac{1}{k}(p_1 + t - 2t\frac{i}{k}) + (\frac{p_1 - p_2 + t}{2t} - \frac{j'}{k})p_2. \end{aligned}$$

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:

$$p_1 = t[1 - \frac{2}{3}\frac{j'}{k} - \frac{4}{3}\frac{j}{k}],$$
$$p_2 = t[1 - \frac{2}{3}\frac{j}{k} - \frac{4}{3}\frac{j'}{k}].$$

Replacing these values in the above demands and prices gives:

$$p_{1i} = 2t - \frac{4}{3}\frac{j't}{k} - \frac{2}{3}\frac{jt}{k} - 2\frac{it}{k},$$
$$p_{2i} = 2t - \frac{4}{3}\frac{jt}{k} - \frac{2}{3}\frac{j't}{k} - 2\frac{it}{k}.$$

and

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

Profits are:

$$\begin{aligned}\pi_1^* &= \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k} - \frac{2}{3}\frac{j'}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j}{k} - \frac{1}{3}\frac{j'}{k})t[1 - \frac{2}{3}\frac{j'}{k} - \frac{4}{3}\frac{j}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{j^2t}{k^2} + \frac{2}{9}\frac{j'^2t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{jt}{k} - \frac{2}{3}\frac{j't}{k} - \frac{jt}{k^2}.\end{aligned}$$

$$\begin{aligned}\pi_2^* &= \sum_{i=1}^{j'} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j'}{k} - \frac{2}{3}\frac{j}{k}] + (\frac{1}{2} - \frac{2}{3}\frac{j'}{k} - \frac{1}{3}\frac{j}{k})t[1 - \frac{2}{3}\frac{j}{k} - \frac{4}{3}\frac{j'}{k}] \\ &= \frac{t}{2} - \frac{7}{9}\frac{j'^2t}{k^2} + \frac{2}{9}\frac{j^2t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{j't}{k} - \frac{2}{3}\frac{jt}{k} - \frac{j't}{k^2}.\end{aligned}$$

The data intermediary maximizes the following profit function:

$$\begin{aligned}\Pi_2(j,j') &= (\pi_1^{I,I}(j,j') - \pi_1^{NI,I}(\emptyset,j')) + (\pi_2^{I,I}(j,j') - \pi_2^{NI,I}(\emptyset,j)) \\ &= -\frac{7}{9}\frac{j'^2t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{j't}{k} - \frac{j't}{k^2} - \frac{7}{9}\frac{j^2t}{k^2} - \frac{4}{9}\frac{jj't}{k^2} + \frac{2}{3}\frac{jt}{k} - \frac{jt}{k^2}.\end{aligned}$$

At this stage, straightforward FOCs with respect to j and j' confirm that, in equilibrium, $j = j'$. 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 $j = j' \in [0, \frac{k}{2}]$ is:

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

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

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

We can write the profit of the data intermediary in the corner solution where all information is sold by replacing j, j' by $\frac{k}{2}$ to obtain firms' profits when both firms are informed ($\pi^{I,I}(k,k) = \frac{t}{4} - \frac{t}{2k}$), and by considering the profits of an uninformed firm facing a competitor informed with all data, given in Liu and Serfes (2004) ($\pi^{NI,I}(\emptyset, k) = \frac{t}{8} + \frac{t}{4k} + \frac{t}{8k^2}$).

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

Profits are higher with the corner solution where all information is sold than with the interior solution, and the data intermediary sells all information to both firms. The overall profits of the data intermediary are:

$$\Pi_{both}(k) = \frac{t}{4} - \frac{3t}{2k} - \frac{t}{4k^2}.$$

and the first-degree derivative of the profit function (including the data collection cost) with respect to k is:

$$\frac{3t}{2k^2} + \frac{2t}{4k^3} - c'(k).$$

Finally, consumer surplus in this case is

$$V - \frac{19t}{36} + \frac{t}{2k}.$$

□

L Proof of Example 10

We focus on information partitions where the data intermediary sells to each firm all consumer segments closest to its location, up to a cutoff point after which no consumer segment is sold. Equivalently, we could directly assume that the optimal partition has the same structure than when the data intermediary sells information to only one firm. We show that the four selling mechanisms are equivalent when the data intermediary sells information to both firms.

Under the first-price auction mechanism, the data intermediary simultaneously auctions partitions j_1^{both} customized for Firm 1 in auction 1, and j_2^{both} customized for Firm 2 in auction 2. Firm 1 (Firm 2) can bid in the two auctions but is only interested in partition j_1^{both} (j_2^{both}). Since both firms are guaranteed to obtain their preferred partition, they will underbid in both auctions from their true valuation. To avoid underbidding, the data intermediary respectively sets reserve prices w_1 and w_2 that correspond to the willingness to pay of Firm 1 for j_1^{both} and of Firm 2 for j_2^{both}. Since partition j_2^{both} is optimal for Firm 2, Firm 1 will not bid above w_2 in the auction for j_2^{both} and similarly Firm 2 will not bid above w_1 in the auction for j_2^{both}. Thus, the subgame perfect equilibrium is characterized by the following strategies: Firm 1 bids the reserve price w_1 for

j_1^{both}, and Firm 2 bids the reserve price w_2 for j_2^{both}. We will show in Appendix K that in equilibrium partitions are symmetric: $j_1 = j_2$. The data intermediary will set in the two auctions reserve prices equal to the willingness to pay of each firm $p_{both} = w_1 = w_2$. This analysis also applies to second-price auctions.

Under sequential bargaining, the problem is simplified by the fact that there is no discount factor, and no first mover advantage since the data intermediary sells to both firms. Thus the data intermediary has no incentive to favour one firm instead of the other, and will choose identical partitions. In this situation, the data intermediary sequentially proposes to Firm 1 partition j_1^{both} at price p_{both}, and to Firm 2 partition j_2^{both} at price p_{both}. Thus, in equilibrium, both firms purchase information at price p_{both}.

With posted prices, the data intermediary posts two partitions tailored to each firm, composed of j^{both} segments of information at price p_{both}. A firm, say Firm 1 (the reasoning will be similar for Firm 2), thus either purchases information and makes profits equal to $\pi_1(j^{both})$. Or it remains uninformed, competes with Firm 2 informed with j^{both} segments, and makes profits equal to $\bar{\pi}_1(j^{both})$. In the only subgame perfect equilibrium of this game, it is easy to show that both firms purchase information at price $p_{both} = \pi_1(j^{both}) - \bar{\pi}_1(j^{both})$. Thus the sum of prices charged by the data intermediary when selling information to both firms is $\Pi_{both}(k) = 2p_{both}$. □