



**HAL**  
open science

## Market for Information and Selling Mechanisms

David Bounie, Antoine Dubus, Patrick Waelbroeck

► **To cite this version:**

David Bounie, Antoine Dubus, Patrick Waelbroeck. Market for Information and Selling Mechanisms. 2021. hal-02288708v4

**HAL Id: hal-02288708**

**<https://hal.science/hal-02288708v4>**

Preprint submitted on 21 Sep 2021 (v4), last revised 4 Nov 2022 (v6)

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

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

# Market for Information and Selling Mechanisms\*

David Bounie,<sup>†</sup> Antoine Dubus<sup>‡</sup> and Patrick Waelbroeck<sup>§</sup>

September 20, 2021

## Abstract

A monopolist data intermediary strategically collects and sells consumer information to competing firms in a product market for price discrimination purposes. He charges a price of information and chooses the optimal information partition that maximizes the willingness to pay for information of firms. Different selling mechanisms are compared: take it or leave it offers, sequential bargaining, and auctions. The intermediary always prefers to sell information through auctions, whereas consumer surplus is maximized with sequential bargaining and take it or leave it offers. We discuss regulatory tools to reconcile the interests of the data intermediary and of regulators over which selling mechanism to use.

**Keywords:** Selling mechanisms; market for information; data intermediaries; competition policy.

---

\*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 the 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 on this article were done while as Antoine Dubus was receiving financial support from the FNRS Grant PDR T.01.47.19. This research has also been conducted within the 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.

<sup>†</sup>i3, CNRS, Télécom Paris, Institut Polytechnique de Paris; [david.bounie@telecom-paris.fr](mailto:david.bounie@telecom-paris.fr).

<sup>‡</sup>Department of Management, Technology and Economics, ETH Zurich Leonhardstrasse 21 Switzerland – 8092 Zurich, Switzerland; [antoine1dubus@gmail.com](mailto:antoine1dubus@gmail.com).

<sup>§</sup>i3, CNRS, Télécom Paris, Institut Polytechnique de Paris; [patrick.waelbroeck@telecom-paris.fr](mailto:patrick.waelbroeck@telecom-paris.fr).

# 1 Introduction

## 1.1 Motivations

Numerous companies around the world spend a lot of money buying information on potential customers they are trying to reach. Data intermediaries are the leaders of this new market for information. They collect consumer information, they 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. The mechanisms used to sell consumer information greatly vary according to data intermediaries and platforms. Some platforms may use first-price or second-price auctions,<sup>1</sup> while others may use list prices or take it or leave it offers,<sup>2</sup> or again bargain sequentially with prospective buyers.<sup>3</sup>

In this article, we develop a framework to analyze how the selling mechanism of information used by a strategic data intermediary determines the amount of data collected from consumers and sold to firms, but also competition in product markets and consumer welfare. We consider a monopolist data intermediary that strategically collects and sells consumer information for price discrimination purposes to firms competing in a product market. Data collected divides consumer demand into segments: thinner segments give more precise information, but are more costly to collect. The data intermediary charges a price of information and chooses the optimal information partitions (a collection of segments) that maximizes the willingness to pay of firms for information. To sell consumer information, the data intermediary can use different selling mechanisms, namely take it or leave it offers, sequential bargaining, and auctions. We also discuss second-price auctions and symmetric offers in an extension.

The ability of the intermediary to choose optimal partitions, and thus to charge a high price for information, depends on the selling mechanism. The main reason is that

---

<sup>1</sup>See for instance the data platform [BIG](#) that enables companies to bid on user data for digital advertisement targeting purposes, like commercial emails or website banners.

<sup>2</sup>See the data marketplace [Advaneo](#) for an illustration.

<sup>3</sup>Facebook has also been trading data through repeated interactions leading to sequential bargaining (<https://adage.com/article/digital/facebook-blocks-valuable-ad-data-privacy-update-its-marketing-partner-program/2238451>).

different selling mechanisms have different outside options for prospective buyers: a firm that does not purchase information can face a negative externality if it competes with an informed competitor. Consider a take it or leave it offer: the data intermediary proposes information to a firm, and if the firm declines the offer, all firms on the market remain uninformed. Now consider an auction with a negative externality: if the firm loses the auction, the data intermediary sells information to the winning bidder. Contrary to take it or leave it offers, a losing bidder without information is worse off with auctions since it may face a well-informed competitor. This negative externality<sup>4</sup> relates to the outside option of firms. We then characterize in this article the optimal collecting and selling strategies of a monopolist data intermediary using different selling mechanisms that have different outside options, and analyze how they impact competition and consumer welfare.

Take it or leave it offers, sequential bargaining, and auctions are commonly used in digital markets and have been largely studied in the literature as we will show below. An interesting feature of our analysis is that these selling mechanisms cover a wide range of bargaining power of the data intermediary. With take it or leave it offers, the data intermediary has a relatively low bargaining power: if the negotiation fails, the intermediary does not sell information and makes zero profits. With sequential bargaining, the data intermediary can negotiate with a firm's competitor in case the negotiation fails. Thus, it can exert a threat on a prospective buyer, who may remain uninformed facing an informed competitor if it does not buy information. The bargaining power of the data intermediary is higher than with take it or leave it offers. Finally, the data intermediary can design an auction that penalizes the losing bidder, and thus maximizes the price of information, allowing the intermediary to reach the first-best outcome. The data intermediary has a stronger bargaining power with auctions than with take it or leave it offers and sequential bargaining. We will study how different selling mechanisms with different outside options and levels of bargaining power influence the data strategies of the intermediary.

---

<sup>4</sup>See Jehiel and Moldovanu (2000) for the case of auctions. We extend their result to a large class of selling mechanisms.

## 1.2 Related literature and contributions

The study of selling mechanisms is a central topic in economics that goes back to [Rubinstein \(1982\)](#) and [Binmore et al. \(1986\)](#) among many others. More recently, empirical studies have revisited the question of which mechanism is optimal for a data seller.<sup>5</sup> [Jindal and Newberry \(2018\)](#) study in which case a seller should use bargaining or fixed price to sell a good; [Larsen and Zhang \(2018\)](#) empirically analyze auctions and bargaining; [Coey et al. \(2020\)](#) compare auctions and fixed price. [Milgrom and Tadelis \(2018\)](#) study how machine learning techniques can be used to improve mechanism design. We contribute to this literature by comparing from a theoretical point of view how different selling mechanisms used by a data intermediary impact data collection and selling strategies.<sup>6</sup>

Our article belongs to a growing economic literature analyzing the markets for information. Recent articles investigate how strategic data intermediaries affect competition between firms in digital markets and consumer welfare. Competition between firms is indeed influenced by how much consumer information firms can acquire from data intermediaries. On the one hand, more information allows firms to better target consumers and price discriminate. They can extract more consumer surplus, which increases their profits. On the other hand, more information means that firms will fight more fiercely for consumers that they have identified as belonging to their business segments. This increased level of competition lowers the profits of the firms. Overall, there is an economic trade-off between surplus extraction and increased competition.

[Montes et al. \(2019\)](#) study the strategy of a monopolist data intermediary selling information to one or two firms willing to price-discriminate consumers, à la [Thisse and Vives \(1988\)](#). The data intermediary uses auctions to sell information, and the authors show that it is optimal to sell all available information to only one firm. [Bounie et al. \(2021\)](#) study also the case of a monopolist data intermediary that can strategically choose to withhold information from firms to minimize the competitive effect of infor-

---

<sup>5</sup>[Larsen \(2014\)](#), [Backus et al. \(2018, 2019\)](#), and [Backus et al. \(2020\)](#) study bargaining descriptively using field data.

<sup>6</sup>The selling mechanism is a crucial component of the design of a data marketplace as well as fairness, efficiency, security, privacy and adherence to regulations (see [O'kelley and Pritchard \(2009\)](#), [Babaioff et al. \(2012\)](#), [Banerjee and Ruj \(2018\)](#), [Agarwal et al. \(2019\)](#), and [Abraham et al. \(2020\)](#)).

mation. The data intermediary can strategically sell segments of the consumer demand to firms competing in the product market, and can weaken or strengthen the intensity of competition by determining the quantity of information available on the market. In other words, the data intermediary has the choice to sell information on all available consumer segments, on a subset of consumer segments, or no information at all. By acquiring information from the data intermediary, firms can identify the most profitable consumer segments, on which they set specific prices. Using auctions, they find that it is also optimal for the intermediary to sell information to only one firm, but not to sell all available consumer segments. By doing so, the data intermediary allows firms to extract surplus from the most valuable consumers while minimizing the competitive pressure of information.

We adopt a framework similar to [Bounie et al. \(2021\)](#), which we apply to three selling mechanisms – auctions, sequential bargaining and take it or leave it offers. We show that they belong to a class of mechanisms for which all results can be generalized. Moreover, the selling mechanism changes the ability of the data intermediary to internalize the competitive effect of information, and most importantly, changes the intensity of competition between firms in the product market. We find that it is optimal for the intermediary to sell information to one firm with auctions and two firms with sequential bargaining and take it or leave it offers. Using an auction, 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 sequential bargaining and with take it or leave it offers, as in this last case, the data intermediary cannot threaten a firm if it declines the offer. Therefore, the data intermediary prefers to sell information to two firms using sequential bargaining and take it or leave it offers, while it only sells information to one firm with auctions. Thus, the selling mechanism has an impact on the number of firms that are informed on a market, and therefore on the intensity of competition and consumer surplus.

A second strand of the literature related to our contribution deals with the sale of information to decision makers. Following [Admati and Pfleiderer \(1986, 1990\)](#), [Bergemann et al. \(2018\)](#) analyze a monopolist data intermediary selling data to a single buyer. The data buyer seeks to augment his initial private information by obtaining additional information from the data seller in order to improve the quality of his decisions. The

authors investigate the revenue-maximizing information policy, i.e. how much information the data seller should provide and how she should price access to the data given that the data intermediary has no information on the initial private information of the data buyer. In this static framework, the data seller offers a menu of statistical signals, and the authors show that the optimal menu includes, in general, both the fully informative signal and partially informative, distorted signals. In our article, we compare different selling mechanisms, including static selling mechanisms such as take it or leave it offers but also sequential mechanisms such as sequential bargaining, and we allow the data intermediary to sell information to one or two firms, thus giving it the ability to modify the intensity of competition in the product market. When analyzing the sale of information to a single buyer, a similar setting to [Bergemann et al. \(2018\)](#), we show that the precision of the information sold to a firm by the data intermediary varies across different selling mechanisms, especially between static mechanisms such as take it or leave it offers and dynamic mechanisms such as sequential bargaining.

A second contribution to this literature is that we explicitly consider that data is costly to collect, and that the selling mechanism, which determines the price of information, will influence the data collection strategies of the data intermediary, and consumer surplus. This analysis is new in the literature: studying mechanisms by simply looking at how much information is sold to firms does not take into account how different selling mechanisms change the value of the threat of the outside option through more precise information partitions. We show that collecting more data increases the profits of information buyers through better consumer surplus extraction. This rent extraction effect of information increases the willingness to pay of firms, which also increases the incentives of the intermediary to collect consumer data. We find that the marginal gain from collecting data is the highest when information is sold to both firms, and that data collection and consumer surplus are the highest with sequential bargaining and take it or leave it offers.

Overall, we argue in this article that the selling mechanism is essential to analyze the amount of information collected on the market, the competitiveness of product markets, and consumer surplus. Among the three mechanisms that we focus on – take it or leave it offers, sequential bargaining, and first-price auctions –, we find that the data intermediary always prefers to sell information through auctions, which is the worst-

case scenario for consumers. These mechanisms lead to less consumer information sold, which softens the competitive effect of information and minimizes consumer surplus. Our results suggest that the data intermediary and regulators might have conflicting views over which selling mechanism to use. Indeed, a competition authority concerned with consumer surplus may prefer a situation where all market participants are informed, which is achieved with sequential bargaining and take it or leave it offers, while we have shown that a data intermediary prefers to sell information to only one firm using auctions. We discuss two tools that allow the regulator to reach the market outcomes of take it or leave it offers, therefore increasing consumer surplus. The first one is a limit on the amount of data collected, and the second one is the enforcement of a level playing field that can be achieved through a non-discriminatory pricing clause, and a cap on the price of information.

The remainder of the article is organized as follows. We describe the model in Section 2. We study in Section 3 the selling strategies of the data intermediary and of the firms. We assume in the baseline model that the data intermediary sells information to only one firm. We first define the price of information, then the three selling mechanisms, and finally the optimal information structure. In Section 4, we characterize the number of consumer segments sold and collected at the equilibrium by the data intermediary for any selling mechanism, and we examine how consumer surplus changes. In Section 5, we investigate how the number of consumer segments sold and collected by the data intermediary as well as consumer surplus vary with each selling mechanism, and we also analyze whether it is more profitable for the intermediary to sell information to one firm only or to two firms on the market. We extend the model in Section 6 to the study of other selling mechanisms such as second-price auctions and symmetric offers. We show that our main results hold. We discuss regulatory implications, and how to use a limit on data collection, non-discriminatory clauses, and a price cap as regulatory tools in Section 7. Section 8 concludes.

## 2 Model

We consider a model of competition à la Hotelling on the product market. 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 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 entire line.

## 2.1 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. A consumer located at  $x \in [0, 1]$  receives a utility  $V$  from purchasing the product, but incurs a cost  $t > 0$  of consuming a product that does not perfectly fit his 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:<sup>7</sup>

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

## 2.2 Data intermediary

We consider a data intermediary that collects consumer information that allows firms to distinguish consumer segments on the unit line. The data intermediary then chooses the optimal information partition by selling more or less information to firms.

The data intermediary collects  $k$  consumer segments at a cost  $c(k)$ .<sup>8</sup> The cost of collecting information encompasses various dimensions of the activity of the data intermediary such as installing trackers, or storing and handling data (see [Varian \(2018\)](#) for

---

<sup>7</sup>We assume that the market is covered, so that all consumers buy at least one product from the firms. This assumption is common in the literature. See for instance [Bounie et al. \(2021\)](#) and [Montes et al. \(2018\)](#).

<sup>8</sup>See [Appendix A](#) for a characterization of the cost function.

a detailed discussion on the structure of the costs associated with data collection). The data collection cost  $c(\cdot)$  captures the sum of the costs associated with these activities.

Collecting data is costly for the intermediary but provides more information on consumers. Data collected divides the unit line into  $k$  segments of equal size, that 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, and that we will denote by partition  $k$  for simplicity.



Figure 1: Reference partition  $\mathcal{P}_k$

The data intermediary then sells a combination of consumer segments to firms. In the baseline model, we assume that the data intermediary sells information to only one firm, say Firm 1, and we study in Section 5.3 the case where the data intermediary sells information to both firms. We denote by  $\mathcal{P}_1$  the partition sold to Firm 1. We expose in detail the choice of partitions by the data intermediary as well as the different selling mechanisms in Section 3.

## 2.3 Firms

Firms may purchase information to price discriminate consumers. If Firm 1 purchases information, it can price discriminate identified consumers. If Firm 1 remains uninformed, consumers are unidentified, and Firm 1 only knows that they are uniformly distributed on the unit line.

In order to compute the profits of the firms, we need to determine demands and prices on 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 on 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.<sup>9</sup> 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  $i$ th segment. Firm 1 is informed and maximizes the following profit function with respect to  $p_{11}, \dots, p_{1i}, \dots, p_{1n}$ :

$$\pi_1 = \sum_{i=1}^n d_{1i} p_{1i}$$

Firm 2 is uninformed and maximizes  $\pi_2 = \sum_{i=1}^n d_{2i} p_2$  with respect to  $p_2$ .

## 2.4 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.

---

<sup>9</sup>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 [Lam 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.

We describe in Section 3 the mechanisms used by the intermediary to sell information and the optimal partition chosen by the intermediary.

### 3 Selling information

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, and the incentive of the data intermediary to collect information. We first define the price of information for any selling mechanism. We then present the three mechanisms that we analyze in this article: take it or leave it offers (*tol*), sequential bargaining (*seq*), and auctions (*a*).

#### 3.1 Price of information

We assume that the intermediary can charge a price of information that corresponds to the willingness to pay of Firm 1. Hence, the data intermediary chooses an information partition that maximizes the willingness to pay for information of Firm 1.

We introduce notations that simplify the exposition of the model. Let's denote by  $\mathcal{P}_1$  the partition sold to Firm 1 if it purchases information, and by  $\mathcal{P}_2$  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  $\mathcal{P}_2$ .

Using these notations, when the data intermediary has collected  $k$  segments, the price of information for any selling mechanism can be written:

$$p_1(\mathcal{P}_1, \mathcal{P}_2, k) = \pi_1(\mathcal{P}_1, k) - \bar{\pi}_1(\mathcal{P}_2, k). \quad (1)$$

The data intermediary chooses partitions  $\mathcal{P}_1$  and  $\mathcal{P}_2$  to maximize the price of information  $p_1(\mathcal{P}_1, \mathcal{P}_2, k)$ . The ability of the intermediary to choose partitions  $\mathcal{P}_1$  and  $\mathcal{P}_2$ , and thus to charge a high price for information, will depend on the selling mechanism.

## 3.2 Selling mechanisms

We present the three selling mechanisms that we compare in this analysis. First, with take it or leave it offers, 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 third selling mechanism is an auction with a negative externality: a firm that loses the auction faces an informed competitor, similarly to sequential bargaining.

The three mechanisms cover a wide range of bargaining power of the data intermediary. With take it or leave it offers, the data intermediary has a relatively low bargaining power: if the negotiation fails, the intermediary does not sell information and makes zero profits. With sequential bargaining, the data intermediary can negotiate with a firm's competitor in case the negotiation fails. Thus, it can exert a threat on a prospective buyer, who may remain uninformed facing an informed competitor if it does not buy information. The bargaining power of the data intermediary is higher than with take it or leave it offers. Finally, the data intermediary can design an auction that penalizes the losing bidder, and thus maximizes the price of information, allowing the intermediary to reach the first-best outcome. The data intermediary has a stronger bargaining power with auctions than with take it or leave it offers and sequential bargaining.

### 3.2.1 Take it or leave it offers

Take it or leave it offers characterize over the counter negotiations, which are used by many information intermediaries such as Nielsen.<sup>10</sup> They are also classically used in theoretical models (Binmore et al., 1986), in particular for the sale of information (Bergemann and Bonatti, 2019). Take it or leave it offers yield the lowest profit for the data intermediary compared to the two other mechanisms, as there is no possibility of renegotiation for the intermediary. The insights that we will gain from analyzing take

---

<sup>10</sup>Details are provided on the [Nielsen's website](#).

it or leave it offers can be applied to all mechanisms where renegotiation is not possible, such as Nash bargaining and menu pricing.

The data intermediary proposes information to Firm 1 that accepts or declines the offer. If Firm 1 declines the offer, the data intermediary does not propose information to Firm 2, and Firm 1 and Firm 2 remain uninformed. This selling mechanism rules out the possibility for the data intermediary to renegotiate the deal if no selling agreement is found, contrary to the sequential bargaining mechanism that we describe in Section 3.2.2.

We focus our analysis on pure strategy Nash equilibrium where the data intermediary makes an offer to Firm 1 that consists of an information partition  $\mathcal{P}_1^{tol}$ , and a price of information  $p_{tol}$ . Firm 1 can either accept the offer and make profits  $\pi_1(\mathcal{P}_1^{tol}) - p_{tol}$ , or reject the offer and make profits  $\pi$ . The partitions are therefore  $(\mathcal{P}_1^{tol}, \emptyset)$ . Thus, the willingness to pay of Firm 1 for information is  $\pi_1(\mathcal{P}_1^{tol}) - \pi$ . The price of information is found by equalizing the profits of Firm 1 with and without information, which yields:

$$p_{tol}(\mathcal{P}_1^{tol}) = \pi_1(\mathcal{P}_1^{tol}) - \pi. \quad (2)$$

### 3.2.2 Sequential bargaining

Selling information through sequential bargaining allows the data intermediary to propose information to Firm 2 in case Firm 1 declines the offer. This dynamic interaction thus introduces the ability for the data intermediary to exert a threat on Firm 1. Such a threat is commonly used by data intermediaries that leverage on the willingness to pay of firms by making offers to their competitors.<sup>11</sup>

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  $\theta$  ( $\theta = 1, 2$ ), and no information to Firm  $-\theta$ .

Firm 1 can acquire information  $\mathcal{P}_1^{seq}$  and make profits  $\pi_1(\mathcal{P}_1^{seq})$ , or decline the offer,

---

<sup>11</sup>Facebook blocks valuable ad data in privacy update to its marketing partner program, AdAge, February 21 2020.

and the data intermediary proposes information  $\mathcal{P}_2^{seq}$  to Firm 2. If Firm 2 acquires information, the profits of Firm 1 are  $\bar{\pi}_1(\mathcal{P}_2^{seq})$ . 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 Firm 1 purchases information. The data intermediary will propose a price  $p_{seq}(\mathcal{P}_1^{seq})$  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 the offer, and the profit of Firm 1 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. The price of information in sequential bargaining is:<sup>12</sup>

$$p_{seq}(\mathcal{P}_1^{seq}) = \pi_1(\mathcal{P}_1^{seq}) - \bar{\pi}_1(\mathcal{P}_2^{seq}). \quad (3)$$

### 3.2.3 Auctions

The data intermediary can also sell information through first-price auctions, which are important to consider for two reasons. First, using auctions allows the data intermediary to reach the maximal price of information. Thus, first-price auctions can be considered as a benchmark to characterize the first best scenario where the data intermediary has the highest bargaining power.<sup>13</sup> Secondly, auctions are used more and more frequently by major data intermediaries such as Google,<sup>14</sup> and in data marketplaces (O'kelley and Pritchard, 2009).

Nevertheless, designing 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

<sup>12</sup>We show that this price is the only Nash equilibrium in Appendix C.

<sup>13</sup>Several articles study auction design (Vickrey, 1961; Klemperer, 1999). Auctions are particularly well suited to the sale of information with a negative externality (Jehiel and Moldovanu, 2000; Figueroa and Skreta, 2009).

<sup>14</sup>First-price Auction, Second-price, and the Header-Bidding, Smartyads, February 2018.

of information achieved through auctions. We solve this problem by setting a reserve price that corresponds to the true valuation of the highest bidder.<sup>15</sup> We describe such strategy in the following paragraph that a reader uninterested in technical details can skip. Let  $\pi_1(k), \pi_2(k)$  and  $\bar{\pi}_1(k), \bar{\pi}_2(k)$  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 Equation 4.

*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  $p_a$  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  $\pi_2(k) - \bar{\pi}_2(k)$  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  $\pi_1(k) - \bar{\pi}_1(k)$ , and make profits  $\bar{\pi}_1(k)$ . 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  $p_a = \pi_1(\mathcal{P}_1^a) - \bar{\pi}_1(k) - \epsilon$ , where  $\epsilon$  is an arbitrary small positive number. Thus,  $\pi_1(\mathcal{P}_1^a) - p_a > \bar{\pi}_1(k)$ , 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  $\pi_2(k) - \bar{\pi}_2(k)$ . The partitions are therefore  $(\mathcal{P}_1^a, k)$ .

---

<sup>15</sup>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. Underbidding practices are discussed by [Calvano et al. \(2019\)](#) who show that algorithmic pricing by competing firms lead to collusive outcomes even without information transmission.



The price paid by Firm 1 for information is:

$$p_a(\mathcal{P}_1^a) = \pi_1(\mathcal{P}_1^a) - \bar{\pi}_1(k). \quad (4)$$

We have just described how to implement auctions using simultaneous auctions in order to reach the first best price for the data intermediary.<sup>16</sup> 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 auctions.<sup>17</sup> We characterize in the next section the partition that maximizes the profit of Firm 1.

### 3.3 Information structure in equilibrium

We characterize the optimal information structure for the three selling mechanisms. We first show that the objective function of the data intermediary choosing partition  $\mathcal{P}_1$  is to maximize the profit of Firm 1 with information. We then characterize the optimal partition for Firm 1.

#### 3.3.1 Objective function of the data intermediary

The data intermediary maximizes the willingness to pay of Firm 1 for information. Consider the price of information in Equation 5: the data intermediary chooses  $\mathcal{P}_1$  that maximizes the profit of Firm 1, and  $\mathcal{P}_2$  that minimizes its profits if it remains uninformed. However, depending on the selling mechanism, the data intermediary cannot choose these information partitions. For instance, with take it or leave it offers the data intermediary cannot sell information to Firm 2 in case the offer is declined. In this case,  $\mathcal{P}_2$  is the partition that contains no information.

It will be useful to define the set  $\mathcal{M}$  of selling mechanisms for which the data intermediary sells partition  $\mathcal{P}_1$  that maximizes the profit of Firm 1. Many selling mechanisms

---

<sup>16</sup>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.

<sup>17</sup>For instance, sequential bargaining with commitment to sell the reference partition to a competitor would lead to the same result.

satisfy this property, such as the three selling mechanisms studied in this article. In particular, if  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are chosen independently,  $\mathcal{P}_1$  will be chosen in order to maximize  $\pi_1(\mathcal{P}_1, k)$ , and the same partition will maximize the profit of Firm 1 and its willingness to pay for information. However,  $\mathcal{P}_1$  can also change the value of  $\mathcal{P}_2$ , changing in turn the value of  $\bar{\pi}_1(\mathcal{P}_2, k)$ . In this case, the partition that maximizes the profit of Firm 1 can differ from the one that maximizes its willingness to pay. In Proposition 1, we will analyze how the equilibrium depends on whether  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are chosen independently or not.

### 3.3.2 Optimal information structure

Finding the optimal partition without any restriction is a complex task given the high dimensionality of the optimization problem, since the data intermediary can potentially recombine consumer segments in any arbitrary fashion.<sup>18</sup>

Nevertheless, for selling mechanism belonging to  $\mathcal{M}$ , and for which the data intermediary sells partition  $\mathcal{P}_1$  that maximizes the profits of Firm 1, we can show that 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  $j_1$  an integer 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.<sup>19</sup> 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

*For any selling mechanism 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.*

---

<sup>18</sup>For instance, the data intermediary could sell a partition starting with one segment of size  $\frac{1}{k}$ , and another segment of size  $\frac{2}{k}$ , and so on.

<sup>19</sup>Thus  $\frac{j_1}{k} \in [0, 1]$ .

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. Theorem 1 generalizes the result of Bounie et al. (2021) established for auctions to any selling mechanisms where the data intermediary sells an information partition that maximizes the profits of Firm 1.

The three mechanisms that we analyze in this article belong to  $\mathcal{M}$ : the data intermediary chooses  $\mathcal{P}_1$  that maximizes the profit of Firm 1. We will show in Proposition 1 that the three mechanisms belong to a class of independent offers, among which all mechanisms satisfy the conditions of Theorem 1.

Proving that the partition described in Theorem 1 is optimal for any arbitrary general selling mechanism is beyond the scope of this article. For the tractability of the model, we assume for the remainder of the article that the intermediary can only sell information structures satisfying Assumption 1.

### **Assumption 1**

*The set of feasible information partitions are given by Theorem 1.*

The data intermediary can sell any partition of the unit line among the class of partitions that satisfy Assumption 1. The data intermediary will find the optimal partition by choosing a single value  $j_1$  that corresponds to the number of consumer segments sold to Firm 1. We drop subscript  $k$  when there is no confusion. Partition  $\mathcal{P}_1$  is described in Figure 2.

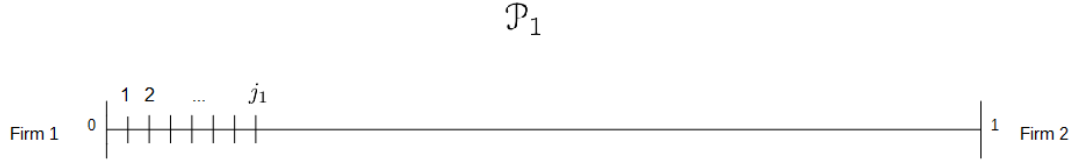


Figure 2: Selling partition  $\mathcal{P}_1$  to Firm 1

A partition that satisfies Assumption 1 is optimal for Firm 1 as it balances the rent extraction effect of information while limiting the competitive effect of information. On the one hand, by identifying consumers close to Firm 1, this partition allows Firm 1 to extract surplus from consumers who have a high willingness to pay. Indeed, selling segments of size larger than  $\frac{1}{k}$  on  $[0, \frac{j_1}{k}]$  is not optimal since Firm 1 could extract more surplus with thinner segments. On the other hand, by keeping unidentified consumers located far away from Firm 1, an optimal information partition softens competition on the market. Any optimal partition must be similar to partition  $\mathcal{P}_1$ , and the optimization problem for the data intermediary boils down to choosing a single value  $j_1(k)$  in partition  $\mathcal{P}_1$ .<sup>20</sup>

### 3.3.3 Price of information in equilibrium

In the remaining of the analysis, we denote by  $\pi_1(j_1)$  the profit of Firm 1 when it has information on the  $j_1$  consumer segments closest to its location (Firm 2 is uninformed). We denote by  $\bar{\pi}_1(j_2)$  the profit of Firm 1 when it is uninformed and faces Firm 2 that has information on the  $j_2$  consumer segments closest to its location.

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

$$p_1(j_1, j_2, k) = \pi_1(j_1, k) - \bar{\pi}_1(j_2, k). \quad (5)$$

Replacing  $\mathcal{P}_1$  by  $j_1$  and  $\mathcal{P}_1$  by  $j_2$  in the three selling mechanisms, we can write the price of information in the following way:

<sup>20</sup>Similarly, we focus our analysis on information structures that are optimal for Firm 2, such that they identify all consumer segments closest to its location up to a cutoff point.

- Take it or leave it offers:

$$p_{tol}(j_1^{tol}) = \pi_1(j_1^{tol}) - \pi.$$

- Sequential bargaining:

$$p_{seq}(j_1^{seq}) = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq}).$$

- Auctions:

$$p_a(j_1^a) = \pi_1(j_1^a) - \bar{\pi}_1(k).$$

Finally, let  $j_2(j_1) : [0; k] \rightarrow [0; k]$  be the number of consumer segments proposed to Firm 2 by the data intermediary for a given  $k$ , as a function of  $j_1$ .<sup>21</sup> We will use this specification in Section 4 to characterize the equilibrium strategies of the intermediary.

In the next section, we characterize the data collection and selling strategies of the intermediary for a large class of selling mechanisms, and we analyze their impact on market competition and consumer surplus. We then focus on the three selling mechanisms in Section 5.

## 4 Characterization of the equilibrium

We solve the game by backward induction, and we characterize the number of consumer segments sold and collected by the data intermediary for any selling mechanisms in Sections 4.1 and 4.2. We then discuss how consumer surplus changes with the number of consumer segments collected and sold in equilibrium in Section 4.3.

### 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

---

<sup>21</sup>In order to find the optimal integer value of  $j_2(\cdot)$ , we consider  $j_2$  as a continuous variable, differentiable with respect to  $j_1$ . This is verified in particular for the three selling mechanisms on which we focus.

in the product market, as well as consumer surplus. As we will show in Section 4.3, consumer surplus increases with the number of segments sold to Firm 1 for a given  $k$ . Different selling mechanisms will thus yield different levels of consumer surplus through this competitive effect of information.

The price of information  $p_1(j_1, j_2, k)$  is defined in Equation 5. The optimal number of consumer segments sold to Firm 1 satisfies the following first-order condition:<sup>22</sup>

$$\frac{\partial p(j_1, j_2, k)}{\partial j_1} = \frac{\partial \pi_1(j_1, k)}{\partial j_1} - \frac{\partial \bar{\pi}_1(j_2, k)}{\partial j_2} \frac{\partial j_2(j_1)}{\partial j_1} = 0 \quad (6)$$

In order to characterize the price of information, we study the signs of the two terms of Equation 6. The sign of the first term depends on the value of  $j_1$  that maximizes the profit of Firm 1 ( $\hat{j}_1$ ). The sign of the second term will be shown to depend only on the sign of  $\frac{\partial j_2(j_1)}{\partial j_1}$ .

**Lemma 1** *The number of consumer segments sold to Firm 1 for a given precision  $k$  has the following properties:*

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

$$(b) \quad \forall j_1 \in [0, \hat{j}_1] : \frac{\partial \pi_1(j_1, k)}{\partial j_1} \geq 0,$$

$$(c) \quad \forall j_1 \in [\hat{j}_1, 1] : \frac{\partial \pi_1(j_1, k)}{\partial j_1} \leq 0,$$

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

Proof: see Appendix D.

Lemma 1 (a) characterizes the partition that maximizes the profit of Firm 1, and Lemma 1 (b) and Lemma 1 (c) guarantee that this maximum is unique. We will see that it can be profitable for the data intermediary to sell a different partition, depending on the

---

<sup>22</sup>We focus on mechanisms for which the second-order condition is locally satisfied:  $\frac{\partial^2 p(j_1, j_2, k)}{\partial j_1^2} \Big|_{j_1^*, j_2^*} < 0$ .

outside option of Firm 1. Lemma 1 (d) shows that the profits of Firm 1 always decrease when more segments are sold to Firm 2. As Firm 2 can better target consumers, the competitive pressure is increased and Firm 1 makes lower profits. We will show how this negative externality impacts the willingness to pay of Firm 1 for information depending on the selling mechanism.

We now state Proposition 1 that characterizes the optimal information structure sold to Firm 1:

**Proposition 1**

*When  $j_2(\cdot)$  is independent from  $j_1$ , the data intermediary sells a partition that maximizes the profit of Firm 1:*

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

Proof: see Appendix E.

Following Proposition 1, we can identify a specific class of information partitions that we refer to as independent offers. The latter have the property that 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. Thus, independent partitions lead to the same number of consumer segments sold to Firm 1 ( $\hat{j}_1$ ). 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 the three selling mechanisms studied in this article. Hence, Proposition 1 generalizes the result of Bounie et al. (2021) to any selling mechanism where  $j_1$  and  $j_2$  are independent.

The fact that the data intermediary chooses the same number of segments with independent information structures is mathematically straightforward, but is far from being trivial from an economic point of view. Indeed, the outside options in take it or leave it offers, sequential bargaining, and auctions reflect different levels of threats. For example, with take it or leave it offers, there is no threat to Firm 1 if it declines the offer. On the contrary, if Firm 1 declines the offer with auctions, the data intermediary sells to Firm 2 the partitions that minimize the profits of Firm 1. Thus, the strength of the threat of the outside option greatly varies between the different selling mechanisms. This

result opens the door to new further research characterizing the properties of different classes of selling mechanisms.

This equivalence does not hold in general as there are many selling mechanisms that do not satisfy independence between information structure, and for which the number of consumer segments sold will be different. For instance, the data intermediary can simultaneously auction symmetric partitions to Firm 1 and Firm 2. In this case, the information partition proposed to Firm 1 appears in its outside option if it does not acquire information:  $p_{alt} = \pi_1(j_1^{alt}) - \bar{\pi}_1(j_1^{alt})$ . Thus in the notations of Proposition 1,  $j_1$  and  $j_2$  are dependent, and the number of segments chosen by the data intermediary does not maximize the profit of Firm 1 anymore. We characterize these mechanisms in Appendix G. Note that there are partitions that are symmetric in equilibrium and that are independent. For instance, with sequential bargaining, the optimal partitions  $j_1^{seq}$  and  $j_2^{seq}$  are chosen independently, and symmetry is not imposed but is a consequence of the analysis.

When partitions are not independent, the data intermediary will sell more or less information to Firm 1 depending on  $j_2(j_1)$ , as characterized in Corollary 1.

**Corollary 1**

*The amount of information sold to Firm 1 when  $j_2$  increases or decreases with  $j_1$  satisfies:*

$$(a) \quad \frac{\partial j_2(j_1)}{\partial j_1} > 0 \implies j_1^* > \frac{6k - 9}{14},$$

$$(b) \quad \frac{\partial j_2(j_1)}{\partial j_1} < 0 \implies j_1^* < \frac{6k - 9}{14}.$$

Proof: see Appendix E.

Corollary 1 characterizes the amount of information sold to Firm 1 when  $j_2$  increases or decreases with  $j_1$ . Corollary 1 and Proposition 1 are critical to understand the impacts of selling mechanisms on consumer surplus. As we will see in Section 4.3, surplus increases with the number of consumer segments sold to Firm 1, and Corollary 1 allows us to compare surplus under different selling mechanisms, considering only the relations between  $j_1$  and  $j_2$ .



Corollary 1 and Proposition 1 have theoretical and practical implications. When offers are independent, the data intermediary maximizes the profits of Firm 1. This is not necessarily the case with other types of selling mechanisms. For instance, selling mechanisms characterized in Corollary 1 (a) lead to a higher number of segments sold to Firm 1, increasing market competition and consumer surplus. Second-price auctions examined in Section 6 belong to this set of mechanisms: the data intermediary maximizes the willingness to pay of the second highest bidder, which is different from maximizing the profits of Firm 1. On the contrary, less information is sold with selling mechanisms characterized by Corollary 1 (b), leading to a lower intensity of competition and a lower consumer surplus. For instance, a total cap  $J$  on the number of consumer segments sold on the market introduces a negative correlation between  $j_1$  and  $j_2$ :  $j_1 + j_2 = J$ .

Secondly, two selling mechanisms that belong to the class of partitions of Proposition 1 will always lead to the same number of consumer segments sold to Firm 1. Thus, a competition authority can analyze the properties of the couple of partitions to determine whether an action is required to limit the amount of information sold on a market.

## 4.2 Consumer data collection in equilibrium

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

The intermediary maximizes its profit by collecting  $k$  consumer segments. In the outside option  $\bar{\pi}_1$ , the profit of Firm 1 without information only depends on  $k$  through  $\frac{j_2^*(k)}{k}$ . Therefore, we can write  $\bar{\pi}_1(j_2, k) = \bar{\pi}_1(j_2(k))$ , which depends on  $j_2^*(k)$ . Thus, the data intermediary maximizes the following profits with respect to  $k$ :

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

The first-order condition for an optimum is:

$$\frac{\partial p(j_1^*(k), j_2^*(k), k) - c(k)}{\partial k} = \frac{\partial \pi_1(j_1^*(k), k)}{\partial k} - \frac{\partial \bar{\pi}_1(j_2^*(k))}{\partial k} - \frac{\partial c(k)}{\partial k} = 0. \quad (7)$$

The amount of data collected impacts the willingness to pay of Firm 1 through two dimensions: a rent extraction effect and a change of profits in the outside option. First,

Lemma 2 characterizes the rent extraction effect by showing that the profits of Firm 1 always increase with data collection. Secondly, Lemma 3 shows that the profits of Firm 1 when it remains uninformed decrease as the intermediary collects more data. Thus the threat of the outside option is stronger when the intermediary collects more data, which in turn gives more incentives to collect information.

**Lemma 2** *The profits of Firm 1 increases with  $k$ :*

$$\frac{\partial \pi_1(j_1^*(k), k)}{\partial k} \geq 0.$$

The proof of Lemma 2 is straightforward and states that profits increase with more precise information. Better information provides Firm 1 with finer segments, which allows it to extract more consumer surplus. This is the rent extraction effect that gives incentives for the intermediary to collect consumer data. Different selling mechanisms may lead to different profits of Firm 1 depending on the number of segments sold, impacting in turn the incentives of the intermediary to collect consumer data.

We characterize in Lemma 3 how information precision 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 3** *The impact of data collection on the outside option*

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

Proof: Straightforward from Lemma 1.

Consider mechanisms that satisfy Lemma 3. The profits of Firm 1 decrease when it remains uninformed since it faces Firm 2 with information on more consumers. To summarize, there are two positive effects for the data intermediary from having more precise information: the rent extraction effect described in Lemma 2; and the effect on the outside option, described in Lemma 3. These two effects go in the same direction, and we can state Proposition 2 that shows that the price of information increases with  $k$ .

**Proposition 2**

When  $\frac{\partial}{\partial k} \left( \frac{j_2^*(k)}{k} \right) \geq 0$ , the price of information always increases with  $k$ .

To prove this result, we show that the first term of Equation 7 always increases with  $k$  given Lemma 2, and the second term also increases with  $k$  using Lemma 3. 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  $k$ .

There are mechanisms that do not satisfy Lemma 3, 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 will see in Section 5 that take it or leave it offers, sequential bargaining and auctions satisfy Lemma 3, and that the price of information always increases with the precision of information.

**4.3 Consumer surplus**

Finally, we provide a welfare analysis of the strategies of the data intermediary. First, for a given precision  $k$ , we consider two selling mechanisms for which different amounts of consumer segments are sold to Firm 1:  $j_1$  and  $j_1'$ . 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 through better rent extraction. On the other hand, all consumers benefit from the increased competitive pressure. Overall, the competition effect always dominates the rent extraction effect, regardless of the size of the segment of newly identified consumers.

Secondly, if two selling mechanisms identify the same number of consumers  $x$  so that  $x = \frac{j_1^*}{k} = \frac{j_1'^*}{k'}$ , consumer surplus decreases with information precision. In this case, there is no competitive effect of having more information on the market, as the location of the last consumer identified remains the same. There is thus only a rent extraction effect, and consumer surplus decreases with more precise information. This discussion is summarized in Proposition 3.

**Proposition 3**

*Consumer surplus varies with data collection and data selling:*

$$(a) \quad \forall k, j_1 > j'_1 : CS(j_1, k) > CS(j'_1, k),$$

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

Proof: see Appendix F.

Proposition 3 offers a convenient way to assess the welfare implications of a selling mechanism. Proposition 3 (a) implies that when more consumers are identified, consumer surplus increases. Following Proposition 3 (b), we show in Section 5.2 that consumer data collection drives consumer surplus for the three selling mechanisms that are the focus of our study.

## 5 Application to take it or leave it offers, sequential bargaining and first-price auctions

For the three selling mechanisms that we focus on, we solve the game by backward induction, and we characterize the number of consumer segments sold and collected by the data intermediary in Sections 5.1 and 5.2. We then analyze in Section 5.3 whether it is more profitable for the intermediary to sell information to one firm only or to two firms on the market.

### 5.1 Number of consumer segments sold in equilibrium

We characterize in Proposition 4 the number of consumer segments sold to Firm 1 in equilibrium with take it or leave it offers, sequential bargaining and auctions.

**Proposition 4**

*The number of consumer segments sold in equilibrium is:*

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

Proof: see Appendix G.

The proof of Proposition 4 is based on the independence of the choice of  $j_1$  and  $j_2$ . In other words, the information proposed to Firm 1 ( $j_1$ ) is independent of the information proposed to Firm 2 ( $j_2$ ) if Firm 1 does not acquire information. With take it or leave it offers, Firm 1 has no information when it declines the offer, and thus its outside option is independent of the information partition proposed by the data intermediary to Firm 2. With 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 affected 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. Regardless of the selling mechanism, when the outside option does not depend on  $j_1$ , the data intermediary simply maximizes the profit of Firm 1 with respect to  $j_1$ . In other words, the three selling mechanisms belong to the class characterized in Proposition 1 (a), for which  $\frac{\partial j_2(j_1)}{\partial j_1} = 0$ , and that leads to  $\hat{j}_1$  consumer segments sold to firms. The integer value of  $j_1$  that maximizes the profits of the data intermediary is chosen by comparing  $\pi(|j_1|)$  and  $\pi(|j_1| + 1)$ :  $\max(\pi(|j_1|), \pi(|j_1| + 1))$ .

## 5.2 Consumer data collection

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. Even though the data intermediary sells the same information partitions to Firm 1 with the different selling mechanisms, we will show that the number of segments collected in the first stage of the game changes with the selling mechanism,<sup>23</sup> as the outside option changes with different selling mechanisms.

The profit of the data intermediary  $\Pi \in \{\Pi_{tol}, \Pi_{seq}, \Pi_a\}$  is given by the price of information  $p \in \{p_{tol}, p_{seq}, p_a\}$ , net of the cost of data collection  $c(k)$ :  $\Pi(k) = p(k) -$

---

<sup>23</sup>We assume that the cost of collecting data does not depend on the selling mechanism.

$c(k)$ .<sup>24</sup>

We have established in Proposition 4 that the number of segments sold by the data intermediary in the second stage of the model is the same for the three selling mechanisms:  $j_1^*(k) = \frac{6k-9}{14}$ . Thus, selling mechanisms will only impact the strategies of the data intermediary through the number of consumer segments collected  $k$ . Indeed, different selling mechanisms will lead to different prices for information, and thus to different amounts of data collected by the data intermediary.

Proposition 5 compares the number of segments collected by the data intermediary and consumer surplus with the three selling mechanisms.

### Proposition 5

*The number of consumer segments collected  $k$  and consumer surplus  $CS$  are inversely related:*

$$(a) \quad k_{seq} > k_a > k_{tol},$$

$$(b) \quad CS_{tol} > CS_a > CS_{seq}.$$

Proof: see Appendix H.

Proposition 5 shows that the number of consumer segments collected is minimized with take it or leave it offers. The optimal level of data collected depends on the marginal gain from increasing information precision. The marginal gain is the lowest in the take it or leave it offers mechanism since no firm is informed in the outside option of Firm 1, and the profits of Firm 1 do not depend on the precision of information if it remains uninformed. Thus, information collection is minimized with this selling mechanism, the rent extraction effect is the lowest, and consumer surplus is maximized. In sequential bargaining and auctions, an increase in the precision of information has two positive effects on the price of information. First, more precise information increases the profit of Firm 1 through a better targeting of consumers, which increases the rent extraction effect. Secondly, the negative externality for an uninformed firm that faces an informed competitor is stronger with more precise information. The data intermediary chooses

---

<sup>24</sup>We make the assumption that  $\Pi$  is concave, and reaches a unique maximum on  $\mathbb{R}^+$ . See Appendix A for a mathematical expression of this assumption.

the value of  $k$  according to these two effects. As the profit functions of an informed firm are equal in sequential bargaining and auctions (Proposition 4), the amount of data collected ( $k$ ) is only driven by the outside option. The marginal gain of more precise information is higher with the sequential bargaining mechanism than with auctions. Indeed, the marginal effect of more precise information on the outside option is higher with sequential bargaining than with auctions where the outside option is already the harshest. Thus, information collection is maximized, and consumer surplus minimized with sequential bargaining.

Proposition 6 shows that the data intermediary prefers auctions, and that take it or leave it offers is the least profitable selling mechanism.

**Proposition 6**

*The profits of the data intermediary are maximized with auctions and minimized with take it or leave it offers:*

$$\Pi_a > \Pi_{seq} > \Pi_{tol}.$$

Proof: see Appendix I.

With auctions, the data intermediary can maximize the value of the threat of the outside option, and maximizes the willingness to pay of Firm 1. On the contrary, with take it or leave it offers, both firms are uninformed when a firm rejects the offer of the data intermediary, resulting in a lower willingness to pay for information. The fact that profit is inversely related to consumer surplus is in itself not surprising, but raises issues about models that take it for granted that auctions are welfare maximizing (Braulín and Valletti, 2016; Montes et al., 2019). This result calls for more research on the impact of selling mechanisms on consumer surplus.

### 5.3 Selling information to one or two firms

We have focused our analysis on cases where the data intermediary sells information to only one firm, and keeps the other firm uninformed. In this section, we allow the data intermediary to sell information to two firms, and we compare profits for the three

selling mechanisms to find the optimal selling strategy. We first establish that profits for the data intermediary are identical with the three selling mechanisms when selling information to two firms. Next, we show that the data intermediary sells information to two firms only with sequential bargaining and take it or leave it offers, and to only one firm with auctions. Finally, we compare the equilibrium outcomes with the auction mechanism, where the intermediary sells information to only one firm, and with sequential bargaining and take it or leave it offers when the intermediary sells information to both firms.

We show in Proposition 7 that profits are identical with the three selling mechanisms when the data intermediary sells information to two firms.

**Proposition 7**

*The three selling mechanisms lead to the same profit for the data intermediary:*

$$\Pi_{both}^{seq} = \Pi_{both}^a = \Pi_{both}^{tol} = \Pi_{both}.$$

Proof: see Appendix J.

The data intermediary maximizes the sum of the prices of information paid by each firm. Each price is defined by the difference between the profit of a firm when both firms are informed, and profits when a firm is uninformed facing an informed competitor. The proof first establishes that the optimal partitions with the three selling mechanisms are identical – and contains 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 three selling mechanisms.

We characterize in Proposition 8 whether the data intermediary sells information to one or two firms with the three selling mechanisms.

**Proposition 8**

*The data intermediary sells information:*

- *To Firm 1 only with auctions.*
- *To both firms with sequential bargaining and take it or leave it offers.*



Proof: see Appendix K.

The intuition behind Proposition 8 is the following. When selling information to only one firm 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. On the contrary, with sequential bargaining, the partition sold to Firm 2 in case Firm 1 declines the offer is chosen in order to maximize the profit of Firm 2, and not to exert a maximal negative externality on Firm 1. With take it or leave it offers, the data intermediary cannot threaten Firm 1 if it declines the offer. Therefore, the data intermediary prefers to sell information to both firms using sequential bargaining and take it or leave it offers, while it only sells information to one firm with auctions. Thus, the selling mechanism has an impact on the number of firms that are informed on a market, and then on the intensity of competition and consumer surplus.

We compare in Proposition 9 the number of consumer segments collected by the intermediary using auctions, and when selling information to both firms in equilibrium, as well as consumer surplus.

**Proposition 9**

*The equilibrium when the data intermediary sells all information to both firms is characterized by:*

$$(a) \quad k_{both} > k_a,$$

$$(b) \quad CS_{both} > CS_a.$$

Proof: see Appendix K.

Proposition 9 (a) shows that the number of consumer segments collected when selling information to both firms is higher than with auctions. The marginal benefits from collecting consumer data are higher when selling information to both firms, as all available segments are sold. Moreover, in this case all consumers are identified, and consumer surplus is higher than in the auction mechanism where the data intermediary internalizes the competitive effect of information, as stated in Proposition 9 (b).

## 6 Second-price auctions and symmetric offers

We consider in this section an alternative mechanism used to sell information to firms: second-price auctions. Second-price auctions allow us to identify another class of selling mechanisms where information partitions proposed to firms are not independent, as characterized in Proposition 1 (b). Studying second-price auctions will allow us to shed light on the ongoing debate in the online advertising industry on the use of first or second-price auctions.<sup>25</sup>

There are three additional reasons that motivate our analysis of second-price auctions. First, characterizing market equilibrium with this mechanism allows to better understand the data intermediary's strategies of consumer data collection and selling. Secondly, second-price auctions prevent underbidding from auction participants. Finally, second-price auctions allow the data intermediary to extract surplus from firms, even when it has no information about their willingness to pay. Indeed, in second-price auctions, firms compete fiercely for the acquisition of information. Thus, second-price auctions are useful when the data intermediary has a low bargaining power.

With second-price auctions,<sup>26</sup> the data intermediary auctions partitions  $j_1^{a2}$  and  $j_2^{a2}$ , and Firm 1 (the highest bidder) pays the price corresponding to the bid of Firm 2 (the lowest bidder) for partition  $j_2^{a2}$ . We first characterize two main properties of second-price auctions when the intermediary sells information to Firm 1 only.

First, when information is sold only to Firm 1 with second-price auctions, partitions auctioned to Firms 1 and 2 are symmetric. Thus, second-price auctions belong to the class of mechanisms characterized in Proposition 1 (b), for which  $\frac{\partial j_2(j_1)}{\partial j_1} = 1 > 0$ , and that lead to more information sold compared with independent offers  $\hat{j}_1$ . Indeed, consider second-price auctions where the winning bidder, Firm 1, has to pay the valuation of the second highest bidder, Firm 2. There are two cases to consider in which the data intermediary auctions partitions with different numbers of segments. First, if Firm 1 is proposed more segments of information than Firm 2,  $j_1^{a2*} > j_2^{a2*}$ , the data intermediary

---

<sup>25</sup>Google's adoption of first-party auction creates migration headaches for buyers, Digiday, March 8 2019.

<sup>26</sup>We focus on information partitions that satisfy Assumption 1. We do not show that such partition is optimal for second price auctions, as the proof of optimality is challenging and beyond the scope of our analysis.

can increase the willingness to pay of Firm 2 by increasing  $j_2^{a_2^*}$ . Secondly, if Firm 1 is proposed less segments of information than Firm 2, the data intermediary can increase the willingness to pay of Firm 2 by increasing  $j_1^{a_2^*}$ , which will worsen its outside option. In both cases, the data intermediary has interest to equalize the number of segments auctioned in both partitions, and the equilibrium is reached when the two partitions are symmetric:  $j_1^{a_2^*} = j_2^{a_2^*}$ .

Secondly, we show in Proposition 10 that the number of segments proposed to Firm 2 in the outside option of Firm 1 is equal to  $j_2^{a_2^*}(k) = \frac{4k-3}{6}$ , and thus the equilibrium satisfies Proposition 2:  $\frac{\partial j_2^{a_2^*}(k)}{\partial k} > 0$ . This means that the price of information increases with the number of segments collected because more consumer segments reduces the profits of Firm 1 if it remains uninformed.

When the data intermediary sells information to both firms through second-price auctions, the mechanism is identical to first-price auctions, and thus is equivalent to sequential bargaining and take it or leave it offers. We compare profits when selling information to Firm 1 with second-price auction  $\Pi_{a_2}$ , consumer surplus  $CS_{a_2}$ , and the amount of data collected  $k_{a_2}$ , with the outcomes of first-price auctions and when selling information to Firm 1 and Firm 2.

**Proposition 10**

*The equilibrium with second-price auctions has the following properties:*

- (a)  $j_1^{a_2^*} = \frac{4k-3}{6}$ ,
- (b)  $\Pi_a > \Pi_{a_2} > \Pi_{both}$ ,
- (c)  $k_{both} > k_a > k_{a_2}$ ,
- (d)  $CS_{both} > CS_{a_2} > CS_a$ .

Proof: see Appendix L.

The data intermediary sells information to Firm 1 only with second-price auctions rather than to both firms. Comparing second-price auctions with first-price auctions, we see that the data intermediary prefers first-price auctions that lead to the highest willingness

to pay of Firm 1. Moreover, the amount of consumer data collected is higher, and consumer surplus lower with first-price auctions than with second-price auctions. Thus, this result contributes to the debate on the design of the optimal auctions for online advertising: second-price auctions reduce the amount of data collected, but first-price auctions maximize the price of information.

## 7 Regulatory implications and policy guidelines

We analyze in this section the implications of our results for the regulation of the market for consumer information. We focus on take it or leave it offers, sequential bargaining and auctions. All results can be generalized to selling mechanisms satisfying the assumption of Theorem 1. Our results suggest that a data intermediary and regulators might have conflicting views over which selling mechanism to use for two reasons. First, Propositions 5 and 6 show that the data intermediary prefers auctions that maximize its profits but lead to a lower consumer surplus than take it or leave it offers. However, a competition authority concerned with consumer surplus prefers take it or leave it offers.

Secondly, a competition authority may prefer a market situation where all market participants are informed, which is achieved with take it or leave it offers, while we have shown that a data intermediary prefers to sell information to only one firm using first-price auctions and sequential bargaining. Access to data is indeed scrutinized by competition authorities who want to guarantee fair and equal access to information for firms. Market practices have revealed that data intermediaries play a significant role in shaping competition, which can cause important harm to other companies and consumer welfare. For instance, Facebook offered companies such as Netflix, Lyft, or Airbnb special access to data, while denying its access to other companies such as Vine.<sup>27</sup>

While enforcing a specific selling mechanism is a particularly hard task to do for a regulator, we consider two regulatory tools that allow to reach the market outcomes of take it or leave it offers, therefore increasing consumer surplus. The first one is a limit on the amount of data collected  $k$ . The second regulatory tool is the enforcement of a level playing field. This can be achieved in two different ways. On the one hand, a

---

<sup>27</sup>Facebook gave Lyft and others special access to user data; engadget, May 12th, 2018.

non-discriminatory pricing clause stipulates that the data intermediary cannot charge an arbitrarily high price to a prospective buyer. As a consequence, the data intermediary will sell symmetric partitions to both firms. On the other hand, equal access to information can also be reached by setting a cap on the price of information.

## 7.1 Limiting the amount of data collected

A data protection agency can set a limit  $\bar{k}$  over the amount of consumer data collected by a data intermediary. The aim of such regulation is to protect consumer privacy by forcing firms to collect as little data as possible (see for instance the European [General Data Protection Regulation](#)). Proposition 11 provides the implications for market equilibrium of a change in the maximal amount of consumer data that the intermediary can collect.

### Proposition 11

- (a) *The ranking of profits in Proposition 10 does not change with  $\bar{k}$ .*
- (b) *Consumer surplus decreases with  $\bar{k}$ .*

Proof: see Appendix M.

The results of Proposition 6 still hold, and the data intermediary prefers to sell information through auctions. Indeed, surplus extraction from Firm 1 depends on the threat for a firm of being uninformed, which is the highest with auctions, and the lowest with the take it or leave it offers. Proposition 11 (b) shows that reducing the amount of consumer information collected by the data intermediary will increase consumer surplus. The rent extraction effect is weaker when the data intermediary collects less information. Any regulation that limits the value of  $k$  will thus benefit consumers. For instance, the European GDPR enforces such a data minimization principle ([General Data Protection Regulation](#)). Thus, data protection regulations are complementary to standard competition policy tools to protect consumer surplus. We study two such policies in the following section.

## 7.2 Enforcing a level playing field

In this section, we show how non-discriminatory clauses and price caps can be used to force a data intermediary to sell information to both firms, thus allowing fair competition between firms.

### 7.2.1 Non-discriminatory clauses

We have seen in Proposition 8 that the intermediary has interest to sell information to Firm 1 only under auctions, which raises concerns for the resulting dominance of Firm 1 over Firm 2 that remains uninformed. Such exclusionary practices have been criticized by the U.S. Congress in its recent report,<sup>28</sup> and by Crémer et al. (2019) and Tirole (2020) among others.

A non-discriminatory clause will force the data intermediary to sell information to both firms. It is easy to show that the data intermediary will sell symmetric information to firms, leading to the equilibrium described in Proposition 9.

### 7.2.2 Price cap

We analyze the impacts of a price cap on the strategies of the data intermediary. Such policies have been recently advocated by Rey and Tirole (2019). By imposing a price cap, a regulator can lower the profits of the data intermediary who will then sell information to both firms. As a consequence, the amount  $k$  of consumer data collected will change.<sup>29</sup>

Regardless of the selling mechanism, the amount of data collected by the data intermediary decreases with the value of the price cap. This property results from the log concavity of the price of information with respect to  $k$ , meaning that the rent extraction effect is always stronger than the competition effect that is internalized by the data intermediary. Moreover, there is a price below which the data intermediary will prefer to sell information to both firms, thus resulting in an equilibrium similar to sequential bargaining and take it or leave it offers.

---

<sup>28</sup>Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

<sup>29</sup>Proofs are provided in Appendix N.

## 8 Conclusion

The dominance of data intermediaries is today the source of intense debates between economists regarding the ability of competition authorities to protect consumer welfare. Our article contributes to this debate by emphasizing how the design of the market for information, and especially the mechanism used by a data intermediary to sell information, can reduce consumer surplus by increasing the amount of data collected, and by limiting competition between firms on the markets. These results have implications for competition policy and for the regulation of data marketplaces.

First, the selling mechanism impacts competition by changing the number of firms informed on the market, potentially leading to differentiated access to data. Indeed, the data intermediary prefers to sell information to only one firm with auctions, but not with sequential bargaining and take it or leave it offers. 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 starting to acknowledge the strategic role of information on competition. As Crémer et al. (2019) emphasize, 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, concerned with potential anti-competitive practices, to increase their scrutiny of the activity of big-tech companies and data brokers.<sup>30</sup>

Secondly, our results show that the price of information influences the amount of data collected. Indeed, selling information to both firms using sequential bargaining or take it or leave it offers mechanism results in a higher level of data collected compared to auctions. The amount of consumer data collected in equilibrium is driven by the price of information, which depends on the amount of data sold and on the profit of

---

<sup>30</sup>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.

the firm that purchases information, and on what happens if the firm declines the offer. The data intermediary can then leverage out on this threat by increasing the precision of information, i.e. by collecting more data, which will increase firms' willingness to pay for information. We find that the amount of consumer data collected is the highest when information is sold to both firms, where the number of consumer segments sold is the highest, and the outside option is the strongest.

Further research could investigate the relationship between privacy and competition. Our results question the traditional view that regulators cannot protect privacy and consumer surplus at the same time (see for instance [Thisse and Vives \(1988\)](#)). A strategic data intermediary internalizes the negative competitive effect of information so that more information on the market does not necessarily increase consumer surplus: when information is sold to one firm, take it or leave it offers minimize data collection and maximize consumer surplus, while sequential bargaining maximizes data collection and minimizes consumer surplus. Understanding the theoretical properties of selling mechanisms is therefore essential to promote a competitive digital economy.

## References

- Abraham, I., S. Athey, M. Babaioff, and M. D. Grubb (2020). Peaches, lemons, and cookies: Designing auction markets with dispersed information. *Games and Economic Behavior* 124, 454–477.
- Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science* 24(3), 367–381.
- Admati, A. R. and P. Pfleiderer (1986). A monopolistic market for information. *Journal of Economic Theory* 39(2), 400–438.
- Admati, A. R. and P. Pfleiderer (1990). Direct and indirect sale of information. *Econometrica: Journal of the Econometric Society*, 901–928.
- Agarwal, A., M. Dahleh, and T. Sarkar (2019). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 701–726.
- Babaioff, M., R. Kleinberg, and R. Paes Leme (2012). Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 92–109. ACM.



- Backus, M., T. Blake, B. Larsen, and S. Tadelis (2020). Sequential bargaining in the field: Evidence from millions of online bargaining interactions. *The Quarterly Journal of Economics* 135(3), 1319–1361.
- Backus, M., T. Blake, and S. Tadelis (2018). Communication and bargaining breakdown: An empirical analysis. *NBER Working Paper*.
- Backus, M., T. Blake, and S. Tadelis (2019). On the empirical content of cheap-talk signaling: An application to bargaining. *Journal of Political Economy* 127(4), 000–000.
- Banerjee, P. and S. Ruj (2018). Blockchain enabled data marketplace—design and challenges. *arXiv preprint arXiv:1811.11462*.
- Bergemann, D. and A. Bonatti (2019). Markets for information: An introduction. *Annual Review of Economics* 11.
- Bergemann, D., A. Bonatti, and A. Smolin (2018). The design and price of information. *American economic review* 108(1), 1–48.
- Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 176–188.
- Bounie, D., A. Dubus, and P. Waelbroeck (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics* 52(2), 283–313.
- Braulín, F. C. and T. Valletti (2016). Selling customer information to competing firms. *Economics Letters* 149, 10–14.
- Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2019). Artificial intelligence, algorithmic pricing and collusion. *Algorithmic Pricing and Collusion*.
- Coey, D., B. J. Larsen, and B. C. Platt (2020). Discounts and deadlines in consumer search.
- Crémer, J., Y.-A. de Montjoye, and H. Schweitzer (2019). Competition policy for the digital era. *Report for the European Commission*.
- Figuroa, N. and V. Skreta (2009). The role of optimal threats in auction design. *Journal of Economic Theory* 144(2), 884–897.
- Fudenberg, D. and J. M. Villas-Boas (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems* 1, 377–436.
- Jehiel, P. and B. Moldovanu (2000). Auctions with downstream interaction among buyers. *Rand journal of economics*, 768–791.
- 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.
- Klemperer, P. (1999). Auction theory: A guide to the literature. *Journal of economic surveys* 13(3), 227–286.
- Lam, W., P. Belleflamme, and W. Vergote (2020). Competitive imperfect price discrimination and market power. *Marketing Science*.
- Larsen, B. and A. L. Zhang (2018). A mechanism design approach to identification and estimation. Technical report, National Bureau of Economic Research.
- Larsen, B. J. (2014). The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions. *The Review of Economic Studies*.
- 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.
- Milgrom, P. R. and S. Tadelis (2018). How artificial intelligence and machine learning can impact market design. Technical report, National Bureau of Economic Research.
- Montes, R., W. Sand-Zantman, and T. Valletti (2018). The value of personal information in online markets with endogenous privacy. *Management Science*.
- 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.
- O’kelly, C. B. and A. R. Pritchard (2009, January 8). Data marketplace and broker fees. US Patent App. 11/772,965.
- Rey, P. and J. Tirole (2019). Price caps as welfare-enhancing competition. *Journal of Political Economy*.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 97–109.
- 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.
- Varian, H. (2018). Artificial intelligence, economics, and industrial organization. Technical report, National Bureau of Economic Research.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16(1), 8–37.

## A Mathematical assumptions

We denote by  $p \in \{p_{tol}, p_{seq}, p_a\}$  the price of information, defined in Section 3 for the three selling mechanisms. The cost function is defined such that:

$$\begin{cases} \frac{\partial^2[p(k)-c(k)]}{\partial k^2} < 0 \text{ and } \exists! k^* \text{ s.t. } \frac{\partial[p(k)-c(k)]}{\partial k} = 0 \\ \exists! k^* \text{ s.t. } \frac{\partial \Pi}{\partial k} = 0 \text{ and } \Pi(k^*) \geq 0 \\ c(0) = 0; c'(0) > 0. \end{cases}$$

These technical hypothesis are common in the literature. It allows profits to be maximized in a unique point, which is usually true for linear and convex cost functions. The cost of collecting information encompasses various dimensions of the activity of the data intermediary such as installing trackers, or storing and handling data. For instance [Varian \(2018\)](#) describes the various costs associated with collecting and handling data.

Finally, the marginal production costs of firms on the product market are normalized to zero.

## B Proof: partitions in Assumption 1 are optimal for Firm 1

The data intermediary can choose any partition in the sigma-field  $\mathbb{P}$  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 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 1.

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.

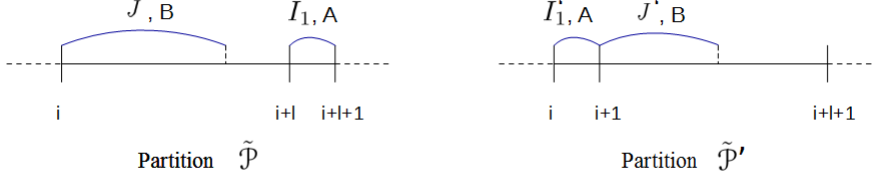


Figure 4: Step 2: relative position of type A and type B segments

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:

$$\forall i, l \in \mathbb{N} \text{ s.t. } 0 \leq i \leq k-1 \text{ and } 1 \leq l \leq k-i-1, \quad (8)$$

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

In particular, we use the relation that Eq. 8 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  $\frac{l_i}{k}$ .<sup>31</sup> There are  $h \in \mathbb{N}$  segments of type A, of size  $\frac{1}{k}$ , where prices are noted  $\tilde{p}_{1_i}^A$ . On each of these segments, the demand is  $\frac{1}{k}$ . There are  $n \in \mathbb{N}$  segments of type B, where prices are noted  $\tilde{p}_{1_i}^B$ . We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

<sup>31</sup>With  $u_i$  and  $s_i$  integers below  $k$ .

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

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

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

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

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

Profits of Firm 2 can be written therefore as:

$$\pi_2(\tilde{\mathcal{P}}) = \sum_{i=1}^n \tilde{p}_2 \left[ \frac{\tilde{p}_{1i}^B - \tilde{p}_2 - t}{2t} + \frac{s_i + l_i}{k} \right]. \quad (9)$$

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

Equilibrium prices are:

$$\begin{aligned} \tilde{p}_{1i}^A &= t + \tilde{p}_2 - 2\frac{u_i t}{k} \\ \tilde{p}_{1i}^B &= \frac{\tilde{p}_2 + t}{2} - \frac{s_i t}{k} = \frac{t}{3} + \frac{2t}{3n} \left[ \sum_{i=1}^n \left[ \frac{s_i}{2k} + \frac{l_i}{k} \right] \right] - \frac{s_i t}{k} \\ \tilde{p}_2 &= -\frac{t}{3} + \frac{4t}{3n} \sum_{i=1}^n \left[ \frac{s_i}{2k} + \frac{l_i}{k} \right]. \end{aligned} \quad (10)$$

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. 8:  $\frac{\tilde{p}_2 + t}{2t} - \frac{l_i}{k} \leq \frac{s_i + l_i}{k}$  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 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:

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

for segments of type B where inequalities in Eq. 8 hold:

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

for segments of type B where inequalities in Eq. 8 do not hold:

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

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

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

By definition,  $\tilde{s}_i$  verifies the inequalities in Eq. 8, thus  $\frac{\tilde{s}_i}{k} \leq \frac{\tilde{p}_2+t}{2t}$ , which allows us to establish that  $\frac{4t}{3(n-m)} \left[ \frac{3m\tilde{p}_2}{4t} + \frac{1}{2k} + \frac{m}{4} - \sum_{i=1}^m \frac{\tilde{s}_i}{2k} \right] \geq \frac{2t}{3nk}$ . It is then immediate to show that:

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

Also, by assumption, firms compete on  $J = [\frac{i}{k}, \frac{i+l}{k}]$  with  $\tilde{\mathcal{P}}$ , which implies that inequalities in Eq. 8 hold, and in particular,  $\frac{\tilde{p}_2+t}{4t} - \frac{i}{2k} \leq \frac{l}{k}$ .

Thus:

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

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

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

For segments of type A:

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

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

For segments of type B:

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

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. 8.

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

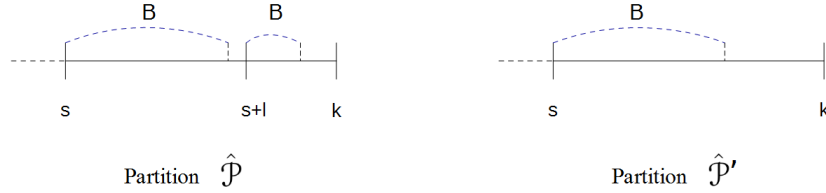


Figure 5: Step 3: demands of Firm 1 on segments of type B (dashed line)

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:

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

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

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

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

Inequalities in Eq. 8 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:

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

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

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

higher profits:

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

We consider the terms separately. First,

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

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

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

On these  $m$  segments, inequalities in Eq. 8 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:

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

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

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

Getting back to the profits difference, we obtain:

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

The first bracket of Equation 11 is positive given Eq. 8. 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,  $\hat{p}_2 \geq \frac{t}{2}$ .<sup>32</sup> And as this price is greater than  $\frac{1}{6}$ , the second bracket of Equation 11 is positive. This proves that  $\Delta\pi_1 \geq 0$ .

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

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

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

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.

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

<sup>32</sup>As shown in Liu and Serfes (2004).

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 established 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 Proofs of Equation 3

We propose a candidate equilibrium function. We consider partitions  $j_1^{seq} = j_2^{seq}$  that maximize respectively the profit of Firm 1 and Firm 2 and that are symmetric. We show that  $p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$  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:

$$\begin{cases} V_1 + \pi_1(j_1^{seq}) - p_{seq} & \text{if Firm 1 accepts the offer,} \\ \bar{\pi}_1(j_2^{seq}) & \text{if Firm 1 declines the offer and Firm 2 accepts the offer,} \\ V_1 & \text{if Firm 2 declines the offer.} \end{cases}$$

Thus, the overall value of Firm 1 is:

$$V_1 + \pi_1(j_1^{seq}) - p_{seq} - \bar{\pi}_1(j_2^{seq}) - V_1 = \pi_1(j_1^{seq}) - p_{seq} - \bar{\pi}_1(j_2^{seq})$$

Thus:

$$p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$$

The data intermediary has no interest in deviating from this price, as lowering  $p_{seq}$  would decrease its profits, and increasing  $p_{seq}$  would have Firm 1 rejecting the offer. Thus  $p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$  is the unique equilibrium of this game.

Moreover, the data intermediary has no interest in deviating from partitions  $j_1^{seq} = j_2^{seq}$ . Indeed, consider  $j_1 \neq j_1^{seq}$ . Necessarily,  $\pi_1(j_1) \leq \pi_1(j_1^{seq})$  as  $j_1^{seq}$  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  $j_2 \neq j_2^{seq}$ . 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  $j_1^{seq}$  and  $j_2^{seq}$ , and are symmetric. ■

## D 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  $i = 1, \dots, j$  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,  $\frac{i}{k}$ :<sup>33</sup>

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

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

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

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:

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

<sup>33</sup>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.

**Step 3: prices, demands and profits in equilibrium.**

We solve prices and profits in equilibrium. First-order conditions on  $\pi_\theta$  with respect to  $p_\theta$  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:<sup>34</sup>

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

Thus, first-order conditions on  $\pi_1$  gives us

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

## E Proof of Proposition 1 and Corollary 1

Proposition 1 comes directly from the expressions  $\pi_1(j_1) = \frac{t}{2} + \frac{2j_1 t}{3k} - \frac{7t}{9} \frac{j_1^2}{k^2} - \frac{tj_1}{k^2}$ , which is clearly concave with a unique maximum and  $\bar{\pi}_1(j_2) = \frac{t}{2} + \frac{2t}{9} \frac{j_2^2}{k^2} - \frac{2}{3} \frac{j_2 t}{k}$ , which is always decreasing on  $[0, 1]$ .

More generally, we have the following equivalence:

$$\begin{aligned}(a) \quad \left. \frac{\partial j_2(j_1)}{\partial j_1} \right|_{\hat{j}_1} = 0 &\iff j_1^* = \frac{6k - 9}{14}, \\ (b) \quad \left. \frac{\partial j_2(j_1)}{\partial j_1} \right|_{\hat{j}_1} > 0 &\iff j_1^* > \frac{6k - 9}{14}, \\ (c) \quad \left. \frac{\partial j_2(j_1)}{\partial j_1} \right|_{\hat{j}_1} < 0 &\iff j_1^* < \frac{6k - 9}{14}.\end{aligned}$$

## F Proof of Proposition 3

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

---

<sup>34</sup>For  $p_{1i} \geq 0 \implies \frac{j}{k} \leq \frac{3}{4}$ . Profits are equal whatever  $\frac{j}{k} \geq \frac{3}{4}$ .

$$\begin{aligned}
CS(j_1, j_2, k) &= \sum_{i=1}^{j_1} \left[ \int_0^{\frac{1}{k}} V - 2t \left[ 1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} - \frac{i}{k} \right] - txdx \right] \\
&+ \int_{\frac{j_1}{k}}^{\frac{1}{2} + \frac{j_1}{3k} - \frac{j_2}{3k}} V - t \left[ 1 - \frac{4}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] - txdx + \int_{\frac{1}{2} + \frac{j_1}{3k} - \frac{j_2}{3k}}^{1 - \frac{j_2}{k}} V - t \left[ 1 - \frac{2}{3} \frac{j_1}{k} - \frac{4}{3} \frac{j_2}{k} \right] - txdx \\
&+ \sum_{i=1}^{j_2} \left[ \int_0^{\frac{1}{k}} V - 2t \left[ 1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} - \frac{i}{k} \right] - txdx \right] \\
&= \sum_{i=1}^{j_1} \frac{1}{k} \left( V - 2t \left[ 1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} - \frac{i}{k} \right] \right) - \frac{j_1 t}{2k^2} \\
&+ \sum_{i=1}^{j_2} \frac{1}{k} \left( V - 2t \left[ 1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} - \frac{i}{k} \right] \right) - \frac{j_2 t}{2k^2} \\
&+ V \left[ 1 - \frac{j_2}{k} - \frac{j_1}{k} \right] - \left[ \frac{1}{2} - \frac{2j_1}{3k} - \frac{j_2}{3k} \right] t \left[ 1 - \frac{4}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] \\
&- \left[ \frac{1}{2} - \frac{2j_2}{3k} - \frac{j_1}{3k} \right] t \left[ 1 - \frac{4}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} \right] - t \left[ \frac{1}{4} - \frac{1}{9} \frac{j_1 j_2}{k^2} - \frac{7}{18} \frac{j_2^2}{k^2} - \frac{7}{18} \frac{j_1^2}{k^2} \right] \\
&= \frac{j_1}{k} \left[ V - 2t \left[ 1 - \frac{1}{3} \frac{j_1}{k} - \frac{2}{3} \frac{j_2}{k} \right] \right] + \frac{j_1(j_1 + 1)t}{k^2} - \frac{j_1 t}{2k^2} \\
&+ \frac{j_2}{k} \left[ V - 2t \left[ 1 - \frac{1}{3} \frac{j_2}{k} - \frac{2}{3} \frac{j_1}{k} \right] \right] + \frac{j_2(j_2 + 1)t}{k^2} - \frac{j_2 t}{2k^2} \\
&+ V \left[ 1 - \frac{j_2}{k} - \frac{j_1}{k} \right] + t \left[ -\frac{5}{4} + \frac{1}{3} \frac{j_1}{k} + \frac{1}{3} \frac{j_2}{k} + \frac{5}{6} \frac{j_1^2}{k^2} + \frac{5}{6} \frac{j_2^2}{k^2} - 2 \frac{j_1 j_2}{k^2} \right] \\
&= V + t \left[ -\frac{5}{4} + \frac{17}{18} \frac{j_1^2}{k^2} + \frac{17}{18} \frac{j_2^2}{k^2} + \frac{j_1 j_2}{k^2} \right] + \frac{1}{2} \frac{j_1 t}{k^2} + \frac{1}{2} \frac{j_2 t}{k^2}
\end{aligned} \tag{13}$$

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

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

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

## G Proof of Proposition 4

We prove that the optimal partition in equilibrium does not depend on the selling mechanism.

The prices of information under the three selling mechanisms are:

$$p_a(\mathcal{P}_1, \mathcal{P}_2) = \pi_1^{I, NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI, I}(\emptyset, \mathcal{P}_{ref})$$

$$p_{tol} = \pi_1^{I, NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI, NI}$$

$$p_{seq} = \pi_1^{I,NI}(\mathcal{P}_1, \emptyset) - \pi_1^{NI,I}(\emptyset, \mathcal{P}_2)$$

It is immediate to see that in each mechanism, the data intermediary chooses  $\mathcal{P}_1$  in order to maximize the profits of Firm 1. Thus, the optimal information partition in equilibrium  $\mathcal{P}_1^*$  does not depend on the selling mechanism.

## G.1 Characterization of selling mechanisms with dependent partitions

The price of information can be written as:

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

Suppose that there exist a function  $f$  such that  $j_1$  and  $j_2$  can be written:  $j_2 = f(j_1)$  and  $j_1 = f^{-1}(j_2)$ . (for the sake of simplicity we restrict our discussion to functions that are continuous and differentiable).

The price of information can be equivalently written as:

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

$$p(j_2) = \pi_1(f^{-1}(j_2)) - \bar{\pi}_1(j_2).$$

Thus, solving for the optimal values of  $j_1$  we have:

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

The optimal values of  $j_1$  will thus depend on functions  $f$  that vary with the selling mechanism. Solving for the optimal value of  $j_2$  depends on the selling mechanism considered.

The three selling mechanisms belong to a class for which

$$\frac{\partial f(j_1)}{\partial j_1} = \frac{\partial f^{-1}(j_2)}{\partial j_2} = 0$$

## G.2 Example of selling mechanisms where partitions are not independent and yet that lead to the same number of consumer segments sold

There exists selling mechanisms 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  $j_1^*(k) = \frac{6k-9}{14}$ . ■

## H Proof of Proposition 5

### *Data collection*

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.

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

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

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

Comparing the derivatives gives us:

$$\frac{\partial p_{seq}^*}{\partial k} > \frac{\partial p_a^*}{\partial k} > \frac{\partial p_{tol}^*}{\partial k}.$$

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

$$k_{seq} > k_a > k_{tol}$$

### *Consumer surplus*

Prices when the data intermediary sells  $j$  segments of information to Firm 1 are provided in Appendix G:

- *Firm 1 captures all demand on each segment  $i = 1, \dots, j$ , and:*

$$p_{1i} = 2t\left[1 - \frac{i}{k} - \frac{1}{3} \frac{j}{k}\right].$$

- *Firms compete on the segment of unidentified consumers, and the prices are:*

$$p_1 = t\left[1 - \frac{4}{3} \frac{j}{k}\right], \quad \text{and} \quad p_2 = t\left[1 - \frac{2}{3} \frac{j}{k}\right].$$

We need to compute demands in order to find consumer surplus. Firm 1 is a monopolist on the first  $j$  segments of size  $\frac{1}{k}$ . Demand is  $\frac{1}{k}$  on each segment.

On the segment of unidentified consumers, where firms compete, the indifferent consumer is characterized by

$$\tilde{x} = \frac{p_2 - p_1 + t}{2t} + \frac{j}{k} \implies \tilde{x} = \frac{4}{3} \frac{j}{k}$$

As  $j^* = \frac{6k-9}{14}$ ,  $\tilde{x}^* = \frac{4k-12}{7k}$ .

We can write consumer surplus in equilibrium:

$$\begin{aligned}
CS(k) &= \sum_{i=1}^{j^*} \left[ \int_0^{\frac{1}{k}} V - 2t \left[ 1 - \frac{1}{3} \frac{j}{k} \right] + \frac{t}{k} + \frac{it}{k} - tx dx \right] \\
&+ \int_{\frac{j^*}{k}}^{\frac{1}{2} + \frac{j^*}{3k}} V - t \left[ 1 - \frac{4}{3} \frac{j^*}{k} \right] - tx dx + \int_0^{\frac{1}{2} - \frac{j^*}{3k}} V - t \left[ 1 - \frac{2}{3} \frac{j^*}{k} \right] - tx dx \\
&= \sum_{i=0}^{j^*-1} \frac{1}{k} \left[ V - 2t \left[ 1 - \frac{1}{3} \frac{j^*}{k} \right] + \frac{t}{k} + \frac{it}{k} \right] - \frac{j^* t}{2k^2} \\
&+ V \left[ 1 - \frac{j^*}{k} \right] - \left[ \frac{1}{2} - \frac{2j^*}{3k} \right] \left[ t - \frac{4}{3} \frac{j^* t}{k} \right] - \frac{t}{2} \left[ \frac{1}{4} - \frac{8}{9} \frac{j^{*2}}{k^2} + \frac{j^*}{3k} \right] \\
&- \left[ \frac{1}{2} - \frac{j^*}{3k} \right] \left[ t - \frac{2}{3} \frac{j^* t}{k} \right] - \frac{t}{2} \left[ \frac{1}{2} - \frac{1}{3} \frac{j^*}{k} \right]^2 \\
&= \frac{j^*}{k} \left[ V - 2t \left[ 1 - \frac{1}{3} \frac{j^*}{k} \right] + \frac{t}{k} \right] + \frac{j^*(j^*-1)t}{k^2} - \frac{j^* t}{2k^2} \\
&+ V \left[ 1 - \frac{j^*}{k} \right] - \frac{t}{2} \left[ 1 + \frac{16j^{*2}}{9k^2} - \frac{8j^*}{3k} \right] - \frac{t}{2} \left[ \frac{1}{4} - \frac{8}{9} \frac{j^{*2}}{k^2} + \frac{j^*}{3k} \right] \\
&- \frac{t}{2} \left[ 1 + \frac{4}{9} \frac{j^{*2}}{k^2} - \frac{4j^*}{3k} \right] - \frac{t}{2} \left[ \frac{1}{4} - \frac{1}{3} \frac{j^*}{k} + \frac{j^{*2}}{9k^2} \right] \\
&= V - \frac{2j^* t}{k} - \frac{j^* t}{2k^2} + \frac{2j^{*2} t}{3k^2} \\
&- \frac{5t}{4} + 2t \frac{j^*}{k} - \frac{13t j^{*2}}{18 k^2} \\
&= V - \frac{5t}{4} - \frac{j^* t}{2k^2} - \frac{7j^{*2} t}{18k^2} \\
&= - \frac{(170k^2 - 144k - 9)t - 56V k^2}{56k^2}
\end{aligned} \tag{14}$$

Consider now the first degree derivative of consumer surplus with respect to  $k$ :

$$\frac{\partial CS(k)}{\partial k} = - \frac{4032k + 9t}{28k^3}$$

This is always negative for  $k \geq 0$ , and thus consumer surplus decreases with information precision. ■

## I Proof of Proposition 6

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:

$$\pi^{NI,NI} = \frac{t}{2}.$$

- 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:

$$\pi^{I,NI}(j_1^*, \emptyset) = \frac{(18k^2 - 12k + 9)t}{28k^2}.$$

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

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

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

$$\pi^{NI,I}(\emptyset, j_1^*) = \frac{(25k^2 + 30k + 9)t}{98k^2}.$$

- Finally, the profit of an uninformed firm facing a competitor informed with  $k$  information segments is provided in [Liu and Serfes \(2004\)](#):

$$\pi^{NI,I}(\emptyset, \mathcal{P}_{ref}) = \frac{(k^2 + 2k + 1)t}{8k^2}.$$

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

$$p_a^* = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,I}(\emptyset, \mathcal{P}_{ref}) = \frac{(29k^2 - 38k + 11)t}{56k^2}$$

$$p_{tol}^* = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,NI} = \frac{(4k^2 - 12k + 9)t}{28k^2}$$

$$p_{seq} = \pi^{I,NI}(j_1^*, \emptyset) - \pi^{NI,I}(\emptyset, j_1^*) = \frac{(76k^2 - 144k + 45)t}{196k^2}$$

Direct comparison of the profits provides the ranking of Proposition 6. ■

## J Proof of Proposition 7

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 three selling mechanisms are equivalent when the data intermediary sells information to both firms.

Under the 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  $j_1^{both}$ . 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  $p_{both} = w_1 = w_2$ .

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 take it or leave it offers, the data intermediary proposes to each firm  $j^{both}$  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 declines the offers, 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 profit of the data intermediary when selling information to both firms is  $\Pi_{both}(k) = 2p_{both} - c(k)$ . ■

## K Proofs of Propositions 8 and 9

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  $p_{\theta}$  on the rest of the unit line. The demand for Firm  $\theta$  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  $\pi_\theta$  with respect to  $p_\theta$ . Prices in equilibrium are:

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

Replacing these values in the above demands and prices gives:

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

and

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

Profits are:

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

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

The data intermediary maximizes the following profit function:

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

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:

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

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, and by  $j = \frac{3k}{4}$  and  $j' = 0$  to obtain profits when Firm 1 is informed with all segments and firm 2 has no information.

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

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:

$$\Pi_{both}(k) = \frac{t}{4} - \frac{t}{k} - c(k).$$

and the first-degree derivative of the profit function with respect to  $k$  is:

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

Finally, consumer surplus in this case is

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

Straightforward comparisons with the values in Appendix I lead to the rankings in Proposition 8 and 9 ■

## L Proof of Proposition 10

We characterize the equilibrium under second-price auctions.

The willingness to pay of firms when the data intermediary auctions information  $j_1^{a2}$  to Firm 1 and  $j_2^{a2}$  to Firm 2 are:

$$\begin{cases} \pi_1(j_1^{a2}) - \bar{\pi}_1(j_2^{a2}), \\ \pi_2(j_2^{a2}) - \bar{\pi}_2(j_1^{a2}) \end{cases}$$

We show that in equilibrium  $j_1^{a2} = j_2^{a2}$ .

Assume  $\pi_1(j_1^{a2}) - \bar{\pi}_1(j_2^{a2}) > \pi_2(j_2^{a2}) - \bar{\pi}_2(j_1^{a2})$  (the other case is solved similarly).

- If  $j_1^{a2} > j_2^{a2}$ :  $\pi_2(j_2^{a2}) - \bar{\pi}_2(j_1^{a2})$  increases when  $j_2^{a2}$  increases.
- If  $j_1^{a2} < j_2^{a2}$ :  $\pi_2(j_2^{a2}) - \bar{\pi}_2(j_1^{a2})$  increases when  $j_1^{a2}$  increases

Thus the data intermediary chooses  $j_1^{a2} = j_2^{a2}$ .

This implies that

$$p_{a_2} = -\frac{((3j_1^{alt2} - 4j_1^{a2})k + 3j_1^{a2})t}{3k}$$

Maximizing  $p_{a_2}$  with respect to  $j_1^{a2}$  and using the FOC give:

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

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

and

$$\frac{\partial p_{a_2}^*}{\partial k} = \frac{(6k - 2)t}{9k^3}.$$

The equality of profits, surplus, and optimal data collection, as well as their relative value with other selling mechanisms is then straightforward. ■



## M Proof of Proposition 11

See the proofs of Propositions 5 and 6.

## N Characterization of the equilibrium with price caps

We prove that data collection decreases when the price cap decreases. We note  $\bar{p}$  the highest price of information allowed by the regulator. The claims are the following:

- (a) Regardless of the selling mechanism, the amount of data collected by the data intermediary decreases with the value of the price cap  $\bar{p}$ .
- (b) The data intermediary will sell information to both firms if  $\bar{p} \leq 2p_{both}$ .

Consider a binding price cap. Then the profits of the data intermediary are:

$$\Pi(k) = \bar{p} - c(k)$$

The optimal value of  $k$  is such that  $p(k^*) = \bar{p}$ . Indeed, if  $k > k^*$ , then costs increase but the price of information does not change as the price cap is binding.

If  $k < k^*$  profits are below the constrained optimal as the data intermediary can increase  $\Pi$  by increasing  $k$ .

As  $p(k)$  increases in  $k$  (see Appendix H), the lower the  $\bar{p}$  the lower the  $k$ .

Consider now a binding price cap  $\bar{p}$ .

If  $\bar{p} \geq 2p_{both}$ , the data intermediary uses auction as it is the only selling mechanism allowing to reach the highest profit possible:  $\max\{p_a, \bar{p}\}$ .

If  $\bar{p} \leq 2p_{both}$ , selling information to both firms is always more profitable because twice the maximal value of  $\bar{p}$  can always be sold.