Market for Information and Selling Mechanisms*

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

December 23, 2020

Abstract

This article investigates how selling mechanisms influence the data collection strategies of intermediaries selling information to competing firms. We provide a general framework that we apply to three selling mechanisms commonly studied in the literature: take it or leave it offers, sequential bargaining, and auctions. We highlight conflicting interests between data intermediaries, data protection agencies and competition authorities, and we discuss regulatory implications.

*An earlier version of this paper has circulated under the title "Data Intermediaries and Selling Mechanisms for Customized Consumer Information". We would like to thank Andreea Cosnita-Langlais, Axel Gautier, Victor Ginsburgh, Doh Shin Jeon, Brad Larsen, Patrick Legros and Wilfried Sand-Zantman, as well as the participants at the 46th EARIE conference, and the CESifo Area Conference on the Economics of Digitization 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. Antoine Dubus acknowledges financial support from the FNRS grant 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, Telecom Paris and University of Paris 2 Panthéon-Assas.

† Université Libre de Bruxelles, ECARES; antoine1dubus@gmail.com.

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

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

1 Introduction

Big-tech companies such as Alphabet, Apple, Facebook, Amazon and Microsoft are today the largest companies in the world with an aggregate market value of more than 5 Trillion USD.[1] They hold a dominant position in multiple sectors of the digital economy such as online search, e-commerce, and social networks, and are also active on multiple other markets (e.g. online advertising and payments). Their success is largely built upon the collection, use, sharing and sale of huge amounts of consumer data.

Acting as data intermediaries, they organize a new market for information by selling information to firms seeking to improve their business practices through better analysis of markets, forecasting trends, and personalized ads, products, and prices (Varian, 1989; Bergemann et al., 2015; Bergemann and Bonatti, 2019).

The promise of the digital revolution to improve the efficiency of markets has recently given way to questionings by economists, data protection and competition authorities (Crémer et al., 2019; Furman et al., 2019; Scott Morton et al., 2019; Tirole, 2020). Recent market practices have indeed shown that data intermediaries can refuse to grant firms access to data, to share or to sell data, and cause important harms to companies and competition.[2]

Another important issue at stake is the enormous amount of personal data that these data intermediaries possess. A recent FT articles[3] assesses that large data brokers have access to more than 3000 data points on more than 200 million Americans, and the FTC recently fined Facebook \$5 billion for not respecting consumer privacy,[4] while in France, Google was fined EUR 50 million for violating the European GDPR.[5] There is thus a pressing societal need to better understand

[1]Stock Market Warning: 6 Mega Stocks Dominate S&P 500's \$21.4 Trillion Cap; CCN, April 27, 2020.

[2]Facebook gave Lyft and others special access to user data; engadget, May 12th, 2018.

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

[4]FTC Imposes \$5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook, Federal Trade Commission, July 24, 2019.

[5]The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC, January 21, 2019.

why companies operating the market for information collect so much data on each consumer. On the one hand, a more detailed consumers profile allows firms purchasing data to propose better targeted offers and this increases the willingness to pay for information. On the other hand, if more companies purchase data, they will fight more fiercely for consumers who have no strong preferences for a particular brand or store, and this will decrease their willingness to pay for information.

This article shows that this fundamental trade-off in the price of information depends on the design of the market for information and especially on how information is sold. On the one hand, Google matches consumer information with specific keywords through Google ad auctions. Similarly, platforms financed by ads sell consumer profiles through targeted ads in Real Time Auction. On the other hand, Facebook also sells user information through direct negotiation with third-parties in take it or leave it offers, and so do most data brokers such as Equifax and other credit rating agencies. Overall, take it or leave it offers are widely used in the data economy.[6] The price of data depends of course on the segment of consumers requested and the number of data points defining each consumer profile, but more importantly it depends on the difference between the direct benefit of purchasing information and the opportunity cost of not purchasing the information. Both components fundamentally depend on the selling mechanism.

Indeed, the selling mechanism determines the price of information, which in turn drives the data collection strategy of the intermediary. On the one hand, firms are ready to pay more for high quality relevant data that will increase profits through better consumer surplus extraction. On the other hand, without information a firm might have to compete against a firm that has acquired information. When deciding whether to purchase information or not, a firm may face a negative externality: if a firm does not purchase information from the data intermediary, its competitors might. Depending on the selling mechanism, this negative externality increases with the quality or precision of information. Indeed, consider a take it or leave it offer: the data intermediary proposes information to a firm, and

[6] Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

if the firm declines the offer, all firms on the market remain uninformed. Now consider an auction with negative externality: if the firm loses the auction, the data intermediary sells information to the winning bidder. It is clear that the value of the threat in the take it or leave it offer does not depend on the precision of information since no firm is informed. In the auction mechanism, however, the value of the threat increases with the precision of information: a firm makes lower profits when it has to compete with a better informed competitor. A data intermediary thus may have interest to collect more information with the auction mechanism where there is a negative externality than with the take it or leave it mechanism. Thus, different selling mechanisms will change how much information will be collected and sold to firms.

The objective of this article is to analyze how the data collection and selling strategies of intermediaries crucially depend on the mechanism through which information is sold. Our results have important consequences for competition policy and personal data protection regulations and give new insights on the subtle interplay between these two dimensions of the data strategies of large data intermediaries. More specifically, we provide a general framework to study how different selling mechanisms influence the price of information and data collection strategies. We apply our results to three selling mechanisms commonly studied in the literature: take it or leave it offers, sequential bargaining, and auctions.

First, data intermediaries can sell information through take it or leave it offers documented by Nielsen,[7] and studied by Binmore et al. (1986) and Bergemann and Bonatti (2019). Secondly, repeated interactions leading to sequential bargaining are implemented by data intermediaries like Facebook.[8] This selling mechanism has the advantage of increasing the bargaining power of the data intermediary compared with take it or leave it offers. Indeed, by making sequential offers to different buyers, a data intermediary may exert a threat on a prospective buyer, as information may be sold to its competitor. This increases the value of information, as shown by Sobel and Takahashi (1983). Finally, auctions are also extensively used

[7] For more detail, see the Nielsen website https://www.nielsen.com/us/en/.

[8] Facebook blocks valuable ad data in privacy update to its marketing partner program, AdAge, February 21 2020.

in data marketplaces (Sheehan and Yalif, 2001; O'kelley and Pritchard, 2009).[9] Auctions also allow intermediaries to exert a strong market power, and extract a large share of surplus from an information buyer.[10]

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 seller.[11] 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) analyze 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, thus highlighting a new interplay between competition and privacy.

Our article also relates to a growing theoretical literature that analyzes the selling strategies of data intermediaries. Bergemann and Bonatti (2015) and Bergemann et al. (2018) study how a data intermediary chooses the precision of information to maximize surplus extraction from firms.[12] Bergemann et al. (2020) also examine the impact of collecting consumer data on the signal sold by an intermediary to a single firm through take it or leave it offers when there is a data externality. Montes et al. (2019) consider a data broker selling to firms all information on consumers on a Hotelling line: firms compete for each consumer on [0,1], and the competitive effect of information is maximal, à la Thisse and Vives (1988). Because all information is sold in this model, the only strategic space of the data broker is to sell to one or to two firms.

We build a model where a data intermediary strategically collects and sells

[9] First price Auction, Second price, and the Header-Bidding, Smartyads, February 2018.

[10] Vickrey (1961), Klemperer (1999), Jehiel and Moldovanu (2000), Figueroa and Skreta (2009) among others analyze auction design.

[11] Larsen (2014); Backus et al. (2018, 2019), and Backus et al. (2020) study bargaining descriptively using field data.

[12] A related literature studies consumer privacy concerns with exogenous information acquisition (Shy and Stenbacka, 2016; Casadesus-Masanell and Hervas-Drane, 2015; Gal-Or et al., 2018).

consumer data. Data collected divides consumer demand into segments of arbitrary size. Thinner segments give more precise information but are more costly to collect. The intermediary sells recombined partitions of the consumer demand to firms for price discrimination purposes. In equilibrium the data intermediary does not sell all information collected to firms, as it would increase the competitive effect of information: firms fight more fiercely for consumers identified as belonging to their core segment. A data intermediary collecting more information will have thinner segments, and thus an increased precision of information. We will show how the mechanism used by the data intermediary to sell information to firms will impact the price of information, and its data collection strategies.

By relating the data collection strategies to the mechanism used to sell consumer information, we contribute to the literature on three points.

First, we find that the selling mechanism used in the market for information has a strong economic impact on the data strategies of data intermediaries. Moreover, among the three mechanisms that we focus on (take it or leave it, sequential bargaining and first price auctions), the data intermediary always prefers to sell information through sequential bargaining or through auctions, which is the worst case scenario for consumers, as it minimizes their surplus and maximizes data collection, reducing their privacy. There are thus conflicting interests between data intermediaries, data protection agencies and competition authorities, on how to design the market for information.

Secondly, we show that the optimal strategy of the data intermediary is to sell information to only one firm with auctions and sequential bargaining, and to two firms with the take it or leave it mechanism. Thus, selling mechanisms will also have an impact on competition through the number of firms that buy information. This has strong implications for competition authorities that want all firms to have equal access to information: both firms only purchase information and compete on a level playing field with take it or leave it offers.

Finally, we show that there is subtle yet important relationship between competition policy and data protection laws that depends on the price of information determined in the market for information. Indeed, when information is sold through auctions, consumer surplus is low, only one firm is informed and the amount of data collected is high; when the data intermediary uses take-it or leave-it offers, consumer surplus is high, both firms are informed and the amount of data collected is low. Thus protecting privacy (through a data minimization principle for instance) can preserve competition and consumer surplus. Conversely protecting competition (by enforcing a level playing field such as an open data policy or a cap on the price of information for instance) can preserve privacy.

The remainder of the article is organized as follows. We describe the model in Section 2. We provide in Section 3 a characterization of selling mechanisms, and we present the three selling mechanisms on which we focus: take it or leave it, sequential bargaining and auctions. We derive general properties of mechanisms in Section 4. We then characterize in Section 5 the equilibria under the three selling mechanisms. We show in Section 5.1 that the three mechanisms belong to a broader class, which we refer to as independent offers. In Section 5.2, we analyze how the price of information is related to the amount of data collected. We examine whether it is more profitable for the data intermediary to sell information to one or to both firms in Section 5.3. We extend the model to second price auctions that do not belong to the class of independent offers in Section 6. We show that nevertheless our main results hold. We discuss regulatory implications, and how to use a data minimization principle, open data, 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.[13] The data intermediary collects and sells data that segment consumers on the Hotelling line. A firm that acquires an information partition, i.e. an informed firm, can set a price on each consumer segment. On the contrary, a firm that does not purchase consumer segments, i.e. that is uninformed,

[13]The marginal production costs are also normalized to zero.

cannot distinguish consumers, and sets a single price on the entire line. This simple model of horizontal differentiation can be used to analyze the impact of information acquisition on the profits of firms (Thisse and Vives, 1988).

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:[14]

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

2.2 Data intermediary

The data intermediary collects information on consumers that allows firms to distinguish consumer segments on the unit line. The data intermediary has therefore to choose the optimal information partition to sell to firms to maximize its profits.[15]

In the baseline model we assume that the data intermediary sells information to only one firm, say Firm 1, and study in Section 5.3 the case where the data intermediary sells information to both firms. We consider a general set-up of information selling by the intermediary, then we focus our analysis on three selling mechanisms: take it or leave it (tol), sequential bargaining (seq), and auctions (a).

[14] 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. (2018) or Montes et al. (2018).

[15] Previous research has assumed that the data intermediary sells all available information (Montes et al., 2018). Bounie et al. (2018) show that this assumption is not valid.

2.2.1 Collecting consumer data

We consider a data intermediary that collects k consumer segments at a cost $c(k)$.[16] 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 a detailed discussion on the structure of the costs associated with data collection). The data collection cost $c(.)$ captures the sum of the costs associated with these activities.

Collecting data is costly for the intermediary but provides more information on consumers. Data collected allows the intermediary to increase the value of information, as a firm can now 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.

Figure 1: Partition $\mathcal{P}_k$

2.2.2 Selling information

Selling information consists for the data intermediary of selling any subset of segments collected in the partition depicted in Figure 1. For instance, the data intermediary can sell a partition starting with one segment of size $\frac{1}{k}$, and another segment of size $\frac{2}{k}$, and so on. Thinner segments in the partition allow a firm to extract more surplus from consumers. This is the rent extraction effect that increases the value of information. It is easy to understand that selling all consumer

[16] See Appendix A for a characterization of the cost function.

segments is not optimal for the data intermediary. Indeed, 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). For instance, if the data intermediary sells all consumer segments, Firm 1 can set lower prices on consumer segments that are closest to Firm 2. This is the competition effect that lowers the value of information.

Finding the optimal partition without any restriction is a complex task given the high dimensionality of the optimization problem. Indeed the data intermediary can potentially recombine consumer segments in any arbitrary fashion. Nevertheless, Bounie et al. (2018) have shown that, under weak conditions, a data intermediary selling information using auctions chooses an optimal information structure that balances a competition and a rent extraction effect of information.[17] The optimal partition for auctions is described in Figure 2.

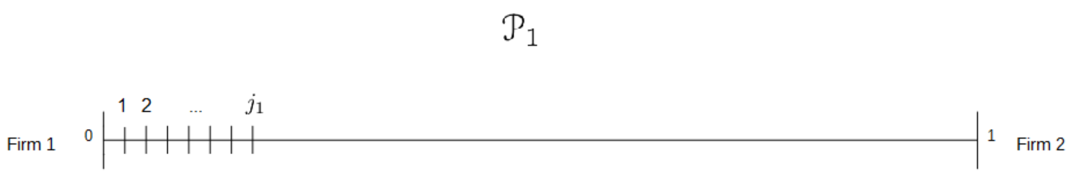

Figure 2: Selling partition $\mathcal{P}_1$ to Firm 1

Consider partition $\mathcal{P}_1$ represented in Figure 2. 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 $[0, \frac{j_1}{k}]$, that Firm 1 can price discriminate. We refer to this interval as the share of identified consumers.[18] 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. The number of segments of identified consumers j_1 depends on the total number of segments on the market k. We denote by $j_1(k)$ the number of segments as a function of k.

[17]See Appendix B for a proof.
[18]Thus $\frac{j_1}{k} \in [0, 1]$.

The partition described in Figure 2 balances the rent extraction effect of information while limiting the competitive effect of information, thus maximizing the willingness to pay for information of Firm 1. 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 coarser than $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$ is not optimal as Firm 1 could always extract more surplus by selling segments of size $\frac{1}{k}$. On the other hand, by keeping consumers far away from Firm 1 unidentified, an optimal information partition softens the competitive pressure due to information on Firm 2. In turn, Firm 2 will keep a relatively high price, and the competitive pressure on Firm 1 will remain low. Any optimal partition must be similar to partition $\mathcal{P}_1$, and the optimization problem for the data intermediary boils down to choosing the number of segments $j_1(k)$ in partition $\mathcal{P}_1$.[19]

It is straightforward to show that partition $\mathcal{P}_1$ is also optimal for take it or leave it offers and sequential bargaining. Indeed, for all three selling mechanisms the data intermediary finds the optimal price for information by maximizing the profits of Firm 1, regardless of the outside option. Therefore, any selling mechanisms for which the outside option is not impacted by the choice of the optimal partition sold to Firm 1, will lead to a partition of type $\mathcal{P}_1$, in particular independent offers that we study in Section 4.1.

As a matter of fact, we conjecture that the optimal information structure for any selling mechanisms must be of type $\mathcal{P}_1$ for the following reason. The rent extraction effect is strongest for segments close to Firm 1, and the competition effect is strongest for segments located far away from Firm 1. Therefore, the price of information is highest when the rent extraction effect is maximized and when the competition effect is minimized by withholding information on far away consumers. However, proving a general result for any selling mechanisms is beyond the scope of this article, and we make the following assumption.

Assumption 1

[19]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.

A permissible 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.*

Assumption 1 will allow us to derive the properties of selling mechanisms based on comparative statics with respect to j_1.

2.3 Firms

A firm may decide to remain uninformed, and in this case it only knows that consumers are uniformly distributed on the unit line. When Firm 1 acquires $j_1(k)$ segments of information, it can price discriminate consumers on these segments. Firms set prices in two stages.[20] First, Firm 1 and Firm 2 simultaneously set homogeneous prices p_1 and p_2 on the whole unit line. Secondly, Firm 1 sets a personalized price on each consumer segment on $[0, \frac{j_1}{k}]$, with p_{1i} being the price on the *ith* segment from the origin. Then consumers observe prices. When setting the competitive price p_1, Firm 1 already knows which consumers it can identify, and thus charges p_1 accordingly. Firm 2 has no information but can observe price p_1 set by Firm 1 on the competitive segment, and thus sets price p_2 as a simultaneous best response. Competition in homogeneous prices p_1 and p_2 is thus similar to standard the Hotelling framework without information.

[20]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 intertemporal 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.

Using this setting, we denote by $d_{\theta i}$ the demand of Firm $\theta = \{1, 2\}$ on the ith segment. Firm 1 is informed and maximizes the following profit function with respect to $p_{11}, .., p_{1j_1}, p_1$:

$$\pi_1 = \sum_{i=1}^{j_1+1} d_{1i} p_{1i} = \sum_{i=1}^{j_1} \frac{1}{k} p_{1i} + d_1 p_1.$$

Firm 2 is uninformed and maximizes $\pi_2 = d_2 p_2$ with respect to p_2.

2.4 Timing

The data intermediary first collects data and sells partition $\mathcal{P}_1$ to Firm 1. Then Firms 1 and 2 set homogeneous prices on the whole unit line. Finally, Firm 1 sets personalized prices on the segments where it has information. Then consumers see prices and buy the product.

- Stage 1: the data intermediary collects data on k consumer segments at cost $c(k)$.
- Stage 2: the data intermediary sells information partition $\mathcal{P}_1$ by choosing the number of segments $j_1(k)$ to include in the partition.
- Stage 3: firms set prices p_1 and p_2 on the competitive segments.
- Stage 4: Firm 1 price discriminates consumers on whom it has information by setting p_{1i}, $i \in [1, j_1(k)]$.

We describe in Section 3 the mechanisms used by the intermediary to sell information, and we show in Sections 4 and 4.2 how the data collection and information selling strategies of the data intermediary are affected by the mechanism.

3 Selling mechanisms

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 define the price of information as the difference between the profits of Firm 1 with information and its profits without information. We then analyze three mechanisms: take it or leave it, sequential bargaining, and auctions. First, with the take it or leave it selling mechanism, the data intermediary proposes an information partition to Firm 1. Following the offer, there is no possibility for the data intermediary to sell information to Firm 2, even if Firm 1 discards the offer. 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 negative externality where the data intermediary auctions simultaneously two information partitions that are potentially different. Firm 1 and Firm 2 can bid in the two auctions, however only the partition with the highest bid will be sold. Thus a firm that remains uninformed will face an informed competitor, similarly to sequential bargaining.

We focus on these three selling mechanisms for two main reasons. First, they are extensively used by data intermediaries,[21] and they have been widely studied in the theoretical literature. Take it or leave it has been studied by Binmore et al. (1986), and used by Admati and Pfleiderer (1986) and Bergemann and Bonatti (2019) to model markets for information. Sequential games have been analyzed for instance by Sobel and Takahashi (1983). Auctions have been studied by Vickrey (1961); Klemperer (1999); Jehiel and Moldovanu (2000); Figueroa and Skreta (2009) among others, and used more recently by Montes et al. (2018) and by Bounie et al. (2018) for the sale of consumer information. Secondly, the three mechanisms cover a wide range of bargaining power of the data intermediary. With the take it or leave it mechanism, the data intermediary has a relatively low bargaining power, as if the negotiation fails, it 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

[21] Nielsen.;
Facebook blocks valuable ad data in privacy update to its marketing partner program, AdAge, February 21 2020.;
First price Auction, Second price, and the Header-Bidding, Smartyads, February 2018.

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 the take it or leave it mechanism. 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 the strongest bargaining power with auctions among the three mechanisms that we consider in this article. In the remainder of this section, we compute the price of information paid by Firm 1 with the three selling mechanisms.

3.1 General set-up

The intermediary sets a price of information that corresponds to the willingness to pay of Firm 1. Considering such situation where the intermediary has strong bargaining power is in line with the observations of the U.S. congress who has accused DIs of abuse of dominant position.[22]

We introduce notations that simplify the presentation of the model. 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). In the take it or leave it mechanism, if Firm 1 declines the offer, Firm 2 is not informed either, and both firms are uninformed. In this case, they set a single price on the unit line and make profit π. In the sequential bargaining and auction formats, Firm 2 can purchase information when Firm 1 declines the offer. We denote by $\bar{\pi}_1(j_2)$ the profit of Firm 1 when Firm 2 has information on the j_2 consumer segments closest to its location.

Moreover, we define a couple of information partitions as the pair (j_1, j_2), where j_1 is the information proposed to Firm 1, and j_2 is the information proposed to Firm 2 (which can include the empty set in the take it or leave it for instance).

The price of information can be written:

$$p_1(j_1, j_2, k) = \pi_1(j_1, k) - \bar{\pi}_1(j_2, k) \tag{1}$$

[22] Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

To simplify notations, we drop index 1 from the price of information when there is no confusion. 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, proposed to Firm 1.[23] We will use this specification in Section 4 to characterize the equilibrium strategies of the intermediary. Moreover we assume that p_1 is a concave function with respect to j_1, with a unique maximum. In the remaining of the section we characterize the three selling mechanisms for a given information precision k, and we drop variable k from the notations.

3.2 Take it or leave it

Take it or leave it offers characterize over the counter negotiations, which are used by many information intermediaries such as Nielsen.[24] 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 has recently been under the critics of the U.S. congress for allowing data intermediaries to extract large shares of profits from information buyers.[25] However, as we show in this article, take it or leave it corresponds in fact to situations where the data intermediary has a low bargaining power compared with the two other mechanisms, as there is no possibility for renegotiation for the intermediary. The insights that we derive from the analysis of this mechanisms can be applied to all mechanisms where renegotiation is not possible, such as Nash bargaining with any level of bargaining power 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 both Firm 1 and Firm 2 remain uninformed. This selling mechanism rules out the possibility for the data intermediary to renegotiate if no selling agreement is found, contrary to the sequential bargaining mechanism

[23] In order to find the optimal integer value of $j_2(.)$, 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.

[24] Nielsen.

[25] Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

that we describe in Section 3.3.

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 j_1^{tol}, and a price of information p_{tol}. Firm 1 can either accept the offer and make profits $\pi_1(j_1^{tol}) - p_{tol}$, or reject the offer and make profits π. The partitions are therefore $(j_1^{tol}, \emptyset)$. Thus, the willingness to pay of Firm 1 for information is $\pi_1(j_1^{tol}) - \pi$. The data intermediary sets the price of information to:

$$p_{tol}(j_1^{tol}) = \pi_1(j_1^{tol}) - \pi. \tag{2}$$

3.3 Sequential bargaining

Selling information through sequential bargaining extends take it or leave it: in case the negotiation with Firm 1 fails, the data intermediary can now propose information to Firm 2. 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 interacting with their competitors.[26] Considering sequential bargaining thus offers insights on over the counter negotiations where data intermediaries have a stronger bargaining power than with take it or leave it.

With the sequential bargaining mechanism, the data intermediary proposes information to each firm sequentially, in a potentially infinite bargaining game. There is no discount factor and the game stops when one firm acquires information. At each stage, the data intermediary proposes information j_θ^{seq} to Firm θ and no information to Firm $-\theta$.

Firm 1 can acquire information j_1^{seq} and make profits $\pi_1(j_1^{seq})$, or decline the offer, and the data intermediary proposes information j_2^{seq} to Firm 2. If Firm 2 acquires information, the profits of Firm 1 are $\bar{\pi}_1(j_2^{seq})$. If Firm 2 declines the offer, the two previous stages are repeated. The partitions are therefore (j_1^{seq}, j_2^{seq}).

[26] Facebook blocks valuable ad data in privacy update to its marketing partner program, AdAge, February 21 2020.

To compute the value 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 $p_{seq}(j_1^{seq})$ of information 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 $\pi_1(j_1^{seq})$. 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. It is clear that, in order to form a stopping time equilibrium of this infinitely repeated game, the two partitions must be symmetric.

Lemma 1

Partitions j_1^{seq} proposed to Firm 1, and j_2^{seq} proposed to Firm 2 are symmetric with respect to $\frac{1}{2}$.

Proof: see Appendix C.

Thus, to find the equilibrium, it is enough to characterize j_1^{seq}. We look for a pure strategy Nash equilibrium in this infinitely repeated game with a stopping time. Consider the equilibrium at the stopping time where Firm 1 purchases information (without loss of generality), we show in Appendix C that the data intermediary sets the price of information to:

$$p_{seq}(j_1^{seq}) = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq}). \tag{3}$$

3.4 Auctions

Finally, the data intermediary can sell information through first price auctions in which Firm 1 and Firm 2 bid for partitions proposed by the data intermediary. Auctions have three main benefits. 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.[27] Secondly, auctions are well designed when a seller wants to sell a unique product that is differently valued by bidders. Thirdly, auctions are used more and more frequently by major data intermediaries such as Google,[28] and in data marketplaces (Sheehan and Yalif, 2001; O'kelley and Pritchard, 2009).

Selling information through auctions in our setup is challenging, as auctions are traditionally used to reveal the willingness to pay of potential bidders. In our model, both firms and the data intermediary know the willingness to pay of all bidders. This raises an underbidding problem:[29] the firm with the highest willingness to pay knows the bid of the other firm. Thus, it can bid just above the willingness to pay of its competitor and win the auction. The data intermediary loses from this underbidding strategy as the firm with the highest willingness to pay wins the auction even though it bids below its valuation. Nevertheless, 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.[30]

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, j_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 j_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 since Firm 2 is at least as well off with partition $\mathcal{P}_k$

[27] Several papers study auction design (Vickrey, 1961; Klemperer, 1999). Auctions are particularly well suited to the sale of information with negative externality (Jehiel and Moldovanu, 2000; Figueroa and Skreta, 2009).

[28] First price Auction, Second price, and the Header-Bidding, Smartyads, February 2018.

[29] These issues arise with auctions as they have been used in previous literature for the sale of information (Montes et al., 2018).

[30] Underbidding practices are in line with the results of Calvano et al. (2019), who show that algorithmic pricing by competing firms lead to collusive outcome even without information transmission.

as in a situation without information and facing Firm 1 informed with k. However, Firm 2 will never bid above the reserve price j_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 j_1^a, win the auction by bidding the reserve price p_a, and make profits $\pi_1(j_1^a) - p_a$. The data intermediary will set a reserve price $p_a = \pi_1(j_1^a) - \bar{\pi}_1(k) - \epsilon$, where ϵ is an arbitrary small positive number. Thus, $\pi_1(j_1^a) - p_a > \bar{\pi}_1(k)$, and since only one partition is sold, it will be j_1^a. In equilibrium, Firm 1 bids p_a for j_1^a, and Firm 2 bids $\pi_2(k) - \bar{\pi}_2(k)$. The partitions are therefore (j_1^a, k). The price paid by Firm 1 for information is:

$$p_a(j_1^a) = \pi_1(j_1^a) - \bar{\pi}_1(k). \tag{4}$$

We have described how to implement auctions using this simultaneous auctions set up, in order to reach the first best price for the data intermediary.[31] 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 in auctions.[32]

4 Characterization of the equilibrium

We solve the game by backward induction and we characterize selling mechanisms with respect to the number of consumer segments sold and collected by the data intermediary in Sections 4.1 and 4.2. We discuss the impact of the number of consumer segments sold in equilibrium and of consumer data collection on consumer surplus in Section 4.3.

[31]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.

[32]For instance, direct offers with this threat, or sequential bargaining with commitment to sell the reference partition to a competitor, would lead to the same result.

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. We show in particular that the number of consumer segments sold by the data intermediary is the same for a class of mechanisms that we refer to as independent offers, that we characterize in Proposition 1 (a). The selling mechanisms among this class have this property that the information proposed to Firm 2 is independent of the information proposed to Firm 1.

The optimal number of consumer segments sold to Firm 1 satisfies the following first order condition:

$$\frac{\partial p(j_1, j_2, k)}{\partial j_1} = \frac{\partial \pi_1(j_1, k)}{\partial j_1} - \frac{\partial \bar{\pi}_1(j_2, k)}{\partial j_2} \frac{\partial j_2(j_1)}{\partial j_1} = 0 \tag{5}$$

It is useful to characterize the information partition that maximizes the profit of Firm 1 ($\hat{j}_1$), as well as general properties of the profits function:

Lemma 2

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

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

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

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

Proof: see Appendix D.

Having highlighted these properties of the profit functions, we can now state Proposition 1 that characterizes the optimal information structure sold to Firm 1:

Proposition 1

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

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

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

Proof: See Appendix E.

Following Proposition 1 (a), we can identify a specific class of information partitions. 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 and infinite sequential bargaining with discount factors, but also the three selling mechanisms studied in this article. For instance, with a Nash bargaining selling mechanism, the data intermediary maximizes with respect to j_1 a share of the joint profits with Firm 1, and does not propose information to Firm 2 if the negotiation breaks down.

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. The properties of take it or leave it, sequential bargaining, and auctions, are indeed radically different: their outside options cover a wide range of values from the absence of threat on Firm 1 if it declines the offer in the take it or leave it, to the maximal feasible threat reached in the benchmark scenario with first price auctions. This result opens the doors to new theoretical approaches focusing on classes of 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: $p_{alt} = \pi_1(j_1^{alt}) - \bar{\pi}_1(j_1^{alt})$. Consequently, the number of segments chosen by the data intermediary affects both the profit of Firm 1 and its outside option, and will not maximize the profit of Firm 1. 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 result of the equilibrium.

Proposition 1 characterizes the properties of information partitions based on the amount of information sold in equilibrium. It has theoretical and practical implications. First, when offers are independent, the data intermediary maximizes the profits of Firm 1. Thus, the joint profits of the data intermediary and Firm 1 are maximized. This collusive behavior benefits Firm 1 to the detriment of Firm 2. This is not necessarily the case with other types of contracts. For instance, with second price auctions, which are equivalent to symmetric offers analyzed in Section 6, the data intermediary maximizes the willingness to pay of the second highest bidder, and the objectives of Firm 1 and of the data intermediary are not aligned.

Secondly, Proposition 1 offers a convenient criterion to assess the impact of a selling mechanism on the amount of information sold on the market. Two selling mechanisms that belong to the class of partitions of Proposition 1 (a) 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

We analyze in this section how different selling mechanisms impact the profits of the data intermediary, the number of consumer segments collected (k), and consumer surplus.

The intermediary collects k consumer segments to maximize its profit $p(j_1^*, j_2^*, k) - c(k) = \pi_1(j_1^*, k) - \bar{\pi}_1(j_2^*, k) - c(k)$:

$$\frac{\partial p(j_1^*, j_2^*, k) - c(k)}{\partial k} = \frac{\partial \pi_1(j_1^*, k)}{\partial k} - \frac{\partial \bar{\pi}_1(j_2^*, k)}{\partial k} - \frac{\partial c(k)}{\partial k} \tag{6}$$

In order to identify the forces at stake when collecting information, we characterize the variations of the profit function with respect to k:

Lemma 3

$$\frac{\partial \pi_1(\hat{j}_1, k)}{\partial k} \geq 0.$$

The proof of this Lemma is straightforward, and stems from the fact that with more precise information, the data intermediary can always do at least as good as with information with a lower precision.

On the contrary, the variations of $\bar{\pi}_1$ depend only on the location of the last consumer identified by $j_2^*(k)$. We can indeed directly write $\bar{\pi}_1(j_2, k) = \bar{\pi}_1(j_2(k))$, which depends on the variations of j_2^* with respect to k. As stated in Lemma 2 (d), $\bar{\pi}_1(j_2, k)$ varies negatively with j_2. Lemma 4 formalizes this relation:

Lemma 4

$$(a) \quad \frac{\partial j_2^*(k)}{\partial k} \geq 0 \implies \frac{\partial \bar{\pi}_1(\hat{j}_2^*, k)}{\partial k} \leq 0.$$

$$(b) \quad \frac{\partial j_2^*(k)}{\partial k} \leq 0 \implies \frac{\partial \bar{\pi}_1(\hat{j}_2^*, k)}{\partial k} \geq 0.$$

Proof: Straightforward from Lemma 2.

Consider mechanisms that satisfy Lemma 4 (a), there are two positive effects for the data intermediary from having more precise information. A first effect increases the profit of Firm 1 when it purchase information $\pi_1(j_1, k)$ through better consumer surplus extraction. A second effect lowers the profits of Firm 1 if it remains uninformed, as it faces Firm 2 with information on more consumers.

Different selling mechanisms will lead to different levels of consumer data collected according to these two strengths.

On the opposite, for mechanisms under which Lemma 4 (b) is verified, more precise information has conflicting effects for the intermediary. On the one hand, more precise information still increases the profits of Firm 1 when it purchases information $\pi_1(j_1, k)$. On the other hand, a higher k now increases the profits of Firm 1 if it remains uninformed, as it faces Firm 2 with information on fewer consumers.

We will see that the three mechanisms considered in this article satisfy Lemma 4 (a), and have j_2^* that are non-decreasing with k. In this case, more precise information will allow the intermediary to sell consumer segments located further away from a firm's location, as surplus extraction dominates competition on a larger share of consumers when information is more precise.

4.3 Consumer surplus

We show how consumer surplus varies with the number of consumer segments sold and with consumer data collection: even if data collection increases, surplus always increases with the number of consumer segments sold to Firm 1. Thus the main driver of consumer surplus is the number of consumers that Firm 1 can identify. However, for any pair of mechanisms in which the number of consumers identified is the same, consumer surplus is driven by data collection:

Proposition 2

$$(a) \;\; \forall \;\; k, k' : \quad j_1 \geq j_1' \quad \Longrightarrow \quad CS(j_1, k) \geq CS(j_1', k')$$

$$(b) \;\; \forall \;\; k \geq k' : \quad j_1 = j_1' \quad \Longrightarrow \quad CS(j_1, k) \leq CS(j_1', k')$$

Proof: see Appendix F.

Proposition 2 offers a convenient way to assess the welfare implications of a selling mechanisms. Mechanisms characterized by Proposition 1 (b) are those for which consumer welfare will be the highest, while consumer welfare will be the lowest under mechanisms that satisfy Proposition 1 (c). These results call for better scrutiny over selling mechanisms, as knowing which mechanism is used by the industry reveals useful information on how information is provided on a market, and how consumer surplus is affected.

Following Proposition 2 (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. This is because they are characterized by independent offers under which the same number of consumer segments are sold for a given precision k.

5 Application to take it or leave it, sequential bargaining and first price auction mechanisms

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 both firms on the market.

5.1 Number of consumer segments sold in equilibrium

We characterize in Proposition 3 the number of consumer segments sold to Firm 1 in equilibrium with the take it or leave it, sequential bargaining and auction mechanisms.

Proposition 3
The number of consumer segments sold in equilibrium is:

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

Proof: see Appendix G.

The proof of Proposition 3 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 the take it or leave it mechanism, 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 the auction mechanism, 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 which 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. 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 value of information, which is determined by the outside option that varies with the selling mechanism. Even though the data intermediary sells the same information partitions to firms 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,[33] 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 $c(k)$: $\Pi(k) =$

[33]We assume that the cost of collecting data does not depend on the selling mechanism.

$p(k) - c(k)$.[34]

We have established in Proposition 3 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 4 compares the number of segments collected by the data intermediary and consumer surplus with the three selling mechanisms.

Proposition 4

The number of consumer segments collected k and consumer surplus CS are inversely correlated:

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

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

Proof: see Appendix H.

Proposition 4 shows that the number of consumer segments collected is minimized with the take it or leave it mechanism. 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 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 profits 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 in

[34] We make the assumption that Π is concave and reaches a unique maximum on $\mathbb{R}^+$. See Appendix A for a mathematical expression of this assumption.

stronger with more precise information. The data intermediary chooses 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 3), 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, and thus is less sensitive to an increase of precision. Thus information collection is maximized, and consumer surplus minimized with sequential bargaining. Proposition 4 sharply contrasts with the existing literature that argues that more information leads to higher consumer surplus due to the competitive effect of information (Thisse and Vives, 1988; Stole, 2007). Here, more information collected by the data intermediary allows firms to extract more consumer surplus, while at the same time the data intermediary can reduce the intensity of competition on the product market by selling an appropriate partition to Firm 1. The data intermediary thus maximizes rent extraction and minimizes the competitive effect of information.

Proposition 5 shows that the data intermediary prefers auctions, and that the take it or leave it is the least profitable selling mechanism.

Proposition 5

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

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

Proof: see Appendix I.

With the auction selling mechanism, 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 the take it or leave it mechanism, both firms are uninformed when a firm rejects the offer of the data intermediary, resulting in a lower willingness to pay of firms for information.

5.3 Selling information to one or to 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 both firms, and for the three selling mechanisms, we compare profits 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 both firms. Next, we show that the data intermediary sells information to both firms only with the take it or leave it mechanism, and to only one firm with auctions and sequential bargaining. Finally, we compare the equilibrium outcomes with the three selling mechanisms, acknowledging the fact that the data intermediary only sells information to both firms with take it or leave it.

We show in Proposition 6 that profits are identical with the three selling mechanisms when the data intermediary sells information to both firms.

Proposition 6

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

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

Proof: see Appendix J.

The data intermediary maximizes the sum of the prices of information paid by each firm. Each price is 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 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 7 whether the data intermediary sells information to one or to both firms with the three selling mechanisms.

Proposition 7

The data intermediary sells information:

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

Proof: see Appendix K.

The intuition behind Proposition 7 is the following. For the auctions and the sequential bargaining mechanisms, 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 the take it or leave it mechanism, the data intermediary cannot threaten Firm 1 if it declines the offer. Therefore the data intermediary prefers to sell information to both firms using the take it or leave it mechanism, while it only sells information to one firm in the auction and sequential bargaining mechanisms. Thus the selling mechanism has an impact on the number of firms that are informed on a market, and thus on the intensity of competition and on consumer surplus. This result is in line with the observations of the report of the U.S. congress on antitrust for the digital economy[35], that condemns practices of refusal to deal by data intermediaries, thus excluding certain firms from information acquisition.

Accounting for the optimal selling strategy of the data broker, we rank profits with the three selling mechanisms in Proposition 8.

Proposition 8

$$\Pi_a > \Pi_{seq} > \Pi_{both}^{tol} = \Pi_{both}$$

Proof: see Appendix K.

The data intermediary can exert a threat on Firm 1 with auctions and sequential bargaining, which increases its willingness to pay for information. Selling information to both firms intensifies competition between firms, lowers their surplus,

[35] Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

and in turn lowers the price of information. Thus selling information to both firms results in lower profit than selling information to Firm 1 only with auctions and sequential bargaining. On the contrary, when the data intermediary sells information to only one firm with take it or leave it, surplus extraction is relatively low as there is no threat on Firm 1 is it declines the offer: both firms remain uninformed. Thus selling information to both firms is more profitable in this case.

The ranking of profits is identical to Proposition 5. However, as the data intermediary sells information to both firms with take it or leave it, equilibrium values are changed. We characterize in Proposition 9 the number of consumer segments collected and sold when selling information to both firms in equilibrium, as well as consumer surplus. We compare these values with equilibrium with auctions and sequential bargaining. Similarly to Proposition 4, we show that there is a negative relation between consumer surplus and the amount of data collected.

Proposition 9

$$(a) \quad j^{both*} = \frac{6k-9}{22}$$

$$(b) \quad k_{seq} > k_a > k_{both}$$

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

Proof: see Appendix K.

Proposition 9 confirms the results obtained in Proposition 4. The number of consumer segments collected when selling information to both firms with take it or leave it is lower than with auctions and sequential bargaining, where the data intermediary extracts a large share of profits from Firm 1 by preventing Firm 2 from acquiring information. The number of consumer segments sold to firms is lower than, $j_1^*(k) = \frac{6k-9}{14}$, when the data intermediary sells information to Firm 1 only. By selling fewer segments to both Firm 1 and Firm 2, the data intermediary internalizes the competitive effect of information, which increases the profits of firms, and their willingness to pay for information. When both firms are informed, more consumers are identified and consumer surplus is higher.

All results of Sections 4 and 4.2 hold when the data intermediary chooses whether to sell information to both firms. The take it or leave it mechanism is still optimal for consumers: the data intermediary chooses to sell information to both firms, which minimizes the number of consumer segments collected, and maximizes consumer surplus compared to the sequential bargaining and auction mechanisms.

6 Second price auctions and symmetric offers

We consider in this section an alternative mechanism used to sell information to firms: second price auctions. There are four main reasons that make second price auctions an interesting mechanism to analyze. First, second price auctions prevent underbidding from auction participants. Secondly, 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 data intermediaries have a low bargaining power. Thirdly, focusing on second price auctions will allow us to shed light on the ongoing debate in the online ads industry, on the use of first or second price auctions.[36] Finally, second price auctions allows us to identify another class of selling mechanisms where information partitions proposed to firms are perfectly correlated.

With second price auctions,[37] the data intermediary auctions partitions $j_1^{a_2}$ and $j_2^{a_2}$, and Firm 1 (the highest bidder) pays the price corresponding to the bid of Firm 2 (the lowest bidder) for partition $j_2^{a_2}$. We compare profits Π_{a_2}, consumer surplus CS_{a_2}, and the amount of data collected k_{a_2} with second price auctions, with the outcomes of the three other selling mechanisms.

Proposition 10

[36] Google's adoption of first-party auction creates migration headaches for buyers, Digiday, March 8 2019.

[37] 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.

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

$$(a) \quad j_1^{a_2*} = j_2^{a_2*} = \frac{4k-3}{6}$$

$$(b) \quad \Pi_a > \Pi_{a_2} > \Pi_{seq} > \Pi_{both}^{tol}$$

$$(c) \quad k_{seq} > k_a > k_{a_2} > k_{both}^{tol}$$

$$(d) \quad CS_{both}^{tol} > CS_{a_2} > CS_a > CS_{seq}.$$

Proof: see Appendix L.

Introducing the possibility for the data intermediary to sell information with second price auctions does not change the comparison between auctions and take it or leave it. The take it or leave it mechanism still minimizes the number of consumer segments collected and maximizes consumer surplus. The data intermediary would prefer the first price auction mechanism as it leads to the highest willingness to pay of Firm 1. Thus this result contributes to the debate on the design of the optimal auctions for online advertisement: second price auctions reduce the amount of data collected, but first price auctions maximize the price of information.

Comparing second price auctions with first price auctions, we see that the amount of consumer data collected is higher, and consumer surplus lower with first price auctions than with second price auctions. First price auctions are preferred by the data intermediary as they maximize its profits. Moreover, the data intermediary auctions an information partition that is optimal for Firm 1 with first price auction, while both firms have access to symmetric partitions with second price auctions. Thus, second price auctions guarantee fair and equal access to data, and ensures competition on a level playing field. For these reasons, second price auctions are preferred by data protection agencies and by competition authorities.

Finally, partitions proposed to Firms 1 and 2 in second price auctions are symmetric. 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, $j_1^{a_2*} > j_2^{a_2*}$, the data intermediary 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: $j_1^{a_2*} = j_2^{a_2*}$. 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 the value under independent offers $\hat{j}_1$.

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 derive policy implications with a special focus on take it or leave it, sequential bargaining and auctions as they are extensively used by the industry, and thus they allow us to capture concrete effects of regulatory measures. The data intermediary and regulators have conflicting views over which selling mechanism to use for two reasons. First, Propositions 4 and 5 show that the data intermediary prefers the auction mechanism that maximizes its profits but leads to a lower consumer surplus than the take it or leave it mechanism. However, a competition authority, concerned with consumer surplus, and a data protection agency, concerned with the amount of consumer data collected by the data intermediary, prefer the take it or leave it mechanism.

Secondly, access to data is scrutinized by competition authorities who want to guarantee a 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 harms to other companies and to 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.[38] A competition authority may prefer a market situation where all market participants are informed, while we have shown that a data intermediary prefers to sell information to only one firm using first price auctions.

While enforcing a specific selling mechanism is a particularly hard task to do for a regulator, we consider three regulatory tools that allow to reach the market outcomes of a take it or leave it mechanism, therefore lowering data collection and increasing consumer surplus. The first one is a data minimization principle: data protection agencies may change the data collection strategy of a data intermediary by setting a limit over the amount of data collected k. For instance the European GDPR enforces a data minimization principle, purpose of data processing, and informed consent (General Data Protection Regulation). The second regulatory tool is the enforcement of open data, under which a data intermediary does not have the right to refuse selling information to Firm 2. Finally we show that this outcome can also be reached by setting a cap on the price of information. Thus these last two regulatory tools allow to ensure competition on a level playing field.

7.1 Data minimization principle

A data protection agency can set a limit $\bar{k}$ over the amount of consumer data collected by a data intermediary. The aim of a data minimization principle is to protect consumer privacy, by forcing firms to collect as few data as possible. This regulatory tool, enacted for instance in the European General Data Protection Regulation (General Data Protection Regulation), ignores the potential benefits for consumers of customization of services and product with their data, which appear in our model since consumer surplus is always higher when firms price discriminate than in the standard Hotelling model without information. 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

[38]Facebook gave Lyft and others special access to user data; engadget, May 12th, 2018.

- *(a) The ranking of profits of Propositions 10 is independent of $\overline{k}$.*
- *(b) Consumer surplus decreases with $\overline{k}$.*

Proof: See Appendix M.

Proposition 11 shows that reducing the amount of consumer information collected by the data intermediary will increase consumer surplus. With less precise information, firms can identify consumers less precisely and there is less surplus extraction from consumers. The results of Proposition 5 still hold, and the data intermediary prefers to sell information through the auction mechanism. Indeed, surplus extraction from Firm 1 depends on the threat of being uninformed, which is the highest with auctions, and the lowest in the take it or leave it mechanism.

In the next section we show how open data regulation can be used to force the data intermediary to sell information to both firms, thus allowing fair competition between firms.

7.2 Enforcing a level playing field

7.2.1 Open data

We have seen in Proposition 7 that the intermediary has interest to sell information to Firm 1 only under auctions and sequential bargaining, which raises concern for the resulting dominance of Firm 1 over Firm 2 that remains uninformed. Such exclusionary practices have been assessed by the U.S. congress in its recent report,[39] and by Crémer et al. (2019) and Tirole (2020) among other. These reports propose several regulatory measures, among which open data has caught special attention from scholars and practitioners. They propose that a regulator forces the data intermediary to offer fair and equal access to both firms which leads to the equilibrium where both firms are informed, as characterized in Section 5.3. Proposition 9 describes the implications of such measure: consumer surplus would increase, while the number of consumer segments collected by the intermediary

[39] Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations, last accessed, November 28 2020.

would decrease. Open data would thus have a positive effect on consumer surplus while protecting consumer privacy.

It is hard to enforce open data without compulsory licenses or FRAND. This topic is currently debated among experts (Crémer et al., 2019; Tirole, 2020).

In the next section we show how a price cap can be used to force the data intermediary to sell information to both firms, thus allowing fair competition between firms.

7.2.2 Price cap

Setting a price cap is another tool for competition authorities to protect consumers purchasing power (see recently Rey and Tirole (2019)). We analyze the impacts of a price cap over the strategies of the data intermediary: 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 result, the amount k of consumer data collected will change.

First, regardless of the selling mechanism, the amount of data collected by the data intermediary decreases with the value of the price cap.[40] This results from the log concavity of the price with respect to k, meaning that the rent extraction effect is always stronger than the competition effect that is internalized by the data intermediary. This relationship was noticed by Varian (2018), who shows that the performance of artificial intelligence algorithms displays a decreasing return to scale with respect to the amount of data used. Moreover, a price cap can also be of interest for data protection agencies since the amount of data collected increases with the value of the cap.

Secondly, there exists a price below which the DI will prefers to sell information to both firms, thus resulting in an equilibrium similar to take it or leave it.[41] This result can be used by competition authorities to ensure a level playing field, by setting the price cap such that the data intermediary sells information to both firms. In this case the DI sells (symmetric) information to all firms, regardless of

[40] See Appendix N.
[41] See Appendix N.

the selling mechanism. In other words, lowering the price cap reduces the amount of consumer data collected, and setting the price cap low enough increases market competition and consumer surplus, and guarantees fair competition between firms.

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 way data intermediaries sell information can harm consumer welfare by increasing the amount of data collected, and by limiting competition between firms on the markets. Our model of data intermediary that collects and sells consumer information has therefore implications for competition policy, personal data protection and emphasizes the interplay between both regulatory frameworks.

First, the selling mechanism can impact competition on markets by encouraging data intermediaries to offer firms differentiated access to data. Indeed, the data intermediary prefers to sell information to only one firm with sequential bargaining and auction but not with take it or leave it offers. Consumer surplus when information is sold to only one firm, is lower than when both firms are informed. More information on the market could be enforced by regulation to guarantee a level playing field, which can be achieved by enforcing open data regulations or using 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). These new insights can fuel 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.[42]

Secondly, our results show that the price established on the market for information will influence the amount of data collected, and thus how well consumer privacy is protected. Indeed, the take it or leave it mechanism results in a lower level of data collected compared to auction or sequential bargaining mechanisms. The amount of consumer data collected in equilibrium is driven by the price of information, which depends positively on the profit of the firm that purchases information, and negatively 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 lowest with the take it or leave it mechanism, where the outside option does not change with the data collection strategy. Information collection is maximized, and consumer surplus minimized with sequential bargaining. These new results can be of interest for data protection agencies concerned with the amount of personal data collected by firms.

Finally, our model sheds light on the subtle interplay between data protection regulations and competition policy. According to the economic literature, there is a tradeoff between data protection and competition, as increasing the amount of information on markets increases consumer surplus (Thisse and Vives, 1988) but at the cost of consumer privacy. We challenge this view by showing that when data intermediaries behave strategically, they internalize the negative competitive effect of information so that more information on the market does not necessarily increase consumer surplus. The three selling mechanisms that we have analyzed – take it or leave it, sequential bargaining and first price auctions – are characterized by an inverse relationship between data collection (less privacy protection) and consumer surplus: more data collected means less consumer surplus. Among the three selling mechanisms, the take it or leave it mechanism is the only one to achieve both goals of data protection agencies willing to minimize data collection,

[42] 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.

and of competition authorities who want to maximize consumer surplus. Understanding the theoretical properties of selling mechanisms is therefore essential to promote a competitive digital economy that preserves consumer data protection.

References

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.

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.

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

Bergemann, D. and A. Bonatti (2019). Markets for information: An introduction. *Annual Review of Economics 11*.

Bergemann, D., A. Bonatti, and T. Gan (2020). The economics of social data.

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

Bergemann, D., B. Brooks, and S. Morris (2015). The limits of price discrimination. *American Economic Review 105*(3), 921–57.

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 (2018). Selling strategic information in digital competitive markets. Technical report, CESifo Working Paper.

Calvano, E., G. Calzolari, V. Denicolò, and S. Pastorello (2019). Artificial intelligence, algorithmic pricing and collusion. *Algorithmic Pricing and Collusion*.

Casadesus-Masanell, R. and A. Hervas-Drane (2015). Competing with privacy. *Management Science 61*(1), 229–246.

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*.

Figueroa, 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.

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

Gal-Or, E., R. Gal-Or, and N. Penmetsa (2018). The role of user privacy concerns in shaping competition among platforms. *Information Systems Research*.

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'kelley, 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 coopetition. *Journal of Political Economy*.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 97–109.

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

Sheehan, A. and G. Yalif (2001, December 6). Multi-round auction and internet marketplace. US Patent App. 09/746,022.

Shy, O. and R. Stenbacka (2016). Customer privacy and competition. *Journal of Economics & Management Strategy 25*(3), 539–562.

Sobel, J. and I. Takahashi (1983). A multistage model of bargaining. *The Review of Economic Studies 50*(3), 411–426.

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

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

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

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

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

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:

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

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.

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 I.

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

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

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

We write the profit function for any p_2, replacing p_{1i} and d_1:

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

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

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

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

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

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

Going from left to right on the Hotelling line, look for the first time where a type B interval, $J = [\frac{i}{k}; \frac{i+l}{k}]$ of length $\frac{l}{k}$, is followed by an interval $I_1 = [\frac{i+l}{k}, \frac{i+l+1}{k}]$ of type A, shown to be of size $\frac{1}{k}$ in step 1. Consider a reordering of the overall interval $J \cup I_1 = [\frac{i}{k}, \frac{i+l+1}{k}]$ in two intervals $I_1' = [\frac{i}{k}; \frac{i+1}{k}]$ and $J' = [\frac{i+1}{k}, \frac{i+l+1}{k}]$. We show in this step that such a transformation increases the profits of Firm 1.

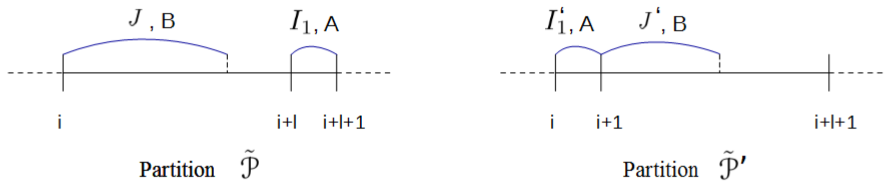

Figure 4: Step 2: relative position of type A and type B segments

The two cases are shown in Figure 4 and correspond respectively to the partitions $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. The curved line represents the demand of Firm 1, which does not cover type B segments. In partition $\tilde{\mathcal{P}}$, a segment of type B of size $\frac{l}{k}$, J, is followed by a segment of type A of size $\frac{1}{k}$, I_1. We show that segments of type A are always located closest to Firm 1 by proving that it is always optimal to change partition starting with segments of type B with a partition starting with segments of type A like in partition $\tilde{\mathcal{P}}'$. To show this claim, we compare the profits of the informed firm with J, I_1 under partition $\tilde{\mathcal{P}}$ and with I_1', J' under partition $\tilde{\mathcal{P}}'$, and we show that the latter is always higher than the former. The other segments of the partition remain unchanged.

To compare the profits of the informed firm under both partition, we first characterize type B segments. Segment J of type B is non null (has a size greater than $\frac{1}{k}$), if the following restrictions imposed by the structure of the model, are met: respectively positive demand and the existence of competition on segments of type B. In order to characterize type A and type B segments, it is useful to consider the following inequality:

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

In particular, we use the relation that Eq. 7 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 $\frac{l_i}{k}$.[43] There are $h \in \mathbb{N}$ segments of type A, of size $\frac{1}{k}$, where prices are noted $\tilde{p}_{1i}^A$. On each of these segments, the demand is $\frac{1}{k}$. There are $n \in \mathbb{N}$ segments of type B, where prices are noted $\tilde{p}_{1i}^B$. We find the demand for Firm 1 on these segments using the location of the indifferent consumer:

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

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

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

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

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

Profits of Firm 2 can be written therefore as:

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

Firm 1 maximizes profits $\pi_1(\tilde{\mathcal{P}})$ with respect to $\tilde{p}_{1i}^A$ 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:

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

We can now compare profits with $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{P}}'$. When we move segments of type B from the left of segments of type A to the right of segment of type A, it is important to check that Firm 1 is still competing with Firm 2 on each segment of type B, and that Firm 1 is still in constrained monopoly on segments of type A. The second condition is met by the fact that price $\tilde{p}_2$ is higher in $\tilde{\mathcal{P}}'$ than in $\tilde{\mathcal{P}}$. The first condition is guaranteed by Eq. 7: $\frac{\tilde{p}_2 + t}{2t} - \frac{l_i}{k} \leq \frac{s_i + l_i}{k}$ for some segments located at s_i of size l_i. By abuse of notation, let s_i denote the segment located at

[43] With u_i and s_i integers below k.

$[\frac{s_i}{k}, \frac{s_i+l_i}{k}]$, 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:

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

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

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

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

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

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

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

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

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

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

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

Thus:

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

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

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

For segments of type A:

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

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

For segments of type B:

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

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

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

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

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

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

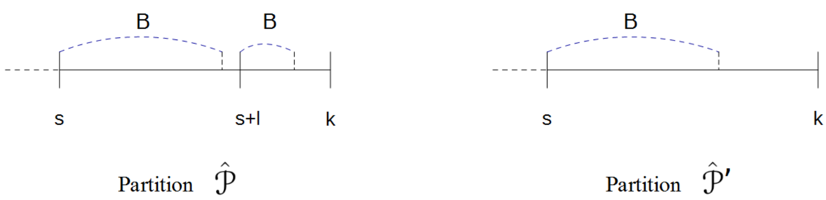

Figure 5: Step 3: demands of Firm 1 on segments of type B (dashed line)

Figure 5 depicts partition $\hat{\mathcal{P}}$ on the left panel, and partition $\hat{\mathcal{P}}'$ on the right panel (on each segment the dashed line represents the demand for Firm 1). Partition $\hat{\mathcal{P}}$ divides the interval $[\frac{i}{k}, 1]$ in two segments $[\frac{i}{k}, \frac{i+l}{k}]$ and $[\frac{i+l}{k}, 1]$, whereas $\hat{\mathcal{P}}'$ only includes segment $[\frac{i}{k}, 1]$. We compare the profits of the firm on the segments where firms compete and we show that $\hat{\mathcal{P}}'$ induces higher profits for Firm 1. There are three types of segments to consider:

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

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

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

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

2. There are m segments which were type B in partition $\hat{\mathcal{P}}$ are no longer necessarily of type B in partition $\hat{\mathcal{P}}$ (and are therefore of type A).

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

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

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

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

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

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

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

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

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

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

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

We consider the terms separately. First,

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

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

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

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

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

thus:

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

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

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

Getting back to the profits difference, we obtain:

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

The first bracket of Equation 10 is positive given Eq. 7. The second bracket is positive if $\frac{\hat{p}_2}{2t}+\frac{s+l}{3k}\geq\frac{1}{6}$. A necessary condition for this result to hold is $\hat{p}_2\geq\frac{1}{6}$. We now show that $\hat{p}_2\geq\frac{t}{2}$

We show in Equation 9 that $\hat{p}_2=-\frac{t}{3}+\frac{4t}{3(n+1)}\sum_{i=1}^{n+1}[\frac{s_i}{2k}+\frac{l_i}{k}]$. We now show that p_2 is minimal when the data intermediary sells the reference partition $\mathcal{P}_{ref}$ to Firm 1, which consists of segments of size $\frac{1}{k}$. Indeed, it is immediate to see that, p_2 always decreases when $\mathcal{P}$ becomes finer. It is thus immediate that p_2 is minimal with the reference partition and $p_2 \geq \frac{t}{2}$[44]. And as this price is greater than $\frac{1}{6}$, the second bracket of Equation 10 is positive. This proves that $\Delta\pi_1 \geq 0$.

We have just 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.

The profits of Firm 1 are minimized when Firm 2 acquires $\mathcal{P}_{ref}$.

This claim is straightforward to establish, as we have shown in step 3 that the price set by an uninformed Firm is minimized when its competitor acquires the reference partition. Thus, demand and profit are also minimized for this partition and the data intermediary sells $\mathcal{P}_{ref}$ to Firm 2.

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 Proof of Lemma 1 and Equation 3

We propose a candidate equilibrium function. We consider $j_1^{seq} = j_2^{seq}$ described in Section 2.2.2, 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:

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

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

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 2

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

Step 1: prices and demands.

Segments of identified consumers are of size $\frac{1}{k}$, and the last one is located at $\frac{j-1}{k}$. Firm 1 sets a price p_{1i} for each segment $i = 1, .., j$ and where it is in constrained monopoly: $d_{1i} = \frac{1}{k}$. Prices on each segment are determined by the indifferent consumer of each segment located at its right extremity, $\frac{i}{k}$:[45]

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

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

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

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

Step 2: profits.

The profits of both firms can be written as follows:

[45] 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.

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

Step 3: prices, demands and profits in equilibrium.

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

Profits are:[46]

$$\begin{aligned} \pi_1^* &= \sum_{i=1}^{j} \frac{2t}{k}[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}] + \frac{t}{2}(1 - \frac{4}{3}\frac{j}{k})^2 \\ &= \frac{t}{2} + \frac{2jt}{3k} - \frac{7t}{9}\frac{j^2}{k^2} - \frac{tj}{k^2} \qquad (11) \\ \pi_2^* &= \frac{t}{2} + \frac{2t}{9}\frac{j^2}{k^2} - \frac{2}{3}\frac{jt}{k}. \end{aligned}$$

Thus, first order conditions on π_1 gives us

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

E Proof of Proposition 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:

$$(a) \quad \left.\frac{\partial j_2(j_1)}{\partial j_1}\right|_{\hat{j}_1} = 0 \iff j_1^* = \frac{6k - 9}{14},$$

$$(b) \quad \left.\frac{\partial j_2(j_1)}{\partial j_1}\right|_{\hat{j}_1} > 0 \iff j_1^* > \frac{6k - 9}{14},$$

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

$$(c) \quad \frac{\partial j_2(j_1)}{\partial j_1}\bigg|_{\hat{j}_1} < 0 \iff j_1^* < \frac{6k-9}{14}.$$

F Proof of Proposition 2

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

$$\begin{aligned}
CS(j_1, j_2, k) &= \sum_{i=1}^{j_1}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
&+ \int_{\frac{j_1}{k}}^{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}} V - t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] - tx\mathrm{d}x + \int_{\frac{1}{2}+\frac{j_1}{3k}-\frac{j_2}{3k}}^{1-\frac{j_2}{k}} V - t[1 - \frac{2}{3}\frac{j_1}{k} - \frac{4}{3}\frac{j_2}{k}] - tx\mathrm{d}x \\
&+ \sum_{i=1}^{j_2}[\int_0^{\frac{1}{k}} V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}] - tx\mathrm{d}x] \\
&= \sum_{i=1}^{j_1}\frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k} - \frac{i}{k}]) - \frac{j_1 t}{2k^2} \\
&+ \sum_{i=1}^{j_2}\frac{1}{k}(V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k} - \frac{i}{k}]) - \frac{j_2 t}{2k^2} \\
&+ V[1 - \frac{j_2}{k} - \frac{j_1}{k}] - [\frac{1}{2} - \frac{2j_1}{3k} - \frac{j_2}{3k}]t[1 - \frac{4}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] \\
&- [\frac{1}{2} - \frac{2j_2}{3k} - \frac{j_1}{3k}]t[1 - \frac{4}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] - t[\frac{1}{4} - \frac{1}{9}\frac{j_1 j_2}{k^2} - \frac{7}{18}\frac{j_2^2}{k^2} - \frac{7}{18}\frac{j_1^2}{k^2}] \\
&= \frac{j_1}{k}[V - 2t[1 - \frac{1}{3}\frac{j_1}{k} - \frac{2}{3}\frac{j_2}{k}] + \frac{j_1(j_1+1)t}{k^2} - \frac{j_1 t}{2k^2} \\
&+ \frac{j_2}{k}[V - 2t[1 - \frac{1}{3}\frac{j_2}{k} - \frac{2}{3}\frac{j_1}{k}] + \frac{j_2(j_2+1)t}{k^2} - \frac{j_2 t}{2k^2} \\
&+ V[1 - \frac{j_2}{k} - \frac{j_1}{k}] + t[-\frac{5}{4} + \frac{1}{3}\frac{j_1}{k} + \frac{1}{3}\frac{j_2}{k} + \frac{5}{6}\frac{j_1^2}{k^2} + \frac{5}{6}\frac{j_2^2}{k^2} - 2\frac{j_1 j_2}{k^2}] \\
&= V + t[-\frac{5}{4} + \frac{17}{18}\frac{j_1^2}{k^2} + \frac{17}{18}\frac{j_2^2}{k^2} + \frac{j_1 j_2}{k^2}] + \frac{1}{2}\frac{j_1 t}{k^2} + \frac{1}{2}\frac{j_2 t}{k^2}
\end{aligned} \tag{12}$$

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

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

Consider $j_1 > j_1'$.

$$
\begin{aligned}
CS(j_1, k) - CS(j_1', k') &> \frac{25}{18}\frac{j_1}{k'^2} - \frac{4}{9k'^2} \\
&> \frac{17}{18k'^2}
\end{aligned}
$$

Which leads us to $CS(j_1, k) \geq CS(j_1', k') \iff k' \geq 1$, which is always verified.
$\blacksquare$

G Proof of Proposition 3

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.

Characterization of selling mechanisms with dependent partitions

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 two functions f: $j_2 = f(j_1)$ and g: $j_1 = g(j_2)$. (for the sake of simplicity we restrict our discussion to functions that are continuous and differentiable).

We can write the price of information in two ways:

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

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

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

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

Solving for the optimal values of j_1 will thus accounts for functions f that depends on the selling mechanism, and thus characterize the relation between j_1 and j_2. Solving for the optimal value of j_2 depends on the selling mechanism considered.

The three selling mechanisms belong to a class for which

$$\frac{\partial f(j_1)}{\partial j_1} = \frac{\partial g(j_2)}{\partial j_2} = 0$$

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

There exists however 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).

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, $\left.\frac{\partial \bar{\pi}_1(f(j_1))}{\partial f(j_1)}\right|_{j_1=\frac{6k-9}{14}} \neq 0$ as this function has no interior solution.

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 4

Data collection

We compare the first derivative of the profits of the data intermediary in the different mechanisms in order to compare the optimal amounts 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 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 $i = 1, .., j$, and:*

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

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

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

We need to compute demands in order to find consumer surplus. On the j segments of size $\frac{1}{k}$ where Firm 1 has information, it is a monopolist and 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:

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

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 5

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:

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

- When 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 5. ■

J Proof of Proposition 6

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_2^{both}. Thus, the subgame perfect equilibrium is characterized by the following strategies: Firm 1 bids the reserve price w_1 for j_1^{both}, and Firm 2 bids the reserve price w_2 for j_2^{both}. We will show in Appendix K that in equilibrium partitions are symmetric: $j_1 = j_2$. The data intermediary will set in the two auctions reserve prices equal to the willingness to pay of each firm $p_{both} = w_1 = w_2$.

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}.

Under the take it or leave it mechanism, the data intermediary proposes to each firm j_1^{both} segments of information at price p_{both}. Let $\bar{\pi}_1(j_1^{both})$ denote the profit of Firm 1 without information but facing Firm 2 informed with j_1^{both}. The only subgame perfect equilibrium is a situation in which both firms purchase information at price $p_{both} = \pi_1(j_1^{both}) - \bar{\pi}_1(j_1^{both})$ (firms have no incentives to deviate from this equilibrium since, by doing so, they would become uninformed but facing an informed competitor). Thus the profit of the data intermediary when selling information to both firms is $\Pi_{both}(k) = 2p_{both} - c(k)$. ■

K Proof of Propositions 7, 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 write in step 1 prices and demands, in step 2 we give the profits, and solve for prices and profits in equilibrium in step 3.

Step 1: prices and demands.

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

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

p_2 is the price set by Firm 2 on interval $[0, \frac{j'}{k}]$ 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:

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

Step 3: prices, demands and profits in equilibrium.

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

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

Replacing these values in the above demands and prices gives:

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

and

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

Profits are:

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

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

The data intermediary maximizes the following profit function:

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

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

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

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

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

$$\Pi_{both}(k) = \frac{2t}{11} - \frac{6t}{11k} + \frac{9t}{22k^2} - c(k),$$

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

$$\frac{(6k-9)}{11k^3} - c'(k).$$

Finally, consumer surplus in this case is

$$\frac{(445k^2 + 216k + 36)t + 484Vk^2}{484k^2}.$$

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

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 $j_1^{a_2}$ to Firm 1 and $j_2^{a_2}$ to Firm 2 are:

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

We show that in equilibrium $j_1^{a_2} = j_2^{a_2}$.

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

- Either $j_1^{a_2} > j_2^{a_2}$, and $\pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2})$ increases when $j_2^{a_2}$ increases.

- Or $j_1^{a_2} < j_2^{a_2}$, and $\pi_2(j_2^{a_2}) - \bar{\pi}_2(j_1^{a_2})$ increases when $j_1^{a_2}$ increases

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

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

FOC on p_{a_2} with respect to $j_1^{a_2}$ gives us:

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

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

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.

We can now derive profits, consumer surplus and data collection in equilibrium. The price of information can be written

$$p_{a_2} = \pi_1(j_1^{a_2}) - \bar{\pi}_1(j_1^{a_2}).$$

FOC on p_{a_2} with respect to $j_1^{a_2}$ gives us:

$$\frac{4k-3}{6},$$

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

and

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

The ranking of profits, surplus, and optimal data collection is then straightforward. ■

M Proof of Proposition 11

See the proofs of Propositions 4 and 5.

N Characterization of the equilibrium with price caps

We prove that data collection decreases when the price cap decreases. We note $\overline{p}$ 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 $\overline{p}$.
- (b) The data intermediary will sell information to both firms if $\overline{p} \leq 2p_{both}$.

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

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

The optimal value of k is such that $p(k^*) = \overline{p}$. 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 Π by increasing k.

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

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

If $\overline{p} \in [p_a, p_{seq}[$, the data intermediary uses auction as it is the only selling mechanism allowing to reach the highest price possible, $\overline{p}$.

If $\overline{p} \in [p_{seq}, p_{tol}[$, auction and sequential bargaining both allow to set the highest price possible, and the data intermediary will chose either mechanism indifferently.

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