ECARES

Market for Information and Selling Mechanisms

David Bounie
Telecom Paris

Antoine Dubus
ECARES, Université libre de Bruxelles

Patrick Waelbroeck
Telecom Paris

Market for Information and Selling Mechanisms[*]

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

April 3, 2020

Abstract

We investigate the strategies of a data intermediary selling consumer information to firms for price discrimination purpose. We analyze how the mechanism through which the data intermediary sells information influences how much consumer information she will collect and sell to firms, and how it impacts consumer surplus. We consider three selling mechanisms tailored to sell consumer information: take it or leave it, sequential bargaining, and auctions. We show that the more information the intermediary collects, the lower consumer surplus. Consumer information collection is minimized, and consumer surplus maximized under the take it or leave it mechanism, which is the least profitable mechanism for the intermediary. We discuss two regulatory tools – a data minimization principle and a price cap – that can be used by data protection agencies and competition authorities to limit consumer information collection, increase consumer surplus, and ensure a fair access to information to firms.

1 Introduction

Economists have for a long time argued that markets are more efficient with more information (Hayek, 1945). Today, with the use of digital technologies and the organization of a market for information dominated by data intermediaries, the amount of information collected on consumers and sold to firms has considerably increased (Bergemann and Bonatti, 2018). Yet, more and more economists and legal scholars are now keen to reconsider the beneficial effects of information (Crémer et al., 2019), and in just a few years, many regulators around the world have come to oversee information-collection strategies to protect consumers.[1]

The objective of this paper is to investigate the role of the selling mechanism on the strategies of information collection by data intermediaries, a central but overlooked topic in the economics literature. Data intermediaries are major actors of the market for information. They collect information on consumers from various sources, and sell recombined information to firms seeking to improve their business practices through better analysis of markets, forecasting trends, and personalized ads, products, and prices. In a study of nine data brokers in 2014, the Federal Trade Commission found that data brokers have information "on almost every U.S. household and commercial transaction. [One] data broker's database has information on 1.4 billion consumer transactions and over 700 billion aggregated data elements; another data broker's database covers one trillion dollars in consumer transactions; and yet another data broker adds three billion new records each month to its databases." (Federal Trade Commission, 2014, Data brokers: A Call for Transparency and Accountability).

Data intermediaries sell information that differs radically from traditional goods on three dimensions. First, information can be recombined and repackaged by data intermediaries to fit with each firm on a market. These strategies have been analyzed by Iyer et al. (2005) for targeted advertising, by Bounie et al. (2018) for the sale of consumer data, and by Shapiro and Varian (1998) for the

---

[1]For example, the California Consumer Privacy Act provides a detailed list of safeguards to protect personal data. Similarly, a (personal) data minimization principle is enacted in the Health Insurance Portability and Accountability Act in the US, and in the General Data Protection Regulation in Europe.

sale of information goods. Secondly, data intermediaries strategically choose the precision of information that they collect and sell to firms. Bergemann and Bonatti (2015) and Bergemann et al. (2018) analyze how a data intermediary chooses the precision of information to maximize surplus extraction from firms. Bergemann and Bonatti (2019) also show that data intermediaries may have interest not to sell perfect signals to firm. In a competitive set up, Casadesus-Masanell and Hervas-Drane (2015) and Gal-Or et al. (2018) analyze how data intermediaries choose the amount of consumer data that they use when consumers are concerned with their privacy. Bergemann et al. (2020) also analyze the impact of consumer's privacy on the selling strategies of a data intermediary. Thirdly, there is a negative externality for a firm not to acquire information, as she can face an informed competitor, which increases competition on the market. Montes et al. (2018) analyze this negative externality and show that data intermediaries have interest to deal exclusively with only certain firms in the market.

Beyond the characteristics of information, the way data intermediaries sell information is also of utmost importance, and can be related to the central question of how to sell a product (Riley and Zeckhauser, 1983; Katz and Shapiro, 1986; Bulow and Klemperer, 1994). This literature has focused in particular on three selling mechanisms which are take it or leave it (Binmore et al., 1986; Bergemann and Bonatti, 2019), sequential bargaining (Rubinstein, 1982; Sobel and Takahashi, 1983), and auctions (Vickrey, 1961; Klemperer, 1999; Jehiel and Moldovanu, 2000; Figueroa and Skreta, 2009). Recently, Backus et al. (2018a), Backus et al. (2018b) and Backus et al. (2019) have empirically revisited the question of the optimal selling mechanism. Milgrom and Tadelis (2018) analyze how machine learning can be used to improve mechanism design. Jindal and Newberry (2018) study in which case it is optimal for a seller to use bargaining or fixed price to sell a good.

Building on the previous literature, we analyze how the three selling mechanisms examined in the literature – auctions, take it or leave it and sequential bargaining – and commonly used in the industry,[2] are critical to understand the

---

[2]For further details about the implementation of the three selling mechanisms, the reader will consult for auctions (First price Auction, Second price, and the Header-Bidding, Smartyads, February 2018.), for take it or leave it offers (Nielsen), and for repeated interactions leading to sequential bargaining (Facebook blocks valuable ad data in privacy update to its marketing

amount of information collected on consumers. Selling mechanisms have indeed an impact on the data collection strategies of the data intermediary since they will impact the price of information, i.e. what firms are willing to pay for information. The price of information has two components: the profits of a firm with and without information. These two elements will drive the strategy of data collection of the data intermediary. First, firms are ready to pay more for high quality relevant data that will increase profits through better consumer surplus extraction. This will be referred to as the rent extraction effect. Secondly, without information a firm might have to compete against a firm that has acquired information. Thus when deciding whether to purchase information or not, a firm may face a negative externality depending on the selling mechanism, and this negative externality increases with the precision of information. Indeed, consider a take it or leave it: the data intermediary proposes information to a firm, and in case the firm declines, all firms on the market remain uninformed. Now consider an auction with negative externality: if the firm declines the offer, the data intermediary sells information to the competing firms. 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 she has to compete with a better informed competitor. A data intermediary thus may have interest to collect more information under the auction mechanism where there is a negative externality than under the take it or leave it mechanism where more precise information does not change the outside option. Thus, different selling mechanisms will change how much information will be collected and sold to firms. This article investigates this important issue.

By relating the data collection strategies of data intermediaries to the mechanism used by data intermediaries to sell consumer information, we contribute to the literature on two aspects. First, we show that the three mechanisms – sequential bargaining, take it or leave it and auctions – share the same theoretical properties. They belong to a class of partition in which the same number of consumer segments is sold in equilibrium. However, different selling mechanisms in

partner program, AdAge, February 21 2020.).

this class will lead to different amounts of data collected: the willingness to pay of a firm for information depends on the threat of being uninformed, while the competitor is potentially informed depending on the selling mechanism. Consider again an auction mechanism: the losing bidder faces an informed competitor, and is thus willing to pay much more than with a take it or leave it, where a firm declining the offer faces an uninformed competitor. Knowing this, the data intermediary will increase the rent extraction effect with auctions by increasing the amount of consumer data collected. Thus the selling mechanism will influence the data collection strategy of the data intermediary.

Secondly, by comparing the amount of consumer data that intermediaries collect under different selling mechanisms, our results shed new lights on divergences between private actors of the market for information and regulators, especially data protection agencies and competition authorities. We show that the data intermediary prefers the selling mechanism that maximizes its profits, but that does not necessarily minimize the amount of data collected, nor maximizes consumer surplus. On the contrary, a data protection agency prefers a mechanism that minimizes data collection, while a competition authority prefers a selling mechanism that maximizes competition on the market, and ensures a fair access of information to firms. We show that the data intermediary prefers selling information through auctions or sequential bargaining, which increases data collection, and in turn lowers consumer surplus. On the contrary, data protection agencies and competition authorities prefer the take it or leave it mechanism that maximizes consumer surplus and minimizes data collection.

The remainder of the article is organized as follows. We describe the model in Section 2. Three selling mechanisms – take it or leave it, sequential bargaining and auctions – have been used extensively by the industry and in the theoretical literature, and we will show that they share similar properties; we present them in Section 3. We characterize the three selling mechanisms and show that they belong to a broader class of selling mechanisms, defined as independent offers in Section 4. In Section 5, we analyze how the price of information is related to the amount of data collected. We then analyze another class of mechanisms in Section 6: the data intermediary proposes to firms information partitions that

are symmetric, and we prove that this mechanism is equivalent to second price auctions. Following, in Section 7 we analyze how selling information to two firms modifies market outcomes. We discuss regulatory implications, and how to use a data minimization principle and a price cap as regulatory tools in Section 8. Section 9 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.[3]   The data intermediary collects and sells data that segment consumers on the Hotelling line. A firm that acquires consumer segments, i.e. an informed firm, can set a price on each consumer segment. On the contrary, a firm that cannot distinguish consumer segments is uninformed 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:[4]

---

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

[4]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).

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

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 can then choose the optimal information partition to sell to firms.[5]

2.2.1 Collecting consumer data

We consider a data intermediary that collects k consumer segments at a cost $c(k)$.[6] The cost of collecting information encompasses various dimensions of the activity of the data intermediary such as installing trackers, or storing and handling data. Collecting more information, by increasing the number of segments, allows a firm to locate consumers more precisely. We will show that thinner segments increase the value of information. 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.

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

2.2.2 Selling information

We assume that the data intermediary sells information to only one firm, say Firm 1.[7] We consider three mechanisms that can be used by the data intermediary to sell information: take it or leave it, sequential bargaining, and auctions. We focus on these three mechanisms as they have been extensively studied in the literature, and they are well suited for the sale of information, which is a central characteristic of the data intermediary industry.

The data intermediary can potentially sell 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. It is easy to understand that selling all consumer segments is not optimal for the data intermediary. On the one hand, thinner segments in the partition allows a firm to extract more surplus from consumers. This is the surplus extraction effect that increases the value of information. On the other hand, selling more consumer segments also 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.

We focus our analysis on partitions that maximize the profits of 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 $[0, \frac{j_1}{k}]$ where consumers are identified so that Firm 1 can price discriminate them.[8] The data intermediary does not sell information on consumers in the second interval of size $1 - \frac{j_1}{k}$, who remain unidentified, and firms charge a uniform price on this second interval. The number of segments of identified consumers j_1 depends on the total number of segments on the market k. We denote by $j_1(k)$

---

[7]Selling information to both firms is in general not optimal because it increases the competitive pressure on the product market (Montes et al., 2018; Bounie et al., 2018), and thus lowers the profits of the data intermediary, who extracts part of the surplus of the firms. We characterize the equilibrium when the data intermediary sells information to both firms in Section 7.

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

the number of segments as a function of k. This partition is optimal as it balances the surplus extraction effect of information while limiting the competitive effect of information. On the one hand, by identifying consumers close to Firm 1, such partition allows Firm 1 to extract surplus from consumers who have a high willingness to pay. Indeed, selling segments greater 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 $\frac{1}{k}$.[9] 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 $\mathcal{P}_1$.

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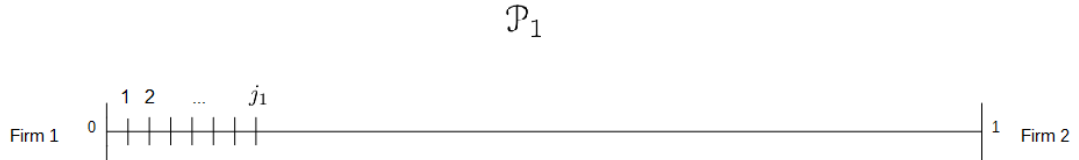
Figure 2: Selling partition $\mathcal{P}_1$ to Firm 1

This model of segmentation of consumer demand has to main benefits. First, we can analyze the strategies of data collection of the data intermediary by analyzing the changes of the precision of information k in equilibrium. Secondly, this specification allows the data intermediary to sell only partial information up to $\frac{j_1}{k}$.[10]

2.3 Firms

A firm may decide not to acquire information, and in this case she only knows that consumers are uniformly distributed on the unit line. When Firm 1 acquires $j_1(k)$ segments of information, she can price discriminate consumers on these segments.

---

[9]The formal proof is however quite involved. See Bounie et al. (2018) for more details.

[10]This contrasts with Montes et al. (2018) and others who assume that all information is sold or no information at all.

Firm 1 sets prices in two stages.[11] First, Firm 1 sets price p_1 on the segment where she competes directly with Firm 2 (the competitive segment). Secondly, Firm 1 sets a price on each segment where she is in a monopoly position, with p_{1i} being the price on the ith segment from the origin. Firm 2 is uninformed but, similarly to the Hotelling framework without information, Firm 2 knows the price p_1 set by Firm 1 on the competitive segment, and sets a price p_2 on the whole unit line.

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

We summarize the timing of the game. The data intermediary first collects data and sells the partition $\mathcal{P}_1$ to Firm 1. Then Firms 1 and 2 set prices on segments where they compete. Finally, Firm 1 sets prices on the segments where she is a monopolist.

- Stage 1: the data intermediary collects data on k consumer segments.

- 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.

---

[11]Sequential pricing decision avoids the nonexistence of Nash equilibrium in pure strategies, and is common in the literature supported by managerial practices. For instance, Acquisti and Varian (2005) use sequential pricing to analyze intertemporal price discrimination with incomplete information on consumer demand. Jentzsch et al. (2013) and Belleflamme et al. (2019) 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.

- Stage 3: firms set prices p_1 and p_2 on the competitive segments.

- Stage 4: Firm 1 price discriminates consumers where she is in a monopoly position by setting p_{1i}, $i \in [1, j_1(k)]$.

The subgame perfect equilibrium is solved by backward induction. In stage 4, Firm 1 sets prices $p_{11}, .., p_{1j_1}$ on segments where she is in a monopoly position. In stage 3, Firm 1 and Firm 2 set prices p_1 and p_2 on the competitive segments. In stage 2, we characterize the strategies of the data intermediary regarding how much consumer information to sell to Firm 1. In stage 1, we determine how much data the intermediary collects in equilibrium.

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 willingness to pay of the firms for information. We describe in Section 3 the three selling mechanisms that we analyze in this article, and we show in Sections 4 and 5 how the data collection and information selling strategies of the data intermediary are affected by the selling mechanism.

3   Selling mechanisms

We analyze three mechanisms that have been extensively studied in the literature: take it or leave it, sequential bargaining, and auctions. First, under 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. This approach has been studied by Binmore et al. (1986), and used by Admati and Pfleiderer (1986) to model markets for financial information. 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. This type of dynamic games has been studied for instance by Rubinstein (1982) or Sobel and Takahashi (1983). The third selling mechanism is an auction with negative externality used by Montes et al. (2018) and by Bounie et al. (2018).

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, while both firms are uninformed in the take it or leave it mechanism.

The three selling mechanisms have a major impact on the strategies of the data intermediary and on the value of information. Indeed, for each selling mechanism, what a firm is ready to pay for information depends on its outside option if she does not purchase information. In the remainder of this section, we compute the value of information of Firm 1 under the three selling mechanisms, and we show that the outside option can be used as a threat by the data intermediary to extract more surplus from Firm 1.

We introduce further notations. We denote by $\pi_1(j_1)$ the profit of Firm 1 when she 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 profits π. In the sequential bargaining and auction formats, Firm 2 has information when Firm 1 is uninformed. 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.

Finally, 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).

Definition 1 : *A couple of information partitions is a pair (j_1, j_2).*

We will show in Theorem 1 that the three selling mechanisms belong to a specific class of information partitions.

3.1 Take it or leave it

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 that we describe in Section 3.2.[12]

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:

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

3.2   Sequential bargaining

Under the sequential bargaining mechanism, the data intermediary proposes information to each firm sequentially, in an 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}).

To compute the value of information under the sequential bargaining mechanism, we characterize a stationary equilibrium of this game where Firm 1 is making profit $\pi_1(j_1^{seq})$ if she accepts the offer, but makes profits $\bar{\pi}_1(j_2^{seq})$ if she declines the offer and Firm 2 purchases information. It is important to stress that when Firm 1 declines the offer, Firm 1 will compete with Firm 2 to which the data intermediary will optimally propose the symmetric partition (j_2^{seq} is the symmetric of j_1^{seq} with respect to $\frac{1}{2}$). Thus the outside option for Firm 1 is a situation where Firm 2 is

---

[12]The take it or leave it mechanism includes in fact many such mechanisms where there is no possibility for renegotiation, including Nash bargaining and menu pricing.

informed with its profit maximizing information structure. We show in Appendix B that the data intermediary sets the price of information to:

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

3.3 Auctions

The data intermediary auctions simultaneously two information partitions: j_1^a is auctioned in auction 1 with a reserve price p_a[13], and the reference partition $\mathcal{P}_k$ that includes all k information segments is auctioned in auction 2.[14] The two partitions are auctioned simultaneously but only the partition with the highest bid will be sold.[15] The data intermediary can design partitions that maximizes the willingness to pay of Firm 1 by leveraging on its outside option. 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 $\mathcal{P}_k$ 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 $(j_1^a, \mathcal{P}_k)$. The data intermediary sets the price of information to:

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

---

[13]Without reserve price firms have incentives to underbid.

[14]There is no reserve price in auction 2 to satisfy the participation constraint

[15]This selling mechanism is introduced in Bounie et al. (2018). It allows the data intermediary to reach first-best profits. The role of optimal threat in auctions is analyzed in Figueroa and Skreta (2009).

4 Number of segments sold in equilibrium

In this section, we characterize the number of consumer segments sold to Firm 1 for each of the three selling mechanisms. We first establish that for a given value of k, i.e. the same number of consumer segments collected, the number of consumer segments sold by the data intermediary is the same for the three selling mechanisms (Proposition 1). The three selling mechanisms have the property that the information proposed to Firm 2 is independent of the information proposed to Firm 1. We then show that take it or leave it, sequential bargaining and auctions belong to a class of mechanisms that we refer to as independent offers. Theorem 1 generalizes Proposition 1 for independent offers.

4.1 Number of segments sold in equilibrium

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

Proposition 1

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 C.

The proof of Proposition 1 is based on the property that the data intermediary optimizes j_1 and j_2 independently. 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. It is the case for the take it or leave it and the auction mechanisms. Under the take it or leave it mechanism, Firm 1 has no information when she declines the offer, and thus its outside option is independent of the information structure proposed by the data intermediary to Firm 2. Under 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

15

is affected by the partition proposed to Firm 2 is independent of the partition proposed to Firm 1. Under 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 structure 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.

4.2 Independent offers

Using the intuition developed in the previous section, we can generalize Proposition 1 to 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. Theorem 1 shows that for a given amount of data collected k, selling mechanisms characterized by such independent offers lead to the same number of consumer segments sold to Firm 1 (j_1^*). Let (j_1, j_2) be the couple of partitions proposed to Firm 1.

Definition 2 **(Independent offers)**

A couple of information partitions (j_1, j_2) is independent if:

$$\frac{\partial j_1}{\partial j_2} = \frac{\partial j_2}{\partial j_1} = 0.$$

Definition 2 includes a large set of selling mechanisms 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, under 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.

Theorem 1 shows that for a given k, all selling mechanisms satisfying Definition 2 lead to the same number of consumer segments sold by the data intermediary.

Theorem 1

Consider s and s', two selling mechanisms that satisfy Definition 2:

$$\forall \ k, \quad j_1^{s*}(k) = j_1^{s'*}(k).$$

Proof: See Appendix C.

Theorem 1 has theoretical and practical implications. First, Theorem 1 is a new result in the literature and characterizes the properties of information partitions based on the amount of information sold in equilibrium.

Secondly, 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, under 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.

Thirdly, Theorem 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 Theorem 1 will always lead to the same number of consumer segments sold to Firm 1. Thus a competition authority can analyze the properties of the couple of partitions to determine whether an action is required to limit the amount of information sold on a market.

Finally, we will show in Section 6 that there are selling mechanisms that do not satisfy Definition 2. For instance, the data intermediary can auction symmetric partitions to Firm 1 and Firm 2. The information partition proposed to Firm 1 thus appears in the outside option of Firm 1 if she 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.

To conclude, we have shown in this section that the number of consumer segments sold to Firm 1 does not vary with the three selling mechanisms satisfying

Definition 2. In the next section, we analyze how the amount of data collected varies under the three selling mechanisms.

5   Amount of data collected in equilibrium

We analyze in this section how the different selling mechanisms impact the profits of the data intermediary, the number of consumer segments collected (k), and consumer surplus. 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 structure to firms under 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.[16] This is because the outside option is different under different selling mechanisms.

The profit of the data intermediary $\Pi \in \{\Pi_{tol}, \Pi_{seq}, \Pi_a\}$ is given by the price of information $p \in \{p_{tol}, p_{seq}, p_a\}$, net of the cost of data collection $c(k)$: $\Pi(k) = p(k) - c(k)$.[17]

We have established in Proposition 1 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 2 compares the number of segments collected by the data intermediary and consumer surplus under the three selling mechanisms.

---

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

[17]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.

Proposition 2

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 D.

Proposition 2 shows that the number of consumer segments collected is minimized under 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 she remains uninformed. Thus, information collection is minimized under this selling mechanism, the surplus 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 with information. Secondly, the negative externality for an uninformed firm that faces an informed competitor in 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, following Proposition 1, the difference of information precision in the two selling mechanisms is driven only by the outside option. The marginal gain of more precise information is higher under the sequential bargaining mechanism than under auctions. This is because the marginal effect of more precise information on the outside option is higher under sequential bargaining than under 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 under sequential bargaining. Proposition 2 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 surplus extraction and minimizes the competitive effect of information.

Proposition 3 shows that the data intermediary prefers auctions, and that the take it or leave it is the least profitable selling mechanism.

Proposition 3

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

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

Proof: see Appendix E.

Under 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, under 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.

6 Alternative selling mechanism: symmetric partitions

In the previous sections, we have analyzed three selling mechanisms for which information partitions j_1 and j_2 are chosen independently by the data intermediary, and satisfy Definition 2. In this section, we study an alternative selling mechanism named symmetric partitions that does not satisfy Definition 2: the information partitions sold by the data intermediary are not chosen independently, they are symmetric with respect to $\frac{1}{2}$ and thus j_1 and j_2 are equal. Analyzing such a selling mechanisms is particularly interesting as symmetric partitions could be required by a regulatory framework to enforce fair and equal access to data for firms (see

Section 8). Furthermore, as we will show in this section, symmetric partitions are equivalent to second price auctions, a selling mechanism that has attracted a lot of interest and debates in the online ad industry (see Section 8).

In symmetric partitions, the data intermediary simultaneously proposes partition j_1^{sym} to Firm 1 and a symmetric partition j_2^{sym} to Firm 2. The price of information p_{sym} can be written as follows: $p_{sym} = \pi_1(j_1^{sym}) - \bar{\pi}_1(j_2^{sym} = j_1^{sym})$. The partitions do not satisfy Definition 2 since j_1^{sym} appears in the outside option of Firm 1. The data intermediary is constrained in the choice of its optimal threat to Firm 1 when she declines the offer, since the data intermediary can only choose symmetric partitions.

We establish in Proposition 4 that symmetric partitions and second price auctions are equivalent in the sense that they lead to the same market outcomes. In second price auctions, 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 $j_2^{a_2}$.

Proposition 4

Symmetric partitions and second price auctions are equivalent, and lead to the same market outcomes:

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

$$(b) \quad \Pi_{sym} = \Pi_{a_2},$$

$$(c) \quad k_{sym} = k_{a_2},$$

$$(d) \quad CS_{sym} = CS_{a_2}.$$

Proof: See Appendix F.

To prove Proposition 4, we show that the 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: $j_1^{a_2*} = j_2^{a_2*}$. Since there is only one optimal value for j_1^{sym*}, Proposition 4 is established.

We can now compare profits, consumer surplus, and the amount of data collected with symmetric partitions (or second price auctions), with the outcomes of the three other selling mechanisms.

Proposition 5

The equilibrium with the symmetric offers mechanism has the following properties:

$$(a) \quad \Pi_a > \Pi_{sym} > \Pi_{seq} > \Pi_{tol}$$

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

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

Proof: see Appendix G.

First, the take it or leave it mechanism still minimizes the number of consumer segments collected. Secondly, the data intermediary would prefer the auction mechanism as it leads to the highest willingness to pay of Firm 1. Thirdly, consumer surplus is now maximized in the symmetric offers mechanism. This is because the data intermediary sells more consumers segments under second price auctions compared with the three other selling mechanisms: $j_1^{sym*} > \frac{6k-9}{14}$.

Proposition 5 also sheds light on the ongoing debate in the online ads industry, on the use of first or second price auctions.[18] The amount of consumer data

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

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 only for Firm 1 under first price auction, while both firms have access to symmetric partitions under 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.

To sum-up, we have identified another class of selling mechanism where partitions proposed to both firms are perfectly correlated and symmetric, and that does not call into question the results established in the previous section.

7 Selling information to both firms

Up to now, we have focused 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. The three selling mechanisms that we analyze are identical 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 prove in Appendix H that partitions are symmetric, and

that the data intermediary sets reserve price $p_{both} = w_1 = w_2$ in the two auctions.

Under sequential bargaining, the problem is simplified by the fact that there is no discount factor, and therefore no first mover advantage. In this situation, the data intermediary 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} (we prove that partitions are symmetric in Appendix H). 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 $\Pi_{both}(k) = 2p_{both} - c(k)$.

We characterize in Proposition 6 the profit of the data intermediary, the number of consumer segments collected, and consumer surplus when the data intermediary sells information to both firms.

Proposition 6

When the data intermediary sells information to both firms, the equilibrium has the following properties:

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

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

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

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

Proof: see Appendix I.

Proposition 6 shows that the data intermediary prefers to sell information to one firm under the sequential bargaining and the auction mechanisms, but prefers to

sell information to both firms under the take it or leave it mechanism. The intuition behind this result is simple. 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, in the take it or leave it mechanism, the data intermediary cannot threaten the prospective buyer. Therefore the data intermediary prefers to sell information to both firms using the take it or leave it mechanism, while he only sells information to one firm in the auctions and sequential bargaining mechanisms. All results of Sections 4 and 5 hold when the data intermediary is allowed 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 sequential bargaining and auction mechanisms.

8   Regulatory implications and policy guidelines

We analyze in this section the implications of our results for the regulation of the market for consumer information. The data intermediary and regulators have conflicting views over which selling mechanism to use for two reasons. First, Propositions 2 and 3 highlight a divergence of interests between the private sector and the regulator. Indeed, 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. While enforcing a specific selling mechanism is a particularly hard task to do for a regulator, we propose two regulatory tools that allow a regulator to reach the market outcomes of a take it or leave it mechanism, therefore minimizing data collection and maximizing 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 a price cap that has been recently proposed by Rey and Tirole (2019).

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.[19] 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. We show in Section 8.2 that price caps can force the data intermediary to sell information to both firms, and thus ensure a fair and equal access to data.

8.1 Data minimization principle

A data protection agency can set a limit $\bar{k}$ over the amount of consumer data collected by a data intermediary. Proposition 7 provides the implications for market equilibrium of a change in the maximal amount of consumer data that the intermediary can collect.

Proposition 7

- (a) The selling mechanism does not change with $\bar{k}$.

- (b) The lower the value of $\bar{k}$ the higher consumer surplus.

Proof: See Appendix J

Proposition 7 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

---

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

extraction from consumers. The results of Proposition 3 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 a price cap can be used to force the data intermediary to sell information to both firms, thus allowing fair competition between firms.

8.2 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. We note $\overline{p}$ the highest price of information allowed by the regulator.

Proposition 8

- *(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 $\overline{p} \leq 2p_{both}$.*

Proof: See Appendix K

Proposition 8 (a) 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 of data protection agencies since the amount of data collected increases with the value of the cap. Proposition 8 (b) 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. When the price cap is below $2p_{both}$, the data intermediary sells (symmetric) information to all firms, regardless of the selling mechanism. In other words, lowering the price cap reduces the amount of consumer data collected, and setting the price cap below $2p_{both}$ increases market competition and consumer surplus, and guarantees fair competition between firms.

9   Conclusion

Our model of data intermediary that collects and sells consumer information has 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 for all selling mechanisms to the exception of the take it or leave it mechanism. The fact that only one firm has information in equilibrium decreases consumer surplus compared to a situation where both firms are informed. More information on the market could be enforced by regulation to guarantee a level playing field, for instance price caps and symmetric sales. 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 web giants and data brokers.[20]

Secondly, our results show that the selling mechanism used by the data in-

---

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

termediary will lead to different levels of data protection and privacy protection. 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 vary with the precision of information since no firm is informed. Information collection is maximized, and consumer surplus minimized under sequential bargaining. These new results can be of interest for data protection agencies concerned with the amount of personal data collected by firms. The data collection strategies will depend on the enforceability of the outside option, which depends on the competitive and the regulatory environments in which data intermediaries operate. If intermediaries have more market power (in the data brokerage industry, Bluekai and Datalogix were acquired by Oracle,[21] and Equifax and FICO have agreed to share information on consumer financial data),[22] there could be an increase in the price of information and in the amount of consumer data collected.[23]

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 competition (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 such that more information on the market means less competition. Therefore, the three selling mechanisms that we analyzed – take it or leave it,

---

[21]AdExchanger, How Datalogix Made Oracle's BlueKai Acquisition Even Smarter, January 19, 2015.

[22]The Wall Street Journal, Equifax, FICO Team Up to Sell Consumer Data to Banks, March 27, 2019.

[23]Regulators and legislators have recently analyzed the impacts of data brokers on markets (Crain, 2018).

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 achieves both goals of data protection agencies willing to minimize data collection, and of competition authorities who want to maximize consumer surplus. However, this inverse relationship is not verified with other selling mechanisms such second price auctions, and the missions of data protection agencies and of competition authorities are not aligned: second price auctions maximize consumer surplus but do not minimize data collection. Understanding the theoretical properties of selling mechanisms is therefore essential to establish whether there is a conflict of interest between data protection agencies and competition authorities.

References

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

Anat R Admati and Paul Pfleiderer. A monopolistic market for information. *Journal of Economic Theory*, 39(2):400–438, 1986.

Matthew Backus, Thomas Blake, Bradley Larsen, and Steven Tadelis. Sequential bargaining in the field: Evidence from millions of online bargaining interactions. Technical report, National Bureau of Economic Research, 2018a.

Matthew Backus, Thomas Blake, and Steven Tadelis. Communication and bargaining breakdown: An empirical analysis. *NBER Working Paper*, 2018b.

Matthew Backus, Thomas Blake, and Steven Tadelis. On the empirical content of cheaptalk signaling: An application to bargaining. *Journal of Political Economy*, 127(4): 000–000, 2019.

Paul Belleflamme, Wynne Lam, Wing Man, and Wouter Vergote. Competitive imperfect price discrimination and market power. 2019.

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

Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. 2018.

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

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

Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. 2020.

Ken Binmore, Ariel Rubinstein, and Asher Wolinsky. The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, pages 176–188, 1986.

David Bounie, Antoine Dubus, and Patrick Waelbroeck. Selling strategic information in digital competitive markets. Technical report, CESifo Working Paper, 2018.

Jeremy Bulow and Paul Klemperer. Auctions vs. negotiations. Technical report, National Bureau of Economic Research, 1994.

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

Matthew Crain. The limits of transparency: Data brokers and commodification. *New Media & Society*, 20(1):88–104, 2018.

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

Nicolás Figueroa and Vasiliki Skreta. The role of optimal threats in auction design. *Journal of Economic Theory*, 144(2):884–897, 2009.

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

Esther Gal-Or, Ronen Gal-Or, and Nabita Penmetsa. The role of user privacy concerns in shaping competition among platforms. *Information Systems Research*, 2018.

Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.

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

Philippe Jehiel and Benny Moldovanu. Auctions with downstream interaction among buyers. *Rand journal of economics*, pages 768–791, 2000.

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

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

Michael L Katz and Carl Shapiro. How to license intangible property. *The Quarterly Journal of Economics*, 101(3):567–589, 1986.

Paul Klemperer. Auction theory: A guide to the literature. *Journal of economic surveys*, 13(3):227–286, 1999.

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

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

Paul R Milgrom and Steven Tadelis. How artificial intelligence and machine learning can impact market design. Technical report, National Bureau of Economic Research, 2018.

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

Patrick Rey and Jean Tirole. Price caps as welfare-enhancing coopetition. *Journal of Political Economy*, 2019.

John Riley and Richard Zeckhauser. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, 98(2):267–289, 1983.

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

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

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

Lars A Stole. Price discrimination and competition. *Handbook of industrial organization*, 3:2221–2299, 2007.

Jacques-Francois Thisse and Xavier Vives. On the strategic choice of spatial price policy. *The American Economic Review*, pages 122–137, 1988.

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

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

A Mathematical interpretation of Assumption 1

The cost function is defined such that:

$$\begin{cases} \frac{\partial^2 [p(k)-c(k)]}{\partial k^2} < 0 \ \text{ and } \ \exists! \ k^* \ \text{ s.t. } \ \frac{\partial [p(k)-c(k)]}{\partial k} = 0 \\ \exists! \ k^* \ \text{ s.t. } \ \frac{\partial \Pi}{\partial k} = 0 \ \text{ and } \ \Pi(k^*) \geq 0 \\ c(0) = 0 \end{cases}$$

These technical hypothesis are common in the literature. It allows profits to be maximized in a unique point, which is usually true for linear cost functions.

B Proof of optimal prices in sequential bargaining

We propose a candidate equilibrium policy function. We show that $p_{seq} = \pi_1(j_1^{seq}) - \bar{\pi}_1(j_2^{seq})$ is an SPE. As only the data intermediary has a non binary choice, uniqueness will result naturally.

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

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

Thus, the overall value of Firm 1 is:

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

Thus:

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

The data intermediary has no interest in deviating from this value, 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 SPE of this game.

C Proof of Theorem 1 and Proposition 1

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

The data intermediary profit functions in the different timings 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 structure in equilibrium $\mathcal{P}_1^*$ does not depend on the selling mechanism.

Prices and demands on the unit line are identical to Bounie et al. (2018) and can be written as follow:

$p_1 = t[1 - \frac{4}{3}\frac{j}{k}]$; $p_{1i} = 2t[1 - \frac{i}{k} - \frac{1}{3}\frac{j}{k}]$; $d_1 = \frac{1}{2} - \frac{2}{3}\frac{j}{k}$.

Profits are:[24]

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

Thus, first order conditions on π_1 gives us

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

There exists however selling mechanisms that do not satisfy Definition 2 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 satisfies Definition 2, that is, there exist j_1 and j_2 that are not 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:

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

Thus, either

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

or

---

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

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

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 $j_2 = \frac{j_1^2}{2} - j_1$, and the number of segments sold in equilibrium is $j_1^*(k) = \frac{6k-9}{14}$.

D   Proof of Proposition 2

Data collection

We compare the first derivative of the profits of the data intermediary in the different mechanisms in order to compare the optimal precisions 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 given in Bounie et al. (2018) and are as follow:

- *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, she 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:

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

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

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

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

information precision.

E Proof of Proposition 3

We compare the profits of the data intermediary in the different selling mechanisms. The profits of the firms depending on the information structure are provided in Bounie et al. (2018):

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

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

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

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

Profits 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 3.

F Proof of Proposition 4

We prove that information structures are identical in both alternative mechanisms. It is then straightforward that prices, profits, and data collection in equilibrium are identical. In the symmetric offers mechanism, the price of information can be written

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

In the second price auction mechanism, the willingness to pay of firms when the data intermediary proposes information j_1^{alt} to Firm 1 and j_2^{alt} to Firm 2 are:

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

We show that in equilibrium $j_1^{alt} = j_2^{alt}$ and thus that the optimization problem is identical to the one in symmetric offer mechanism.

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

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

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

Thus the data intermediary chooses $j_1^{alt} = j_2^{alt}$.
FOC on p_{alt} with respect to j_1^{alt} gives us:

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

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

and

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

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

G   Proof of Proposition 5

In the alternative mechanism, the price of information can be written

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

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

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

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

and

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

The ranking of profits, surplus, and optimal data collection is then straight-forward.

H Proof of the symmetry of the partitions when the data intermediary sells information to both firms

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

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

Replacing these values in the above demands and prices gives:

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

and

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

Profits are:

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

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

The data intermediary maximizes the following profit function:

$$\Pi_2(j_1, j_2) = (\pi_1^{I,I}(j_1, j_2) - \pi_1^{NI,I}(\emptyset, j_2)) + (\pi_2^{I,I}(j_1, j_2) - \pi_2^{NI,I}(\emptyset, j_1))$$

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

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

I Proof of Proposition 6

The profit of the data intermediary when selling information to both firms in a take it or leave it offer is provided in Bounie et al. (2018) and has the following value:

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

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 E lead to the rankings in Proposition 6.

J Proof of Proposition 7

See the proofs of Propositions 3 and 2.

K Proof of Proposition 8

We prove that data collection decreases when the price cap decreases.

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 D), the lower the $\bar{p}$ the lower the k.

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

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

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

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