

Data Intermediaries and Selling Mechanisms for Customized Consumer Information *

David Bounie, Antoine Dubus, Patrick Waelbroeck

► **To cite this version:**

David Bounie, Antoine Dubus, Patrick Waelbroeck. Data Intermediaries and Selling Mechanisms for Customized Consumer Information *. 2019. hal-02288708

HAL Id: hal-02288708

<https://hal.archives-ouvertes.fr/hal-02288708>

Submitted on 15 Sep 2019

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

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

Data Intermediaries and Selling Mechanisms for Customized Consumer Information*

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

September 15, 2019

Abstract

We investigate the strategies of a data intermediary selling customized consumer information to firms for price discrimination purpose. We analyze how the mechanism through which the data intermediary sells information influences how much consumer data he will collect and sell to firms, and how it impacts consumer surplus. We consider three selling mechanisms tailored to sell customized consumer information: take it or leave it offers, sequential bargaining, and simultaneous offers. We show that the more data the intermediary collects, the lower consumer surplus. Consumer data 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 argue that selling mechanisms can be used as a regulatory tool by data protection agencies and competition authorities to limit consumer information collection and increase consumer surplus.

*We would like to thank Wilfried Sand-Zantman as well as the participants at the 46th EARIE conference for useful remarks and comments. Antoine Dubus acknowledges financial support from the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris. Patrick Waelbroeck thanks for insightful discussions the members of the Chair Values and Policies of Personal Information of Institut Mines Télécom, Paris.

[†]i3, CNRS, Telecom Paris, Institut polytechnique de Paris, 46 rue Barrault, 75634 Paris Cedex 13, France; david.bounie@telecom-paris.fr.

[‡]i3, CNRS, Telecom Paris, Institut polytechnique de Paris, 46 rue Barrault, 75634 Paris Cedex 13, France; antoine.dubus@telecom-paris.fr.

[§]i3, CNRS, Telecom Paris, Institut polytechnique de Paris, 46 rue Barrault, 75634 Paris Cedex 13, France; patrick.waelbroeck@telecom-paris.fr.

1 Introduction

Since the seminal works of [Hayek \(1945\)](#) and [Marschak \(1974\)](#), scholars and policy makers have acknowledged that information greatly enhances the efficiency of markets. In the last two decades markets have however considerably evolved with the rising of new intermediaries that supply consumer information to firms willing to improve their business practices. Companies of a new type - data intermediaries - have specialized in collecting data from different sources, and selling customized datasets to firms.¹ The emergence of this new market for data intermediation raises two main concerns related to data protection and competition policies.

First, data intermediaries have become major actors of the economy, up to a point where, in 2014, the market for consumer data was valued around USD 156 billion per year ([Pasquale, 2015](#)). Recent scandals of data breaches and violation of consumer privacy have revealed the huge amount of information possessed by data intermediaries.² For instance, in a study of nine data brokers from 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*). The sheer volume of personal data collected by these data intermediaries raises concerns for data protection agencies. Indeed, new regulations try to limit the amount of personal data collected by data intermediaries. 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

¹For instance, data brokers such as Equifax or Transunion, sell specific consumer segments to firms willing to personalize their advertising campaigns.

²[Huge data breach reveals hundreds of millions of emails and passwords from across the Internet.](#)

the [Health Insurance Portability and Accountability Act](#) in the US, and in the [General Data Protection Regulation](#) in Europe.

Secondly, 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.³ There is a risk that more precise consumer information could lead to more consumer surplus extraction and to increased market power in the data intermediaries' industry.⁴ There is thus a pressing need to analyze the strategies of data intermediaries.

The study of data intermediaries and their role on the data strategies of firms is the subject of a new field of the literature. Two topics are of a particular interest. The first is related to the impact of the selling mechanism on market equilibrium. Economists have for long acknowledged that the way a product is sold has a profound effect on the organization of markets ([Riley and Zeckhauser, 1983](#)). This literature has focused in particular on take it or leave it offers ([Binmore et al., 1986](#)), and sequential bargaining ([Rubinstein, 1982](#); [Sobel and Takahashi, 1983](#)). Recently, [Backus et al. \(2018a\)](#), [Backus et al. \(2018b\)](#) and [Backus et al. \(2019\)](#) empirically analyze patterns of bargaining on e-bay and how bargaining environments are affected by information asymmetries, bargaining power, and private characteristics on the buyer and on the seller's side. [Milgrom and Tadelis \(2018\)](#) study how machine learning is 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.

The second topic deals with the strategies of a data intermediary selling consumer information for a given selling mechanism. [Montes et al. \(2018\)](#) consider a data broker selling information to firms that allows them to first-degree price

³Facebook gave Lyft and others special access to user data; [engadget, May 12th, 2018](#).

⁴See the recent debate on a potential breakup of major data intermediaries. ([Is Big Tech Too Big Or Not Big Enough?; Forbes, June 20th, 2019.](#); [Warren Wants To Break Up Amazon, Facebook, Google; Forbes, March 8th, 2019.](#))

discriminate different consumer segments. The data intermediary auctions the same consumer information to all active firms in the market. This assumption has been recently challenged by [Bounie et al. \(2018\)](#) who allow the data broker to sell customized consumer segments to firms, and show that selling all available information is not optimal for the data broker. An important conclusion of that research is that with the recent advances in information technologies, data intermediaries will propose customized information to firms. Auctions are typically used to sell a standard product for which there are a large number of bidders with different unknown valuations, but are not well suited to sell customized information that fits the need of only one bidder.

In this paper, we complement the literature on information selling in two respects. We first analyze three alternative mechanisms to auctions to sell customized information to firms: a take it or leave it offer, a sequential bargaining, and simultaneous offers. Secondly, we model a central activity of data intermediaries that is ignored in the literature, data collection, and we analyze how strategies of data collection affect the selling strategies of data intermediaries. We contribute to the existing literature on two points.

Our first contribution is to propose a characterization of selling mechanisms using the general notion of data contract. We define a class of independent data contracts for which the choice of information sold to one firm is independent of the choice of information sold to another firm. We show that the three main mechanisms (sequential bargaining, take it or leave it and simultaneous offers) belong to this class of independent data contracts and that they all lead to the same number of consumer segments sold in equilibrium. We also show that the key element that will determine how much information will be collected and sold in equilibrium is the threat for a firm of being uninformed. This threat will determine the willingness to pay of a firm for information, and thus the price of information. This conclusion could not be reached in models where the data intermediary is not selling consumer segments strategically.

Secondly, we endogenize the strategies of data collection of the data intermediary, and we compare the amount of consumer data that the data intermediary collects under the different selling mechanisms. We show that the take it or leave it offer maximizes consumer surplus and minimizes data collection, but that the data intermediary would prefer simultaneous offers or sequential bargaining that lead to a more intense collection of information, and in turn to lower consumer surplus. We discuss the regulatory implications of these results in the conclusion.

The remainder of the article is organized as follows. We describe the model in Section 2. Section 3 describes the three selling mechanisms. We solve the game in Sections 4 and 5. We discuss an alternative selling mechanism in Section 6. Section 7 concludes.

2 Model

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. The data intermediary collects and sells data on consumer segments. An informed firm can set a price on each consumer segment. An uninformed firm cannot distinguish consumer segments and sets a single price on the entire line.

2.1 Consumers

Consumers buy one product at a price p_1 from Firm 1 located at 0, or at a price p_2 from Firm 2 located at 1. A consumer located at $x \in [0, 1]$ receives a utility V from purchasing the product, but incurs a cost $t > 0$ of consuming a product that does not perfectly fit his taste x . Therefore, buying from Firm 1 (resp. from Firm 2), incurs a cost tx (resp. $t(1 - x)$). Consumers choose the product that gives the highest level of utility:⁵

⁵We 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. \(2018\)](#).

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

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).

2.2 Data intermediary

The data intermediary collects information on consumers that allows firms to distinguish k consumer segments on the unit line. The data intermediary can decide to sell all segments collected or only a subset of these segments. We will show that the data intermediary never sells all available consumer segments.⁶

2.2.1 Collecting data

The data intermediary collects k consumer segments at a cost $c(k)$. 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, and thus increases the value of information. For instance, when $k = 2$, the 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 broker 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} , illustrated in Figure 1.



Figure 1: Partition \mathcal{P}

⁶Previous research has assumed that the data intermediary sells all available information (Montes et al., 2018). We show that this assumption is not valid.

2.2.2 Selling information

To present our argument in the simplest way, we assume that the data intermediary only sells information to Firm 1⁷ using one of the three following selling mechanisms: a take it or leave it offer, a sequential bargaining, and simultaneous offers. We show in Section 6.2 that our results are robust to this assumption, i.e. they hold when the data intermediary can sell information to both firms.

The data intermediary can potentially sell any subset of segments collected in the partition depicted in Figure 1. 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 allow a firm to extract more surplus from consumers. 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 can poach them (Thisse and Vives, 1988). For instance, if the data intermediary sells all consumer segments, Firm 1 can set a price on the consumer segment that is the closest to Firm 2.

Thus, an optimal partition must balance the competition and surplus extraction effects. 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 of size $\frac{1}{k}$ on $[0, \frac{j_1}{k}]$ where consumers are identified so that Firm 1 can price discriminate them. 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 $j_1(k)$ the number of segments as a function of k . 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 .⁸

⁷Selling 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.

⁸See Bounie et al. (2018) for a more detailed discussion.

- 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.
- Stage 3: firms set prices p_1 and p_2 on the competitive segments.
- Stage 4: Firm 1 price discriminates consumers where it is in a monopoly position by setting p_{1i} , $i \in [1, j_1(k)]$.

The game is solved by backward induction. In stage 4, Firm 1 sets prices p_{11}, \dots, p_{1j_1} on segments where it 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 Section 4. In stage 1, we determine how much data the intermediary collects in equilibrium in Section 5. 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.

3 Selling mechanisms

We analyze three mechanisms that have been extensively studied in the literature for final goods, that we apply to information goods: customized information can be sold through a take it or leave it offer, a sequential bargaining, and simultaneous offers. First, under the take it or leave it selling mechanism, the data intermediary proposes an information partition to Firm 1. After the offer is made, there is no possibility for the data intermediary to sell information to Firm 2, even if Firm 1 refuses the offer. This approach has been studied for instance by [Binmore et al. \(1986\)](#). The second mechanism, the 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 \(1983\)](#). Thirdly, the data intermediary can sell information to firms through simultaneous offers as in [Bounie et al. \(2018\)](#). The data intermediary proposes to each firm an information partition that can be different for Firm 1 and Firm 2 at different prices.

The three selling mechanisms have a major impact on the strategies of the data intermediary and on the value of information. We compute for each selling mechanism what a firm is ready to pay for information, and determine its outside option if it does not purchase information. In the remainder of this section, 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 it has information on the j_1 consumer segments closest to its location (Firm 2 is uninformed). In the take it or leave it format, 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 and simultaneous offers formats, Firm 2 has information when Firm 1 is uninformed. Let $\bar{\pi}_1(j_2)$ denote the profit of Firm 1 when Firm 2 has information on the j_2 consumer segments closest to its location.

It is also useful to define a data contract as a couple (j_1, j_2) where j_1 is the information proposed to Firm 1, and j_2 is the information sold to Firm 2 if Firm 1 does not acquire information, which can include the empty set, for instance in the take it or leave it offer.

Definition 1 (Data contract). *A data contract is a couple (j_1, j_2) .*

We will show in [Definition 2](#) and [Theorem 1](#) that the three selling mechanisms belong to a specific class of data contracts.

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 format that we analyze in Section 3.2.¹¹

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 data contract is 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 data contract is therefore (j_1^{seq}, j_2^{seq}) .

To compute the value of information under the sequential bargaining format, we characterize a stationary equilibrium of this game where Firm 1 is making profit $\pi_1(j_1^{seq})$ if it accepts the offer, but makes profits $\bar{\pi}_1(j_2^{seq})$ if it declines the offer and Firm 2 purchases information. It is important to stress that when

¹¹The take it or leave it format includes in fact many such mechanisms where there is no possibility for renegotiation, including Nash bargaining and menu pricing.

Firm 1 declines the offer of the data intermediary, it will compete with Firm 2 that is proposed the symmetric partition (j_2^{seq} is the symmetric of j_1^{seq}). 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 Simultaneous offers

Under the simultaneous offers, the data intermediary proposes simultaneously to each firm an information structure and a price. The data intermediary can design a data contract that maximizes the willingness to pay of Firm 1 (without loss of generality). When Firm 1 acquires information, it makes profits $\pi_1(j_1^{so}) - p_{so}$, where p_{so} is the price of information. To maximize the threat on Firm 1, the data intermediary proposes k segment of information to Firm 2, so that the profit of Firm 1 without information is $\bar{\pi}_1(k)$, which is the worst case scenario. The data contract is therefore (j_1^{so}, k) . The data intermediary sets the price of information to:

$$p_{so}(j_1^{so}, k) = \pi_1(j_1^{so}) - \bar{\pi}_1(k)$$

4 Number of segments sold in equilibrium

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

independent from the information proposed to Firm 1. Theorem 1 generalizes Proposition 1 for independent data contracts.

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 simultaneous offers mechanisms.

Proposition 1

The number of consumer segments sold in equilibrium is:

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

Proof: see Appendix C.

The proof of Proposition 1 is based on the fact that the data intermediary optimizes j_1 and j_2 independently. In other words, the information proposed to Firm 1 (j_1) is independent from 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 simultaneous offers mechanisms. Under the take it or leave it format, Firm 1 has no information when it declines the offer of the data intermediary, and thus its outside option is independent with the information structure proposed by the data intermediary to Firm 1. Under the simultaneous offers format, 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 from the partition proposed to Firm 1. Under sequential bargaining, at each stage of the process, the firm who 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 from 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 data contracts

Using the intuition developed in the previous section, we can generalize Proposition 1 to a specific class of data contracts. These independent data contracts have the property that the information sold to Firm 1 (j_1) is independent from 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 independent data contracts lead to the same number of consumer segments sold to Firm 1 (j_1^*).

Let (j_1, j_2) be the data contract proposed to Firm 1.

Definition 2 (Independent data contract)

A data contract (j_1, j_2) is independent if the data intermediary maximizes profits by choosing j_1 and j_2 independently.

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 of the 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. Also, infinite sequential bargaining with discount factors alternate offers to Firm 1 and to Firm 2 independently. However, there are mechanisms that do not satisfy Definition 2. For instance, the data intermediary can propose a symmetric partition to Firm 1, then to Firm 2 if Firm 1 declines the offer. The information structure proposed to Firm 1 appears in its outside option: $p_{alt} = \pi_1(j_1^{alt}) - \bar{\pi}_1(j_1^{alt})$. Thus, the number of segments chosen by the data intermediary affects both the profit of Firm 1 and its outside option, violating Definition 2.

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:

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

Theorem 1 has theoretical and practical implications. First, Theorem 1 provides a first attempt to characterize data contracts based on their theoretical properties. Other dimensions of interest include the length of the data contract, exclusive sales, renegotiation conditions, or quantity discount.

Secondly, with independent data contracts, 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 favors Firm 1 on the market, to the detriment of Firm 2. This is not necessarily the case with other types of contracts. For instance under symmetric offers analyzed in Section 6 the data intermediary maximizes the willingness to pay of the second highest bidder, and the interest of Firm 1 and the data intermediary are not aligned.

Thirdly, Theorem 1 offers a convenient criteria 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 data contracts 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 data contract to determine if an action is required to limit the amount of information sold on a market.

We have shown in this section that the number of consumer segments sold to Firm 1 does not vary with the selling mechanism. Next, we analyze how the amount of data collected varies under different selling mechanisms.

5 Collecting data in equilibrium

In this section we analyze how the profits of the data intermediary vary with the number of consumer segments collected (k) for the three selling mechanisms considered so far. 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, the number of segments collected in the first stage of the game is not necessarily the same.¹²

The profit of the data intermediary $\Pi \in \{\Pi_{tol}, \Pi_{seq}, \Pi_{so}\}$ is given by the price of information $p \in \{p_{tol}, p_{seq}, p_{so}\}$, net of the cost of data collection $c(k)$:¹³ $\Pi(k) = p(k) - c(k)$.

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 amount 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.

Proposition 2

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

$$k_{seq} > k_{so} > k_{tol}, \quad \text{and} \quad CS_{tol} > CS_{so} > CS_{seq}.$$

Proof: see Appendix D.

¹²We assume that the cost of collecting data does not depend on the selling mechanism.

¹³We make the assumption that Π_{net} is concave and reaches a unique maximum on \mathbb{R}^+ . See Appendix A for a mathematical expression of this assumption.

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. This marginal gain is the lowest in the take it or leave it offer since the outside option of the firm does not depend on any partition proposed by the data broker. Thus, the surplus extraction effect is the weakest under this selling mechanism, and consumer surplus is maximized.¹⁴ Proposition 2 sharply contrasts with the existing literature that argues that information leads to higher consumer surplus due to the competitive effect of information (Thisse and Vives, 1988; Stole, 2007). We show here that more information on the market can decrease consumer surplus, because more information means more surplus extraction but competition on the market remains low with the selling strategy of the data intermediary.

Proposition 3 shows that the data intermediary chooses the simultaneous offers mechanism, 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 simultaneous offers and minimized under the take it or leave it mechanism:

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

Proof: see Appendix E.

Under the simultaneous offers 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.

¹⁴The marginal gain is higher in the simultaneous offers mechanism since the data intermediary threatens the highest bidder with the harshest partition, \mathcal{P} , which includes all consumer segments. The marginal gain is highest under the sequential bargaining mechanism, since collecting more data in that case increases the threat of the outside option the most.

Proposition 3 is relevant for regulators. The data intermediary chooses the simultaneous offers mechanism that maximizes its profits among the three mechanisms that we propose in this article. Thus a data intermediary will never choose the take it or leave it mechanism. However, Proposition 2 shows that a competition authority, concerned with consumer surplus, and a data protection agency, concerned with the amount of consumer data collected, would choose the take it or leave it format. Enforcing specific selling mechanisms is a simple and powerful tool for regulators.

6 Extensions

6.1 Alternative selling mechanism

We analyze an alternative selling mechanism in which the data intermediary simultaneously proposes symmetric partitions to both firms (symmetric offers, indexed by sym). Such selling mechanism therefore does not verify Definition 2. We show that the main results of Sections 4 and 5 hold under this alternative selling mechanism.

In the symmetric offers mechanism, the data intermediary proposes a partition j_1^{sym} to Firm 1. If Firm 1 declines the offer, a symmetric partition is proposed to Firm 2. Such a mechanism can be enforced by a competition authority to guarantee a level playing field. The price of information p_{sym} can be written as follows: $p_{sym} = \pi_1(j_1^{sym}) - \bar{\pi}_1(j_1^{sym})$. The data contract does not satisfy Definition 2 since j_1^{sym} appears in the outside option of Firm 1. The data intermediary will take this negative effect of j_1^{sym} on the profits of Firm 1 when it declines the offer into account.

Proposition 4

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

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

$$(b) \quad \Pi_{so} > \Pi_{sym} > \Pi_{seq} > \Pi_{tol}$$

$$(c) \quad k_{seq} > k_{so} > k_{sym} > k_{tol}$$

$$(d) \quad CS_{sym} > CS_{tol} > CS_{so} > CS_{seq}.$$

Proof: see Appendix F.

First, the take it or leave it mechanism still minimizes the number of consumer segments collected, so that a data protection agency would prefer it to any other selling mechanism. Secondly, the data intermediary still chooses the simultaneous offers mechanism as it leads to the highest willingness to pay of Firm 1. Thus there is still a tension between private and public interests. Thirdly, consumer surplus is now maximized in the symmetric offers mechanism. Thus there is a new tradeoff between data protection agencies and competition authorities. On the one hand, a data protection agency prefers the take it or leave it mechanism that minimizes the amount of personal data collected. On the other hand, a competition authority prefers the symmetric offers mechanism that maximizes consumer surplus.

To sum-up, we have identified another class of selling mechanisms, where partitions proposed to both firms are perfectly correlated and symmetric, under which our main results hold. It remains to show that Propositions 4 also holds for a broader set of classes. This is likely to be true given the fact that simultaneous offers is the selling mechanism that extracts the most surplus from the firm who purchases information.

6.2 Selling information to both 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 in a simultaneous offer.¹⁵

¹⁵We focus on simultaneous offers as it is the only mechanism among the three considered where the data intermediary can sell information to both firms simultaneously.

The data intermediary can sell symmetric information to both firms at the price $p_{both} = \pi_1(j_1^{both}) - \bar{\pi}_1(j_1^{both})$.¹⁶ Thus the profit of the data intermediary in that case is $\Pi_{both}(k) = 2p_{both} - c(k)$.

We characterize in Proposition 5 the profit of the data intermediary, the number of consumer segments collected, and consumer surplus when the data intermediary sells information to both firms, and we rank these values.

Proposition 5

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

- (a) $j^{both*} = \frac{6k - 9}{22}$
- (b) $\Pi_{so} > \Pi_{seq} > \Pi_{both} > \Pi_{tol}$
- (c) $k_{seq} > k_{so} > k_{both} > k_{tol}$
- (d) $CS_{both} > CS_{tol} > CS_{so} > CS_{seq}$.

Proof: see Appendix G.

The data intermediary optimally sells information to one firm under the sequential bargaining and the simultaneous offers mechanisms, but prefers to sell information to both firms rather than selling to one firm under the take it or leave it mechanism. All results of Sections 4 and 5 hold when the data intermediary is allowed to sell information to both firms. First, 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 simultaneous offers. A data protection agency and a competition authority would still prefer the take it or leave offers mechanism to any other selling mechanism. Secondly, the data intermediary still prefers selling information to

¹⁶The symmetry of the information structure when the data intermediary sells information to both firms is shown in Bounie et al. (2018)

one firm under the simultaneous offers mechanism as it leads to the highest willingness to pay of Firm 1. Thus there is still a tension between private and public interests. The tradeoff between data protection agencies and competition authorities remains. On the one hand, a data protection agency would prefer a situation where only one firm is informed since in that case the number of consumer segments collected and sold is minimized. On the other hand, a competition authority would prefer a situation where both firms are informed since consumer surplus is maximized in that case.

7 Conclusion

With the rise of digital giants such as Facebook, Apple, Google and Amazon, access to data and information is now central for competition policy in the digital era. As [Cr mer 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 anticompetitive practices, to increase their scrutiny of the activity of web giants.¹⁷

A first central contribution of this paper is to show that it is not only crucial to take access to data into account, but also to consider *differentiated* access to consumer data. Indeed, we have shown that data intermediaries can influence competition on product markets by selling different information to firms. Our main results indicate that the design of the market for information is of most importance for the data strategies of market participants. We have focused on selling mechanisms, but other dimensions matter, such as for instance resolving information asymmetries between sellers and buyers of information ([Bergemann and Bonatti, 2015](#))

The second main contribution of this paper is to emphasize the importance

¹⁷Congress, [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.](#)

of the mechanism through which consumer information is sold on the strategies of data intermediaries regarding how much consumer information they collect and sell. We argue that policy makers could feel uneasy about leaving the market for information unregulated. First, data intermediaries prefer selling information through simultaneous offers that allow firms to extract more consumer surplus. In turn, data intermediaries can raise the price of information, and gain market power. Competition authorities could be worried about the resulting market dominance. Secondly, data intermediaries will choose a selling mechanism that maximizes the amount of data collected, which can raise privacy concerns. Indeed, recent legislations such as the European GDPR impose a data minimization principle.

Data intermediaries are growing fast, collecting any type of information on huge masses of consumers.¹⁸ It is therefore important to know how much personal information data intermediaries collect to understand their strategies with respect to collecting and selling consumer information. Controlling the mechanism through which consumer information is sold is a simple yet powerful tool to minimize consumer data collection.

¹⁸[Data brokers: regulators try to rein in the 'privacy deathstars'](#).

A Mathematical interpretation of Assumption 1

The cost function is defined such that:

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

This technical hypothesis is 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:

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

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 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:

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

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

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

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:¹⁹

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

Thus, first order conditions on π_1 gives us

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

¹⁹For $p_{1i} \geq 0 \implies \frac{j}{k} \leq \frac{3}{4}$. Profits are equal whatever $\frac{j}{k} \geq \frac{3}{4}$.

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.

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

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

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

Comparing the derivatives gives us:

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

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

$$k_{seq} > k_{so} > k_{tol}$$

Consumer surplus

Prices when the data intermediaries 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 $i = 1, \dots, j$, and:*

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

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

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

We need to compute demands in order to find consumer surplus. 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

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

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

We can write consumer surplus in equilibrium:

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

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:

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

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

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

Direct comparison of the profits provide the ranking of Proposition 2.

F Proof of Proposition 4

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 straightforward.

G Proof of Proposition 5

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 5](#).

References

- 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 cheap-talk signaling: An application to bargaining. *Journal of Political Economy*, 127(4):000–000, 2019.
- Dirk Bergemann and Alessandro Bonatti. Selling cookies. *American Economic Journal: Microeconomics*, 7(3):259–94, 2015.
- 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.
- Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. Competition policy for the digital era. *Report for the European Commission*, 2019.
- Drew Fudenberg and J Miguel Villas-Boas. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.
- Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.
- 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.
- Jacob Marschak. Economic measurements for policy and prediction. *Economic information, decision, and prediction*, pages 293–322, 1974.
- 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.
- Frank Pasquale. *The black box society: The secret algorithms that control money and information*. 2015.
- 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.
- 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.