Learning-by-doing in Data Markets*

Antoine Dubus,† Wilfried Sand-Zantman,‡

March 11, 2024

Abstract

We consider a data broker specialized in the treatment of consumer data, and selling to firms its analytics services. The broker accesses the data of a firm and returns information on the willingness to pay of consumers, allowing the firm to achieve higher profits by price-discriminating consumers. By accessing the data of a firm, the broker can benefit from effects of learning-by-doing reducing its further treatment cost. At first glance, the intuition suggests that this learning effect allows the broker to treat data in a wider range of cases, such as when the treatment cost is high, and induces an expansion of the data market. Yet, our contribution is to show that learning effects can induce a shrinkage of the data market: by anticipating how providing their data to the broker will induce the learning effect, firms will strategically prefer to remain uninformed if doing so implies that their competitor also remains uninformed.

1 Introduction

The digitization of the economy has seen the rise of data markets. Retailers, banks, or real estate companies now routinely acquire the services of data brokers such as Nielsen, Axciom or Equifax, specialized in supplying fine-grained data and business analytics allowing firms to optimize their interactions with customers. In 2020, the data brokerage industry was valued at 200 Billion USD (Tucker and Neumann, 2020), and is expected to rise to more than 462 Billion USD by 2031.[1]

These data brokers spend tremendous efforts to collect the most granular data that they use to fuel machine learning algorithms allowing them to predict at best consumers' purchas-

*We would like to thank

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

‡ESSEC Business School and Thema; sand@essec.edu.

[1] Transparency Market research, Data Broker markets, trends and forecast 2022-2031, last accessed 26.01.2024.

ing behavior. For instance, Nielsen collects data using the National Consumer Panel where consumers report their daily purchasing behavior.[2]

Yet, an important part of their data comes from their clients themselves. An online retailer that purchases the services of Nielsen to optimize its pricing strategy will first provide data on its past consumers' purchase behavior and payment transactions, and Nielsen will then treat the data using its analytics services – such as the Scantrack service for instance – to provide the retailer with recommendations on its pricing strategy. Similarly, in the US, banks provide Equifax, Axciom and Transunion with financial data on their customers, and then they purchase the credit scores computed by these data brokers, which they use to offer consumers with personalized interest rates. As pointed by the FTC (2014), "data brokers report that they obtain data directly from their merchant and financial service company clients". Hence, by selling their analytics services to firms, data brokers learn from the data of their clients and improve the quality of their data sets and algorithms. Data markets are characterized by strong effects of learning-by doing, and as data brokers accumulate experience, they can provide their clients with better and cheaper analytics services.

In light of these effects of learning-by-doing that characterize data markets, this paper examines what tactic firms should adopt with regard to purchasing the services of a data broker and thereby providing it with their data. To structure the setting that we analyze, we consider one data broker and two firms – labelled Firm A and Firm B – competing in a product market. Firms have data on some of their consumers. They can pay the data broker to treat their data and provide them information on the willingness to pay of these customers, allowing firms to charge consumers personalized prices. This requires a firm to provide its data to the broker, which may help it to reduce its treatment cost later on. For simplicity, Firm A decides first whether to deal with the broker, and then Firm B makes its decision.

When the data broker benefits from strong learning-by-doing effects, one would thus expect the data market to expand, with the broker selling its services to Firm B more often. These dynamics inspire us to ask the following questions: May a firm choose in fact *not to purchase* the services of the broker because of the learning-by-doing effect? If so, when and why might such a counter-intuitive strategy make sense? In turn, what are the impacts of such a strategy on profits and consumers?

Our main contribution is to show that learning-by-doing effects in data may induce firms not to purchase the services of a broker and therefore reduce rather than increase the overall use of data. The intuition for this result is the following.

When Firm A knows that providing its data to the broker will make it more efficient, it

[2]National Consumer Panel.

anticipates that its improved efficiency will allow the broker to provide its analytics services to Firm B, the competitor of Firm A. This would result in a sharp increase of the intensity of competition in the product market, that Firm A can sometimes avoid by not purchasing the services of the broker. Such a strategic anticipation does not take place without learning-by-doing, in which case Firm A would know that purchasing the services of the broker does not change its ability to treat the data of Firm B. This anticipation of the competitive effect of learning-by-doing results in a shrinkage of the data market, a novel result in the literature.

Perhaps even more surprisingly, we show that the broker is sometimes even ready to *pay* Firm A to treat its data, anticipating the cost reduction that this induces and the further profits that the broker will realize. In such case, treating the data of Firm A has the value of investments in cost-reducing innovations that benefit the broker later on. Paying to treat its data provides Firm A incentives to share the data, which it would not find profitable otherwise. Indeed, in support of these intuitions, the FTC (2014) report points to practices of data brokers purchasing data from firms, and particularly from retailers.[3]

Our results have both managerial and policy implications. First, managers of the competing firms should consider that their data are valuable inputs that benefit the broker. With this in mind, the bargaining power between buyers and sellers of data may change, as a firm can account for the efficiency gains that its data provides with the broker. Secondly, managers of the data broker company should account for these effects of learning by doing when dealing with their clients. Otherwise, our results will show that the broker may sell data to fewer firms and achieve lower profits. In terms of policy and welfare, our results contribute to policy debates where regulators are increasingly wary of the importance taken by data brokers in our economics, in particular regarding their ability to collect precise sensitive information on consumers, raising a threat to their privacy.[4] By highlighting the importance of effects of learning-by-doing in data markets, our results suggest that the data market may expand as data brokers become even more efficient, but that this expansion may be limited if firms account for the value of their data when dealing with brokers.

The rest of the article if organized as follows. We review the literature on learning-by-doing and on data markets in Section 2. We describe the model in Section 3, and we solve the product-market equilibrium in Section 4. We characterize the value of information in Section

[3]See FTC (2014) p.13: "Data brokers in this study purchase information about individuals from wide-ranging commercial sources. For example, the data brokers obtain detailed, transaction-specific data about purchases from retailers and catalog companies. Such information can include the types of purchases (e.g., high-end shoes, natural food, toothpaste, items related to disabilities or orthopedic conditions), the dollar amount of the purchase, the date of the purchase, and the type of payment used".

[4]See for instance the FTC (2014) report, and more recently, the Digital Platform Services Inquiry – March 2024 report on data brokers of the Office of the Australian Information Commissioner. Last accessed, 26.01.2024.

5 and the selling strategy of the broker in equilibrium in Section 6. Section 7 concludes.

2 Literature

Learning-by-doing Effects of learning-by-doing have been for long identified in several industries, including plane manufacturers (Rosenberg, 1982; Miller and Chen, 1994; Von Hippel and Tyre, 1995) and service organization (Darr et al., 1995). These early evidence have then initiated a rich theoretical literature.

Cabral and Riordan (1994) analyze duopolists benefiting from effects of learning-by-doing, and show that intense price competition can provide firm with a strong competitive advantage allowing it to drive its competitor out of the market.

A rich literature builds on these model considering infinite interaction among firms in markets with learning-by doing. Besanko et al. (2010) consider firms acquiring experience that may decay over time. Besanko et al. (2014) analyze how learning-by-doing impacts competitive pricing. Sweeting et al. (2022) analyze a strategic buyer that anticipates how its consumption will reduce further prices. Deng et al. (2023) build on this last model to look at varying bargaining power in the product market, by considering partial extraction of surplus by the buyer.

These papers consider a rich set of questions including among other issues of entry and exit and product variety, showing that learning-by-doing reduces the number of variety compared to the social optimum. They also highlight the large number of equilibria in such infinite games, for instance as buyers can choose to postpone their consumption decision to benefit from lower prices at later stages.

We contribute to this literature by introducing competition among buyers. We will show that a firm that anticipates how its purchasing of the services of the broker impacts its ability to deal with the competing firm in the following stage can lead to an equilibrium where all firms prefer not to use the data at all, resulting in a shrinkage of the data market.

Data and markets We contribute to a rich literature analyzing the value of data in digital markets. Theoretical contributions in various fields have modelled effects of learning-by-doing with data in static models by focusing on increasing return to scale (Agrawal et al., 2019; Mihet and Philippon, 2019; Jones and Tonetti, 2020; Martens et al., 2021; Farboodi and Veldkamp, 2023).

By considering sequential interactions, we explicitly model learning effects and how they impact further interactions between the broker and the firms. We will show that, even though learning effects in data take place in principle, their implementation may not occur when accounting for firm's behavior. On the contrary, the market value of data will decrease if firms anticipate that the learning effect will allow the broker to serve their competitors.

3 Model

3.1 Consumers

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

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

3.2 Firms

Firm A is located at 0 and has some information on consumers in $[0, \delta_A]$, Firm B is located at 1 and has some information on consumers in $[1 - \delta_B, 1]$.

The information owned by each firm can be generated for instance by customers sharing data as part of a loyalty program. Firms cannot exploit this information on their own, but they can share it with a data broker that will treat the data and return to firms perfect information on these consumers.

3.3 Data Broker

The broker sells to firms the possibility to treat their data and price discriminate consumers. The broker makes sequentially a take-it-or-leave-it offer to each firm. First, the broker makes an offer to Firm A, and then to Firm B. Thereby, it can charge a maximal price of information equal to the willingness to a firm, i.e., equal to the difference of profits with information and those if it remains uninformed. We delve into the details of these offers when solving the game in Section 4.

Inferring consumers' types from data has a cost. If the broker has treated the data of Firm A, it benefits from a cost-reduction that is proportional to δ_A and to a factor α. Hence, at the time it makes an offer to Firm B, the broker will be able to treat Firm B's data at a lower cost if it has treated Firm 's data before. This specification allows us to model effects of learning by doing, and parameter α captures the strength of these effects. For simplicity we use the following functional form for the cost function:

$$\begin{cases} c\delta_A \text{ for Firm A,} \\ c\delta_B \text{ for Firm B without learning effect,} \\ c\delta_B(1-\alpha\delta_A) \text{ for Firm B when Firm A treated } \delta_A. \end{cases}$$

Hence, the broker makes an offer if the gains from selling information dominate the cost to treat the data.

3.4 Timing

The timing of the game is as follows:

- Stage 1: The broker makes an offer to sell information to Firm A.
- Stage 2: The broker makes an offer to sell information to Firm B.
- Stage 3: Firms set a homogeneous price for consumers on whom they do not have information. If they have acquired information on δ consumers, they then personalize prices for these identified consumers.

This timing is commonly used in the literature on learning-by-doing as it allows to simply identify the different effects of the interaction between an upstream seller and one of the firms (Besanko et al., 2010, 2014, 2019; Sweeting et al., 2022).

If a firm declines the offer of the broker, we ignore the possibility of renegotiation later on. For instance, if Firm A refuses the offer of the broker and Firm B accepts it, Firm A cannot reconsider the offer afterwards.

4 Competitive equilibrium in the product market

We compute prices and demands when firms have acquired information from the data broker. Firm A price discriminates consumers on $[0, \delta_A]$, and charges consumers on $[\delta_A, 1]$ a homogeneous price. Similarly, Firm B price discriminates consumers on $[1-\delta_B, 1]$, and charges a homogeneous price to consumers on $[0, 1-\delta_B]$.

Prices and demand. Firm A sets a price $p_A(x)$ for consumers located at $[0, \delta_1]$. Similarly, Firm B sets a price $p_B(x)$ for consumers located at $[1-\delta_B, 1]$. Firm i then sets a unique price p_{i2} on the rest of the unit line. The prices charged to consumers targeted by Firm A and Firm B satisfy:

$$V - tx - p_A(x) = V - t(1-x) - p_B \implies p_A(x) = p_B + t - 2tx,$$
$$V - t(1-x) - p_B(x) = V - tx - p_A \implies p_B(x) = p_A + 2tx - t.$$

Let denote d_A the demand for Firm A (resp. d_B the demand for Firm B) where firms compete. d_A is determined by the indifferent consumer $\tilde{x}$:

$$V - t\tilde{x} - p_A = V - t(1-\tilde{x}) - p_B \implies \tilde{x} = \frac{p_B - p_A + t}{2t},$$

and $d_A = \tilde{x} - \delta_A = \frac{p_B - p_A + t}{2t} - \delta_A$ (resp. $d_B = 1 - \delta_B - \frac{p_B - p_A + t}{2t}$).

Profits of the firms. The profits of the firms are:

$$\pi_A = \int_0^{\delta_A} p_A(x)dx + d_A p_A = \int_0^{\delta_A} (p_B + t - 2tx)dx + \left(\frac{p_B - p_A + t}{2t} - \delta_A\right) p_A,$$
$$\pi_B = \int_{1-\delta_B}^{1} p_B(x)dx + d_B p_B = \int_{1-\delta_B}^{1} (p_A + 2tx - t)dx + \left(\frac{p_A - p_B + t}{2t} - \delta_B\right) p_B.$$

Prices and demands in equilibrium. We now compute the optimal prices and demands, using first-order conditions on π_i with respect to p_i. Prices in equilibrium are:[5]

$$p_A = t[1 - \frac{2}{3}\delta_B - \frac{4}{3}\delta_A], \quad p_B = t[1 - \frac{2}{3}\delta_A - \frac{4}{3}\delta_B].$$

Replacing these values in the above demands and prices gives:

$$p_A(x) = 2t - \frac{4t}{3}\delta_B - \frac{2t}{3}\delta_A - 2tx, \quad p_B(x) = 2tx - \frac{4t}{3}\delta_A - \frac{2t}{3}\delta_B$$

[5] We rule out negative prices from the analysis, and a price is taken equal to zero in case its expression below is negative.

The indifferent consumer is located at:

$$\tilde{x} = \frac{1}{2} + \frac{1}{3}\delta_A - \frac{1}{3}\delta_B$$

Demands in equilibrium are as follows:

$$d_A = \frac{1}{2} - \frac{2}{3}\delta_A - \frac{1}{3}\delta_B, \quad d_B = \frac{1}{2} + \frac{2}{3}\delta_B - \frac{1}{3}\delta_A.$$

Profits in equilibrium. We compute profits by replacing prices and demands by their equilibrium values:

$$\pi_i = \frac{t}{2} - \frac{7}{9}\delta_i^2 t + \frac{2}{9}\delta_{-i}^2 t - \frac{4}{9}\delta_i\delta_{-i}t + \frac{2}{3}\delta_i t - \frac{2}{3}\delta_{-i}t. \tag{1}$$

Profits are strictly concave functions with respect to δ_A and δ_B, and they have a unique maximum.

5 Value of Information

In the rest of the analysis we focus on the symmetric case $\delta_A = \delta_B = \delta$. This implies that, depending on whether firms have their data treated by the broker, firms make the following profits:

$$\begin{cases} \pi_A = \pi_B = \frac{t}{2} \quad \text{if firms do not have their data treated,} \\ \pi_A = \pi_B = \hat{\pi} = \frac{t}{2} - \delta^2 t \quad \text{if both firms have their data treated,} \\ \overline{\pi} = \frac{t}{2} + \frac{\delta t}{3}[2 - \frac{7\delta}{3}] \text{ and } \underline{\pi} = \frac{t}{2} - \frac{2\delta t}{3}[1 - \frac{\delta}{3}] \text{ if only one firm (with profits } \overline{\pi}\text{) treats its data.} \end{cases}$$

When the broker treats the data of a firm, it enhances its ability to extract the surplus of the closest consumers by personalizing prices. This increases the profits of a firm, but also intensifies competition and harms the competitor. Hence, a firm benefits from having its data treated, to the expense of its competitor.

This competitive effect of information is so strong that when both firms have their data treated, they achieve lower profits than if they had remained uninformed. Overall we can rank the profits as follows:

$$\overline{\pi} > \frac{t}{2} > \hat{\pi} > \underline{\pi}$$

5.1 Selling Information to Firm B

A monopolist data broker can charge a maximal price of information equal to the difference of profits of a firm with and without information. For Firm B, this maximal price ρ depends on whether Firm A had its data treated:

$$\rho_B = \begin{cases} \hat{\pi} - \underline{\pi} & \text{if Firm A has acquired data,} \\ \overline{\pi} - \frac{t}{2} & \text{if Firm A is uninformed.} \end{cases}$$

In turn, the broker sells information to Firm B if the price of information is greater than the cost to treat information, equal to $\delta c(1-\alpha\delta)$ if Firm A had its data treated, and to δc otherwise.

Firm A had its data treated. In this case, Firm B will purchase data from the broker if and only if $\hat{\pi} - \underline{\pi} > \delta c(1-\alpha\delta)$. Denoting $\tilde{c} = \frac{c}{t}$, this conditions implies that Firm B has its data treated in this case if the treatment cost is not too high – $\tilde{c} < \frac{2}{3}$ – and if the amount of data to treat is smaller than a threshold: $\delta < \hat{\delta}_{B2} = \frac{\tilde{c}-\frac{2}{3}}{\alpha\tilde{c}-\frac{11}{9}}$.

The threshold $\hat{\delta}_{B2}$ is smaller than $\frac{1}{2}$ if $\tilde{c} > \frac{1}{9(2-\alpha)}$. Hence, Firm B acquires data if

$$\begin{cases} \tilde{c} \in [0, \frac{1}{9(2-\alpha)}] \text{ and } \delta \in [0, \frac{1}{2}], \\ \tilde{c} \in [\frac{1}{9(2-\alpha)}, \frac{2}{3}] \text{ and } \delta < \hat{\delta}_{B2} = \frac{\frac{2}{3}-\tilde{c}}{\frac{11}{9}-\alpha\tilde{c}}. \end{cases}$$

Firm A did not have its data treated. In this case, Firm B will purchase data from the broker if $\overline{\pi} - \frac{t}{2} > \delta c$. This condition requires that the treatment cost is not too high – $\tilde{c} < \frac{2}{3}$ – an that the amount of data to treat is smaller that the following threshold: $\delta < \hat{\delta}_B = \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right)$.

$\hat{\delta}_B$ is smaller than $\frac{1}{2}$ if $\tilde{c} > \frac{5}{18}$. Hence, Firm B acquires data if

$$\begin{cases} \tilde{c} \in [0, \frac{5}{18}] \text{ and } \delta \in [0, \frac{1}{2}], \\ \tilde{c} \in [\frac{5}{18}, \frac{2}{3}] \text{ and } \delta < \hat{\delta}_B = \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right). \end{cases}$$

5.2 Selling Information to Firm A

Firm A anticipates how acquiring information impacts the decision of Firm B to acquire data in the next stage. This determines its willingness to pay and changes the price that the broker can charge for information:

$$\rho_A = \begin{cases} \hat{\pi} - \underline{\pi} & \text{if Firm B purchases data,} \\ \overline{\pi} - \frac{t}{2} & \text{if Firm B remains uninformed,} \\ \overline{\pi} - \underline{\pi} & \text{if Firm B remains uninformed only if Firm A purchases data.} \end{cases}$$

The decision of Firm A to acquire data has two opposite effects on the decision of Firm B to acquire data. On the one hand, we can show that $\overline{\pi} - \frac{t}{2} > \hat{\pi} - \underline{\pi}$ and the gains from having its data treated decrease for Firm B if Firm A treats its data as well. On the other hand, when Firm A deals with the broker, Firm B benefits from a cost reduction that increases its incentives to deal with the broker too. Firm A then balances these strategic anticipations with its own incentives to deal with the broker.

6 Selling Information in Equilibrium

We have seen that the treatment cost is a critical determinant of the decision of firms to have their data treated. In particular, when $\tilde{c} > \frac{2}{3}$, firms do not have their data treated and they compete in the standard Hotelling mode. Hence, we focus of low treatment costs and $\tilde{c} < \frac{2}{3}$.

The decision of Firm A to acquire information depends on the resulting impact on the decision of Firm B, which can be summarized by the respective positions of the thresholds $\hat{\delta}_B$ and $\hat{\delta}_{B2}$.

Hence, the decision of Firm A to acquire information will depend on the strength of the effect of learning-by-doing, captured in our model by the term $\alpha\ c\ \delta$. The effect will be weak when this product will be smaller than a cutoff value, otherwise, the effect of learning-by-doing will be strong.

Definition 1.

The learning effect is said to be weak when:

- *Either the learning parameter is smaller than $\frac{2}{3}$: $\alpha < \frac{2}{3}$,*
- *or the treatment cost is lower than $\frac{4}{9\alpha}$: $\tilde{c} \in [0, \frac{4}{9\alpha}]$,*
- *or the amount of data to treat is small: $\delta \in [0, \min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}]$*

Conversely, the learning effect is said to be strong when:

- *the learning parameter is higher than* $\frac{2}{3}$*, the cost is greater than* $\frac{4}{9\alpha}$ *and the amount of data to treat is high:*

$$\alpha > \frac{2}{3} \quad \& \quad \tilde{c} > \frac{4}{9\alpha} \quad \& \quad \delta \geq \min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}.$$

We analyze the data acquisition decision of the firms in cases of weak and strong learning-by-doing in the rest of the section, and we explain the role of the thresholds provided in the definition in changing the decision of the firms to acquire data.

6.1 Weak Learning Effect

Firm B does not have its data treated. A condition for Firm B not to have its data treated is when the learning parameter is small, $\alpha < \frac{2}{3}$, or when the parameter is high but the cost satisfies $\tilde{c} \in [\frac{1}{9(2-\alpha)}, \frac{4}{9\alpha}]$.

In this case, a necessary and sufficient condition for Firm B not to treat its data is for the amount of data acquired to satisfy $\delta \in [\hat{\delta}_B, \frac{1}{2}]$. Indeed, simple algebra allows us to show that $\hat{\delta}_B > \hat{\delta}_{B2}$ regardless of the value of $\tilde{c}$.

Firm A does not treat its data. When $\delta \in [\hat{\delta}_B, \frac{1}{2}]$, Firm B does not have its data treated. Firm A treats its data if $\overline{\pi} - c\delta > \frac{t}{2}$, which is equivalent to $\delta < \frac{9}{7}(\frac{2}{3} - \tilde{c}) = \hat{\delta}_B$, and Firm A does not treat its data either.

Firm A treats its data. When $\delta \in [\hat{\delta}_{B2}, \hat{\delta}_B]$, Firm A can either remain uninformed and Firm B will treat its data, or Firm A can treat its data, which will discourage Firm B from treating its data. Hence, the decision of Firm B to treat its data is conditional of Firm A remaining uninformed.

We can show that Firm A has its data treated. Indeed, its decision is either getting its data treated and make $\overline{\pi}$, or not and in this case, Firm B has them treated and makes profits equal to $\underline{\pi}$. Hence, Firm A treats its data if $\overline{\pi} - \underline{\pi} \geq \tilde{c}\delta$, which is always true and Firm A only has its data treated.[6]

By doing so, Firm A reduces the gains from data acquisition for Firm 2. Because the learning effect is weak, it is dominated by this discouragement effect and Firm B does not acquire information when Firm A does. On the contrary, because the alternative to remain

[6]Indeed, $\overline{\pi} - \underline{\pi} \geq \tilde{c}\delta \iff \frac{t}{2} + \frac{\delta t}{3}[2 - \frac{7\delta}{3}] - \frac{t}{2} + \frac{2\delta t}{3}[1 - \frac{\delta}{3}] > \tilde{c}\delta \iff \frac{4}{3} - \tilde{c} > \delta$, which is always true as $\tilde{c} < \frac{2}{3}$.

uninformed for Firm A is to face Firm B that acquires information, it increases the willingness to pay of Firm A for information.

Firm B has its data treated regardless of Firm A: $\delta < \min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}$**.** Firm A has the choice to have its data treated and make profits equal to $\hat{\pi}$, or to remain uninformed and make profits equal to $\underline{\pi}$. Hence, Firm A treats its data if $\hat{\pi} - c\delta > \underline{\pi}$.

This condition is satisfied δ is smaller than the threshold $\hat{\delta}_{A2} = \frac{9}{11}\left(\frac{2}{3} - \tilde{c}\right)$. This value is always smaller than $\min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}$.[7]

Hence, when $\delta \in [0, \hat{\delta}_{A2}]$ both firms have their data treated. When $\delta \in [\hat{\delta}_{A2}, \min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}]$, Firm A does not have its data treated, but Firm B does. In this case, Firm A anticipates low profits as Firm B will acquire data, and prefers not to acquire information. For Firm B, even if Firm A acquire data, having its data treated is beneficial as it faces lower treatment costs due to the learning effect.

Lemma 1.

When $\delta \in [\hat{\delta}_{A2}, \min\{\hat{\delta}_B, \hat{\delta}_{B2}, \frac{1}{2}\}]$, Firm A does not acquire data, anticipating low competitive profits as Firm B will acquire information.

Proof. There are two cases to consider:

- When $\alpha < \frac{2}{3}$, or when $\alpha > \frac{2}{3}$ and the cost satisfies $\tilde{c} \in [\frac{1}{9(2-\alpha)}, \frac{2}{3}]$, we have $\hat{\delta}_B < \hat{\delta}_{B2}$.
- When, $\tilde{c} < \frac{1}{9(2-\alpha)}$, $\hat{\delta}_B$ and $\hat{\delta}_{B2}$ are greater than $\frac{1}{2}$ so that Firm B always has its data treated.

□

6.2 Strong Learning Effect

We now consider the case where the learning effect is strong. In this case, when $\delta \in [\hat{\delta}_B, \hat{\delta}_{B2}]$ Firm B will choose to acquire data from the broker only if Firm A has acquired data too, and the broker can benefit from the learning effect, thereby incurring a lower treatment cost when dealing with Firm B. When $\delta > \hat{\delta}_{B2}$ firms do not have their data treated and they remain uninformed.

[7] Indeed, we have $\frac{9}{11}\left(\frac{2}{3} - \tilde{c}\right) < \min\{\frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right), \frac{\frac{2}{3} - \tilde{c}}{\frac{11}{9} - \alpha\tilde{c}}, \frac{1}{2}\}$.

Firm B has its data treated if Firm A does: $\delta \in [\hat{\delta}_B, \hat{\delta}_{B2}]$. Anticipating this effect, Firm A expects profits equal to $\hat{\pi} - c\delta$ if it acquires information, and to $\frac{t}{2}$ otherwise.

Hence, Firm A treats its data if and only if $\hat{\pi} - c\delta - \frac{t}{2} > 0$. This condition is never satisfied, and in this case, Firm A does not have its data treated.

Proposition 1.

When the learning effect is strong, Firm A does not acquire data in equilibrium. By doing so, it prevents the broker to benefit from learning by doing, preventing in turn Firm B from acquiring data.

Hence, in this case, the learning-by-doing effect in data induces firms not to acquire data from the broker. This effects results from a strategic anticipation by Firm A of how the learning effect will enable Firm B to acquire information.

The broker can pay for data. In the previous reasoning, the broker maximizes its profits at each period, and the participation constraint must be satisfied for both firms. This ignores the possibility for the broker to maximizes total profits across periods, thereby transferring a positive amount of money to Firm A to benefit from the learning effect an be able to sell to Firm B in the second period.

Proposition 2 provides conditions for the broker to benefit from purchasing data from Firm A and learn from it.

Proposition 2.

When the learning effect is strong, there exists $\hat{\delta}_{T2} = \frac{\frac{2}{3} - 2\tilde{c}}{\frac{20}{9} - \alpha}$ *with* $\hat{\delta}_{T2} \in [\hat{\delta}_B, \hat{\delta}_{B2}]$ *such that:*

- *When* $\delta \in [\hat{\delta}_B, \hat{\delta}_{T2}]$, *the data broker pays Firm A to acquire its data and benefit from the learning effect.*
- *When* $\delta \in [\hat{\delta}_{T2}, \hat{\delta}_{B2}]$ *Firm A does not acquire information to prevent the broker from learning from data. In this case, none of the firms acquire data from the broker.*

Proof. When transfers are allowed, the broker maximizes the sum of profits in the first stage – $\hat{\pi} - c\delta - \frac{t}{2}$ – and in the second stage – $\hat{\pi} - \underline{\pi} - c\delta(1 - \alpha\delta)$. Hence, there is room for data treatment if

$$\hat{\pi} - c\delta - \frac{t}{2} + \hat{\pi} - \underline{\pi} - c\delta(1 - \alpha\delta) > 0 \iff \delta < \frac{\frac{2}{3} - 2\tilde{c}}{\frac{20}{9} - \alpha} = \hat{\delta}_{T2}.$$

Simple comparison with $\hat{\delta}_B$ and $\hat{\delta}_{B2}$ allows us to show that $\hat{\delta}_{T2} \in [\hat{\delta}_B, \hat{\delta}_{B2}]$. Indeed, on the one hand we have $\frac{\frac{2}{3} - 2\tilde{c}}{\frac{20}{9} - \alpha} - \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right) > 0 \iff \frac{26 + 18\alpha}{27} > \tilde{c}(\alpha - \frac{2}{3})$, which is always true. On the other hand, $\hat{\delta}_{T2} < \hat{\delta}_{B2} \iff \tilde{c} > \frac{\frac{1}{9}+(\frac{1}{81}+\alpha\frac{2}{3})^{1/2}}{\alpha}$. For $\alpha > \frac{2}{3}$, this implies that $\tilde{c} > \frac{2}{3}$, which is outside of the range of values that we consider. □

6.3 Comparative Statics

We summarize these results using Figure 1

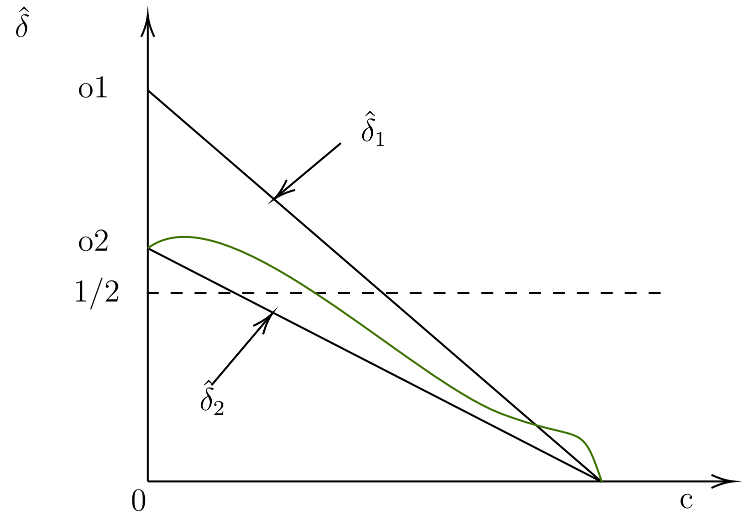

Figure 1: Profits

An increase in the value of the learning effect has ambiguous impact on the incentives of firms to acquire information, and overall on the expansion or reduction of the data market.

On the one hand, when α increases, the threshold $\hat{\delta}_{B2}$ increases, so that the range of values of δ for which Firm B is willing to treat its data if Firm A had its data treated becomes wider. As the learning effect becomes stronger, the treatment cost of Firm B decreases relaxing the condition under which the firm treats its data. This induces an expansion of the data market due the the learning effect.

On the other hand, when α becomes greater than $\frac{2}{3}$, a new zone appears for the cost parameter characterized by $\tilde{c} \in [\frac{4}{9\alpha}, \frac{2}{3}]$, where Firm A anticipate that acquiring information will enable Firm B to acquire it too thanks to the learning effect, and for this reason, Firm A prefers to remain uninformed. When α increases above $\frac{2}{3}$, $\frac{4}{9\alpha}$ decreases, and this new zone becomes wider. This means that a stronger learning effect may reduce the possibility for both firms to acquire information, inducing a shrinkage of the data market.

This effect is softened by the possibility for the broker to pay for the data of Firm A. Indeed, we observe an increases with α of the threshold $\hat{\delta}_{T2}$ below which Firm A accepts to sell its data to the broker, even though this implies that Firm B will also be able to acquire information.

Proposition 3 formalizes these intuitions.

Proposition 3.

An increase in the strength of the learning effect:

- *expands the data market if* $\alpha < \frac{2}{3}$*, as Firm B can have its data treated for for higher treatment costs,*
- *shrinks the data market if* $\alpha > \frac{2}{3}$*, as Firm A anticipates how acquiring information will allow Firm B to acquire it too, and forces its competitor to remain uninformed by remaining uninformed and preventing the learning effect to take place.*

7 Conclusions

TBD

References

Ajay Agrawal, Joshua Gans, and Avi Goldfarb. Economic policy for artificial intelligence. *Innovation policy and the economy*, 19(1):139–159, 2019.

David Besanko, Ulrich Doraszelski, Yaroslav Kryukov, and Mark Satterthwaite. Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica*, 78(2):453–508, 2010.

David Besanko, Ulrich Doraszelski, and Yaroslav Kryukov. The economics of predation: What drives pricing when there is learning-by-doing? *American Economic Review*, 104(3):868–897, 2014.

David Besanko, Ulrich Doraszelski, and Yaroslav Kryukov. How efficient is dynamic competition? the case of price as investment. *American Economic Review*, 109(9):3339–3364, 2019.

Luis MB Cabral and Michael H Riordan. The learning curve, market dominance, and predatory pricing. *Econometrica: Journal of the Econometric Society*, pages 1115–1140, 1994.

Eric D Darr, Linda Argote, and Dennis Epple. The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management science*, 41(11):1750–1762, 1995.

Shanglyu Deng, Dun Jia, Mario Leccese, and Andrew Sweeting. Bargaining and dynamic competition. 2023.

Maryam Farboodi and Laura Veldkamp. Data and markets. *Annual Review of Economics*, 15:23–40, 2023.

FTC. Data brokers: A call for transparency and accountability. *Ramirez, Edith and Brill, Julie and Ohlhausen, Maureen K and Wright, Joshua D and McSweeny, Terrell*, (s 47), 2014.

Charles I Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020.

Bertin Martens, Geoffrey Parker, Georgios Petropoulos, and Marshall W Van Alstyne. Towards efficient information sharing in network markets. 2021.

Roxana Mihet and Thomas Philippon. The economics of big data and artificial intelligence. In *Disruptive Innovation in Business and Finance in the Digital World*, pages 29–43. Emerald Publishing Limited, 2019.

Danny Miller and Ming-Jer Chen. Sources and consequences of competitive inertia: A study of the us airline industry. *Administrative science quarterly*, pages 1–23, 1994.

Nathan Rosenberg. *Inside the black box: technology and economics.* cambridge university press, 1982.

Andrew Sweeting, Dun Jia, Shen Hui, and Xinlu Yao. Dynamic price competition, learning-by-doing, and strategic buyers. *American Economic Review*, 112(4):1311–1333, 2022.

Catherine Tucker and Nico Neumann. Buying consumer data? tread carefully. *Harvard Business Review*, 2020.

Eric Von Hippel and Marcie J Tyre. How learning by doing is done: problem identification in novel process equipment. *Research policy*, 24(1):1–12, 1995.

A Appendix

A.1 Selling Information

A.1.1 Firm A had its data treated

Firm B has its data treated under the following condition:

$$\hat{\pi} - \underline{\pi} - c\delta(1 - \alpha\delta) > 0$$

Denoting $\tilde{c} = \frac{c}{t}$ we obtain the following participation constraint:

$$\frac{t}{2} - \delta^2 t - \left(\frac{t}{2} - \frac{2\delta t}{3}[1 - \frac{\delta}{3}]\right) - c\delta(1 - \alpha\delta) > 0 \implies \delta > \frac{\tilde{c} - \frac{2}{3}}{\alpha\tilde{c} - \frac{11}{9}}.$$

If $\alpha\tilde{c} > \frac{11}{9}$, we have $\delta > \frac{\tilde{c}-\frac{2}{3}}{\alpha\tilde{c}-\frac{11}{9}} > \frac{1}{2}$ which is a contradiction with $\delta < \frac{1}{2}$ and participation is never satisfied.

If $\frac{11}{\alpha 9} > \tilde{c} > \frac{2}{3}$ we have $\delta > 0 > \frac{\tilde{c}-\frac{2}{3}}{\alpha\tilde{c}-\frac{11}{9}}$ and participation is always satisfied.

Hence we focus our analysis on the meaningful case where $\frac{2}{3} > \tilde{c}$. We denote $\hat{\delta}_{B2} = \frac{\tilde{c}-\frac{2}{3}}{\alpha\tilde{c}-\frac{11}{9}}$, the threshold below which participation is ensured (because the sign of the inequality reverses when dividing by the negative term in $\tilde{c}$).

Condition for $\hat{\delta}_{B2} < \frac{1}{2}$: $\tilde{c} > \frac{1}{9(2-\alpha)}$.

$$\begin{cases} \tilde{c} \in [0, \frac{1}{9(2-\alpha)}] : \text{ participation is ensured for } \delta \in [0, \frac{1}{2}], \\ \tilde{c} \in [\frac{1}{9(2-\alpha)}, \frac{2}{3}] : \text{ participation is ensured for } \delta < \hat{\delta}_{B2} = \frac{\frac{2}{3}-\tilde{c}}{\frac{11}{9}-\alpha\tilde{c}}. \end{cases}$$

A.2 Firm A did not have its data treated by the broker

The profits of the firms are equal to:

$$\begin{cases} \pi_B = \overline{\pi} = \frac{t}{2} + \frac{\delta t}{3}[2 - \frac{7\delta}{3}] \text{ and } \pi_A = \underline{\pi} = \frac{t}{2} - \frac{2\delta t}{3}[1 - \frac{\delta}{3}] \text{ if Firm B has its data treated,} \\ \pi_A = \pi_B = \frac{t}{2} \text{ otherwise.} \end{cases}$$

Firm B has its data treated if $\delta < \hat{\delta}_B = \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right)$.

- $\tilde{c} > \frac{2}{3}$: Firm B does not have its data treated and firms make profits equal to $\frac{t}{2}$.
- $\frac{2}{3} > \tilde{c}$: Firm B does not have its data treated if $\delta > \hat{\delta}_B$ and firms make profits equal to $\frac{t}{2}$.
- $\frac{2}{3} > \tilde{c}$: Firm B has its data treated if $\delta < \hat{\delta}_B$ and firms make profits equal to $\pi_A = \underline{\pi}$ and $\pi_B = \overline{\pi}$.

Summary If Firm A has its data treated:

- $\tilde{c} < \frac{1}{9(2-\alpha)}$: $\pi_A = \hat{\pi} - c\delta$ and $\pi_B = \hat{\pi} - c\delta(1 - \alpha\delta)$.
- $\frac{2}{3} > \tilde{c} > \frac{1}{9(2-\alpha)}$:
  - $\delta < \hat{\delta}_2$: $\pi_A = \hat{\pi} - c\delta$ and $\pi_B = \hat{\pi} - c\delta(1 - \alpha\delta)$.
  - $\frac{1}{2} > \delta > \hat{\delta}_2$: $\pi_A = \overline{\pi} - c\delta$ and $\pi_B = \underline{\pi}$.
- $\tilde{c} > \frac{2}{3}$: $\pi_A = \overline{\pi} - c\delta$ and $\pi_B = \underline{\pi}$.

If Firm A does not have its data treated:

- $\frac{2}{3} > \tilde{c}$:
  - $\delta < \hat{\delta}_B$: $\pi_A = \underline{\pi}$ and $\pi_B = \overline{\pi} - c\delta$.
  - $\frac{1}{2} > \delta > \hat{\delta}_B$: $\pi_A = \pi_B = \frac{t}{2}$.
- $\tilde{c} > \frac{2}{3}$: $\pi_A = \pi_B = \frac{t}{2}$.

A.2.1 Firm A

(i) If $\tilde{c} > \frac{2}{3}$: no treatment.

(ii) If $\frac{2}{3} > \tilde{c} > \frac{1}{9(2-\alpha)}$: comparing $\hat{\delta}_{B2} = \frac{\frac{2}{3} - \tilde{c}}{\frac{11}{9} - \alpha\tilde{c}}$ and $\hat{\delta}_B = \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right)$ we show that $\hat{\delta}_B > \hat{\delta}_{B2} \iff \tilde{c} < \frac{4}{9\alpha}$.

- $\alpha < \frac{2}{3} \implies \hat{\delta}_B > \hat{\delta}_{B2}$
- $\alpha > \frac{2}{3}$ and $\tilde{c} \in [\frac{4}{9\alpha}, \frac{2}{3}] \implies \hat{\delta}_{B2} > \hat{\delta}_B$[8]

[8]Note that a simple verification allows us to ensure that $\frac{1}{2} > \hat{\delta}_B, \hat{\delta}_{B2}$.

- $\delta \in [\hat{\delta}_B, \hat{\delta}_{B2}]$ Firm B has its data treated if Firm A does.
  - Either Firm A has its data treated: makes profits $\pi_A = \hat{\pi} - c\delta$
  - Or Firm A does not have its data treated and makes profits $\frac{t}{2}$
  - So Firm A has its data treated iff $\hat{\pi} - c\delta - \frac{t}{2} > 0$.
  - This condition is never satisfied and firms do not have their data collected.
  - Is there room for treatment with transfer to Firm A?
    * Coalition's gains in the first stage: $\hat{\pi} - c\delta - \frac{t}{2}$.
    * Coalition's gains in the second stage: $\hat{\pi} - \underline{\pi} - c\delta(1 - \alpha\delta)$.
    * Room for treatment with transfer iff $\hat{\pi} - c\delta - \frac{t}{2} + \hat{\pi} - \underline{\pi} - c\delta(1 - \alpha\delta) > 0$ $\iff 2\left(\frac{t}{2} - \delta^2 t\right) - \frac{t}{2} - \left(\frac{t}{2} - \frac{2\delta t}{3}[1 - \frac{\delta}{3}]\right) > c\delta(2 - \alpha\delta)$ $\iff \delta < \frac{\frac{2}{3} - 2\tilde{c}}{\frac{20}{9} - \alpha} = \hat{\delta}_{T2}$.
    * $\hat{\delta}_{T2} > \delta_{B2}$? $\iff \alpha\tilde{c}^2 - \frac{2}{9}\tilde{c} - \frac{2}{3} > 0 \iff \tilde{c} > \frac{\frac{1}{9} + (\frac{1}{81} + \alpha\frac{2}{3})^{1/2}}{\alpha}$
    * To have $\hat{\delta}_{T2} > 0$, we need that $\frac{\frac{1}{9} + (\frac{1}{81} + \alpha\frac{2}{3})^{1/2}}{\alpha} < \frac{1}{3}$. Simple algebra show that this is true for $\alpha > \frac{2}{3}(1 + \frac{7^{1/2}}{3})$.

- $\hat{\delta}_B > \delta$ Firm B has its data treated. Firm A either has its data treated and makes profits equal to $\hat{\pi}$ or not and make profits equal to $\underline{\pi}$.
  - Firm A has its data treated iff $\hat{\pi} - \underline{\pi} > \tilde{c}\delta \iff \frac{t}{2} - \delta^2 t - \frac{t}{2} + \frac{2\delta t}{3}[1 - \frac{\delta}{3}] > \tilde{c}\delta \iff \delta < \frac{9}{11}(\frac{2}{3} - \tilde{c}) = \hat{\delta}_{A2}$.
  - $\hat{\delta}_{A2} < \hat{\delta}_B \iff \frac{9}{11}\left(\frac{2}{3} - \tilde{c}\right) < \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right)$, which is always true.
  - $\frac{1}{2} > \hat{\delta}_{A2} \iff \tilde{c} > \frac{1}{18}$, which is true for $\tilde{c} > \frac{4}{9\alpha}$.
  - As $\hat{\delta}_{A2} < \hat{\delta}_B$, Firm B has its data treated regardless of the decision of Firm A for $\delta < \hat{\delta}_B$.

- $\tilde{c} \in [\frac{1}{9(2-\alpha)}, \frac{4}{9\alpha}] \implies \hat{\delta}_B > \hat{\delta}_{B2}$.
  - $\delta \in [\hat{\delta}_B, \frac{1}{2}]$: Firm B does not have its data treated.
  - Firm A has its data treated iff $\overline{\pi} - c\delta > \frac{t}{2} \iff \delta < \frac{9}{7}(\frac{2}{3} - \tilde{c}) = \hat{\delta}_B$. Hence, Firm A does not have its data treated.
  - $\delta \in [\hat{\delta}_{B2}, \hat{\delta}_B]$: only Firm A has its data treated.

    Indeed, its decision is either getting its data treated and make $\overline{\pi}$, or not and in this case, Firm B has them treated and makes profits equal to $\underline{\pi}$. Hence, there is room for treatment iff $\overline{\pi} - \underline{\pi} \geq \tilde{c}\delta \iff \frac{t}{2} + \frac{\delta t}{3}[2 - \frac{7\delta}{3}] - \frac{t}{2} + \frac{2\delta t}{3}[1 - \frac{\delta}{3}] > \tilde{c}\delta \iff \frac{4}{3} - \tilde{c} > \delta$, which is always true as $\tilde{c} < \frac{4}{3\alpha}$, and Firm A only has its data treated.
  - $\delta < \hat{\delta}_{B2}$: both firms have their data treated

- $\tilde{c} < \frac{1}{9(2-\alpha)}$: both firms have their data treated

At $\alpha = \frac{2}{3}$ there is a new zone that appears, where we have $\delta_B > \delta_2$ if $\tilde{c} < \frac{4}{9\alpha}$.