Data, Product Targeting and Competition*

Alexander Guembel
Toulouse School of Economics

Ulrich Hege
Toulouse School of Economics

Preliminary and incomplete

6 January 2021

Abstract

We analyze competition when two firms use data to locate a consumer's preferred product specification, and then compete in customized product offers and prices. We show that competition becomes fiercer when the less well informed firm improves data accuracy, but not when the better informed firm does so. More information improves consumer surplus in a duopoly, but not in a monopoly since the monopolist price-discriminates more effectively. Consumer surplus is higher in duopoly compared to monopoly, but the same is not true for total surplus. An allocational inefficiency arises when the competitors differ in their information quality.

We investigate multi-platform gatekeepers that can use information learned from purchase decisions in one market to compete in a second market. A gatekeeper may then strategically increase or decrease prices to learn more effectively from the consumer's purchase behavior about product preferences.

We find that strategic investments in data quality behave like strategic substitutes because average profits increase with data quality heterogeneity. We show that an incumbent may overinvest in data to deter entry or accelerate exit of competitors.

Key words: Digital platforms, product targeting, product quality, data and competition.

JEL classification: D82, L13, L15.

*We would like to thank Edward Corcoran, Alexandre de Corniére, Doh-Shin Jeon, Bruno Jullien, Alvaro Martin, Alejandro Neut, Fernando Soto, Pablo Urbiola, and Wilfried Sand-Zantman for helpful suggestions and advice. We gratefully acknowledge financial support from BBVA. All errors are our own.

1 Introduction

1.1 Motivation and contribution

In many digital markets, customized product offers are important to attract consumers and increase consumers' willingness-to-pay, for example media platforms in various forms (news, film, music, social media), subscription-based markets, financial services, search markets or retail platforms. At the same time, digital platforms, content producers and service providers collect vast amounts of customer data and develop prediction algorithms to pinpoint customer preferences and to develop customized product offers to their customers. This paper proposes an analytical framework to investigate the role of data precision in the competition among firms that process customer-specific information and can target their product offers. In our model, two firms use data to extract information about consumers' preferred product specification. The firms have noisy observations about consumer preferences; they simultaneously choose a product specification for each customer, modelled as a location choice, and a price.

Our findings are as follows. We show that, when two firms compete, improving the information of the less well informed firm renders price competition fiercer. This is because on average firms will offer more similar products. By contrast, improving the better informed firm's information has an ambiguous effect on competition as the better informed firm increases its price, taking advantage of the fact that it will in expectation offer a product that is more desirable. Overall, more information makes consumers better off. This is in contrast with the case where the consumer faces a monopolist supplier. Although the latter is able to sell a better targeted product when his information improves, the consumer may be worse off as the monopolist can also price discriminate more effectively.

Although consumer surplus is always higher in duopoly compared to monopoly, the same is not true for total surplus. We show that the duopoly suffers from an allocation inefficiency that is particularly severe when the two firms have very different levels of information quality. In this case, the less well informed firm offers a significantly lower price than the better informed firm. This in turn leads the consumer to choose the less desirable product some of the time - which is socially wasteful. Although the aggregate amount of information is higher under duopoly (as the second firm adds a source of information about consumer preferences), monopoly may dominate in terms of total surplus.

We further investigate how a firm's strategy changes if it can use information obtained in one market in order to compete in another, motivated for instance by strategies of gate-keepers of multiple platforms to leverage their information for data-driven entries in new product markets (including media platforms and financial services). Instead of simply assuming learning effects, we explicitly model learning from a consumer's purchase decision and consider a firm which is a monopolist in one market and then competes with another firm in a second market. The only way for the monopolist to learn about consumer preferences, is by observing whether or not the consumer purchases the good offered by the monopolist. When the monopolist fully covers the market (i.e., the price is low enough that the consumer always buys), it cannot learn anything from the consumer's purchase decision. Equally, when price is so high that the consumer (almost) never buys, no learning is possible. The learning motive pushes the monopolist towards offering a "biased" product (i.e., the product specification will not be equal the firm's unbiased expectation of the consumer's preference) at a price that leads to a purchase probability of one half. As a result, a monopolist may increase or decrease its price in order to learn from the consumer's purchase behaviour about product preferences.

We also explore the strategic role of investment in data quality to gain competitive advantage. We extend the model by adding an initial stage where firms make strategic investments in data quality, before they simultaneously choose product specification and price in the second stage. Motivated by competition concerns about platforms with high data capabilities (i.e. capabilities to access and process customer-specific data), we consider in particular dynamic competition involving entry and exit. We show that an incumbent firm can use investment in data quality as a strategic entry deterrent. Since better information by the incumbent forces a potential entrant to choose a lower price, entry can be deterred by overinvesting in data quality. Similarly, a potential entrant can strategically overinvest in data precision to force the exit of an incumbent with lower data capabilities (higher costs to scale up its data quality). We also consider the strategic interaction of investments in data quality, and show an interesting asymmetry in equilibrium: the lower the investment costs and hence the higher the equilibrium data quality of the firm with higher data capabilities, the lower the data investment of the competing firm. This asymmetry arises because average profits will be higher when data qualities are more heterogeneous. These observations give rise to regulatory interventions to limit the use of exclusive data by firms with high data capabilities.

1.2 Relationship to the literature

A number of papers have studied how data affects competition. One strand of the literature associates data with an increased ability to price discriminate. Firms can move from uniform pricing to discriminatory pricing by gaining access to data (see, among others, Gu, Madio and Reggiani, 2019; Montes, Sand-Zantmann and Valetti, 2019; Belleflamme, Lam and Vergote, 2019; or Taylor and Wagman, 2014). These papers differ from ours in that they assume fixed locations of the competing firms and data therefore does not affect the average desirability of the products on offer. Another strand of the literature allows for data to improve the quality of a product. For example, Prüfer and Schottmüller (2017) and Hagiu and Wright (2020) focus on industry dynamics when sales generate data and data improves product quality (or reduces production costs as in Farboodi, Mihet, Philippon and Veldkamp, 2019). Other papers view data as information about the valuation that a given customer has for a given good and focus on the design of an information structure when there is a market for information (see Bergemann and Bonatti, 2019 and the papers discussed there). Those papers focus on the pricing and structure of information by data intermediaries or producers, but are less interested in how data changes the actual products that are being offered, nor how data affects product market competition. Finally, some papers, such as Casadesus-Masanell and Hervas-Drane (2015) or Jullien, Lefouilli and

Riordan (2020) consider the role of data collection by a web-site or platform when data collection affects the quality of a user's experience and the data can be re-sold.

De Cornière and Taylor (2020) provide a more general approach to study the link between data and competition. They remain agnostic as to the precise mechanism via which data affects consumers and focus on modeling competition in utility. They model data as a positive revenue shifter for a firm and explore under which conditions more data increases or decreases consumers' equilibrium utilities. Although that approach can in principle also encompass utility enhancement via improved product design or targeting, like in our paper, our specific modeling of data and competition generates results that are not nested by their model. In particular, we show that more data available to one firm may actually decrease its revenues or profits. This result violates the key assumption of data as a positive revenue shifter in de Cornière and Taylor (2020) and is due to the effect that data has on the equilibrium degree of competition.

Our paper is generally related to the literature of the economics of privacy in the digital age, following the seminal paper by Varian (1997) and summarized in the survey by Acquisti, Taylor, and Wagman (2016). We contribute to this vast and diverse literature by explicitly considering the cost at which firms can improve their data quality. We show that consumers' choice to voluntarily relinquish data will not only have an effect on the fit of customized product offers but also influence competition. There are behavioral dimensions to the importance of product targeting in digital markets: for example in markets for subscriptions and streaming services, high prices can only be sustained if the product offered captures a high level of attention and loyalty from consumers. We add to the literature on the economics of attention that has focused on product features (e.g., Bordalo, Gennaoli, Schleifer 2016; Anderson and DePalma, 2012; de Clippel, Eliaz, Rozen, 2014) the dimension of product targeting in this competition.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 2.1 analyzes the monopoly case, and Section 2.2 the baseline duopoly case. Section 3 undertakes a welfare commparison between these two cases, and Section 4 looks at the endogenous choice of data quality and its role in entry deterrence. Section **??** considers further aspects that are relevant for regulation, in particular information spillovers across markets. Section 6 discusses robustness issues, and Section 7 concludes.

2 The Model

There is one consumer with a taste parameter η drawn from a uniform (either on a large circle or the entire real line) for a good of which he wishes to purchase a single unit.

There are two firms A and B. Each firm receives a signal $x_i = \eta + \tilde{e}_i$, about the consumer's taste parameter, where $\widetilde{e}_i$ is uniformly distributed on the interval $[-\varepsilon_i, \varepsilon_i]$ $(i = A, B)$. ε_i is a parameter measuring the information precision and we assume without loss of generality that $\varepsilon_A \geq \varepsilon_B$. One could think of firm B as being an incumbent who may have informational advantage over entrant A, and we will often refer to firms A and B as entrant and incumbent, resp. Alternatively, B could be a firm that has gathered information about the consumer's preferences from observing the consumer's behavior on another market (or multiple markets) or a platform.

After having received the signal, each firm simultaneously chooses a location l_i and a price p_i for the good. The consumer values the good offered by seller i at $v - |\eta - l_i|$ which expresses the consumer's willingness-to-pay. When the consumer's participation constraint is met, the consumer surplus when buying form seller i is $v - |\eta - l_i| - p_i \geq 0$.

2.1 Equilibrium with a Monopolistic Firm

As a benchmark, we can calculate the monopolist's price and profits. For ease of exposition we omit the subscript i in this Section.

Lemma 1 *The monopolist's optimal pricing strategy and corresponding profits are given* *by*

$$p_M = \begin{cases} v - \varepsilon & \textit{if} \quad \varepsilon \leq \frac{v}{2} \\ \frac{v}{2} & \textit{if} \quad \varepsilon > \frac{v}{2} \end{cases}; \tag{1}$$
$$\pi_M = \begin{cases} v - \varepsilon & \textit{if} \quad \varepsilon \leq \frac{v}{2} \\ \frac{v^2}{4\varepsilon} & \textit{if} \quad \varepsilon > \frac{v}{2} \end{cases}$$

Proof. The monopolist sells one unit if

$$v - p - |\widetilde{e}| \geq 0.$$

Hence, if $p \leq v - \varepsilon$, the monopolist always sells. When $p > v - \varepsilon$ the monopolist only sells if $|\widetilde{e}| \leq v - p$, which happens with probability $\frac{v-p}{\varepsilon}$. The monopolist's expected profits are therefore given by

$$\pi_M(p) = \begin{cases} p & \text{if} \quad p \leq v - \varepsilon \\ p\frac{v-p}{\varepsilon} & \text{if} \quad p > v - \varepsilon \end{cases}$$

On the interval $p \leq v - \varepsilon$ profits are obviously maximized at $p = v - \varepsilon$. On the interval $p > v - \varepsilon$, we can take the first-order condition

$$\frac{v - 2p}{\varepsilon} = 0$$

to yield the optimal price $p = \frac{v}{2}$ and corresponding profits $\pi_M(p > v - \varepsilon) = \frac{v^2}{4\varepsilon}$. The firm therefore sets $p = \frac{v}{2}$ when $\frac{v}{2} > v - \varepsilon$, i.e., when $\varepsilon > \frac{v}{2}$, and sets $p = v - \varepsilon$ otherwise. Profits follow directly. ■

When the monopolist has strong information ($\varepsilon \leq \frac{v}{2}$) he sets the price just low enough to always serve the customer. When the information becomes weaker, serving the customer would require a pretty low price, at which point the firm prefers to maintain a high price and not to sell some of the time (when the product offered is relatively far away from the customer's preferred product).

Given the pricing strategy, we can calculate consumer and total surplus.

Lemma 2 *Consumer surplus CS_M and total surplus TS_M are given by*

$$CS_M = \begin{cases} \frac{\varepsilon}{2} & \text{if} \quad \varepsilon \leq \frac{v}{2} \\ \frac{1}{2}\frac{v^2}{4\varepsilon} & \text{if} \quad \varepsilon > \frac{v}{2} \end{cases}, \tag{2}$$

$$TS_M = \begin{cases} v - \frac{\varepsilon}{2} & \text{if} \quad \varepsilon \leq \frac{v}{2} \\ \frac{3}{2}\frac{v^2}{4\varepsilon} & \text{if} \quad \varepsilon > \frac{v}{2} \end{cases} \tag{3}$$

Proof. For $\varepsilon \leq \frac{v}{2}$ we have $p = v - \varepsilon$ and consumer surplus is given by

$$\begin{aligned} CS_M &= \frac{1}{\varepsilon}\int_0^{\varepsilon} (v - (v - \varepsilon) - \tilde{e})\, d\tilde{e} \\ &= \frac{\varepsilon}{2}. \end{aligned}$$

For $\varepsilon > \frac{v}{2}$, we have $p = \frac{v}{2}$ and the consumer only purchases the good if $\tilde{e} \leq v - p = \frac{v}{2}$. Hence,

$$\begin{aligned} CS_M &= \frac{1}{\varepsilon}\int_0^{\frac{v}{2}} \left(\frac{v}{2} - \tilde{e}\right) d\tilde{e} \\ &= \frac{1}{2}\frac{v^2}{4\varepsilon}. \end{aligned}$$

Total surplus follows directly by taking $TS_M = CS_M + \pi_M$. ■

From Lemma 2 we can see that total welfare always decreases in the monopolist's information quality. This makes intuitive sense, as worse information implies that, on average, the firm offers a less well targeted product. More interestingly, consumer surplus is not monotonic in information quality. For high levels of information quality (low ε), the consumer would benefit from providing less information to the monopolist. This is because worse information induces the monopolist to lower the price in an attempt to ensure that the consumer purchases even a less well targeted product. Both price and average quality thus drop in information quality. Since the lower price applies to all product varieties that may be offered (infra-marginal effect), the price effect dominates. This is true up to the point $\varepsilon = \frac{v}{2}$. When information worsens beyond that point, the monopolist prefers not to drop the price any further and instead accepts that the consumer sometimes makes no purchase at all. Hence, the market is no longer fully covered. Worsening information quality in this region reduces consumer surplus because the consumer only experiences its negative side, that is, an increasing likelihood of making no purchase at all.

2.2 Equilibrium with Duopoly

We start by characterizing each firm's best response to a given strategy by the other firm. We conjecture and prove below that in equilibrium each firm chooses as a location the unbiased expectation of η, i.e., each firm chooses $l_i = x_i$. Given these location choices, firm A sells its good if it is more attractive to the consumer than B's offer, i.e., when

$$v - p_A - |\tilde{e}_A| \geq v - p_B - |\tilde{e}_B|\,, \tag{4}$$

and when the purchase dominates no purchase, i.e., when

$$v - p_A - |\tilde{e}_A| \geq 0. \tag{5}$$

Firm B sells if inequality (4) is reversed and when

$$v - p_B - |\tilde{e}_B| \geq 0. \tag{6}$$

This allows us to calculate the probabilities of A or B selling the good, as follows. We conjecture (and later prove) that the better informed firm (A) uses its informational advantage to charge at least as high a price as its rival, i.e., $p_A \geq p_B$. Moreover, we conjecture (and later prove) that the price difference is bounded by the informational difference; specifically, $p_B - p_A \leq \varepsilon_A - \varepsilon_B$. Given these conjectures, we can calculate the probability that either firm sells its product as a function of both firms' prices. The shape of this function depends on whether there are realizations of $\widetilde{e}_A, \widetilde{e}_B$ such that the consumer does not purchase the good at all. (6) is binding for some $\widetilde{e}_B$, if $p_B > v - \varepsilon_B$. Under the condition $p_B - p_A \leq \varepsilon_A - \varepsilon_B$, $p_B > v - \varepsilon_B$ implies $p_A > v - \varepsilon_A$. In other words, if p_B is so high that the consumer would sometimes prefer not to buy the good at all, then the boundary on price differences implies that also firm A cannot sell its product for high enough $\tilde{e}_A$. Moreover, whenever $p_B \leq v - \varepsilon_B$, the constraint (5) becomes irrelevant for firm A, because before it binds, A would already have lost out against B (i.e., (4) is reversed for any $\tilde{e}_A$ and $\tilde{e}_B$ that satisfy $v - p_B - |\tilde{e}_B| \geq 0$ and $v - p_A - |\tilde{e}_A| < 0$.)

$$\Pr\left(Sell_A\right) = \begin{cases} \frac{p_B - p_A}{\varepsilon_A} + \frac{1}{2}\frac{\varepsilon_B}{\varepsilon_A} & \text{if} \quad p_B \leq v - \varepsilon_B \\ \frac{p_B - p_A}{\varepsilon_A} + \frac{1}{2}\frac{\varepsilon_B}{\varepsilon_A} - \frac{\frac{1}{2}(p_B - (v - \varepsilon_B))^2}{\varepsilon_A \varepsilon_B} & \text{if} \quad p_B > v - \varepsilon_B \end{cases} \tag{7}$$

and

$$\Pr\left(Sell_B\right) = \begin{cases} 1 - \frac{p_B - p_A}{\varepsilon_A} - \frac{1}{2}\frac{\varepsilon_B}{\varepsilon_A} & \text{if} \quad p_B \leq v - \varepsilon_B \\ \frac{\frac{1}{2}(v - p_B)^2}{\varepsilon_A \varepsilon_B} + \frac{(p_A - (v - \varepsilon_A))(v - p_B)}{\varepsilon_A \varepsilon_B} & \text{if} \quad p_B > v - \varepsilon_B \end{cases} \tag{8}$$

We say that the "market is fully covered" when there will always be a sale by one of the two firms. It is useful to note that a sale that this is only the case when $p_B \leq v - \varepsilon_B$ is firm B's equilibrium price. When $p_B > v - \varepsilon_B$ then there is no sale becaue the consumérs reservation value is not met by the combination of price and product location offers with $\Pr\left(No\ Sale\right) = \frac{\frac{1}{2}(p_A - (v - \varepsilon_A))(p_B - (v - \varepsilon_B))^2}{\varepsilon_A \varepsilon_B} > 0$. As we will show, whether the market is fully covered or not depends on the ratio between value v and noise ε_B of firm B.

We can express the expected profits as

$$\pi_A = \Pr\left(Sell_A\right) p_A, \tag{9}$$

$$\pi_B = \Pr(Sell_B) p_B. \tag{10}$$

Depending on parameter values, the equilibrium may feature low enough prices, such that the consumer always buys the good, or high prices such that the consumer sometimes does not make any purchase at all. Similar to the monopoly case, we get full market coverage when v is relatively large compared to ε_A and ε_B. For a full description it is useful to introduce the following functions, which turn out to be important thresholds of v required in the description of equilibrium

$$f\left(\varepsilon_A, \varepsilon_B\right) = \varepsilon_A + \frac{7}{4}\varepsilon_B - \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2},$$
$$g\left(\varepsilon_A, \varepsilon_B\right) = \frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B,$$

where $f(\varepsilon_A, \varepsilon_B) \leq g(\varepsilon_A, \varepsilon_B)$.

Equilibrium is characterized as follows:

Proposition 1 *There exists an equilibrium in which each firm chooses $l_i = x_i$, and equilibrium prices depend on thresholds of v as follows:*

(i) If $g(\varepsilon_A, \varepsilon_B) < v$, then

$$p_A = \frac{\varepsilon_A}{3} + \frac{\varepsilon_B}{6} \tag{11}$$

$$p_B = \frac{2\varepsilon_A}{3} - \frac{\varepsilon_B}{6}. \tag{12}$$

(ii) If $v \in [f(\varepsilon_A, \varepsilon_B), g(\varepsilon_A, \varepsilon_B)]$ then

$$p_A = \frac{v}{2} - \frac{\varepsilon_B}{4}, \tag{13}$$

$$p_B = v - \varepsilon_B. \tag{14}$$

(iii) If $v < f(\varepsilon_A, \varepsilon_B)$ then prices are given by the unique solution of the following two equations on the interval $p_B \in \left(v - \varepsilon_B, \frac{v}{2}\right)$:

$$p_A = \frac{v}{2} - \frac{1}{4}\frac{(v - p_B)^2}{\varepsilon_B}, \tag{15}$$

$$p_A = \frac{1}{2}\frac{3p_B - v}{v - 2p_B} + v - \varepsilon_A. \tag{16}$$

Proof see Appendix.

As A's information gets more precise, the market gets more competitive in equilibrium. This is because, on average, firms will produce more similar products, giving each less market power. However an improvement in B's information makes B less competitive but A more competitive. The price effect is symmetric, but since the consumer ends up buying B's product with a higher likelihood, the average price paid actually increases with B's information. These effects will be important when we consider the impact of information changes on firm profit and entry.

For the rest of the paper we will assume that information is relatively precise such that

$$\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq v \tag{17}$$

holds.

3 Welfare in Duopoly

We start by analyzing consumer and total surplus.

Lemma 3 *Consumer surplus CS_D and total surplus TS_D are given by*

$$CS_D = v - \frac{p_A^2}{\varepsilon_A} - \frac{p_B^2}{\varepsilon_A} - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\}, \tag{18}$$

$$TS_D = v - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\} \tag{19}$$

where $p_A = \frac{\varepsilon_A}{3} + \frac{\varepsilon_B}{6}$ and $p_B = \frac{2\varepsilon_A}{3} - \frac{\varepsilon_B}{6}$ (from (11) and (12))

Proof. Since we are focusing on the case $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq v$ where the market is fully covered, consumer surplus can be decomposed into the following effects. First, the consumer always receives value v from consuming the product. Second, there is an "allocational" loss stemming from the fact that the chosen product does not coincide with the consumer's favoured specification. When the consumer buys from A this loss is $|\tilde{e}_A|$ while it is $|\tilde{e}_B|$ when he buys from B. We denote the expected loss by L. Moreover, the consumer pays a price p_A or p_B for the product. Since marginal production costs are zero, the expected price paid also corresponds to the firm's expected profits. We can thus write

$$CS_D = v - \pi_A - \pi_B - L.$$

Using $\Delta p \equiv p_B - p_A$, the expected allocational loss can be calculated as

$$\begin{aligned} L = &\int_0^{\Delta p} \tilde{e}_A \frac{d\tilde{e}_A}{\varepsilon_A} + \int_{\Delta p}^{\Delta p + \varepsilon_B} \int_{\tilde{e}_A - \Delta p}^{\varepsilon_B} \tilde{e}_A \frac{d\tilde{e}_B}{\varepsilon_B}\frac{d\tilde{e}_A}{\varepsilon_A} \\ &+ \int_{\Delta p}^{\Delta p + \varepsilon_B} \int_0^{\tilde{e}_A - \Delta p} \tilde{e}_B \frac{d\tilde{e}_B}{\varepsilon_B}\frac{d\tilde{e}_A}{\varepsilon_A} + \int_{\Delta p + \varepsilon_B}^{\varepsilon_A} \int_0^{\varepsilon_B} \tilde{e}_B \frac{d\tilde{e}_B}{\varepsilon_B}\frac{d\tilde{e}_A}{\varepsilon_A} \\ = &\frac{1}{\varepsilon_A}\left\{\frac{(\Delta p)^2}{2} + \frac{\varepsilon_A\varepsilon_B}{2} - \frac{\varepsilon_B^2}{6}\right\}. \end{aligned} \tag{20}$$

Firm profits are given from (9) and (10). Applying equilibrium prices (11) and (12) we get

$$\pi_A = \frac{p_A^2}{\varepsilon_A}, \pi_B = \frac{p_B^2}{\varepsilon_A}.$$

■

We can then turn to a comparison of the monopoly and duopoly cases. In this comparison we focus on the case where (17) holds and the market is thus fully covered under a duopoly. Moreover, since the two firms in duopoly potentially have different information quality, one needs to make an assumption about how the monopolist's information quality compares to each of the duopolist's. We assume that the better informed firm in duopoly (firm B) has the same information quality as the monopolist. Thus, the duopoly can be thought of as the monopoly firm facing entry from a less well informed rival. Assuming the contrary would stack the comparison in favour of duopoly, simply because the available information of one of the firms improves. We also assume that the monopolist will always make a sale, $\varepsilon_B \leq \frac{v}{2}$ (see eq. (1) and the proof of Lemma 1).

Proposition 2 *Total surplus under monopoly is higher than under duopoly if* $\varepsilon_B < \frac{\sqrt{3}-1}{2}\varepsilon_A$ *and lower otherwise. Consumer surplus is always higher under duopoly.*

Proof see Appendix.

Moving from monopoly to duopoly affects total surplus in two ways. First, there is a second firm with an independent signal about the consumer's preferred product specification. Although the quality of the second firm's information is inferior, its signal does contain additional information. As the consumer can choose between two products under the duopoly, on average, he can thus choose a better suited product. This effect plays in favour of duopoly and is particularly important when the monopolist has fairly noisy information. Second, competition between two firms generates a distortion in the allocation of the good. In order to compete, the less well informed firm (A) sets a lower price. This means that firm A sometimes sells the good, even though the consumer prefers B's specification. This misallocation becomes more pronounced when the informational gap between

A and B increases and hence the difference in prices is larger. Taken together, the two effects make total monopoly surplus higher than duopoly surplus when ε_B is fairly small compared to ε_A. In that case, adding a second (inferior) signal has little value since (i) the monopolists' information is already very good, and (ii) with a much worse informed entrant the two firms charge very different prices, generating a more significant misallocation.

Consumer surplus, on the other hand, is always higher under duopoly as competition ensures that prices are lower.

Let us consider now the consumer's incentives to provide more information. Unlike in the monopoly case, the consumer now always benefits from improving the information available to the incumbent.

Proposition 3 *Total and consumer surplus are increasing in either firms' information quality, i.e.,*

$$\frac{\partial TS_D}{\partial \varepsilon_A} < 0, \ \ \textit{and} \ \ \frac{\partial CS_D}{\partial \varepsilon_A} < 0,$$
$$\frac{\partial TS_D}{\partial \varepsilon_B} < 0, \ \ \textit{and} \ \ \frac{\partial CS_D}{\partial \varepsilon_B} < 0.$$

Proof. Taking the derivative of (19) yields

$$\frac{\partial TS_D}{\partial \varepsilon_A} < 0 \Longleftrightarrow -\frac{4}{9}\varepsilon_A^2 - \frac{2}{9}\varepsilon_B^2 < 0,$$

which is true. Moreover,

$$\frac{\partial TS_D}{\partial \varepsilon_B} < 0 \Longleftrightarrow \varepsilon_B < \frac{7}{4}\varepsilon_A,$$

which is true since $\varepsilon_B \leq \varepsilon_A$.

Taking the derivative of (18) yields

$$\frac{\partial CS_D}{\partial \varepsilon_A} < 0 \Longleftrightarrow -\frac{11}{18}\varepsilon_A^2 - \frac{1}{18}\varepsilon_B^2 < 0,$$

which always true. Similarly,

$$\frac{\partial CS_D}{\partial \varepsilon_B} < 0 \Longleftrightarrow \varepsilon_B < \frac{5}{2}\varepsilon_A,$$

which holds by our assumption $\varepsilon_B \leq \varepsilon_A$. ∎

An improvement in either firms' information leads to an increase in total surplus, because the consumer gets, on average, a more suitable product. This effect also improves consumer surplus. When the incumbent gets better informed (ε_B decreases), he charges a higher price - an effect we already saw at play for the monopolist. However, in duopoly, the "entrant" (firm A), competes more aggressively when B is better informed. This limits the extent to which the incumbent can take advantage of his better information (the incumbent increases prices less strongly than the monopolist). It also dampens the effect on the average price paid, as the consumer sometimes buys from A whose product becomes cheaper.[1] When A's information improves both firms compete more aggressively. This leads to an unambiguous improvement in consumer surplus.

Note that a consumer's incentives to provide information differ drastically depending on whether the supplier is a monopolist or in competition with another firm. Since a monopolist takes advantage of better information by raising prices significantly, this effect is mitigated (if the better informed firm's information improves), or even reversed, as firms compete more aggressively (if the less well informed firm's information improves). Hence, a consumer benefits from making more information available to competing firms, but suffers from doing the same for a monopoly supplier.

4 Choice of Information Quality and Firm Entry

The analysis so far has focused on a comparison between monopoly and duopoly, without asking under which conditions entry would actually occur, and has considered only exogenously given information quality ε_A and ε_B. In this section, we endogenize information quality and address the question of firm entry and exit since the two issues are closely

[1] Since B's equilibrium probability of selling the good increases when B's information improves, the average price increases. However, this effect is dominated by the improved product choices available to the consumer.

related: information quality ε_A and ε_B is the strategic instrument of interest in our model to influence entry and exit of competitors, helping to understand whether advantages in data collection and data processing can be exploited to limit competition, potentially at the detriment of welfare and consumer surplus. Obviously, the question of entry only becomes relevant when there is an entry cost and the question of exit only when there is a fixed cost of continuing operations, which we assume from now on when relevant. The cost of data strategies can be considered either as a fixed cost of entry (or adjustment), for example investments in data acquisition, hardware and algorithms to learn more about consumer tastes. It can also be considered as a flow cost of operating the data infrastructure (or the contribution to cover the cost of the data infrastructure). In our model of static competition in data strategies and products/prices, there is no difference.

We are in particular interested in understanding the strategic choice of information quality ε_i; the choice is strategic since it influences competition, either competition in prices (as already seen), or the entry and exit of competitors. We make the assumption that competition plays out in two sequential decisions: firms first decide on their data strategies determining ε_i, then on their product location and prices. While in the second round firms move simultaneously, in the first round we consider sequential decision. We introduce a cost of investment information precision $c(\varepsilon_i)$, which we assume to be increasing and convex in the quality of information, i.e. to be decreasing in ε_i since a lower ε_i means better information quality.[2] We write for the profits net of information quality costs, $r(\varepsilon_i) = \pi(\varepsilon_i) - c(\varepsilon_i)$.

We will undertake this analysis in several steps. We look in Subsection 4.1 at the reaction of profits to a change in information quality, before looking at one firm's choice of its data quality ε_i while the information quality of the other firm is fixed (one-sided choice), and then looking at both firms making sequential choices on ε_B and ε_A (two-sided choice) in the following two subsections, and at overinvestment incentives in Subsection 4.4

[2] A possible parametrization is a quadratic cost function of the type $c(\varepsilon_i) = k_i (b_i - \varepsilon_i)^2$, where i denotes a low level of precision that is available without investment $c(\varepsilon_i)$.

4.1 Information quality and profits

The previous description of equilibrium goes through unchanged, except that we have to apply the appropriate profit function to a given firm, depending on whether or not it is the better informed of the two. Doing so gives us the following

$$\pi_A = \begin{cases} \frac{1}{\varepsilon_B}\left(\frac{2}{3}\varepsilon_B - \frac{1}{6}\varepsilon_A\right)^2 & \text{if} \quad \varepsilon_A < \varepsilon_B \\ \frac{1}{\varepsilon_A}\left(\frac{1}{3}\varepsilon_A + \frac{1}{6}\varepsilon_B\right)^2 & \text{if} \quad \varepsilon_A \geq \varepsilon_B \end{cases}, \tag{21}$$

and

$$\pi_B = \begin{cases} \frac{1}{\varepsilon_B}\left(\frac{1}{3}\varepsilon_B + \frac{1}{6}\varepsilon_A\right)^2 & \text{if} \quad \varepsilon_A < \varepsilon_B \\ \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2 & \text{if} \quad \varepsilon_A \geq \varepsilon_B \end{cases}. \tag{22}$$

A comparison between (21) and (22) reveals that $\pi_A > \pi_B \Leftrightarrow \varepsilon_A < \varepsilon_B$. This makes it interesting to look at information choices. We consider first a one-sided decision to invest in information, i.e. only one of the firms has the opportunity to do so, say firm $i \in \{A, B\}$. We analyze firm i's incentives to vary information quality, taking as given that both firms are present in the market. This corresponds to the case where entry costs are zero for both firms.

Lemma 4 *Firm i benefits from a marginal improvement in information if and only if it is the better informed firm, i.e.,*

$$\frac{\partial \pi_i}{\partial \varepsilon_i} < 0 \quad \textit{iff} \quad \varepsilon_i < \varepsilon_j.$$
$$\frac{\partial \pi_i}{\partial \varepsilon_i} > 0 \quad \textit{iff} \quad \varepsilon_i > \varepsilon_j.$$

Proof. Follows directy by calculating the derivative of (22). ■

Let us assume, as we did before, that B is the better informed firm ($\varepsilon_B < \varepsilon_A$). Consider A, the less well informed of the two firms . Maybe somewhat surprisingly, A does not benefit from an improvement in its information. Consider the equilibrium prices (11) and (12) to understand why that is. As ε_A falls (A's information improves), the two firms compete more fiercely with one another. Although A's equilibrium probability of selling increases, the damaging effect the information improvement has on prices more than offsets this to render the overall effect negative. When A is the better informed firm, this argument no longer holds. Firm A can now afford to translate better information into a *higher* price resulting in higher equilibrium profits. Hence, incentives for information production are exactly the opposite for a firm that is already the “information leader”.

4.2 One-sided choice of information quality

We move to the analysis of the strategic choice of firms to make an investment in ε_i at cost $c(\varepsilon_i)$, where $i \in \{A, B\}$. We begin with the case where this is a one-sided decision of one firm but not its rival. Without loss of generality, assume that firm B makes that strategic choice whereas the information quality of A is given as ε_A. To perform a full analysis, we no longer assume that one firm (B) always has better information than the other (A), presumably due to lower cost $c(\varepsilon_B)$ of improving information quality. The timing is as follows: first, firm B chooses ε_B while observing firm's A fixed information quality ε_A. Then both firms privately observe their signal x_i, and choose their product locations and prices simultaneously (ε_B and ε_A are common knowledge at this stage). As before we consider a duopoly and disregard for the moment any entry or exit decision - implicitly, entry costs are so low that both firms will always find it profitable to participate in this market, regardless of ε_A. We introduce the notation $\varepsilon_B^{D*} = \arg\max_{\varepsilon_B} r_B(\varepsilon_B)$, where $r_B(\varepsilon_B) = \pi_B(\varepsilon_B) - c(\varepsilon_B)$ to denote an internal solution for the choice of ε_B, i.e. a solution where the optimal ε_B is not located at one of the boundary points of B's profit function π_B where π_B is not continuously differentiable (these boundary points are $\varepsilon_B \in \left\{0, \varepsilon_A, \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right\}$). The most interesting case is when A is sufficiently competitive so that its information ε_A matters for prices. As shown in Proposition 3, in eqn. (11) and (12), this will be the case when $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B < v$, and hence, since ε_B is now endogenous, a necessary condition is that $\varepsilon_A \leq \frac{3}{2}v$.

We say that company B “leapfrogs” when it chooses $\varepsilon_B < \varepsilon_A$ and it does not leapfrog otherwise.

Proposition 4 *For sufficiently large values of ε_A, firm B will leapfrog and choose $\varepsilon_B = 0$ when cost $c(\varepsilon_B)$ is small, and $\varepsilon_B^{D*} \in (0, \varepsilon_A)$ when cost $c(\varepsilon_B)$ is large. For small values of ε_A, firm B will not leapfrog and choose $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$ for small cost $c(\varepsilon_B)$, and $\varepsilon_B^{D*} \geq \frac{6}{5}v - \frac{4}{5}\varepsilon_A$ for large $c(\varepsilon_B)$.*

Proof see Appendix.

The intuition follows closely that of Proposition 4: The profit function of firm B is V-shaped, with a trough when both firms have identical information precision, $\varepsilon_A = \varepsilon_B$. At this symmetrical point, price competition is maximal, meaning that when the firm with lower information quality increases its information quality, both firms' profits decline. When the firm increases its information quality beyond the point of symmetric information precision, $\varepsilon_A = \varepsilon_B$, it becomes the industry leader in information precision, then firms become more heterogeneous and less competitive. Such differentiation in information benefits the information leader, but is detrimental for the less informed firm that now makes a sale less often has to compete in prices more aggressively. Thus, when ε_A is small, firm B benefits from increasing ε_B and hence softening price competition. But only up to the point where the lessening of information precision means that the market is no longer fully covered in equilibrium. From that point onward, the loss of sales hurts more than the continued relaxation of competitiveness, and B's profits decline. When the information cost $c(\varepsilon_B)$ matter, B's optimal equilibrium information quality changes gradually, and internal optima of ε_B become likely, but the structure of equlibrium remains intact.

From this starting point, we develop the analysis in two directions: first, understanding the strategic interaction when both firms choose their information quality. Second, entry and exit.

4.3 Two-sided choice: Differentiation of data quality

We next consider the case where both firms make a strategic choice on ε_i (at cost $c(\varepsilon_i)$) before engaging in competition about location and price. We assume again that firm B

has a Stackelberg leadership position in information choices. The timing is as follows: first, firm B chooses ε_B as the Stackelberg leader. Then firm A observes ε_B and chooses ε_A. Both firms then observe privately their signal x_i, and simultaneously choose their product locations and prices (with ε_B and ε_A being common knowledge). For simplicity, we ignore the information cost $c(\varepsilon_A)$, assuming that it is negligible in our formal analysis. We add an informal discussion at the end about the case where cost $c(\varepsilon_A)$ are non-trivial.

Proposition 5 *Firm B chooses $\varepsilon_B = 0$ when cost $c(\varepsilon_B)$ is small, and $\varepsilon_B^{D*} \in \left(0, \frac{2}{3}v\right)$ when cost $c(\varepsilon_B)$ is large. In either case, firm A chooses $\varepsilon_A = \frac{3}{2}v - \frac{5}{4}\varepsilon_B$, and hence $\varepsilon_B < \varepsilon_A$ in equilibrium.*

Proof see Appendix.

The result shows that there are two local optima for the Stackelberg leader B, but that B always prefers the choice $\varepsilon_B < \varepsilon_A$, i.e. B will "leapfrog": B chooses $\varepsilon_B = 0$ or the closest level $\varepsilon_B > 0$ that maximizes its net profit $r_B(\varepsilon_B) = \pi_B(\varepsilon_B) - c(\varepsilon_B)$. Firm A reacts by choosing a rather low level of information, $\varepsilon_A = \frac{3}{2}v - \frac{5}{4}\varepsilon_B$, more precisely the lowest information precision that still allows the market to be fully covered (recall that if B was a monopoly it would not cover the market even when $\varepsilon_B = 0$, and the same is true when B enjoys a quasi-monopoly because A's information is too imprecise). A plays as an ambiguous role as an opportunistic niche supplier: it maintains uncompetitive prices and hopes to be a "chance" supplier by fishing for customers that are distant from B. This role is reflected in prices and profits: for $\varepsilon_B = 0$, B's price is the monopoly price $p_B = v$, and B makes the sale with probability $\frac{2}{3}$, earning duopoly profit is $\pi_B = \frac{2}{3}v$. Firm A reacts with the equilibrium choice of $\varepsilon_A = \frac{3}{2}v$ and a price $p_A = \frac{v}{2}$ that characterizes its role as fringe competitor for distant customers, allowing A to conclude a sale with probability $\frac{1}{3}$. The consumer surplus is $CS = \frac{1}{12}v$, barely above the monopoly level. This result highlights our result that firms will choose opposite positions of information precision to ease price competition.

We add an informal discussion of the case when cost $c(\varepsilon_A)$ are not negligible. A moderate increase in $c(\varepsilon_A)$ will leave optimal A's best response unchanged at $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v - \frac{5}{4}\varepsilon_B$ since A's profit function is not continuously differentiable around the best response $\varepsilon_A(\varepsilon_B)$. A large increase in $c(\varepsilon_A)$, however, will lead to a decrease in information precision, hence an increase in ε_A. The optimal reaction of ε_B can be derived from applying the implicit function theorem to the net profit function $r_B(\varepsilon_B) = r_B(\varepsilon_B(\varepsilon_A))$ as a function of the increase in $c(\varepsilon_A)$ and increase in ε_A. It can be shown that $\frac{\partial \varepsilon_B(\varepsilon_A)}{\partial \varepsilon_A} > 0$, so the two information precision levels behave like strategic complements.

We conclude with an informal discussion of the case when cost $c(\varepsilon_B)$ are so substantial that B's optimal no longer satisfies $\varepsilon_B \leq \frac{3}{2}v$, and hence the market is no longer always covered. In this case, in principle, firm A is tempted to leapfrog by choosing $\varepsilon_A < \varepsilon_B$ and even, as shown in the best response function in the proof, $\varepsilon_A(\varepsilon_B) = 0$ were $c(\varepsilon_A)$ to be small enough. However, when we introduce the assumption that the Stackelberg leader B has lower cost of information, $c(\varepsilon_B) < c(\varepsilon_A)$, then in equilibrium we will always have $\varepsilon_B < \varepsilon_A$. Both firms still benefit from information differentiation so B's choice of a higher ε_B will induce a higher ε_A, but also decrease consumer's willingness to pay that translates into lower market coverage and hence a need to reduce prices. There will be either an internal equilibrium, balancing B's gain from saving on information cost against the loss form lower willingness to pay, or a boundary solution where B will choose a level of information precision that is just large enough to discourage A from leapfrogging.

4.4 Entry, exit and choice of information quality

We now consider that A's entry into the market is not certain, and will only occur if A's expected net profit is positive, after subtracting entry cost $F_A > 0$. One might think of the entry cost as being linked to the information production necessary in order to compete successfully. This could be a direct cost, for example, payments made to an information intermediary. However, it could also be an indirect cost related to costly activities that allow the generation of information. For example, Google offers some (costly) products free of charge to an end user (e.g., the search engine or maps). Such services generate information for Google about end users, which it can then use, for example to offer well-targeted offers for other searches of the consumer (for example, in Google Shopping which makes life for independent price comparison sites increasingly harder). The choice of ε_B is then Google's precision with which it can push search results that meet the consumer's preferences, and prevent from going to other shopping comparison sites; ε_A is the quality of the competing offer of an independent product comparator. Our analysis of entry of A essentially carries over, with changes only in interpretation rather than formalities, when instead of blocking entry there is an opportunity to accelerate exit of a competitor: suppose A is already in the market, but faces a fixed cost of F_A when it stays in the market rather than leaving it.

While we have analyzed before (in Sections 4.2 and 4.3) the information choice of B in a duopoly, the novelty now is that firm B may enjoy a monopoly when it is able to thwart entry of firm A. We add the superscript D to denote profits in the duopoly situation when A enters the market, hence use $\pi_A^D(\varepsilon_A, \varepsilon_B)$ and $\pi_B^D(\varepsilon_A, \varepsilon_B)$ to denote the two firms' duopoly profit functions in this case. We use the superscript M for monopoly. The relevant consideration for firm B is not just the reaction of duopoly profit $\pi_B^D(\varepsilon_A, \varepsilon_B)$ to a change in ε_B, but also to take into account that an investment into ε_B may deter entry and hence lead to a switch to higher monopoly profits, $\pi_B^M(\varepsilon_B)$, for firm B. We are particularly interested in the question whether the threat of entry can lead to a distortion in the incentive to invest in ε_B, in order to secure a monopoly. We again denote B's optimal information choice in the duopoly case by $\varepsilon_B^{D*} = \arg\max_{\varepsilon_B} r_B^D(\varepsilon_B, \varepsilon_A)$, and introduce the notation $\varepsilon_B^{M*} = \arg\max_{\varepsilon_B} r_B^M(\varepsilon_B)$ for the monopoly case, where $r_B^M(\varepsilon_B) = \pi_B^M(\varepsilon_B) - c(\varepsilon_B)$ is B's net profit.

We focus on a qualitative result on overinvestments. The interesting case, therefore, arises when $c(\varepsilon_B)$ is sufficiently high so that an internal optimum with $\varepsilon_B > 0$ occurs, reflected in our assumption $\varepsilon_B^{M*} > 0$ below.[3] We find that the result is essentially the

[3]If the cost $c(\varepsilon_B)$ is negligible, and provided that ε_A is not too small, then firm B will always choose

same for the situation of one-sided information quality choice (Section 4.2) as for the case of a two-sided information quality choice (Section 4.3). As in Section 4.3, we find that the parameter restrictions on ε_A are less restrictive in the case of a two-sided information quality choice because A's endogenous choice will rule out very imprecise information choice in the case of d fo the

To be specific about the set-up, in the one-sided case we consider that only firm B optimally chooses its information quality ε_B, at cost $c(\varepsilon_B)$, whereas firm A's information precision ε_A is fixed. In the one-sided case, we assume that first firm B and then firm A makes the choice In all other aspects, the set-up is identical to that in Section 4.2. In particular, we abstract again from A's information cost $c(\varepsilon_A)$. We assume $\varepsilon_A < \frac{2}{3}v$ in this case.[4] In the two-sided case, we limit attention to the case when $\varepsilon_B < \frac{3}{2}v$ which implies that cost $c(\varepsilon_B)$ is not too large, and for the remainder the set-up is identical to that in Section 4.2.We find:

Proposition 6 *Suppose $c(\varepsilon_B)$ is sufficiently large so that the incumbent firm B would choose $\varepsilon_B^{M*} > 0$ (internal optimum in the monopoly case). If there is a threat of entry of A (or an option to force exit of A) then there are values of $F_A > 0$ so that B will choose $\hat{\varepsilon}_B < \varepsilon_B^{M*}$ if the choice of $\hat{\varepsilon}_B$ can deter entry (force exit) of A.*

Proof see Appendix.

To see the intuition behind this result, consider a one-sided choice by B when the boundary value of $\varepsilon_B = 0$ is optimal for B, which implies small $c(\varepsilon_B)$ and that ε_A cannot be too small. Assume $\varepsilon_A = \frac{3}{2}v$ which is the best case for A (and A's best response when it can choose ε_A). With the choice $\varepsilon_B = 0$, the Stackelberg leader B maximizes entry deterrence, but still it cannot squeeze A's profits upon entry below $\pi_A^D = \frac{v}{6}$. Hence B can

maximal information precision, $\varepsilon_B = 0$, in a duopoly as shown in Propositions 4. It would do the same as a monopolist since its profit is then even more decreasing in ε_B, at least for small ε_B. This implies that overinvestment in information quality is not possible.

[4] This assumption guarantees that for all choices $\varepsilon_B \leq \varepsilon_A$, the market remains fully covered, i.e. Proposition 3, cases (i) or (ii) apply.

only keep the entrant A out if $F_A > \frac{1}{6}v$. If this is the case, then B enjoys a monopoly with $\pi_B^M = v$. By contrast, if $F_A \leq \frac{v}{6}$, then A enters, and it will choose price $p_A = \frac{v}{2}$. B makes the sale with probability $\frac{2}{3}$, A with probability $\frac{1}{3}$, and we have profits $\pi_B^D = \frac{2}{3}v$ and $\pi_A^D = \frac{1}{6}v$. So there is a substantial drop in profit for B when accomodating entry, and hence B is willing to invest in entry deterrence by having better information. This analysis carries over mutatis mutandis to the case of an internal optimum $\varepsilon_B^{M*} > 0$, and hence explains why B prefers a choice $\hat{\varepsilon}_B < \varepsilon_B^{M*}$ in this case.

Entry is also beneficial for the consumer. In our example, if $F_A < \frac{v}{6}$, then A enters, and the consumer surplus is $CS = \frac{1}{12}v$. But if $F_A > \frac{v}{6}$, then A does not enter, B always makes an efficient sale and captures the entire trading surplus ($CS = 0$). But entry leads to the occasional allocation of the consumer to the niche supplier A hence it decreases total welfare TS. The same is tru for the threat of entry: by inducing a choice $\hat{\varepsilon}_B < \varepsilon_B^{M*}$, it increases consumer surplus but decreases total surplus, the latter being the consequence that the unincumbered monopolist internalizes all welfare effects when choosing ε_B.

5 Across market data spillovers

We now consider the case where there are two markets $j = 1, 2$ and one firm, say firm A is only active in market 1, while firm B is active in both markets. We are interested in the possibility that firm B may learn something from one market that is relevant for the other. With firm B, we have in mind in particular gatekeepers with simultaneous presence on multiple platforms: for example Apple or Google collect information on a consumer from her use of the App Store/Google Play that it can then use to offer well-targeted offers elsewhere, such as in Apple TV or Google Shopping. What we want to highlight here is the possibility that firm B behaves strategically in one market in order to gain better information about the other market. We assume that firm B does not receive any additional exogenous information through its presence in a second market (that channel for spillovers is covered in Section) and show how information spillovers arise endogenously.

The mechanism is related to Taylor (2004) who investigates a monopolist's dynamic pricing strategy and shows that in a first period a monopolist may want to set a high price in order to identify customers with a high willingness to pay, allowing more effective price discrimination in the second period. Acquisti and Varian (2005) also consider conditioning prices on past purchase history. Our setting differs in that the information is used in order to design a better targeted product. Moreover, and unlike Taylor (2004), we find that optimal learning may require a price reduction in the first market (period).

Let market 2 be identical to market 1, i.e., in each market there is a single consumer with a valuation v_j for a unit of a good. Suppose also that the consumer has the same preference η in both markets and firm B receives a single signal $x_B = \eta + \widetilde{e}_B$ about that preference. In order to allow learning across markets, we allow the timing to be such that firm B can first make an offer and observe the consumer's decision (buy or don't buy) and then make an offer in market 1. Assume also that the consumer's purchase behaviour in market 2 is not strategic. This could be justified, for example, in a context where the identity of the consumer in the two markets is not the same, and the firm learns about the preference η of a particular profile of consumers.

Consider firm B's optimal behaviour in market 2. If the firm sets a price $p_2 = v_2 - \varepsilon_B$, the consumer will always purchase the good. The firm will thus not learn anything. If the firm sets a higher price instead, such that for some realizations of $\widetilde{e}_B$ the consumer does not purchase the good, then this allows some inference over η. Suppose firm B chooses a location $l_2 \in [x_B - \varepsilon_B, x_B + \varepsilon_B]$ and sets a price $p_2 > v_2 - \varepsilon_B$. Then there must exist realizations $\widetilde{e}_B$ such that the consumer rejects the good and the firm learns that either $\eta > l_2 + v_2 - p_2$ or $\eta < l_2 - (v_2 - p_2)$. Suppose the firm locates close to the middle such that $l_2 + v_2 - p_2 < x_B + \varepsilon_B$ and $l_2 - (v_2 - p_2) > x_B - \varepsilon_B$. Failing to sell the good then implies that $\eta \in (x_B - \varepsilon_B, l_2 - (v_2 - p_2)) \cup (l_2 + v_2 - p_2, x_B + \varepsilon_B)$, i.e., that the preference realization is at either extreme end of the possible interval. Choosing to locate at either extreme end in market 1 however, leaves firm B with a large average distance to the true preference. The average distance is minimized by choosing location and price such that η

must be in a convex set following no sale. Due to the uniform distribution only the size of the covered interval matters for profits in market 2, but not the precise location. We thus focus on location and price choices such that, following a sale, it can be infered that η is to the right of some cut-off, and following no-sale, η is to its left.

Firm B's strategy in market 2 can thus be described by the probability $\alpha \in [0,1]$ of selling in market 2, which determines a location choice and price as follows. The firm chooses the location $l_2 = x_B + (1-\alpha)\,\varepsilon_B$ and the price $p_2 = v_2 - \alpha\varepsilon_B$. It thus sells when $\eta \in [x_B + \varepsilon_B (1-2\alpha), x_B + \varepsilon_B]$ and does not sell when $\eta \in [x_B - \varepsilon_B, x_B + \varepsilon_B (1-2\alpha)]$. The probability of selling is thus $\frac{x_B+\varepsilon_B-(x_B+\varepsilon_B(1-2\alpha))}{2\varepsilon_B} = \alpha$. Firm B's expected profits in market 2 are thus given by

$$\pi_{2,B} = \alpha\,(v_2 - \alpha\varepsilon_B)\,.$$

If the firm were to maximize simply its profits in market 2, it would choose a value of α which we denote by α_2. Taking the first-order condition

$$v_2 - 2\alpha_2\varepsilon_B = 0,$$

and allowing for the restriction $\alpha_2 \in [0,1]$ the optimum is

$$\alpha_2 = \max\left\{\frac{v_2}{2\varepsilon_B}, 1\right\}.$$

Hence, $p_2 = v_2 - \frac{v_2}{2\varepsilon_B}\varepsilon_B = \frac{v_2}{2}$ if $v_2 < 2\varepsilon_B$, and $p_2 = v_2 - \varepsilon_B$ if $v_2 \geq 2\varepsilon_B$, which corresponds to the monopolist's pricing strategy (1). This case will provide a benchmark for how information spillovers from market 2 to market 1 will distort firm B's behaviour in market 2.

In order to understand learning spillovers, we need to analyze the equilibrium in market 1 given firm B's stochastic information structure obtained from market 2. We assume that firm A cannot observe B's activities in market 2, i.e., A observes neither the price nor product offered, nor whether the consumer purchased the good. Firm A does, however, know that B is active in market 2 and understands the structure of that market. This implies that A knows that B will have a signal that is, with probability α, distributed uniformly on an interval of size $2\alpha\varepsilon_B$ and with probability $1-\alpha$, B's information is on an interval of size $2(1-\alpha)\varepsilon_B$. Firm B can set a price $p^s_{1,B}$ and $p^n_{1,B}$, in market 1 that depends on whether it sold (s) or not (n) in market 2, respectively. Firm A's strategy is a price $p_{1,A}$. For simplicity, suppose that $\varepsilon_B \leq \varepsilon_A$.[5]

Proposition 7 *If $\varepsilon_B \leq \varepsilon_A$, there exists an equilibrium in which the single market firm A sets $l_{1,A} = x_A$. The price $p_{1,A}$ and probability α are given by the solution to the two equations*

$$p_{1,A} = \frac{1}{3}\varepsilon_A + \frac{1}{6}\varepsilon_B\left[\alpha^2 + (1-\alpha)^2\right] \tag{23}$$

$$\alpha = \min\left\{\frac{1}{2}\frac{v_2 + \frac{\varepsilon_B}{\varepsilon_A}\left(\frac{p_{1,A}+\varepsilon_A}{2} - \frac{3}{16}\varepsilon_B\right)}{\varepsilon_B + \frac{\varepsilon_B}{\varepsilon_A}\left(\frac{p_{1,A}+\varepsilon_A}{2} - \frac{3}{16}\varepsilon_B\right)}, 1\right\}. \tag{24}$$

The multi-market firm B chooses location and price in market 2 according to

$$l_2 = x_B + (1-\alpha)\varepsilon_B,$$
$$p_2 = v_2 - \alpha\varepsilon_B,$$

and in market 1

$$l_{1,B} = \begin{cases} l_2 & \text{if} \quad \text{a sale occurred in market 2} \\ l_2 = x_B - \alpha\varepsilon_B & \text{if no sale occurred in market 2} \end{cases},$$

$$p_{1,B} = \begin{cases} p^s_{1,B} = \frac{p_A+\varepsilon_A}{2} - \frac{1}{4}\alpha\varepsilon_B & \text{if} \quad \text{a sale occurred in market 2} \\ p^n_{1,B} = \frac{p_A+\varepsilon_A}{2} - \frac{1}{4}(1-\alpha)\varepsilon_B & \text{if no sale occurred in market 2} \end{cases}.$$

Even without characterizing the explicit solution for α we can say a few things about it from equation (24). Take first the case where the good in market 2 is sufficiently valuable for the monopolist to cover the market fully, i.e., $v_2 > 2\varepsilon_B$ and thus $\alpha_2 = 1$. Then are cases where the desire to learn about market 1 would lead firm B to cover market 2 only partially

[5] If we allow for $\varepsilon_B > \varepsilon_A$, the analysis gets complicated by the fact that we need to distinguish between cases, depending on whether $\alpha\varepsilon_B \gtrless \varepsilon_A$ and $(1-\alpha)\varepsilon_B \gtrless \varepsilon_A$, keeping in mind that α is endogenous. Overall, the benefit of learning would be reduced since $\frac{\partial \pi_B}{\partial \varepsilon_B} > 0$ when $\varepsilon_B > \varepsilon_A$.

($\alpha < 1$). In particular, this would be the case when $2\varepsilon_B < v_2 < 2\varepsilon_B + \frac{\varepsilon_B}{\varepsilon_A}\left(\frac{p_{1,A}+\varepsilon_A}{2} - \frac{3}{16}\varepsilon_B\right)$. When $v_2 \in (\varepsilon_B, 2\varepsilon_B)$ then $\alpha_2 > \frac{1}{2}$ and $\frac{1}{2} < \alpha < \alpha_2$, i.e., market coverage worsens, while when $v_2 \in (0, \varepsilon_B)$ then $\alpha_2 < \frac{1}{2}$ and $\frac{1}{2} > \alpha > \alpha_2$, i.e., market coverage improves. In a nutshell, the information spillover effect moves the optimal amount of market coverage towards $\frac{1}{2}$ which can imply an increase or decrease compared to the monopoly without information spillovers.

Note also, that optimal learning requires the firm to design a biased product. This is because if the product is of a specification that is unbiased, then the refusal to purchase the good, provides little information about what the optimal specification looks like. The firm would only learn that the desired specification is more "extreme", say either more "traditional" or more "modern". This would leave the firm with a high likelihood of offering a poorly specified good in market 1 (e.g., when the firm offers a "modern" good, but the preference is strongly "traditional"). It therefore can be optimal to offer a more extreme specification up front, say "modern" so that at least a failed purchase identifies the consumer's taste quite precisely as being "traditional".

6 Extensions

6.1 Equilibrium when information becomes less precise

Consider total and consumer surplus when information is less precise in the sense that $f(\varepsilon_A, \varepsilon_B) \leq v \leq g(\varepsilon_A, \varepsilon_B)$ and we are in region (ii) of Proposition 1. Interestingly, total surplus may now increase when A's information gets poorer.

Lemma 5 *If* $\varepsilon_B < \frac{18-\sqrt{192}}{11}v$ *then* $\frac{\partial TS_D}{\partial \varepsilon_A} > 0$. *We always get* $\frac{\partial CS_D}{\partial \varepsilon_A} < 0$.

Proof. Since the equilibrium features full coverage, the loss function (20) remains valid, only $\Delta p = \frac{v}{2} - \frac{3}{4}\varepsilon_B$. Calculating the derivative yields

$$\frac{\partial TS_D}{\partial \varepsilon_A} < 0 \Leftrightarrow v^2 - 3v\varepsilon_B + \frac{11}{12}\varepsilon_B^2 < 0.$$

Hence, $\frac{\partial TS_D}{\partial \varepsilon_A} > 0$ if either $v > \frac{11}{18-\sqrt{192}}\varepsilon_B$ or $v < \frac{11}{18+\sqrt{192}}\varepsilon_B$. Note that $v < \frac{11}{18+\sqrt{192}}\varepsilon_B$ is incompatible with $\varepsilon_A + \frac{7}{4}\varepsilon_B - \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2} < v$.

We can calculate expected trading profits as $\pi_A = \Pr\left(Sell_A\right) * p_A$ and similarly for B. Using the probabilities in (7) and (8) and the prices from Lemma **??**, we can calculate the derivatives as

$$\frac{\partial \pi_A}{\partial \varepsilon_A} = -\frac{1}{\varepsilon_A^2}\left(\frac{v}{2} - \frac{\varepsilon_B}{4}\right)^2,$$
$$\frac{\partial \pi_B}{\partial \varepsilon_A} = \frac{1}{\varepsilon_A^2}\left(\frac{v^2}{2} - \frac{3}{4}v\varepsilon_B + \frac{1}{4}\varepsilon_B^2\right),$$

and since $\Delta p = \frac{v}{2} - 3\frac{\varepsilon_B}{4} > 0$ we have

$$\frac{\partial \pi_A}{\partial \varepsilon_A} + \frac{\partial \pi_B}{\partial \varepsilon_A} = \frac{1}{\varepsilon_A^2}\left(\frac{v}{2} - \frac{\varepsilon_B}{4}\right)\left(\frac{v}{2} - 3\frac{\varepsilon_B}{4}\right) > 0.$$

Altogether this yields

$$\frac{\partial CS_D}{\partial \varepsilon_A} = \frac{\partial TS_D}{\partial \varepsilon_A} - \left(\frac{\partial \pi_A}{\partial \varepsilon_A} + \frac{\partial \pi_B}{\partial \varepsilon_A}\right) < 0.$$

■

To understand the intuition for this result consider an increase in ε_A from a point such that $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B$ is close to but below v and hence the equilibrium is described by Proposition 1. Both firms increase their prices with an increase in ε_A. However, at some point (when $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B = v$), the higher of the two prices (p_B) has become so high that the market would no longer be fully covered if B increased its price any further. That is, B would not sell for high realizations of its error. At that point a region begins where further increases in ε_A do not lead to any changes in B's price and the same is true for A. In this region total surplus may actually increase when ε_A increases. That is because with unchanging prices and worsening information for A, B becomes more likely to sell. As B's information is on average better, having B be the seller is good for total surplus. Effectively, the relative distortion from unequal prices decreases.

Overall, this does not translate into an increase in consumer surplus. This is because industry profits are increasing in ε_A (even though prices are unchanged). The reasons is, again, that B is more likely to sell and B charges the higher of the two prices.

7 Conclusion

We introduce an original model set-up to investigate the interaction between data quality, product targeting and price competition, motivated by numerous examples of firms that use customer-specific information to customize their product offers. After In our extracting noisy information about consumer's preferred product specification, two firms compete by simultaneously choosing their consumer-specific product specification and prices.

We find that better data processing by the less well informed firm leads to lower prices by both firms, reflecting that firms are expected to offer products in closer proximity. But if the better informed firm better processes data the competition effect is ambiguous since the better informed firm will charge a higher price and its competitor a lower price. Still, in a duopoly, more information makes consumers better off in both cases since any price increase is more than compensated by better product fit on average. When the consumer faces a monopoly, then there is an optimal data quality level, since the monopolist will exploit any better information to fully extract consumer surplus. By contrast, while consumer surplus always increases with better data processing in a duopoly, the same is not true for total surplus may decrease, reflecting that the allocation may be increasingly inefficient when the less informed firm competes more aggressively on price. Thus, even when in a duopoly the aggregate amount of information is higher, it will often not be used efficiently,leading to lower welfare compared to a monopoly.

We investigate the possibility to invest in data quality before firms choose product location and price. We analyze the strategic interaction of investments in data quality, and find that they behave like strategic substitutes in equilibrium: the higher the equilibrium data quality of the firm with higher data capabilities, the lower the data investment of the competing firm. This outcome is driven by our observation that average profits increase in data quality heterogeneity, intermediated by less aggressive pricing policies. We then consider the use of data quality investments to influence entry and exit of competitors. We show that an incumbent can use overinvestment in data quality as a strategic entry deterrent, or accelerator of exit.

We extend the model to consider strategies of multi-platform gatekeepers. We find that a firm that use insights about a consumer's preferences gleaned from the consumer choices in one market may adopt a pricing strategy that elicit more information - typically by more demand-sensitive and higher prices -, in order to gain an informational advantage that it can use when competing in another market.

Appendix

Proof of Proposition 1. Consider first location choices l_i. Suppose A uses the candidate equilibrium location $l_A = x_A$. Denote by $\lambda_B \geq 0$ the bias relative to its signal that B applies when choosing the location, i.e., $l_B = x_B + \lambda_B$. Thus B's good is preferred over A's, if

$$v - p_B - |\eta - (\lambda_B + \eta + \tilde{e}_B)| > v - p_A - |\eta - (\eta + \tilde{e}_A)|.$$

Note that for B to be able to sell, its price-location choice must meet the consumer's willingness to pay, $v - p_B - \lambda_B - |\tilde{e}_B| > 0$, which becomes less likely to hold the larger is λ_B. For any value of $|\tilde{e}_B| \leq \varepsilon_B$, this condition will hold for any pair of λ_B and p_B where firm has a positive probability of selling when $v - \varepsilon_A - 2\varepsilon_B + p_A \geq 0$.

(a) Suppose $v - \varepsilon_A - 2\varepsilon_B + p_A \geq 0$. We can then calculate B's probability of selling as:

(1) If $0 \leq \lambda_B \leq \varepsilon_A - \varepsilon_B - \Delta p$ then

$$\Pr(Sell_B) = 1 - \frac{\Delta p}{\varepsilon_A} - \frac{1}{2}\frac{\varepsilon_B}{\varepsilon_A} - \frac{1}{2}\frac{\lambda_B^2}{\varepsilon_A \varepsilon_B}.$$

(2) If $\varepsilon_A - \varepsilon_B - \Delta p < \lambda_B \leq \varepsilon_B$ then

$$\Pr(Sell_B) = \frac{\frac{1}{2}(\varepsilon_A - \Delta p)^2}{2\varepsilon_A \varepsilon_B} + \frac{\left(\varepsilon_A - \frac{1}{2}\varepsilon_B - \Delta p\right)\varepsilon_B}{2\varepsilon_A \varepsilon_B} - \frac{\lambda_B^2}{2\varepsilon_A \varepsilon_B} - \frac{\frac{1}{2}(\varepsilon_A - \varepsilon_B - \Delta p)\lambda_B}{2\varepsilon_A \varepsilon_B}.$$

(3) If $\varepsilon_B < \lambda_B \leq \varepsilon_A - \Delta p$ then

$$\Pr(Sell_B) = \frac{\frac{1}{2}(\varepsilon_A - \Delta p - \lambda_B)^2}{2\varepsilon_A \varepsilon_B} + \frac{(\varepsilon_A - \Delta p - \lambda_B)\varepsilon_B}{2\varepsilon_A \varepsilon_B} + \frac{\frac{1}{2}\varepsilon_B^2}{2\varepsilon_A \varepsilon_B}.$$

(4) If $\varepsilon_A - \Delta p < \lambda_B \leq \varepsilon_A + \varepsilon_B - \Delta p$ then

$$\Pr(Sell_B) = \frac{\frac{1}{2}(\varepsilon_A + \varepsilon_B - \Delta p - \lambda_B)^2}{2\varepsilon_A \varepsilon_B}$$

(5) If $\varepsilon_A + \varepsilon_B - \Delta p < \lambda_B$ then

$$\Pr(Sell_B) = 0$$

$\Pr(Sell_B)$ is continuous in λ_B at the boundaries between these five cases, and the probability of selling is strictly decreasing in λ_B in each case as long as $\Pr(Sell_B) > 0$. Hence, for any given price p_B, the probability of selling, as well as firm profit, is highest when firm B chooses $\lambda_B = 0$.

(b) Consider $v - \varepsilon_A - 2\varepsilon_B + p_A < 0$, so that for some choices of λ_B and p_B, the consumer's participation constraint is not met. It can then be shown that the probability of selling is more strongly decreasing in λ_B compard with the case (a) $v - \varepsilon_A - 2\varepsilon_B + p_A \geq 0$ (the rate of decrease is unchanged if $\lambda_B \leq \varepsilon_A - \varepsilon_B - \Delta p$, and strictly larger if $\varepsilon_A - \Delta p < \lambda_B \leq \varepsilon_A + \varepsilon_B - \Delta p$). Taken together, these results imply that any optimal choice of λ_B and p_B must include the choice of $\lambda_B = 0$.

An analogous argument can be developed for firm A, showing that firm A's location choice is optimal at $\lambda_A = 0$.

Proof of part (i). Conjecturing $p_B < v - \varepsilon_B$ we get the best-response functions

$$p_A(p_B) = \frac{p_B}{2} + \frac{1}{4}\varepsilon_B \tag{25}$$

$$p_B(p_A) = \frac{p_A + \varepsilon_A}{2} - \frac{1}{4}\varepsilon_B. \tag{26}$$

Price reactions are determined by three factors (a) the rival's price, (b) the rival's information quality, and (c) the firm's own information quality. Each firm sets a lower price in response to a price drop by the rival. That is, prices are strategic complements, as is standard in models of price competition. Furthermore, each firm reacts to an improvement in its rival's information, by lowering its own price. This is also intuitive as firms try to make up for a better targeted rival product by competing more aggressively on price. Finally, only the better informed firm reacts to improvements in its own information by increasing its price. As eq. (25) shows, the less well informed firm A does not alter its price when its own information changes. This is because an increase in its information precision $\frac{1}{\varepsilon_A}$ will lead to a linear increase in its profit, as eq. (9), and does not affect the optimal price. Solving (25) and (26) yields (11) and (12). Note that these prices satisfy $0 \leq p_B - p_A \leq \varepsilon_A - \varepsilon_B$. Moreover, we require that the solution $p_B < v - \varepsilon_B$ which yields

$g(\varepsilon_A, \varepsilon_B) < v$.

Proof of part (ii):

For $p_B \geq v - \varepsilon_B$, we get the following first-order conditions. A wishes to increase the price p_A as long as

$$\varepsilon_B (v - 2p_A) - \frac{1}{2}(v - p_B)^2 \geq 0. \tag{27}$$

B wishes to increase the price p_B as long as

$$\frac{1}{2}(v - p_B)(v - 3p_B) + [\varepsilon_A - (v - p_A)](v - 2p_B) \geq 0. \tag{28}$$

Conjecture a price $p_B = v - \varepsilon_B$. $A's$ best response to this price is

$$p_A = \frac{v}{2} - \frac{\varepsilon_B}{4},$$

regardless of whether A uses the profit function just above or just below the boundary (so there is no jump in A's best response just around the price chosen by B). Using this price p_A it can be shown that for $p_B < v - \varepsilon_B$ the derivative of π_B is positive if

$$v < \frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B.$$

For $p_B > v - \varepsilon_B$ the derivative of π_B is negative if

$$\varepsilon_A + \frac{7}{4}\varepsilon_B - \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2} < v < \varepsilon_A + \frac{7}{4}\varepsilon_B + \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2}.$$

Moreover, it can be shown that

$$\varepsilon_A + \frac{7}{4}\varepsilon_B - \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2} \leq \frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq \varepsilon_A + \frac{7}{4}\varepsilon_B + \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2}.$$

Hence the binding constraints on v are given by $v \in [f(\varepsilon_A, \varepsilon_B), g(\varepsilon_A, \varepsilon_B)]$. Note that for the symmetric case $\varepsilon_A = \varepsilon_B$, the lower and upper bounds take the same value of $\frac{3}{2}\varepsilon$. I.e., for symmetry this case disappears. Using the price (13) and (14) it can be shown that $v \leq g(\varepsilon_A, \varepsilon_B)$ implies $\Delta p \leq \varepsilon_A - \varepsilon_B$ and $f(\varepsilon_A, \varepsilon_B) \leq v$ implies $\Delta p \geq 0$.

Proof of part (iii): Denote $y(p_B)$ the function (15) and by $z(p_B)$ the function (16). We have the following properties: (a) For $p_B \in \left(v - \varepsilon_B, \frac{v}{2}\right)$ we get $z' > y' > 0$. (b)

$z\left(\frac{v}{2}\right) > y\left(\frac{v}{2}\right)$. (c) $z(v-\varepsilon_B) < y(v-\varepsilon_B)$ if either $v < \varepsilon_A + \frac{7}{4}\varepsilon_B - \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2}$ or $v > \varepsilon_A + \frac{7}{4}\varepsilon_B + \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2}$. For $\left(v - \varepsilon_B, \frac{v}{2}\right)$ to be an open interval, we require $v < 2\varepsilon_B$. Note that $v < f_1(\varepsilon_A, \varepsilon_B)$ implies $v < 2\varepsilon_B$. Moreover, $v < 2\varepsilon_B$ implies $y(0) > 0$, which guarantees that the intersection is in the positive quadrant. Moreover, $v > \varepsilon_A + \frac{7}{4}\varepsilon_B + \sqrt{\left(\varepsilon_A - \frac{1}{4}\varepsilon_B\right)^2 + \varepsilon_B^2}$ and $v < 2\varepsilon_B$ are incompatible. These properties imply that when $v < f(\varepsilon_A, \varepsilon_B)$, there is a unique point $p_B \in \left(v - \varepsilon_B, \frac{v}{2}\right)$ where y and z intersect. ■

Proof of Proposition 2. : In (3) set $\varepsilon = \varepsilon_B$ and compare. If $\varepsilon_B < \frac{4}{7}\varepsilon_A$ then assumption (17) implies that $\varepsilon_B < \frac{v}{2}$, i.e., we are in the case of full coverage under monopoly.

The monopoly has higher total surplus if

$$v - \frac{\varepsilon_B}{2} > v - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\}$$

$$\Leftrightarrow$$

$$\frac{1}{3}\varepsilon_B^2 < \frac{1}{9}(\varepsilon_A - \varepsilon_B)^2$$

$$\Longleftrightarrow$$

$$2\varepsilon_B^2 + 2\varepsilon_A\varepsilon_B - \varepsilon_A^2 < 0.$$

The latter can be simplified to

$$\varepsilon_B < \varepsilon_A\frac{\sqrt{3}-1}{2}.$$

Note that $\frac{\sqrt{3}-1}{2} < \frac{4}{7}$. When there is partial coverage ($\varepsilon_B > \frac{v}{2}$) in the monopoly case, total surplus is higher in duopoly if

$$v - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\} > \frac{3}{2}\frac{v^2}{4\varepsilon_B}$$

Since $\varepsilon_B > \frac{v}{2}$ we have $\frac{3}{2}\frac{v^2}{4\varepsilon_B} < \frac{3}{2}\frac{v}{2}$ and a sufficient condition becomes

$$v - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\} > \frac{3}{4}v$$

$$\Leftrightarrow$$

$$\frac{1}{2}v > \frac{1}{\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\}$$

Since we also have $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq v$ a sufficient condition is

$$\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B > 2\frac{1}{\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\}$$

$$\Leftrightarrow$$

$$4\left(\varepsilon_A - \varepsilon_B\right)^2 + \frac{3}{2}\varepsilon_A\varepsilon_B > 0,$$

which always holds.

Consumer surplus under duopoly is higher than under monopoly if

$$v - \frac{\left(\frac{\varepsilon_A}{3} + \frac{\varepsilon_B}{6}\right)^2}{\varepsilon_A} - \frac{\left(\frac{2\varepsilon_A}{3} - \frac{\varepsilon_B}{6}\right)^2}{\varepsilon_A} - \frac{1}{2\varepsilon_A}\left\{\left(\frac{\varepsilon_A - \varepsilon_B}{3}\right)^2 + \varepsilon_A\varepsilon_B - \frac{\varepsilon_B^2}{3}\right\} > \frac{\varepsilon_B}{2}.$$

This can be re-written as

$$v\varepsilon_A > \frac{\varepsilon_B^2}{2} + \varepsilon_A\varepsilon_B + \frac{\varepsilon_A^2}{2} + \left[\frac{\varepsilon_A^2}{9} - 2\frac{\varepsilon_A\varepsilon_B}{9} + \frac{\varepsilon_B^2}{9}\right] - 2\frac{\varepsilon_B^2}{3}.$$

Since $v > \frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B$ by assumption (17) a sufficient condition is

$$\left(\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B\right)\varepsilon_A > \frac{\varepsilon_B^2}{2} + \varepsilon_A\varepsilon_B + \frac{\varepsilon_A^2}{2} + \left[\frac{\varepsilon_A^2}{9} - 2\frac{\varepsilon_A\varepsilon_B}{9} + \frac{\varepsilon_B^2}{9}\right] - 2\frac{\varepsilon_B^2}{3}$$

After further simplification the inequality can be re-written as

$$\varepsilon_A^2 - 2\varepsilon_A\varepsilon_B + \varepsilon_B^2 > 0,$$

which is always true. ■

Proof of Proposition 4. : We consider first that $c\left(\varepsilon_B\right)$ is negligible, and hence can approximate $r\left(\varepsilon_i\right) \approx \pi\left(\varepsilon_i\right)$. We consider first the case where $\varepsilon_A < \frac{3}{2}v$ which implies that $\varepsilon_A < \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. This is the when A is sufficiently competitive so that ε_A influences prices.

Consider then a choice of $\varepsilon_B < \varepsilon_A$. As shown in Proposition 3(i), B's profit is then $\pi_B = \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2$ and hence $\frac{\partial \pi_B}{\partial \varepsilon_B} < 0$ over the entire range $\varepsilon_B \in (0, \varepsilon_A)$. Thus, with $c(\varepsilon_B)$ sufficiently small, $\varepsilon_B = 0$ is the only candidate outcome in this region, with profit level $\pi_B(\varepsilon_B = 0) = \frac{4}{9}\varepsilon_A$.

Then consider a choice of $\varepsilon_B > \varepsilon_A$. As long as $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq v$, which holds for $\varepsilon_B < \frac{6}{5}v - \frac{4}{5}\varepsilon_A$, the results of Proposition 3(i) apply, but we need to invert the roles of A and of B (since $\varepsilon_B > \varepsilon_A$). We have then: $\pi_B(\varepsilon_B > \varepsilon_A) = \frac{1}{\varepsilon_B}\left(\frac{1}{3}\varepsilon_B + \frac{1}{6}\varepsilon_A\right)^2$, and taking derivatives: $\frac{\partial \pi_B}{\partial \varepsilon_B} = \frac{1}{36\varepsilon_B^2}\left(4\varepsilon_B^2 - \varepsilon_A^2\right) > 0$. Thus, in this case, B wants to expand ε_B until at least $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$, the boundary between the regions of validity of Proposition of 3(i) and the region of Proposition 3(ii).

Consider expanding ε_B beyond $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. In this case, the conditions of Proposition 3(ii) apply, with the roles of A and of B inverted, and we have for B's profit: $\pi_B\left(\varepsilon_B > \varepsilon_A, \varepsilon_B \geq \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right) = \frac{1}{16\varepsilon_B}(2v - \varepsilon_A)^2$, with the derivative: $\frac{\partial \pi_B}{\partial \varepsilon_B} = -\frac{1}{16\varepsilon_B^2}(2v - \varepsilon_A)^2 < 0$. Hence, there is no equilibrium choice of B that maximizes $r(\varepsilon_B)$ inside the parameter region of Proposition 3(ii), and $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$ is the candidate solution (local optimum for $\varepsilon_B > \varepsilon_A$), leading to profit level $\pi_B\left(\varepsilon_B > \varepsilon_A, \varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right) = \frac{1}{40}\frac{(4v - \varepsilon_A)^2}{3v - 2\varepsilon_A}$.

When $\varepsilon_A = 0$, then $\pi_B\left(\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right) = \frac{2}{15}v > \pi_B(\varepsilon_B = 0) = 0$. When $\varepsilon_A \to \frac{2}{3}v$, then $\pi_B(\varepsilon_B = 0) = \frac{8}{27}v > \pi_B\left(\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right) = \frac{1}{6}v$. So profits are increasing in ε_A at both $\varepsilon_B = 0$ and at $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. Moreover, when varying B's profit in ε_A, we get a single crossing point ε_A where the global optimum switches from the two local maximum $\varepsilon_B = 0$ to that at $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. To see this, we observe that $\frac{\partial}{\partial \varepsilon_A}(\pi_B(\varepsilon_B = 0)) = \frac{4}{9}$ which is linear. On the other hand, $\frac{\partial \pi_B\left(\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right)}{\partial \varepsilon_A} = \frac{1}{40}\frac{2}{(2\varepsilon_A - 3v)^2}\left(4v^2 + \varepsilon_A(3v - \varepsilon_A)^2\right) > 0$, so the function is convex in ε_A, showing that the profit levels for $\pi_B\left(\varepsilon_B > \varepsilon_A, \varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A\right)$ and $\pi_B(\varepsilon_B < \varepsilon_A, \varepsilon_B = 0)$ can cross at most once.

We consider then the case where $\varepsilon_A > \frac{2}{3}v$ which implies that $\varepsilon_A > \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. For $\varepsilon_B < \frac{6}{5}v - \frac{4}{5}\varepsilon_A$, Proposition 3(i) applies and $\frac{\partial \pi_B}{\partial \varepsilon_B} < 0$ as above. For $\frac{6}{5}v - \frac{4}{5}\varepsilon_A < \varepsilon_B < \varepsilon_A$, Proposition 3(ii) applies and it can be shown that $\frac{\partial \pi_B}{\partial \varepsilon_B} < 0$. For $\varepsilon_B > \varepsilon_A$, Proposition 3(ii) applies, with the roles of A and of B inverted, and $\frac{\partial \pi_B}{\partial \varepsilon_B} < 0$ as shown above. Thus, when

$\varepsilon_A > \frac{2}{3}v$, B 's profit decreases globally in ε_B, and only a choice of $\varepsilon_B = 0$ can be optimal.

Finally, consider an increase in cost $c(\varepsilon_B)$, leading possibly to an internal solution ε_B^{D*}. Observe from the functional forms stated above that the profit function $\pi_B(\varepsilon_B, \varepsilon_A)$ is piecewise differentiable in ε_B in the intervals where candidate solutions can be located: for large ε_A, in the interval $\varepsilon_B \in (0, \varepsilon_A)$ and for small ε_A, in the interval $\varepsilon_B \geq \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. Thus, the net profit function $r_B(\varepsilon_B, \varepsilon_A) = \pi_B(\varepsilon_B, \varepsilon_A) - c(\varepsilon_B)$ is piecewise differentiable in ε_B in these intervals as well, and concave with an internal optimum when $c(\varepsilon_B)$ is sufficiently large and convex.

Proof of Proposition 5. We initially abstract from costs $c(\varepsilon_B)$ and $c(\varepsilon_A)$, and consider them to be arbitrarily small.

Step 1: Going backwards, we start by characterizing A's best response (BR) function $\varepsilon_A(\varepsilon_B)$. We show that:

$$\varepsilon_A(\varepsilon_B) = \begin{cases} \frac{3}{2}v - \frac{5}{4}\varepsilon_B & \text{if } \varepsilon_B \in \left(0, \frac{2}{3}v\right) \\ 0 & \text{if } \varepsilon_B > \frac{2}{3}v \end{cases}$$

To derive this expression, we need to look at four cases, defined by two separating lines: the straight line $g(\varepsilon_A, \varepsilon_B) = \frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B = v$ that separates Proposition 3(i) ($g(\varepsilon_A, \varepsilon_B) \leq v$) from Proposition 3(ii), and $\varepsilon_A = \varepsilon_B$. We begin with the analysis of the simulateneous choice of locations and precisions where, as we recall, ($g(\varepsilon_A, \varepsilon_B) > v$).

(i) $\varepsilon_A \geq \varepsilon_B$ and $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$ (Proposition 3(i)): This condition holds when there is a value $\varepsilon_A \geq \varepsilon_B$ such that $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B \leq v$. We get: $\pi_A = \frac{(p_A)^2}{\varepsilon_A} = \frac{1}{\varepsilon_A}\left(\frac{1}{3}\varepsilon_A + \frac{1}{6}\varepsilon_B\right)^2$, with derivative: $\frac{\partial \pi_A}{\partial \varepsilon_A} = \frac{1}{36\varepsilon_A^2}\left(4\varepsilon_A^2 - \varepsilon_B^2\right) > 0$. We next show that A will not choose ε_A such that $\frac{2}{3}\varepsilon_A + \frac{5}{6}\varepsilon_B = g(.) > v$. To show this, note that when A choosing ε_A s.t. $g(.) > v$, we need to apply Proposition 3(ii), and hence: $\pi_A = \frac{1}{16\varepsilon_A}(2v - \varepsilon_B)^2$. This expression is decreasing in ε_A. Hence, for ε_A s.t. $g(\varepsilon_A, \varepsilon_B) \leq v$, π_A is increasing in ε_A, and for ε_A s.t. $g(\varepsilon_A, \varepsilon_B) > v$, π_A is decreasing in ε_A. Thus, for all $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$, A's BR function is $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v - \frac{5}{4}\varepsilon_B \geq \frac{2}{3}v$, i.e. $\varepsilon_A(\varepsilon_B)$ coincides with the line $g(.) = v$.

(ii) Next, consider $\varepsilon_B > \frac{2}{3}v$, $\varepsilon_A < \varepsilon_B$ and $g(.) \leq v$. We then have $\varepsilon_A < \varepsilon_B$, so we can use Proposition 3(i) by inverting the roles af A and B, yielding $\pi_A = \frac{1}{\varepsilon_B}\left(\frac{2}{3}\varepsilon_B - \frac{1}{6}\varepsilon_A\right)^2$ and with derivative: $\frac{\partial \pi_A}{\partial \varepsilon_A} = \frac{1}{18\varepsilon_B}(\varepsilon_A - 4\varepsilon_B) < 0$, so we get $\varepsilon_A(\varepsilon_B) = 0$.

(iii) Next, consider $\varepsilon_B > \frac{2}{3}v$, $g(.) > v$, and $\varepsilon_A < \varepsilon_B$. W are in the case of Proposition 3(ii) with inverted roles af A and B, as long as $f < v$ (note that for small enough ε_A, this will hold since, with inverted roles af A and B, $f(\varepsilon_B, 0) = 0$). Hence in this case $\pi_A = \frac{1}{4\varepsilon_B}(4\varepsilon_B + \varepsilon_A - 2v)(v - \varepsilon_A)$, with derivative: $\frac{\partial \pi_A}{\partial \varepsilon_A} < 0$, so A chooses the smallest possible value $\varepsilon_A(\varepsilon_B)$. The smallest candidate value inside the parameter region, $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v - \frac{5}{4}\varepsilon_B$, defined by the boundary condition $g(.) \geq v$, is not the best-reponse function, as seen in case (ii). So there is no BR satisfying $g(.) \geq v$, and the BR is $\varepsilon_A(\varepsilon_B) = 0$.

(iv) In the case $\varepsilon_B > \frac{2}{3}v$, $g(.) > v$ and $\varepsilon_A \geq \varepsilon_B$, we apply Proposition 3(ii), hence $\pi_A = \frac{1}{16\varepsilon_A}(2v - \varepsilon_B)^2$, with derivative:$\frac{\partial \pi_A}{\partial \varepsilon_A} < 0$. As in case (iii), there is no BR satisfying $g(.) > v$, and the BR is $\varepsilon_A(\varepsilon_B) = 0$.

Step 2: We move backwards and check that, given BR function $\varepsilon_A(\varepsilon_B)$, the optimal choice of B is $\varepsilon_B = 0$. Given the characterization above, we need to consider two cases: (i) $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$ and (ii) $\varepsilon_B > \frac{2}{3}v$.

(i) For $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$, we get (using Proposition 3(i)): $\pi_B = \frac{1}{36\varepsilon_A}(4\varepsilon_A - \varepsilon_B)^2$. B anticipates A's best response function $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v - \frac{5}{4}\varepsilon_B$ (see above), and hence we can substitute the BR function in B's objective function so that B maximizes: $\pi_B(\varepsilon_B, \varepsilon_A(\varepsilon_B)) = \frac{4}{6v - 5\varepsilon_B}(v - \varepsilon_B)^2$, with derivative: $\frac{\partial \pi_B}{\partial \varepsilon_B} = -\frac{4\left(7v^2 - 12v\varepsilon_B + 5\varepsilon_B^2\right)}{(6v - 5\varepsilon_B)^2}$. This derivative is negative since $7v^2 - 12v\varepsilon_B + 5\varepsilon_B^2 > (v - \varepsilon_B)(7v - 5\varepsilon_B) > 0$. Hence for $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$, B always prefers the choice $\varepsilon_B = 0$ given A's best response $\varepsilon_A(\varepsilon_B)$.

(ii) For $\varepsilon_B > \frac{2}{3}v$, using Proposition 3(i) by inverting the roles of A and B, we have then: $\pi_B = \frac{1}{\varepsilon_B}\left(\frac{1}{3}\varepsilon_B + \frac{1}{6}\varepsilon_A\right)^2$, and substituting A's BR function $\varepsilon_A(\varepsilon_B) = 0$, this becomes $\pi_B(\varepsilon_B, \varepsilon_A(\varepsilon_B)) = \frac{1}{9}\varepsilon_B$, with derivative:$\frac{\partial \pi_B}{\partial \varepsilon_B} = \frac{1}{9} > 0$. So B wants to increase ε_B at least up to the limit point where the profit function changes since the market is then not fully covered (boundary of Proposition 3(i) A and B are inverted). Since $\varepsilon_B > \varepsilon_A(\varepsilon_B)$, we need to invert ε_A and ε_B in determining this boundary, and hence $g(\varepsilon_A, \varepsilon_B) = \frac{5}{6}\varepsilon_A + \frac{2}{3}\varepsilon_B \leq v$ so $\varepsilon_B = \frac{3}{2}v$ is the limit point; after that point, $g(\varepsilon_A, \varepsilon_B) > v$ and the profit function is in the region of Proposition 3(ii). Consider whether B wants to expand beyond $\varepsilon_B = \frac{3}{2}v$.

Using Proposition 3(ii) by inverting the roles of A and B, we know that B maximizes: $\pi_B = \frac{1}{16\varepsilon_B}\left(2v - \varepsilon_A\right)^2$, with derivative: $\frac{\partial \pi_B}{\partial \varepsilon_B} = -\frac{1}{16\varepsilon_B^2}\left(2v - \varepsilon_A\right)^2 < 0$. Hence there is no equilibrium candidate of B inside the parameter region of Proposition 3(ii), and $\varepsilon_B = \frac{3}{2}v$ is the candidate solution (local optimum for $\varepsilon_B > \frac{2}{3}v$). Note that this solution is identical to A's best response function when $\varepsilon_B = 0$, $\varepsilon_A\left(\varepsilon_B\right) = \frac{3}{2}v$.

We then compare the two local optima, for $\varepsilon_B > \frac{2}{3}v$ and for $\varepsilon_B \leq \frac{2}{3}v$. This allows us to conclude that only $\varepsilon_B = 0$ can be B's optimal choice: B then receives $\pi_B = \frac{2}{3}v$. With B's choice of $\varepsilon_B = \frac{3}{2}v$ and A's reaction $\varepsilon_A\left(\varepsilon_B\right) = 0$, B's profit in this case is: $\pi_B = \frac{1}{6}v$. Thus, $\varepsilon_B = 0$ is B's unique optimal choice.

When we consider an increase in the information costs $c\left(\varepsilon_B\right)$, we can limit attention to the relevant range of B's optimal choice, $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$. We find that the derivative of the profit function after integrating A's best response, $\pi_B\left(\varepsilon_B, \varepsilon_A\left(\varepsilon_B\right)\right) = \frac{4}{6v - 5\varepsilon_B}\left(v - \varepsilon_B\right)^2$, with derivative: $\frac{\partial \pi_B}{\partial \varepsilon_B} = -\frac{4\left(7v^2 - 12v\varepsilon_B + 5\varepsilon_B^2\right)}{\left(6v - 5\varepsilon_B\right)^2}$, varies little over this range, increasing slightly from $\left.\frac{\partial \pi_B}{\partial \varepsilon_B}\right|_{\varepsilon_B = 0} = -\frac{7}{9}v$ to $\left.\frac{\partial \pi_B}{\partial \varepsilon_B}\right|_{\varepsilon_B = \frac{2}{3}v} = -\frac{11}{16}v$. Thus, if $c\left(\varepsilon_B\right)$ is sufficiently convex, then it is possible that there is an internal optimum in the interval $\varepsilon_B \in \left(0, \frac{2}{3}v\right)$ that maximizes net profits $r_B(\varepsilon_B)$. ■ ■

Proof of Proposition 6. Recall B's potential monopoly profit is (Section 2.1):

$$\pi_B^M\left(\varepsilon_B\right) = \begin{cases} v - \varepsilon_B & \text{if} \quad \varepsilon_B \leq \frac{v}{2} \\ \frac{v^2}{4\varepsilon_B} & \text{if} \quad \varepsilon_B > \frac{v}{2} \end{cases}$$

$\pi_B^M\left(\varepsilon_B\right)$ is decreasing throughout, with a max of $\pi^M = v$ for $\varepsilon_B = 0$, then linearly falling to $\pi^M = \frac{v}{2}$ for $\varepsilon_B = \frac{v}{2}$, then falling in a convex fashion. When cost $c\left(\varepsilon_B\right)$ are negligible, then the same pattern also holds for net profits $r_B^M\left(\varepsilon_B\right)$.

One-sided information quality choice of firm B: In this case, the duopoly profit $\pi_B^D\left(\varepsilon_B, \varepsilon_A\right)$ is: $\pi_B^D\left(\varepsilon_B < \varepsilon_A\right) = \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2$ (see Proposition 3(i) and Section 4.2), and hence $\frac{\partial \pi_B^D}{\partial \varepsilon_B} < 0$ over the range $\varepsilon_B \in (0, \varepsilon_A)$. As seen in Subsection 4.2, the duopoly profit of B decreases in ε_B for $\varepsilon_B < \varepsilon_A$, and increases in ε_B for $\varepsilon_B \geq \varepsilon_A$. For the profit function of the potential entrant A, we get $\frac{\partial \pi_A^D}{\partial \varepsilon_B} > 0$ throughout, so the lower is ε_B, the lower is A's profit and hence the more likely is entry deterrence. Moreover, with $c(\varepsilon_B)$ sufficiently small and ε_A sufficiently large, $\varepsilon_B = 0$ is the optimal choice, with profit level $\pi_B^D(\varepsilon_B = 0) = \frac{4}{9}\varepsilon_A$ and A's profit: $\pi_A^D(\varepsilon_B = 0) = \frac{1}{9}\varepsilon_A$, and $\pi_B^M(\varepsilon_B = 0) = v$. $\pi_A^D(\varepsilon_B = 0) = \frac{1}{9}\varepsilon_A$ is the lowest profit level of entrant A that the incumbent B can reach with its choice of ε_B.

Consider the difference between the monopoly and the duopoly profit, $\Delta\pi = \pi_B^M(\varepsilon_B) - \pi_B^D(\varepsilon_B, \varepsilon_A)$. We show that $\Delta\pi > 0$ for all values of ε_A and optimal choices of ε_B when $\varepsilon_A < \frac{2}{3}v$. We consider the points where $\pi_B^D(\varepsilon_B, \varepsilon_A)$ is not continuously differentiable, and then show that $\Delta\pi > 0$ also holds for all intermediate values. Consider $\varepsilon_A < \frac{2}{3}v$ which implies $\varepsilon_A < \frac{6}{5}v - \frac{4}{5}\varepsilon_A$. At $\varepsilon_B = 0$, $\Delta\pi = \pi_B^M(\varepsilon_B) - \pi_B^D(\varepsilon_B, \varepsilon_A) = v - \varepsilon_B - \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2 = v - \frac{4}{9}\varepsilon_A > \frac{1}{3}v$. At $\varepsilon_B = \varepsilon_A$, either $\varepsilon_A < \frac{v}{2}$ and $\Delta\pi = v - \varepsilon_B - \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2 = v - \frac{4}{9}\varepsilon_A > \frac{7}{9}v$ or $\varepsilon_A > \frac{v}{2}$ and $\Delta\pi = \frac{v^2}{4\varepsilon_B} - \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2 = \frac{1}{4\varepsilon_A}\left(v^2 - \varepsilon_A^2\right) > 0$ since we consider $\varepsilon_A < \frac{2}{3}v$. At $\varepsilon_B = \frac{6}{5}v - \frac{4}{5}\varepsilon_A$, $\Delta\pi = \frac{v^2}{4\varepsilon_B} - \frac{1}{16\varepsilon_B}(2v - \varepsilon_A)^2 = \frac{v^2}{4\left(\frac{6}{5}v - \frac{4}{5}\varepsilon_A\right)} - \frac{1}{16\left(\frac{6}{5}v - \frac{4}{5}\varepsilon_A\right)}(2v - \varepsilon_A)^2 = \frac{5}{32}\frac{\varepsilon_A}{3v - 2\varepsilon_A}(4v - \varepsilon_A) > 0$. Moreover, $\Delta\pi$ is continuous and piecewise differentiable between these points, showing that $\Delta\pi > 0$ everywhere. No choice $\varepsilon_B > \frac{6}{5}v - \frac{4}{5}\varepsilon_A$ can be optimal in a duopoly, as shown (in the inverted case, reaction function $\varepsilon_A(\varepsilon_B)$), in Proposition **??**.

To show that there is a generic set of values F_A such that B prefers to decrease $\hat{\varepsilon}_B$ below ε_B^{M*}, consider a situation where ε_B^{M*} is such that it does not deter entry, i.e. $\pi_A^D\left(\varepsilon_B^{M*}, \varepsilon_A\right) > F_A$. Then B prefers $\hat{\varepsilon}_B < \varepsilon_B^{M*}$, at cost $c(\hat{\varepsilon}_B) - c\left(\varepsilon_B^{M*}\right) > 0$, provided that: (i) $\pi_A^D\left(\varepsilon_B^{M*}, \varepsilon_A\right) > F_A > \pi_A^D(\hat{\varepsilon}_B, \varepsilon_A)$, and (ii) $\pi_B^M(\hat{\varepsilon}_B) - c(\hat{\varepsilon}_B) > \pi_B^D\left(\varepsilon_B^{M*}, \varepsilon_A\right) - c\left(\varepsilon_B^{M*}\right)$.

Condition (i) is feasible for some $F_A > 0$ since $\pi_A^D(\varepsilon_B, \varepsilon_A)$ increases in ε_A. Condition (ii) is feasible for some $\hat{\varepsilon}_B < \varepsilon_B^{M*}$ since $\Delta\pi$ is bounded away from zero, and π_B^M and π_B^D are continuous functions in ε_B. Thus, there must exist a generic set of values F_A satisfying both conditions.

Two-sided information quality choice: We next look at the case where firm A can react to B's entry and optimally adjust its knowledge and data quality ε_A before the competition in product location and prices plays out, under the same assumptions are the as in Section **??**. Consider the difference between the monopoly and the duopoly profit, $\Delta\pi = \pi_B^M(\varepsilon_B) - \pi_B^D(\varepsilon_B, \varepsilon_A)$. When A chooses ε_A, then we obtain $\pi_B^D(\varepsilon_B, \varepsilon_A)$ by substituting A's best response function $\varepsilon_A(\varepsilon_B)$ (Proposition 5) into the expression for $\pi_B^D(\varepsilon_B, \varepsilon_A)$ following Proposition 3. This yields, for $\varepsilon_B \le \frac{2}{3}v$, $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B)) = \frac{4}{6v-5\varepsilon_B}(v-\varepsilon_B)^2$, and for $\frac{2}{3}v < \varepsilon_B \le \frac{3}{2}v$, $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B)) = \frac{1}{\varepsilon_B}\left(\frac{1}{3}\varepsilon_B + \frac{1}{6}\varepsilon_A\right)^2$. For $\frac{2}{3}v < \varepsilon_B \le \frac{3}{2}v$, taking into account A's best response $\varepsilon_A(\varepsilon_B) = 0$, $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B)) = \frac{1}{9}\varepsilon_B$.Since $\varepsilon_A(\varepsilon_B) = 0$ drops discontinuously at the point from $\varepsilon_A(\varepsilon_B) = \frac{2}{3}v$ to $\varepsilon_A(\varepsilon_B) = 0$, $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B))$ discontinuously falls at this point from $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B))\big|_{\varepsilon_B=\frac{2}{3}v^-} = \frac{1}{6}v$ to $\pi_B^D(\varepsilon_{B,}\varepsilon_A(\varepsilon_B))\big|_{\varepsilon_B=\frac{2}{3}v^+} = \frac{1}{9}v$.

We show that $\Delta\pi > 0$ for all choices of ε_B and best responses of ε_A. We consider the points where $\pi_B^D(\varepsilon_B, \varepsilon_A)$ is not continuous or not continuously differentiable. At $\varepsilon_B = 0$, $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v$, and hence $\Delta\pi = \pi_B^M(\varepsilon_B) - \pi_B^D(\varepsilon_B, \varepsilon_A) = v - \varepsilon_B - \frac{1}{\varepsilon_A}\left(\frac{2}{3}\varepsilon_A - \frac{1}{6}\varepsilon_B\right)^2 = v - \frac{4}{9}\varepsilon_A = \frac{1}{3}v$. At $\varepsilon_B = \frac{2}{3}v^-$, $\varepsilon_A(\varepsilon_B) = \frac{3}{2}v - \frac{5}{4}\varepsilon_B = \frac{5}{6}v$, hence $\Delta\pi = \frac{v^2}{4\varepsilon_B} - \frac{4}{6v-5\varepsilon_B}(v-\varepsilon_B)^2 = \frac{3}{8}v - \frac{1}{6}v > 0$. At $\varepsilon_B = \frac{2}{3}v^+$, $\varepsilon_A(\varepsilon_B) = 0$, hence $\Delta\pi = \frac{3}{8}v - \frac{1}{9}v > 0$; at $\varepsilon_B > \frac{3}{2}v$, $\varepsilon_A(\varepsilon_B) = 0$, hence $\Delta\pi = \frac{v^2}{4\varepsilon_B} - \frac{1}{9}\varepsilon_B > 0$ (and converging to $\Delta\pi = 0$ for $\varepsilon_B \to \frac{3}{2}v$). Since $\pi_B^M(\varepsilon_B)$ and $\pi_B^D(\varepsilon_B, \varepsilon_A)$ are continuous and continuously differentiable for all intermediate values between these limits points, it follows that $\Delta\pi > 0$ also holds for all intermediate values. ■

Proof of Proposition 7. : Firm A knows that B has a uniformly distributed signal with a range $\alpha\varepsilon_B$ and uses a price $p_{1,B}^s$ with probability α. With probability $1-\alpha$, it has a uniformly distributed signal with range $(1-\alpha)\varepsilon_B$ and sets a price $p_{1,B}^n$ with probability $1-\alpha$. If firm A locates at $l_1 = x_A$ and sets price p_A it sells with probability

$$\Pr(Sell_A) = \frac{\alpha p_{1,B}^s + (1-\alpha)\, p_{1,B}^n - p_A}{\varepsilon_A} + \frac{1}{2}\frac{\alpha^2 + (1-\alpha)^2}{\varepsilon_A}\varepsilon_B,$$

yielding profits $\pi_A = \Pr(Sell_A)\, p_A$. Note that A cannot do better by changing its location as it has no information about where B is likely to locate (the fact that B will move to either side of its original signal is not useful to A because of the uniform distribution).

Maximizing π_A with respect to p_A yields A's best response

$$p_{1,A} = \frac{\alpha p_{1,B}^s + (1-\alpha)\, p_{1,B}^n}{2} + \frac{1}{4}\left(\alpha^2 + (1-\alpha)^2\right)\varepsilon_B.$$

Firm B sets its price in market 1 after having observed either a sale or no sale in market 2. In case of a sale it sets the location in market 1 at the mid-point of the interval on which

η may be located, i.e., at l_2 and sells with probability

$$\Pr\left(Sell_{1,B}\left|Sell_2\right.\right) = 1 - \frac{p^s_{1,B} - p_A}{\varepsilon_A} - \frac{1}{2}\frac{\alpha\varepsilon_B}{\varepsilon_A},$$

yielding a best response

$$p^s_{1,B}(\alpha) = \frac{p_{1,A} + \varepsilon_A}{2} - \frac{1}{4}\alpha\varepsilon_B.$$

Similarly, we get a best-response after a no sale event in market 2:

$$p^n_{1,B}(\alpha) = \frac{p_{1,A} + \varepsilon_A}{2} - \frac{1}{4}(1-\alpha)\varepsilon_B.$$

We highlight that B's choice of price depends on its actual choice of α by writing prices as functions of α. Note that firm A cannot observe α therefore sets a price that depends on a belief about α but not the actual value chosen by firm B. Although the two will be equal in equilibrium, the distinction is needed when determining firm B's optimal choice of α. For this purpose we write B's profits in market 1 following a sale or no sale as a function of B's choice of α:

$$\pi^s_{1,B}(\alpha) = \left(1 - \frac{p^s_{1,B}(\alpha) - p_A}{\varepsilon_A} - \frac{1}{2}\frac{\alpha\varepsilon_B}{\varepsilon_A}\right)p^s_{1,B}(\alpha),$$
$$\pi^n_{1,B}(\alpha) = \left(1 - \frac{p^n_{1,B}(\alpha) - p_A}{\varepsilon_A} - \frac{1}{2}\frac{\alpha\varepsilon_B}{\varepsilon_A}\right)p^n_{1,B}(\alpha).$$

Overall expected profits are given by

$$\begin{aligned}\pi_{1,B} &= \alpha\pi^s_{1,B}(\alpha) + (1-\alpha)\pi^n_{1,B}(\alpha)\\ &= \frac{1}{\varepsilon_A}\left[\left(\frac{p_A + \varepsilon_A}{2}\right)^2 - \frac{1}{2}\varepsilon_B\frac{p_A + \varepsilon_A}{2}\left(\alpha^2 + (1-\alpha)^2\right)\right]\\ &\quad + \frac{1}{16}\frac{\varepsilon_B^2}{\varepsilon_A}\left(1 - 3\alpha(1-\alpha)\right).\end{aligned}$$

Taking the first-order condition of $\pi_{1,B} + \pi_{2,B}$ and solving α yields (24).

Note also that for our previous equilibrium analysis to hold, we require $p^s_{1,B} - p_{1,A} \leq \varepsilon_A - \alpha\varepsilon_B$ and $p^n_{1,B} - p_{1,A} \leq \varepsilon_A - (1-\alpha)\varepsilon_B$. It is sufficient to check one of the two conditions as they are symmetric in α around $\frac{1}{2}$. In order to check the validity of the first inequality we substitute the price $p_{1,B}^{s}$:

$$\frac{p_{1,A}+\varepsilon_A}{2}-\frac{1}{4}\alpha\varepsilon_B-p_{1,A}\leq\varepsilon_A-\alpha\varepsilon_B.$$

Substituting $p_{1,A}$ and re-writing yields

$$\frac{2}{3}\varepsilon_A-\varepsilon_B\left[\frac{3}{4}\alpha-\frac{1}{12}\left[\alpha^2+(1-\alpha)^2\right]\right]\geq 0.$$

The left-hand side reaches its lowest value when $\alpha=1$ so that the inequality holds whenever

$$\frac{2}{3}\varepsilon_A-\varepsilon_B\left[\frac{3}{4}-\frac{1}{12}\right]\geq 0,$$

which holds.

Finally, we prove that (23) and (24) has a unique solution on $\alpha\in[0,1]$. Note that $p_{1,A}(\alpha)$ is decreasing (increasing) for $\alpha<\frac{1}{2}$ ($\alpha>\frac{1}{2}$). Moreover, from (24) it is clear that $\alpha<\frac{1}{2}$ if and only if $v_2<\varepsilon_B$. It is also the case that if α is interior, then $\frac{\partial\alpha}{\partial p_{1,A}}>0$ if and only if $v_2<\varepsilon_B$. Hence, if $v_2<\varepsilon_B$ then $\alpha<\frac{1}{2}$, $p_{1,A}(\alpha)$ is decreasing and $\alpha(p_{1,A})$ is increasing, so there is at most one point of intersection. To show that there exists a point of intersection take the inverse of the function $\alpha(p_{1,A})$ and denote it by $P(\alpha)$. It is straightforward to show that $P(\alpha=0)<0$ and $\lim_{\alpha\to\frac{1}{2}}P(\alpha)=+\infty$. Since $p_{1,A}(\alpha)$ is positive and finite on $\alpha\in[0,1]$, there must be an intersection point. When $v_2>\varepsilon_B$ then $\alpha>\frac{1}{2}$, $p_{1,A}(\alpha)$ is increasing and $\alpha(p_{1,A})$ is decreasing. In this case we either get a unique interior solution, or the corner solution $\alpha=1$. ■

References

[1] Acquisti, A., Taylor, C., and Wagman, L., 2016, "The Economics of Privacy," *Journal of Economic Literature*, 54(2): 442–492.

[2] Acquisti, A., Varian, H.R. , 2005, "Conditioning Prices on Purchase History," *Marketing Science*, 24(3):, 367–381.

[3] Anderson, S.P. and A. DePalma, 2012, "Competition for attention in the Information (overload) Age," *RAND Journal of Economics*, 43(1):1–25.

[4] Belleflamme, P., W. M. Lam and W. Vergote, 2019, "Competitive Imperfect Price Discrimination and Market Power," CESifo working paper #7964.

[5] Bergemann, D. and A. Bonatti, 2015, "Selling Cookies," *American Economic Journal: Microeconomics*, 7:259–294.

[6] Bergemann, D. and A. Bonatti, 2019, "Markets for Information: An Introduction," *Annual Review of Economics*, 11: 1–23.

[7] Bordalo, P., Gennaioli, N., and A Shleifer, 2016, "Competition for Attention," *Review of Economic Studies*, 83:481–513.

[8] Casadesus-Masanell, R., Hervas-Drane A., 2015, "Competing with privacy," *Management Science*, 61(1):229–246.

[9] de Clippel, G., Eliaz, K., and Rozen, K., 2014, "Competing for Consumer Inattention," *Journal of Political Economy*, 6:1203–1234.

[10] de Corniere, A. and G. Taylor, 2020, "Data and Competition: A General Framework with Applications to Mergers, Market Structure and Privacy Policy," unpublished manuscript TSE.

[11] Farboodi, M., R. Mihet, T. Philippon and L. Veldkamp, 2019, "Bid Data and Firm Dynamics," *American Economic Review Papers & Proceedings*, 109, 38 – 42.

[12] Gu, Y. L. Madio and C. Reggiani, 2019, "Exclusive Data, Price Manipulation and Market Leadership," CESifo working paper #7853.

[13] Hagiu, A. and J. Wright, 2020, "Data-enabled learning, network effects and competitive advantage" unpublished manuscript.

[14] Iyer, Soberman, Vilas-Boas, 2005 "The Targeting of Advertising," Marketing Science, 24, 461 – 476.

[15] Jullien, B., Y. Lefouili and M. Riordan, 2020, "Privacy Protection, Security, and Consumer Retention," TSE working paper # 18-947.

[16] Montes, R., Sand-Zantman, W., and Valletti, T. (2019). "The Value of Personal Information in Online Markets with Endogenous Privacy," *Management Science*, 65(3):1342–1362.

[17] Prüfer, J. and C. Schottmüller, 2017, "Competing with Big Data," unpublished manuscript, Tilburg University.

[18] Taylor, C., 2004, "Consumer Privacy and the Market for Customer Information," *RAND Journal of Economics*, 35, pp 631-650.

[19] Taylor, C. and L. Wagman, 2014, "Customer privacy in oligopolistic markets: Winners, losers, and welfare," *International Journal of Industrial Organization*, 34, 80 – 84.

[20] Varian, H. R., 1997, "Economics Aspects of Personal Privacy," in: Privacy and Self-Regulation in the Information Age. Washington, DC: US Department of Commerce, National Telecommunications and Information Administration.