# Data, Product Targeting and Competition*

Alexander Guembel
*Toulouse School of Economics*
*& Toulouse School of Management*

Ulrich Hege
*Toulouse School of Economics*

30 December 2023

**Abstract**

We analyze how access to data affects competition and consumer surplus in a model where more data allows firms to offer products that are better targeted to consumer preferences and at the same time to price discriminate more effectively. We find that consumer surplus in a monopoly market is highest for an intermediate level of data access, while it is increasing in available data when firms compete. The effect of data on competition is asymmetric: Competition becomes fiercer if the more poorly informed firm gets better information, but softens when the better informed firm improves its information. Firm's preferred choice of information is an outcome where they are strongly differentiated by information quality. This preference limits the possibility to create an informational level playing field via data sharing or data brokers, and explains why total surplus may drop following entry. If an entrant can use data gathered in one market in another market, entry does not necessarily improve overall consumer surplus, since it enhances the entrant's ability to price discriminate in the other market.

KEY WORDS: Digital platforms, product targeting, product quality, data and competition.

JEL CLASSIFICATION: D82, L13, L15.

*We would like to thank Gary Biglaiser, Edward Corcoran, Alexandre de Cornière, Doh-Shin Jeon, Bruno Jullien, Alvaro Martin, Alejandro Neut, Patrick Rey, Fernando Soto, Pablo Urbiola, and Wilfried Sand-Zantman for helpful suggestions and advice. This research has benefited from the financial support of BBVA in the context of the research foundation *TSE-Partnerhsip*. Both authors acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010. All errors are our own.

# 1 Introduction

It is by now well understood that data plays a crucial role in digital markets. 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. Product targeting is relevant for instance for digital platforms such as search, retail or subscription-based libraries, where better predictions of consumer tastes put firms in a position to offer customized choice menus, query answers, or blends of content and advertising. For example, in a recent piece on platforms grappling with the impact of generative A.I., the New York Times cites Manish Chandra, the CEO of Poshmark, as "daydreaming during a long flight from India about chatbots building profiles of people's tastes, then recommending and buying clothes or electronics."[1] Given the importance of data access, there is much debate about how to regulate the latter with a view to stimulating competition. In financial services, for example, the European Payments Directive (PSD2) obliges banks to allow access to third party payment service providers if a client so wishes. Since details about payments arguably contain useful information, such access may affect competition in adjacent markets like asset management or lending. It remains an open question how data affects competition when firms can choose both price *and* the featured product variety they offer.

We analyze a stylized model to study the competition effects when firms simultaneously choose prices and targeted products, based on the quality of their data and algorithms. In our model one or two firms serve a single consumer one unit of a product. The firms have noisy information about the consumer's preferred product specification. A firm's signal noise about consumer preferences captures in a reduced form way, the information it can extract from the available quantity and quality of data as well as their ability to process such data. After observing a noisy signal, each firm simultaneously chooses a product specification and a price. The consumer then chooses whether and from which firm to purchase.

We first characterize the optimal pricing strategy of a monopolistic firm as a function of

[1] "The Chatbots Are Here, and the Internet Industry Is in a Tizzy", New York Times, 8 March 2023.

how precise the firm's information about the consumer's tastes is. We show that for fairly precise information, the firm will offer a price just low enough to ensure that the consumer always makes a purchase. As information deteriorates, the monopolist therefore optimally reduces the price. In this region consumer surplus increases as the monopolist's information deteriorates. Although the average product match worsens, from the consumer's perspective the price reduction effect more than off-sets the poorer targeting. As information becomes quite noisy, the monopolist optimally stops lowering the price and instead accepts that the consumer will sometimes not purchase the product, namely when the firm's prediction about product preferences happens to be very poor. This makes consumer surplus hump-shaped in information quality.

We then characterize the equilibrium when two firms compete with each other. Importantly, we allow the two firms to have information of different quality, thereby introducing a difference between a better informed firm, called the incumbent, and a less well informed firm, called the entrant. We show that in equilibrium, the incumbent sets a higher price than the entrant, knowing that in expectation the latter will offer a more poorly matched product specification. Interestingly, the impact of information improvements depends on whether it is the follower's or the incumbent's information that improves. If the entrant's information improves, both firms reduce their price and the market becomes more competitive. Intuitively, both firms know that they will offer more similar product specifications, which intensifies competition. This contrasts with an improvement in the incumbent's information, which leads to an increase in industry profits, i.e., softer competition. When the incumbent's information improves, it increases its price, while the follower decreases it by the same amount. Intuitively, the incumbent faces a captured customer with positive probability, which confers market power. A captured customer is one who strictly prefers even the worst product specification of the incumbent over the follower's product specification. When the incumbent's information improves, the likelihood of facing a captured customer increases, making it optimal for the incumbent to raise its price. By contrast, the follower lowers the price when the incumbent's information improves as it would otherwise significantly lower the likelihood of making a sale. Although the magnitude of the incumbent's and the follower's price changes is equal (but in opposite directions), the customer buys from the incumbent with higher probability and hence an information improvement leads to an increase in industry profits.

We also compare surplus under monopoly and under duopoly. While consumer surplus is always higher under competition, total surplus may not be. Competition increases the range of products to choose from and thereby the average quality of a match. However, since the information follower offers a lower price, the customer sometimes chooses a product, in spite of its poorer match. This allocation inefficiency is more pronounced when the information gap between the two firms is large, in which case total surplus may be higher under monopoly. These findings show that competition and welfare crucially depend on the relative information quality of the two firms.

Next, we analyze endogenous information quality. We first ask whether facilitating access to data can encourage entry. As long as the potential entrant remains an information follower, improving its information would render competition fiercer if entry were to occur. This reduces the entrant's potential profits and thus backfires as a means to encourage entry. An exception to this pattern occurs if data access allows the potential entrant to leap-frog the incumbent in terms of information. That is, if an entrant can become an information leader by getting access to data, then entry may occur.

We further explore the strategic choice of data quality and extend the model by adding an initial stage where both firms strategically choose their data capabilities, before simultaneously choosing product specification and price. We find that firms will then strongly differentiate their data capabilities, with the information leader (incumbent) choosing maximal information precision, and the information follower (entrant) choosing such a low information quality that both firms can charge monopoly-type prices. This asymmetry in equilibrium arises because competition is reduced when data qualities are largely differentiated. We find that market entry in this equilibrium will always reduce total welfare. When there are costs of information acquisition, we find that the lower the investment costs and hence the higher the equilibrium data quality of the information leader, the lower the data investment of the competing firm.

These observations imply caveats for the effectiveness of regulatory interventions: data sharing policies may be toothless since the entrant has no inteest to acquire a comparable data capability to that of the incumbent, which would enhance competition and erode both firm's profits. For the same reason, data sales by information brokers will not mitigate but exacerbate information quality heterogeneity across competitors.

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 how a firm can use information obtained in one market for its pricing strategy in another market. We show that the informational spillover encourages entry, which is good for customers in the market in which entry occurs. On the downside, entry improves information in the entrant's "home" market, where it can price discriminate more effectively using the information obtained. This renders the effect of entry on overall consumer surplus ambiguous.

In an extension we study in more detail the learning problem faced by firm that can learn by experimenting in one market about something that is informative about another market. A firm can learn about consumer preferences, only 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.

Finally, motivated by competition concerns about platforms with high data capabilities (i.e. capabilities to access and process customer-specific data), we consider dynamic competition involving entry and exit. We show that an incumbent 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.

## 1.1 Relationship to the literature

A number of recent papers have studied how data availability affects price discrimination by a monopolist and thereby consumer's willingness to reveal information about themelves (e.g., Ichihashi, 2020), while others consider the effect of data on competition between horizontally differentiated sellers (e.g., Armstrong and Zhou, 2022, and Chen, Choe and Matsushima, 2020, or Nageeb Ali et al., 2022).

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 (2021) 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 (2023) 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. For example, more data available to a monopolist has a non-monotonic impact on consumer utility in our model, but not in de Cornière and Taylor (2023). De Cornière and Montes (2017) study the value of information to a monopolistic firm, when the latter can improve product quality only if it has access to data. Like our paper, they investigate the potential trade-off between price discrimination and product quality implied better data access. The focus of their paper is quite different from ours, in that they only consider a monopoly, and ask whether the firm should optimally charge a uniform price so as to induce customers to allow the firm access to data.

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, Shleifer 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 and analyzes the monopoly and duopoly baseline cases. Section 3 undertakes a welfare comparison between these two cases, and Section 4 looks at the endogenous choice of data quality and its role in entry deterrence. Section 5 considers strategic aspects of information spillovers across markets. Section 6 discusses robustness issues, and Section 7 concludes. Most proofs are relegated to the Appendix.

# 2 The Model

There is one consumer with a taste parameter $\eta$ drawn from an improper uniform distribution on the entire real line for a good of which he wishes to purchase a single unit.[2] There may be one or two firms $I$ and $E$ (for incumbent and entrant). One could think of firm $I$ as an incumbent who holds an informational advantage over entrant $E$, and we will often refer to firms $I$ and $E$ as incumbent and entrant, respectively. Each firm receives a signal $x_i = \eta + \tilde{e}_i$, about the consumer's taste parameter, where $\tilde{e}_i$ is uniformly distributed on the interval $[-\varepsilon_i, \varepsilon_i]$ $(i = I, E)$. $\varepsilon_i$ is a parameter measuring the information precision and we assume without loss of generality that $\varepsilon_I \leq \varepsilon_E$.

After having received the signal, each firm simultaneously chooses a product specification $l_i$ and a price $p_i$ for the good. The consumer values the good offered by seller $i$ at $v - |\eta - l_i|$, i.e., the closer the product is to the consumer's preferred specification $\eta$, the higher the consumer's utility of consuming it. The consumer surplus when buying form seller $i$ is $v - |\eta - l_i| - p_i$. We normalize the utility when no good is purchased to zero. We also normalize the cost of production to zero, so that the social surplus in the event of a sale is $v - |\eta - l_i|$, and zero if there is no sale. While we limit the anlysis to a unit demand for simplicity, the model contains an element of price-sensitive demand since a sale will only occur if $v - |\eta - l_i| \geq p_i$ and hence the probability of a sale is a decreasing function of price $p_i$.

## 2.1 Equilibrium and Welfare with a Monopolistic Firm

We first analyze the case where the incumbent is a monopolist in the market. For ease of exposition we omit the subscript $i$ in this section.

**Lemma 1** *The monopolist's optimal choice of product specification is $l = x$ and pricing*

[2] Alternatively one could assume that preferences are distributed uniformly on a large circle. The disadvantage with the latter specification is that the limiting case of an uninformed firm is more cumbersome to analyze.

*strategy as well as profits are given by*

$$p^M = \begin{cases} v - \varepsilon_I & \text{if} \quad v \geq 2\varepsilon_I \\ \frac{v}{2} & \text{if} \quad v < 2\varepsilon_I \end{cases}; \tag{1}$$

$$\pi^M = \begin{cases} v - \varepsilon_I & \text{if} \quad v \geq 2\varepsilon_I \\ \frac{v^2}{4\varepsilon_I} & \text{if} \quad v < 2\varepsilon_I \end{cases}.$$

*Consumer surplus $CS^M$ and total surplus $TS^M$ are given by*

$$CS^M = \begin{cases} \frac{\varepsilon_I}{2} & \text{if} \quad v \geq 2\varepsilon_I \\ \frac{1}{2}\frac{v^2}{4\varepsilon_I} & \text{if} \quad v < 2\varepsilon_I \end{cases}, \tag{2}$$

$$TS^M = \begin{cases} v - \frac{\varepsilon_I}{2} & \text{if} \quad v \geq 2\varepsilon_I \\ \frac{3}{2}\frac{v^2}{4\varepsilon_I} & \text{if} \quad v < 2\varepsilon_I \end{cases} \tag{3}$$

Proof: see Appendix.

When the monopolist firm has strong information ($\varepsilon_I \leq \frac{v}{2}$) it sets the price just low enough to ensure that the customer makes a purchase, even if the product is poorly targeted ($|\tilde{e}| = \varepsilon_I$). As the firm's information gets worse, it needs to lower the price to ensure that a purchase always occurs. At some point (when $\varepsilon_I > \frac{v}{2}$) the price would have to be so low, that it becomes preferable for the monopolist to maintain the price at $p = \frac{v}{2}$ and accept that the product will not be sold when its specification happens to be less suitable.

From Lemma 1 we can see that total surplus always increases 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 $\varepsilon_I$), 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 $\varepsilon_I = \frac{v}{2}$; up to this point, the monopolist's product is sufficiently well-targeted so that a sale always occurs and the monopolist's high price has no impact on total surplus.

When information worsens beyond point $\varepsilon_I = \frac{v}{2}$, however, the monopolist prefers not to drop the price any further and instead accepts that the consumer sometimes makes no purchase at all. Hence, there is sometimes no sale even though a sale would produce a social value of $v - \varepsilon_I$ (we will denote this outcome as a situation where the market is not fully covered, see the next section). From the point of view of social welfare, therefore, a lower price would be desirable in this region since it would lead to more trade and increase consumer surplus more than it decreases producer surplus. 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. Consumer surplus is thus maximized at the point $\varepsilon_I = \frac{v}{2}$ when the monopolist's information quality is the poorest that still ensures full market coverage. This result is reminiscent of Hidir and Vellodi (2020) who show in a setting where a monopolist seller can offer multiple, horizontally differentiated products, that the buyer-optimal segmentation of a market is the least informative segmentation that guarantees trade.

## 2.2 Equilibrium with Duopoly

We assume that each firm's information quality $\varepsilon_i$ is exogenously given and 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 $\eta$, i.e., each firm chooses $l_i = x_i$. Given these location choices, firm $E$ sells its good if it is more attractive to the consumer than $I$'s offer, i.e., when

$$v - p_E - |\tilde{e}_E| \geq v - p_I - |\tilde{e}_I|, \tag{4}$$

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

$$v - p_E - |\tilde{e}_E| \geq 0. \tag{5}$$

Firm $I$ sells if inequality (4) is reversed and when

$$v - p_I - |\tilde{e}_I| \geq 0. \tag{6}$$

This allows us to calculate the probabilities of $I$ or $E$ selling the good, as follows. We conjecture (and later prove) that the better informed firm ($I$) uses its informational advantage to charge at least as high a price as its rival, i.e., $p_I \geq p_E$. Moreover, we conjecture (and later prove) that the price difference is bounded by the informational difference; specifically, $p_I - p_E \leq \varepsilon_E - \varepsilon_I$. 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 $\tilde{e}_E, \tilde{e}_I$ such that the consumer does not purchase the good at all. This happens when (6) is binding for some realizations of $\tilde{e}_I$, i.e., if $p_I > v - \varepsilon_I$. Given that $p_I - p_E \leq \varepsilon_E - \varepsilon_I$, the inequality $p_I > v - \varepsilon_I$ implies $p_E > v - \varepsilon_E$. In other words, of the two participation constraints, that of the better informed firm $I$ is the more binding. This is intuitive. For poor realizations of $E$'s signal (large realization of $\tilde{e}_E$), $E$ loses out to its competitor even for product specifications that the consumer would be willing to purchase, were firm $E'$s the only offer in the market. Hence, from $E$'s perspective, the binding constraint is not whether the consumer prefers $E$'s product over no purchase (inequality (5)) but whether it prefers $E$'s product over $I$'s.

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 this is only the case when firm $I$'s equilibrium price satisfies $p_I \leq v - \varepsilon_I$. When $p_I > v - \varepsilon_I$, then sometimes there is no sale because the consumer's reservation value is not met by the combination of price and product specification. Whether the market is fully covered in equilibrium depends on the value $v$ compared to firm $E$ and $I$'s information $\varepsilon_E$ and $\varepsilon_I$. For this purpose it is useful to introduce the following functions, which will serve to describe the thresholds of $v$ for which the market is fully or partially covered.

$$
\begin{aligned}
f(\varepsilon_E, \varepsilon_I) &= \varepsilon_E + \frac{7}{4}\varepsilon_I - \sqrt{\left(\varepsilon_E - \frac{1}{4}\varepsilon_I\right)^2 + \varepsilon_I^2}, \\
g(\varepsilon_E, \varepsilon_I) &= \frac{2}{3}\varepsilon_E + \frac{5}{6}\varepsilon_I,
\end{aligned}
$$

where $f(\varepsilon_E, \varepsilon_I) \leq g(\varepsilon_E, \varepsilon_I)$. Figure 1 provides an illustration of the three regions delimited by functions $f(\varepsilon_E, \varepsilon_I)$ and $g(\varepsilon_E, \varepsilon_I)$.

The equilibrium is characterized by the location of maximum surplus $v$ relative to these

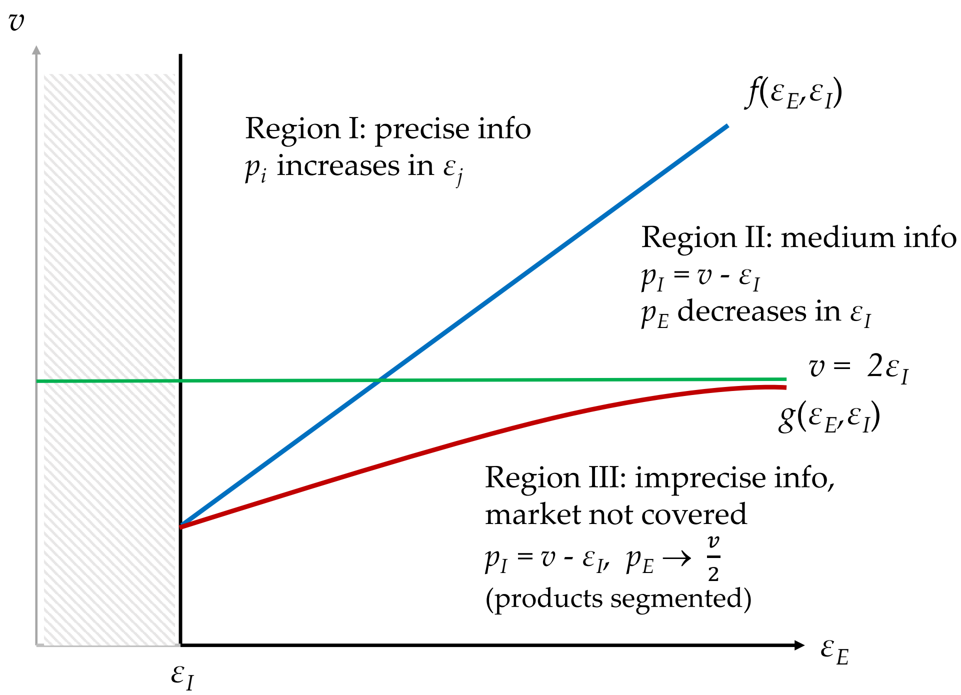


Figure 1: Plot of the boundaries of the Regions I - III as $v$ and $\varepsilon_E$ vary, for a constant value of $\varepsilon_I = 1$.

three different regions, 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:*

*Region I: If $v > g(\varepsilon_E, \varepsilon_I)$, then the the market is fully covered and*

$$p_E = \frac{\varepsilon_E}{3} + \frac{\varepsilon_I}{6} \tag{7}$$

$$p_I = \frac{2\varepsilon_E}{3} - \frac{\varepsilon_I}{6}. \tag{8}$$

*Region II: If $v \in [f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I)]$ then the market is fully covered and*

$$p_E = \frac{v}{2} - \frac{\varepsilon_I}{4}, \tag{9}$$

$$p_I = v - \varepsilon_I. \tag{10}$$

*Region III: If $v < f(\varepsilon_E, \varepsilon_I)$ then the market is not fully covered, and prices are given by* *the unique solution of the following two equations on the interval* $p_I \in \left(v - \varepsilon_I, \frac{v}{2}\right)$*:*

$$p_E = \frac{v}{2} - \frac{1}{4}\frac{(v - p_I)^2}{\varepsilon_I}, \tag{11}$$

$$p_E = \frac{1}{2}\frac{3p_I - v}{v - 2p_I} + v - \varepsilon_E. \tag{12}$$

Proof: see Appendix.

Only in Regions I and II will the consumer always buy the good, i.e., the market is fully covered. In Region I, the entrant (information follower) has relatively precise information, and by implication, the incumbent (information leader) also has pretty good information. Prices are strictly below the consumer's purchase threshold, even for the least well targeted products. This region is relevant when in Region II, the entrant has fairly low quality information and firm $I$ behaves like a monopolist, setting $p_I = v - \varepsilon_I$ so as to ensure that the consumer would just be willing to purchase the less well-targeted product. Note that for $\varepsilon_E = \varepsilon_I$ we get $f(\varepsilon_E, \varepsilon_I) = g(\varepsilon_E, \varepsilon_I)$ so Region II disappears. In Region III the market is no longer fully covered. Note that the market is fully covered in duopoly for a larger set of parameters than for monopoly. If we think of the monopolist corresponding to the better informed firm (with inverse information precision $\varepsilon_I$), we have partial market coverage under monopoly, but full coverage under duopoly for values of $v \in [f(\varepsilon_E, \varepsilon_I), 2\varepsilon_I)$.

To gain some intuition for the way in which information affects competition, consider Region I first. In this region, the firm's best responses to the other firm's price is given by (see Proof of Proposition 1 in the Appendix for the derivation)

$$p_E(p_I) = \frac{p_I}{2} + \frac{1}{4}\varepsilon_I \tag{13}$$

$$p_I(p_E) = \frac{p_E + \varepsilon_E}{2} - \frac{1}{4}\varepsilon_I. \tag{14}$$

Prices are strategic complements, as usual in price competition with differentiated products. Moreover, each firm lowers its price when its rival's information improves (i.e., $p_E$ increases in $\varepsilon_I$ and vice versa). Consider now how each firm reacts to an improvement in its own information. When $\varepsilon_I$ decreases (the incumbent's information is becoming more precise), firm $I$ charges a higher price. This is intuitive as better information corresponds to stronger consumer capture. By this, we mean that there are realizations of $E$'s signal where $|\tilde{e}_E| > \varepsilon_I$, and hence a consumer prefers $I'$s product, no matter what signal $I$ receives. Thus, when $I'$s information improves, the likelihood that a consumer is captured increases, rendering it optimal for $I$ to increase the price (it loses a "marginal" consumer who is indifferent between $E$ and $I$ but gets a higher price on all the infra-marginal consumers who strictly prefer $I$). Note that $E$ does not have captured consumers: for any signal realization of $I$, firm $E$ risks losing out to $I$. Loosely speaking, $E$ has to compete for every consumer, while $I$ has market power over a market segment (of customers who get poor product offers from $E$). When $E$'s information improves marginally, it does not find it optimal to increase the price. $E$'s best response $p_E$ is actually independent of $\varepsilon_E$.

In equilibrium, a reduction in $\varepsilon_E$ therefore unequivocally works to enhance competition: when $\varepsilon_E$ is smaller, $I$ lowers its price, which leads $I$ to also lower its, as prices are strategic complements. On the other hand, with smaller $\varepsilon_I$, $E$ decreases its price, while $I$ increases it. The magnitude of the price changes is symmetric, but since the consumer ends up buying $I$'s product with a higher likelihood, the average price paid actually increases with $I$'s information. To summarize, when the information follower's information improves, competition gets fiercer. The effect of better information for the incumbent is not symmetric: the information follower lowers price, while the incumbent increases it.

Starting from a point in Region I, increasing $\varepsilon_E$ will eventually move us into Region II. Using the expressions for prices it is easy to see that prices are continuous (but exhibit a kink) as we move from Region I to Region II. We can think of Region II as the region in which the entrant is at a sufficiently strong disadvantage that the incumbent can act like a monopolist. That is, $I$'s price is so high that it reaches the corner where the consumer's participation constraint just binds for $I$'s worst possible product specification (relative to the consumer's true preferences). The price $p_I$ is thus pinned down by $p_I = v - \varepsilon_I$. Hence, if $I$ has better information, it will increase its price in this region. Since prices are strategic complements, the information follower also increases its price in response. In this region, an improvement in $I$'s information unambiguously reduces competition. Changing $E$'s information does not affect competition. Although $I$ prices like a monopolist, the second firm, $E$, is not irrelevant. In equilibrium it will steal some business from the incumbent. This disciplines the incumbent $I$ to set a price $p_I = v - \varepsilon_I$ even when this results in a price $p_I < \frac{v}{2}$. In this way, the presence of a second firm ensures the market remains fully covered for a larger parameter set than under monopoly. If $I$ is a monopolist, then the market is only fully covered if $\varepsilon_I \leq \frac{v}{2}$. If there is a entrant firm $E$ with almost equally good information, $\varepsilon_E \approx e_I$, then the market is fully covered even for $\varepsilon_I \in \left\{\frac{1}{2}v, \frac{2}{3}v\right\}$ (see the definition of $f(\varepsilon_E, \varepsilon_I)$ for identical precisions). But as $E$'s information deteriorates, its role gradually fades away: for $\lim_{\varepsilon_E \longrightarrow \infty} f(\varepsilon_E, \varepsilon_I) = 2\varepsilon_I$, the lower bound on being in Region II becomes $v \geq 2\varepsilon_I$, which corresponds to the condition of full market coverage in the monopoly case. That is, the duopoly converges to the monopoly case when $E$'s information becomes useless (when $\varepsilon_E$ goes to infinity).

To summarize, over Regions I and II improvement in the entrant's information (weakly) increases competition, while the opposite is true for improvements in the incumbent's information.

In Region III the market is only partially covered under duopoly. As we will see in Section 4, this situation is unlikely to arise when it is relatively easy or cheap for firms to improve their information. Also, the less precise is the competitors' information, the less likely are firms to overlap in their location choices, and the more likely there will be only one firm offering a product that is sufficiently targeted to offer a positive net value even at a price of zero. That is, the less informed the competitors are, the more likely it is that the market is de facto segmented. For the rest of the paper, therefore, we will assume that information is relatively precise such that the market is fully covered in the duopoly case, i.e.,

$$v \geq f(\varepsilon_E, \varepsilon_I). \tag{15}$$

# 3 Welfare in Duopoly

We start with a technical result concerning consumer and total surplus.

**Lemma 2** *Consumer surplus $CS^D$ and total surplus $TS^D$ are given by*

$$CS^D = TS^D - (\pi_E + \pi_I), \tag{16}$$

$$TS^D = v - \frac{1}{2\varepsilon_E}\left\{(\Delta p)^2 + \varepsilon_E\varepsilon_I - \frac{\varepsilon_I^2}{3}\right\}. \tag{17}$$

*In Region I ($v > g(\varepsilon_E, \varepsilon_I)$) we have*

$$\pi_E = \frac{1}{\varepsilon_E}\left(\frac{\varepsilon_E}{3} + \frac{\varepsilon_I}{6}\right)^2, \tag{18}$$

$$\pi_I = \frac{1}{\varepsilon_E}\left(\frac{2\varepsilon_E}{3} - \frac{\varepsilon_I}{6}\right), \tag{19}$$

$$\Delta p = \frac{\varepsilon_E - \varepsilon_I}{3}. \tag{20}$$

*In Region II ($v \in [f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I)]$) we have*

$$\pi_E = \frac{1}{\varepsilon_E}\left(\frac{v}{2} - \frac{\varepsilon_I}{4}\right)^2, \tag{21}$$

$$\pi_I = \frac{1}{\varepsilon_E}\left(\varepsilon_{EE} - \left\{\frac{v}{2} - \frac{\varepsilon_I}{4}\right\}\right)(v - \varepsilon_I), \tag{22}$$

$$\Delta p = \frac{v}{2} - \frac{3}{4}\varepsilon_I. \tag{23}$$

We can then turn to a comparison of the monopoly and duopoly cases. 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 that of the two duopolists. We assume that the better informed firm in duopoly (firm $I$) has the same information quality as the monopolist. Thus, the duopoly can be thought of as resulting from a monopoly firm that faced entry from a less well informed rival. From a more technical point of view, if one were to assume the opposite (i.e., the entrant is better informed than the monopolist) any improvement in surplus may simply be due to the (hard wired) improvement in information quality stemming from the entrants better information.

**Proposition 2** *(a) If the market is fully covered under monopoly ($v \geq 2\varepsilon_I$), then total surplus under monopoly is higher than under duopoly if and only if $\varepsilon_E$ is sufficiently large.*

*(b) When the market is not fully covered under monopoly ($v < 2\varepsilon_I$), but it is fully covered under duopoly ($v \geq f(\varepsilon_E, \varepsilon_I)$), then total surplus is always higher under a duopoly.*

*(c) 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 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 ($E$) sets a lower price. This means that firm $E$ sometimes sells the good, even though the consumer prefers $I$'s specification. This misallocation becomes more pronounced when the informational gap between $E$ and $I$ increases and hence the difference in prices is larger. Taken together, the two effects make total monopoly surplus higher than duopoly surplus when $\varepsilon_I$ is fairly small compared to $\varepsilon_E$. 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 now consider how a change in the information environment affects total surplus, industry profits and consumer surplus.

**Proposition 3** *(a) Total surplus decreases as either firms' information quality worsens, i.e.,*

$$\frac{\partial TS^D}{\partial \varepsilon_E} < 0, \text{ and } \frac{\partial TS^D}{\partial \varepsilon_I} < 0.$$

*(b) Industry profits $\pi_E + \pi_I$ increase when the information follower's information quality worsens and decreases when the incumbent's information quality worsens, i.e.,*

$$\frac{\partial \left(\pi_E + \pi_I\right)}{\partial \varepsilon_E} > 0, \text{ and } \frac{\partial \left(\pi_E + \pi_I\right)}{\partial \varepsilon_I} < 0.$$

*(c) Consumer surplus decreases when the information follower's information quality wors-*

*ens and is U-shaped in the incumbent's information quality, in particular*

$$\frac{\partial CS^D}{\partial \varepsilon_E} < 0, \; \textit{and} \; \frac{\partial CS^D}{\partial \varepsilon_I} \begin{cases} < 0 \textit{ for } \varepsilon_I \leq \frac{6}{5}v - \frac{4}{5}\varepsilon_E \textit{ (Region I)} \\ > 0 \textit{ for } \varepsilon_I > \frac{6}{5}v - \frac{4}{5}\varepsilon_E \textit{ (Region II)} \end{cases}.$$

Proof: see Appendix.

An improvement in either firms' information leads to an increase in total surplus, because the consumer gets, on average, a more suitable product. When the information follower becomes better informed, competition intensifies and industry profits fall. The consumer thus benefits from better information for the follower in two ways: better product offerings and lower prices. Consumer surplus therefore unambiguously increases when $E$'s information improves.

An improvement in the incumbent's information has more equivocal implications. In Region I, even the information follower has fairly good information and competition is correspondingly fierce. While better information for the incumbent ends up increasing the incumbent's price, it reduces the follower's price by an equal amount. Since the incumbent sells more often, this leads to a modest increase in industry profits. This increase, however, is not enough to reverse the positive effect of better information that consumers enjoy through an improved product offer. Consumer surplus therefore increases when $I$'s information improves.

In Region II, the information follower has relatively poor information, inducing the incumbent to price like a monopolist. The latter raises price more strongly in response to better information, compared to the previous case. This allows the follower to also increase price, leading to a strong increase in industry profits. Indeed, prices increase so strongly that consumers are worse off from better information in spite of the improvement in product offers.

If $\varepsilon_E$ is relatively small compared to $v$ (namely when $v > g(\varepsilon_E, \varepsilon_I = \varepsilon_E) = \frac{3}{2}\varepsilon_E$), then any value of $\varepsilon_I \in [0, \varepsilon_E]$ will leave us in Region I. In this case, consumer surplus is increasing whenever the incumbent's information improves. For large values of $\varepsilon_E$ (namely when $v < g(\varepsilon_E, \varepsilon_I = 0) = \frac{2}{3}\varepsilon_E$), we are in Region II for all admissible values of $\varepsilon_{Ii}$ and hence consumer surplus is always decreasing when the incumbent's information improves.

For intermediate values of $\varepsilon_E$, (when $v \in (\frac{2}{3}\varepsilon_E, \frac{3}{2}\varepsilon_E)$), we move from Region I into Region II as $\varepsilon_I$ increases. In this case, consumer surplus reaches a minimum at the point $\varepsilon_I^*$ defined by $v = g(\varepsilon_E, \varepsilon_I^*)$.

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. When the supplier is a monopolist, the consumer benefits from an intermediate level of information: low enough to reduce price discrimination, but high enough to ensure the market remains fully covered. In duopoly, if the follower is quite well informed, the consumer benefits from making a maximum amount of information available. Competition is in any case fierce and an improvement in the incumbent's information is good for consumers via the effect on product targeting. If the follower is less well informed, the consumer's incentives to provide information are closer to those when he faces a monopolist, i.e., less information is better since it reduces the incumbent's ability to price discriminate.

# 4 Choice of Information Quality: Entry, Data Sharing and Data Sales

The analysis so far reveals that competition, social and consumer welfare in a duopoly crucially depend on the relative information quality of the two firms (represented by $\varepsilon_i$, the inverse of quality). We further explore this crucial relationship by analyzing questions related to the choice of information quality: we look at the consequences for firm entry and then characterize the outcome when firms can freely choose their information quality. The insights of these analyses help us to understand the limits on policies or market activities that in theory could a level playing field in information competition, data sharing and data sales.

## 4.1 Information quality and profits

We first analyze how profits react to changes in information quality, building on expected profits derived in Lemma 2. In order to focus the analysis we restrict attention to Region

I. We slightly generalize the set-up and now allow either incumbent or entrant to be the better informed party but we maintain the firm labels $I$ and $E$. Using (18) and (19) this gives us the profit functions

$$\pi_I = \begin{cases} \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2 & \text{if} \quad \varepsilon_I < \varepsilon_E \\ \frac{1}{\varepsilon_I}\left(\frac{1}{3}\varepsilon_I + \frac{1}{6}\varepsilon_E\right)^2 & \text{if} \quad \varepsilon_I \geq \varepsilon_E \end{cases}, \tag{24}$$

and

$$\pi_E = \begin{cases} \frac{1}{\varepsilon_E}\left(\frac{1}{3}\varepsilon_E + \frac{1}{6}\varepsilon_I\right)^2 & \text{if} \quad \varepsilon_I < \varepsilon_E \\ \frac{1}{\varepsilon_I}\left(\frac{2}{3}\varepsilon_I - \frac{1}{6}\varepsilon_E\right)^2 & \text{if} \quad \varepsilon_I \geq \varepsilon_E \end{cases}. \tag{25}$$

A comparison between (24) and (25) reveals that $\pi_I > \pi_E \Leftrightarrow \varepsilon_I < \varepsilon_E$.

We can then analyze the reaction of profits in this second stage to a change in information quality. The following lemma provides a key insight about the forces behind our subsequent results:

**Lemma 3** *In Region I, 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 directly from calculating the derivatives of (24) and (25). ■

Since we assume that $I$ is the better informed firm ($\varepsilon_I < \varepsilon_E$), it follows, perhaps somewhat surprisingly, that $E$ does not benefit from an improvement in its information. Consider the equilibrium prices (7) and (8) to understand why that is. As $\varepsilon_E$ falls ($E$'s information improves), the two firms compete more fiercely with one another. Although $E$'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. By contrast, the better informed firm, can afford to translate better information into a *higher* price resulting in higher equilibrium profits. Hence, the incentives for the information quality choice are exactly the opposite for the incumbent (firm $I$) and the information follower (firm $E$).

## 4.2 Entry and leap-frogging

The observation of the V-shaped pattern of profits allow us to investigate the incentives for a new entrant to emerge in a market previously dominated by a monopolist. We assume that the potential entrant $E$ faces an entry cost $F_E$,[3] and that entry has not yet occurred. That is, the information qualities of the incumbent ($\varepsilon_I$) and the entrant ($\varepsilon_E$) are such that entry is not profitable ($\pi_E(\varepsilon_E, \varepsilon_I) - F_E < 0$). Suppose also that the entrant is at an informational disadvantage to the incumbent ($\varepsilon_I < \varepsilon_E$). Consider now a regulator who aims to improve the potential entrant's information, for example, by mandating data sharing. For simplicity, we model this as a reduction in $\varepsilon_E$.[4] We can then state the following:

**Proposition 4** *An improvement in the entrant's information to $\varepsilon'_E < \varepsilon_E$ does not trigger entry as long as the entrant remains the information follower ($\varepsilon'_E > \varepsilon_I$). A necessary condition for entry is that the entrant leap-frogs the incumbent in information quality ($\varepsilon'_E < \varepsilon_I$).*

**Proof.** From (25) we can see that the entrant's profits are increasing in $\varepsilon_E$ for $\varepsilon_E > \varepsilon_I$. Hence, if $\pi_E(\varepsilon_E, \varepsilon_I) - F_E < 0$ then for any $\varepsilon'_E \in [\varepsilon_I, \varepsilon_E]$ we also have $\pi_E(\varepsilon'_E, \varepsilon_I) - F_E < 0$. For $\varepsilon_I > \varepsilon_E$ the entrant's profit is decreasing in $\varepsilon_E$. ■

[3]The entry cost 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 $\varepsilon_I$ 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; $\varepsilon_E$ is the quality of the competing offer of an independent product comparator.

[4]One could think of data sharing as literally having the entrant observe the incumbent's information directly. However, it is not realistic to assume that under data sharing two firms have literally identical information. Firstly, the incumbent may have soft information from client relationships which fall outside the remit of a data sharing agreement. Second, to the extent that two firms employ different algorithms to learn from data, they will draw different inferences even if they share the same raw data. Our signals $\widetilde{e}_I$ and $\widetilde{e}_E$ can be thought of as the inference at which each firm arrives after having applied its algorithm.

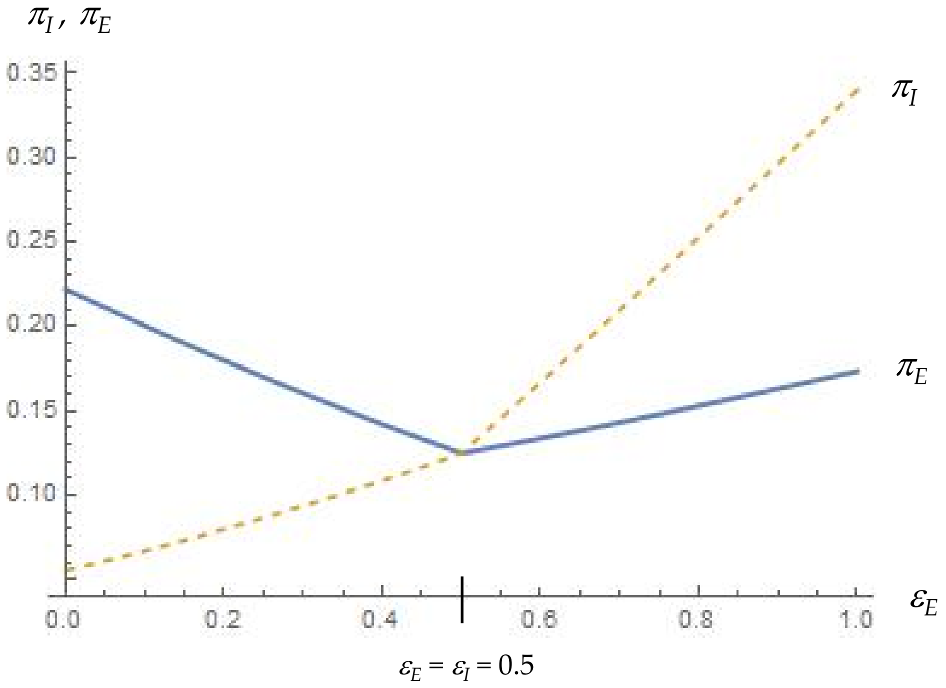


Figure 2: Plot of the boundaries of the Regions I - III as $v$ and $\varepsilon_E$ vary, for a constant value of $\varepsilon_I = 1$.

In the previous section we saw that an improvement in the follower's information makes the market more competitive and reduces both firms' profits. Improving the entrant's information by a small amount therefore will not induce entry. The entrant anticipates that the incumbent will compete more fiercely if entry occurred following, say a data sharing agreement. This undermines incentives to enter in the first place. Things are different if the informational advantage conveyed on the entrant is so large that it would leap-frog the incumbent in information terms. In that case, the entrant's profits increase with an improvement in the information. If the informational advantage is sufficiently large, the entrant's profits may increase enough to justify paying the entry cost $c$.

In an extension in Section 6.2.1, we show that the analysis can be extended to the case when there are explicit costs of improvements in data quality.

## 4.3 Choice of information quality

To further understand the forces that guide the equilibrium of the information quality choice, we now endogenize information quality and consider that firms can freely choose the quality of their information $\varepsilon_i$ at no cost. This benchmark case allows for a simple characterization of the equilibrium; (we consider the case of costly information acquisition in Section 6.2.2, and show that the main insights carry over). The cases of data sharing and data sales, considered in more detail later in this section, offer leading examples why in practice firms have a choice in the quality of the information used to target products. If data sharing is mandated by regulators for instance, the cost of improving information quality might be small. Also, firms may strategically opt to use less precise information than is available to them to limit competition, in which case the assumption of no cost of varying information quality is plausible.

Specifically, we consider the following expansion of the game in which competition now plays out in two sequential decisions: firms first decide on their data strategies determining $\varepsilon_i$, then on their product location and prices. While in the second round firms move simultaneously, in the first round we consider sequential decisions, with the incumbent choosing $\varepsilon_I$ first as the Stackelberg leader. The entrant then observes the incumbent's choice and chooses $\varepsilon_E$ as the Stackelberg follower, and $E$'s choice is also observed by both firms. In the second stage, the same game is played as in the baseline model, with both firms choosing location and price simultaneously.

Before analyzing the equilibrium it is useful to consider the case of a monopolist able to choose information quality $\varepsilon$ at no cost. From Lemma 1, it is easy to see that the monopolist would choose the best possible information quality $\varepsilon_I = 0$ since $\varepsilon_I = 0$ allows it to maximizes its profit: with a monopoly price of $p_M = v$ and a sale with probability 1, it will earn a profit of $\pi^M = v$.

We find the following equilibrium outcome in the duopoly case:

**Proposition 5** *(a) In the unique equilibrium, Firm I chooses $\varepsilon_I = 0$ and Firm E chooses $\varepsilon_E = \frac{3}{2}v$. In the second stage, Firm I offers a price of $p_I = v$, and Firm E a price of $p_E = \frac{v}{2}$. Firm I makes a sale with probability $\frac{2}{3}$, earning a duopoly profit $\pi_I = \frac{2}{3}v$, and*

*Firm $E$ earns a profit of $\pi_E = \frac{1}{6}v$.*

*(b) The Total Surplus is $TS^I = \frac{11}{12}v$ and hence lower than in the monopoly case when $\varepsilon_I = 0$, whereas consumer surplus $CS^I = \frac{1}{12}v$ is higher than in the monopoly case when $\varepsilon_I = 0$.*

Proof: see Appendix.

The result shows that the Stackelberg leader $I$ will in fact continue to choose the maximal information precision $\varepsilon_I = 0$, the same it would choose as a monopolist. Firm $E$ reacts to the incumbent's choice with a choice of $\varepsilon_E$ that is maximally distant in precision, consequence of our observation in Lemma 3 that the information follower's profit *increases* when it has *less* precise information, in other words that larger differentiation in information relaxes competition. In fact, Firm $E$ chooses an information precision that is sufficiently differentiated from that of the incumbent that it chooses a price $p_E = \frac{v}{2}$, the price choice of an uninformed monopolist (monopoly outcome when $\varepsilon \geq \frac{v}{2}$). Therefore, any further decrease in the entrant's information quality would not lead to a concomitant increase in its price but would result in a loss in market share, i.e. firm $E$ would make a sale less often as a result of its noisier product targeting. Hence any further differentiation will reduce $E$ 's profit and Firm $E$ will stop the information differentiation at $\varepsilon_E = \frac{3}{2}v$. In fact, $E$ plays the role of an opportunistic niche supplier: it maintains an uncompetitive price and hopes to be a "chance" supplier by fishing for customers that are accidentally close its noisy product location choice.

This outcome of maximal informational differentiation is reflected in high prices and profits: the aggregate producer surplus of $\pi_I + \pi_E = \frac{5}{6}v$ is quite close to the monopoly profit ($\pi^M = v$), whereas the consumer surplus of $CS = \frac{1}{12}v$, barely above the monopoly level. This result highlights one of the main insights of our paper: firms will choose opposite positions of information precision whenever they can to ease price competition. Perhaps the most intriguing result is the welfare comparison with the monopoly case: entry of a firm in a monopolistic market will lower social welfare that can freely choose its information quality (but increase consumer surplus). This observation shows that regulatory responses need to account for the fundamental driving force of information differentiation in the context of information-based product targeting. Regulatory policies intended to give entrants an informational head start might be counterproductive, as we discuss in the next section.

It is also interesting to locate the equilibrium outcome within the regions that we have discussed in Proposition 1. In fact, with $\varepsilon_I = 0$, the function $f(\varepsilon_E, \varepsilon_I) = 0$, so that Region III cannot occur (the market is always covered). Also, with $\varepsilon_E = \frac{3}{2}v$, we find that $g(\varepsilon_E, \varepsilon_I) = v$ so that the outcome is located exactly at the boundary of Regions I and II, reflecting the fact that Firm $E$ chooses a degree of informational differentiation up to the point where its own price choice will no longer benefit from further differentiation. This analysis also shows that the case of competitors with sufficient data quality is the most interesting one, because in this case firms will compete in information.

While the outcome discussed in this section hinges on the assumption of no information cost, we show in Section 6.2.2 that the structure of the equilibrium outcome generalizes and that we generally get an outcome with $\varepsilon_I < \varepsilon_E$, and where Firm $E$ reacts to the incumbent's choice $\varepsilon_I > 0$ with a choice of $\varepsilon_E$ that is decreasing in $\varepsilon_I$ (the choice will be, $\varepsilon_E = \frac{3}{2}v - \frac{5}{4}\varepsilon_I$).

The choice of information quality is related to the question of firm entry since information quality is a strategic instrument to facilitate or discourage entry and exit of competitors. We discuss this question next and also in Section .

## 4.4 Data sharing

By data sharing, we refer to regulatory or voluntary initiatives that give entrants access to the information available to the incumbent. The two preceding subsections offer valuable insights in this regard.

First, the analysis on entry and leap-frogging throws new light on the challenge faced by regulators if they wish to encourage entry by facilitating access to information. While it is true that an entrant who has an informational disadvantage may be discouraged from entering a market, it does not follow that reducing the informational disadvantage encourages entry. As we can see, the information differential between two firms determines how fierce competition will be following entry. Competition is fiercer, the more similar both firms' information quality. Entry incentives are therefore minimized when the entrant has the same information quality as the incumbent. In order to induce entry, the regulator would have to ensure that the entrant obtains a significant informational advantage over the incumbent. It is hard to see how a regulator could achieve this, although we have more to say about this in the following section.

Second, our analysis of a fully endogenized choice of information quality shows that market participants tend to opt for an outcome where their data capabilities are strongly differentiated. This result sheds light on important obstacles for regulatory policies. The fact that regulators may force incumbents to give entrants access to their data does not mean that entrants will actually make full use of the data access offer. If the incumbent monopolist has fairly precise information capabilities ($\varepsilon_I \approx 0$), then an entrant has no interest to opt for full data access and acquire similar information capabilities, since both firms would then offer very accurately targeted products and will fiercely compete in prices and eliminate profits (with the Bertrand equilibrium outcome in the limit of $\varepsilon_I \to 0$ and $\varepsilon_E \to 0$). Perversely, an entrant with insufficient information to launch a successful entry may exploit data sharing policies for a niche entry strategy as very inaccurately targeting yet profitable fringe supplier.

## 4.5 Data brokers

Access to data may be facilitated by third party data brokers. There is a significant literature that investigates how such information intermediaries should optimally package and sell information (see Bergemann and Bonatti, 2019, for a review). In the context of our model, we wish to explore whether the possibility to improve information by acquiring data can help spur competition. That would be the case, if an information follower could acquire information. We first consider a situation with two firms that compete and a data broker who comits to an exclusive sale of data to the firm that pays the highest price for it. We then discuss the implications for market contestability and incentives for a data broker to make a non-exclusive sale of information.

We model the acquisition of additional data by assuming that it reduces signal noise by a factor $\delta < 1$. That is, a firm $i$ whose signal noise is $\varepsilon_i$ can reduce noise to $\delta\varepsilon_i$ by acquiring information.[5] We now denote the two firms by $i = L$ for the incumbent and $i = F$ for information follower, prior to the acquisition of information, that is $\varepsilon_L \leq \varepsilon_F$. We make now restriction on which of the two firms is incumbent after the sale of information.

**Proposition 6** *The incumbent always outbids the information follower in an exclusive data sale.*

Proof: see Appendix.

In order to better understand Proposition 6 it is useful to distinguish two cases, depending on whether the acquisition of data does or does not allow the follower to leap-frog the leader (i.e., $\delta\varepsilon_F \lesseqgtr \varepsilon_L$). Consider first the case where data acquisition does not allow leap-frogging ($\delta\varepsilon_F > \varepsilon_L$). From the previous section we saw that an information improvement for the follower reduces profits. This, however, does not imply that the follower would be unwilling to pay to acquire the information. This is because the willingness to pay for the data is not only affected by the value of the data to the firm, but also by the value of denying access to the data to the competitor.[6] If the follower does not acquire the information, the leader will and this reduces the follower's profits. The question is then whether the follower's profits suffer more if the information is in the follower's or the leader's possession.[7] The answer is that the follower suffers more from data access being given to the leader. The follower would therefore be willing to pay a strictly positive price to have exclusive data access, even if this does not allow it to leap-frog the leader. The follower's willingness to pay for the data needs to be compared to the leader's, whose profits increase when it

[5] Alternatively, we could model the reduction in noise additively, i.e., noise drops rom $\varepsilon_i$ to $\varepsilon_i - \delta$. This would not actually change the main conclusion. The disadvantage of additive noise, is that we run into the boundary problem $\varepsilon_i - \delta \geq 0$. Moreover, our multiplicative formulation captures the realistic feature that a less well informed firm experiences a stronger reduction in signal noise (in absolute terms) than a better informed firm. The multiplicative formulation therefore makes it less attractive for the leader to acquire information than it would be in an additive formulation.

[6] This effect is typical in settings with bidding for exclusive access to data (see de Corniere and Montes (2019)) or another resource (e.g., Bedre-Defolie and Biglaiser (2022)).

[7] We assume that the follower cannot credibly commit to not using any information that is available ex post.

acquires the data, and drop when the follower gains data access. This gives the leader a stronger incentive to pay for data than the leader.

If data access allows the information follower to leap-frog the leader ($\delta\varepsilon_F < \varepsilon_L$), this obviously increases the follower's willingness to pay for it. Rather than see its profits decline from an information improvement, the follower may see its profits increase when leap-frogging the leader. However, the prospect of losing its informational leadership also raises the stakes for the incumbent. Again, the overall willingness to pay remains higher for the leader than for the follower.

The deeper logic as to why the leader's willingness to pay is higher can be understood by looking at overall industry profits. The allocation of data that we obtain when the two firms bid for exclusive data access is the one that generates higher industry profits. To see this, denote by $\pi_i^j$ firm $i$'s profits when $j$ gets data access, where $i \in \{F, L\}$ and $j \in \{F, L\}$. Firm $i$'s willingness to pay is given by $\pi_i^i - \pi_i^j$, while $j$'s willingness to pay is $\pi_j^j - \pi_j^i$. Firm $i$'s willingness to pay is higher than $j$'s if

$$\pi_i^i - \pi_i^j > \pi_j^j - \pi_j^i$$

$$\Longleftrightarrow$$

$$\pi_i^i + \pi_j^i > \pi_j^j + \pi_i^j,$$

where the last line just gives industry profits depending on whether $i$ or $j$ gain data access. Industry profits decrease when the allocation of data is more homogeneous as this generates fiercer competition. If the follower acquires data, information quality becomes more similar, than when the leader's already stronger information improves even further. This gives the leader a stronger willingness to pay for the data.

An exclusive sale of data therefore serves to reinforce the informational leadership, effectively reducing competition. Notice that the analysis can be applied to a situation where the informational follower is a potential entrant: its willingness to pay for data is zero if it remains the information follower even after data acquisition (if it was sub-optimal to enter before, it remains sub-optimal after). If the acquisition of data allows a potential entrant to leap-frog the leader, this might induce entry. However, a monopolistic incumbent's willingness to pay for data so as to preserve its monopoly position is always higher than that of a potential entrant. This follows from the logic of maximum industry profits. Hence, the incumbent will always outbid a potential entrant.

The above discussion also raises the question what might happen if an information sale is not exclusive. This is akin to a problem where two firms can invest in costly information production (whether the cost is on collecting own information or acquiring it from a third party does not matter in that case). We deal with this case in an extension.

# 5 Firm Entry and Cross-Market Data Spillovers

The previous analysis was based on the premise that a potential entrant considered the relevant market in isolation. This makes sense if the entrant operates only in that one market (if at all), or if there are no spillovers to other markets in which the entrant already operates. Arguably, a BigTech firm as a potential entrant does not fit that description well, as they may be able to use information they acquire in one market in their pricing and product offering decisions in another market. If that is the case, our previous entry game may change: If the entrant is a firm that is also active in another market, and there are data spillovers, then learning may motivate entry. We now analyse this case with a view to understanding how data driven entry affects consumer surplus, taking into account the impact of entry and learning in *both* markets.

Consider the case of two markets $j = a, b$ and denote by $v_a$ and $v_b$ the consumer's valuation of the good in markets $a$ and $b$, respectively. Suppose each firm costlessly observes a signal in one market but not the other. Suppose market $a$ is firm $E$'s home market and market $b$ is its foreign market, while firm $I$'s home market is market $b$ and market $a$ is its foreign market. Denote by $\widehat{x}_{E,a}$ ($\widehat{x}_{I,b}$) the signal $E$ ($I$) costlessly observes in its home market, and denote by $\varepsilon_{E,a}$ and $\varepsilon_{Ii,b}$ the corresponding error ranges of each firm's signal. Suppose also that each firm can acquire at a cost $c_{E,b}$ and $c_{I,a}$ a signal $\widehat{x}_{E,b}$ and $\widehat{x}_{I,a}$ about its respective foreign market. Assume that each firm retains an informational advantage in its home market, even if the other firm pays the information production cost, i.e., $\varepsilon_{E,a} \leq \varepsilon_{I,a}$ and $\varepsilon_{E,b} \geq \varepsilon_{I,b}$. We can think of the information acquisition cost as replacing the entry cost in Section 4.

Suppose in the status quo, the information acquisition cost is large enough so that both firms refrain from cross-entry in each other's market. Hence, we start from a situation where each firm is a monopolist in its home market. Consider now the case where one of the firms, say firm $I$ can use in its home market information obtained in the foreign market. As an example, one could think of financial services as market $a$ and firm $E$ a bank. Market $b$ could be an on-line market place and firm $I$ a platform operating in that market. The European Payment Services Directive (PSD2) may enable an entrant into financial services to obtain information from the payment history, which it could then use to improve the targeted display of a product in the market place.

In order to analyse this case, we assume that, asymmetrically, firm $I$ can use information obtained in market $a$ (its foreign market) in market $b$ (its home market), but not vice versa. This is quite realistic in many circumstances: it is hard to see, for example, how a commercial bank could start competing with Amazon as an on-line market place.[8] We model the information spillover by assuming that a given consumer has an identical preference $\eta$ in both markets and firm $I$'s observation of the signal $\widetilde{x}_{I,a} = \eta + \tilde{e}_{I,a}$ can be treated like an additional signal. Assume $\tilde{e}_{I,a}\tilde{}U\left[-\varepsilon_{I,a}, \varepsilon_{I,a}\right]$ and $\varepsilon_{I,a} > \varepsilon_{I,b}$, i.e., although firm $I$ gets a second signal, its quality is less good than its original information. Assume, moreover, that the signal errors $\tilde{e}_{I,a}$ and $\tilde{e}_{I,b}$ are uncorrelated across markets.

In the status quo, the profits, total and consumer surplus are simply described by the analysis carried out in Section 2.1. Assume information is sufficiently strong to ensure full market coverage, i.e., $v_a > 2\varepsilon_{E,a}$ and $v_b > 2\varepsilon_{I,b}$. We then get

$$\pi_{E,a} = v_a - \varepsilon_{E,a},$$
$$\pi_{I,b} = v_b - \varepsilon_{I,b}$$

and

$$CS_a = \frac{\varepsilon_{E,a}}{2},$$
$$CS_b = \frac{\varepsilon_{I,b}}{2}.$$

[8] We do not to take a stand on whether firm $A$ is not capable of using information from the foreign market, or whether the cost of obtaining it is so large that $A$ would refrain from entering, no matter what.

If firm $I$ enters $E$'s home market, profits and consumer surplus in market $a$ are described by the analysis of Section 2.2. Assume that $v_a \geq \frac{2}{3}\varepsilon_{I,a} + \frac{5}{6}\varepsilon_{E,a}$ so we are in Region I of Proposition 1 and obtain

$$\widehat{\pi}_{E,a} = \frac{1}{\varepsilon_{I,a}}\left(\frac{2}{3}\varepsilon_{I,a} - \frac{1}{6}\varepsilon_{E,a}\right)^2,$$
$$\widehat{\pi}_{I,a} = \frac{1}{\varepsilon_{I,a}}\left(\frac{1}{3}\varepsilon_{I,a} + \frac{1}{6}\varepsilon_{E,a}\right)^2,$$

and

$$\widehat{CS}_a = v_a - \widehat{\pi}_{E,a} - \widehat{\pi}_{I,a} - \frac{1}{2\varepsilon_{I,a}}\left[\left(\frac{\varepsilon_{I,a} - \varepsilon_{E,a}}{3}\right)^2 + \varepsilon_{I,a}\varepsilon_{E,a} - \frac{\varepsilon_{E,a}^2}{3}\right].$$

In market $b$, firm $I$ remains a monopolist but now has two signals. We thus need to take on board explicitly how learning affects the optimal price setting and product targeting decisions. Note that the second signal $\widetilde{x}_{I,a}$ does not improve firm $I$'s inference about $\eta$, if it is so close to the first signal $\widetilde{x}I$ that no additional values of $\eta$ can be ruled out by observing $\widetilde{x}_{I,a}$. This happens when $\widetilde{x}_{I,a} + \varepsilon_{I,a} > \widetilde{x}_{I,b} + \varepsilon_{I,b}$ and $\widetilde{x}_{I,a} - \varepsilon_{I,a} < \widetilde{x}_{I,b} - \varepsilon_{I,b}$, i.e., when $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} \in [-(\varepsilon_{I,a} - \varepsilon_{I,b}), \varepsilon_{I,a} - \varepsilon_{I,b}]$. Therefore any value of $\eta$ that is consistent with the observation $\widetilde{x}_{I,b}$ is considered possible and equally likely, given the uniform distribution of the error, i.e., $\eta \in [\widetilde{x}_{I,a} - \varepsilon_{I,b}, \widetilde{x}_{I,b} + \varepsilon_{I,b}]$. If $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} < -(\varepsilon_{I,a} - \varepsilon_{I,b})$, i.e., the second signal is much smaller than the first, then firm $I$ revises downwards the upper bound on $\eta$, such that $\eta \in [\widetilde{x}_{I,b} - \varepsilon_{I,b}, \widetilde{x}_{I,a} + \varepsilon_{I,a}]$. The opposite happens when $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b}$ so that the lower bound is revised upwards: $\eta \in [\widetilde{x}_{I,a} - \varepsilon_{I,a}, \widetilde{x}_{I,b} + \varepsilon_{I,b}]$. This leads to three regions, each with a different location and price choice as summarized in the following Lemma.

**Lemma 4** *The optimal location and price choices in market $b$ are given by*

$$l_b = \begin{cases} \frac{\widetilde{x}_{I,a} + \widetilde{x}_{I,b} + \varepsilon_{I,a} - \varepsilon_{I,b}}{2} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} < -(\varepsilon_{I,a} - \varepsilon_{I,b}) \\ \widetilde{x}_{I,b} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} \in [-(\varepsilon_{I,a} - \varepsilon_{I,b}), \varepsilon_{I,a} - \varepsilon_{I,b}] \\ \frac{\widetilde{x}_{I,a} + \widetilde{x}_{I,b} - (\varepsilon_{I,a} - \varepsilon_{I,b})}{2} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b} \end{cases} \tag{26}$$

$$p_b = \begin{cases} v_b - \frac{\varepsilon_{I,b} + \varepsilon_{I,a} + (\widetilde{x}_{I,a} - \widetilde{x}_{I,b})}{2} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} < -(\varepsilon_{I,a} - \varepsilon_{I,b}) \\ v_b - \varepsilon_{I,b} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} \in [-(\varepsilon_{I,a} - \varepsilon_{I,b}), \varepsilon_{I,a} - \varepsilon_{I,b}] \\ v_b - \frac{\varepsilon_{I,b} + \varepsilon_{I,a} - (\widetilde{x}_{I,a} - \widetilde{x}_{I,b})}{2} & \textit{if} \quad \widetilde{x}_{I,a} - \widetilde{x}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b} \end{cases} \tag{27}$$

Proof: see Appendix.

In the standard monopoly case of Section 2.1, the price was a constant, regardless of the realization of the location signal. This is no longer the case when the monopolist receives two signals, since the sharpness of the monopolist's inferences depend on the realization of the signals. We can therefore no longer apply the previous results on profits and consumer surplus. The following Proposition provides the details.

**Proposition 7** *Suppose firm $I$ is a monopolist in market $b$ with two signals $\widetilde{x}_{I,b}$ and $\widetilde{x}_{I,a}$ with uniform, independent errors $\tilde{e}_{I,b}\tilde{\ }U\left[-\varepsilon_{I,b},\varepsilon_{I,b}\right]$ and $\tilde{e}_{I,a}\tilde{\ }U\left[-\varepsilon_{I,a},\varepsilon_{I,a}\right]$ and $\varepsilon_{I,a}\geq\varepsilon_{I}$. Expected firm profits and consumer surplus in this market are given by*

$$\widehat{\pi}_{I,b}=v_b-\varepsilon_{I,b}\left(1-\frac{1}{3}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\right), \tag{28}$$

$$\widehat{CS}_b=\frac{\varepsilon_{I,b}}{2}\left(1-\frac{1}{3}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\right). \tag{29}$$

Proof: see Appendix.

We can see that the firm's expected profit is higher when it benefits from the informational spillover. This is intuitive, as the firm can price discriminate more effectively. Moreover, the increase in profits is more pronounced when the information in $I$'s home market is not too strong to begin with. In that case, obtaining an additional signal confers a stronger informational improvement on the monopolist. Correspondingly, consumer surplus suffers most when the monopolist has a moderate amount of information to begin with and learns significantly upon entering the foreign market. Equipped with the formal results above, we can analyse under what circumstances firm $I$ would wish to enter $E$'s home market and what implications this has for *overall* consumer surplus across the two markets.

**Corollary 1** *Firm $I$ enters $E$'s home market if*

$$\widehat{\pi}_{I,a}-c_I\geq-\frac{1}{3}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\varepsilon_{I,b}.$$

In other words, firm $I$ may be willing to enter market $a$, even if the entry cost (in the form of an information acquisition cost) exceeds the profits it can reap in that market. The reason is that entry increases $I$'s profits in its home market by an amount $\frac{1}{3}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\varepsilon_{I,b}$ via the information spillover.

Consider consumer surplus next. Clearly, the consumer in market $a$ is better off if firm $I$ enters, since this adds a product on offer and thus targeting is improved on average. Moreover, competition reduces prices, so the consumer is better off on both accounts. This result was formally shown in Proposition 2. The consumer in market $b$, however, is worse off. Although better information also improves the average product targeting in that market, this effect is more than off-set by the monopolist's price increase. The question is what happens to overall consumer surplus?

**Corollary 2** *Total consumer surplus is higher when both firms are monopolists in their home market, compared to when firm $I$ enters market $a$ (that is $CS_a + CS_b \geq \widehat{CS}_a + \widehat{CS}_b$) if*

$$\frac{1}{6}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\varepsilon_{I,b} \geq v_a - \frac{2}{3}\varepsilon_{I,a} - \frac{5}{6}\varepsilon_{E,a} + \frac{1}{18\varepsilon_{I,a}}\left(\varepsilon_{I,a}^2 + \varepsilon_{I,a}\varepsilon_{E,a} + \varepsilon_{E,a}^2\right) - \frac{1}{2}\varepsilon_{E,a}. \tag{30}$$

The left-hand side captures the loss in consumer surplus in market $b$ when firm $I$ enters market $a$ and reaps the benefit of the informational spillover. The right-hand side reflects the increase in consumer surplus in market $a$, following entry by firm $I$. When entry in market $a$ only generates soft competition ($v_a - \frac{2}{3}\varepsilon_{I,a} - \frac{5}{6}\varepsilon_{E,a}$ close to 0), consumers in that market benefit relatively little from entry and the loss in consumer surplus in market $b$ may therefore dominate. This might happen if $\varepsilon_{I,a}$ is well above $\varepsilon_{E,a}$ and therefore firm $I$'s entry does not increase competition by that much. In this case consumers may be better off without entry. On the other hand, if firm $I$ is already very well informed ($\varepsilon_{I,b}$ low), or learns little from entering market $a$ ($\varepsilon_{I,a}$ high), then entry in market $a$ increases overall consumer surplus. Finally, the absolute valuations attributed to the products in the two markets also play a role. For example, if $v_a$ is small compared to $v_b$ this tends to favour the status quo.[9] This is because entry reduces consumer surplus in market $b$ which is the market for the high value product, but increases surplus in market $a$.

[9] Although $v_2$ does not appear directly in inequality (30), it affects it indirectly by increasing the range of values that $\varepsilon_{B,2}$ can take.

# 6 Extensions

## 6.1 Experimentation and learning across markets

We now study the role of experimentation in one market when there are information spillovers to another market. We modify the previous cross-market spillover case in two ways. First, learning occurs from a purchase decision, rather than from the observation of an independent additional signal. This requires a small change in timing, whereby the firm that is active in two markets first offers a product in one market, observes whether it is sold, and then makes an offer in the second market. Moreover, we allow the firm that is active in two markets (firm $I$) to observe consumer behaviour in the market where it is a monopolist, and then use that information in the market where it competes with another firm.

Consider two markets $j = 1, 2$ and one firm, say firm $E$ is only active in market 1, while firm $I$ is active in both markets. We are interested in the possibility that firm $I$ may learn something from one market that is relevant for the other. With firm $I$, 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 $I$ behaves strategically in one market in order to gain better information about the other market. We assume that firm $I$ does not receive any additional exogenous information through its presence in a second market (that channel for spillovers is covered in Section 5) 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 $\eta$ in both markets and firm $I$ receives a single signal $x_I = \eta + \tilde{e}_I$ about that preference. In order to allow learning across markets, we allow the timing to be such that firm $I$ 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 $\eta$ of a particular profile of consumers.

Consider firm $I$'s optimal behavior in market 2. If the firm sets a price $p_2 = v_2 - \varepsilon_I$, 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 $\tilde{e}_I$ the consumer does not purchase the good, then this allows some inference over $\eta$. Suppose firm $I$ chooses a location $l_2 \in [x_I - \varepsilon_I, x_I + \varepsilon_I]$ and sets a price $p_2 > v_2 - \varepsilon_I$. Then there must exist realizations $\tilde{e}_I$ 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 $l_2 + v_2 - p_2 < x_I + \varepsilon_I$ and $l_2 - (v_2 - p_2) > x_I - \varepsilon_I$. Failing to sell the good then implies that $\eta \in (x_I - \varepsilon_I, l_2 - (v_2 - p_2)) \cup (l_2 + v_2 - p_2, x_I + \varepsilon_I)$, 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 $I$ with a large average distance to the true preference. The average distance is minimized by choosing location and price such that $\eta$ 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 inferred that $\eta$ is to the right of some cut-off, and following no-sale, $\eta$ is to its left.

Firm $I$'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 $l_2 = x_I + (1 - \alpha)\varepsilon_I$ and the price $p_2 = v_2 - \alpha\varepsilon_I$. It thus sells when $\eta \in [x_I + \varepsilon_I (1 - 2\alpha), x_I + \varepsilon_I]$ and does not sell when $\eta \in [x_I - \varepsilon_I, x_I + \varepsilon_I (1 - 2\alpha)]$. The probability of selling is thus $\frac{x_I+\varepsilon_I-(x_I+\varepsilon_I(1-2\alpha))}{2\varepsilon_I}=\alpha$. Firm $I$'s expected profits in market 2 are thus given by

$$\pi_{2,I}=\alpha\left(v_2-\alpha\varepsilon_I\right).$$

If the firm were to maximize simply its profits in market 2, it would choose a value of $\alpha$ which we denote by $\alpha_2$. Taking the first-order condition

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

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

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

Hence, $p_2=v_2-\frac{v_2}{2\varepsilon_I}\varepsilon_I=\frac{v_2}{2}$ if $v_2<2\varepsilon_I$, and $p_2=v_2-\varepsilon_I$ if $v_2\geq 2\varepsilon_I$, 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 $I$'s behaviour in market 2.

In order to understand learning spillovers, we need to analyze the equilibrium in market 1 given firm $I$'s stochastic information structure obtained from market 2. We assume that firm $E$ cannot observe $I$'s activities in market 2, i.e., $E$ observes neither the price nor product offered, nor whether the consumer purchased the good. Firm $E$ does, however, know that $I$ is active in market 2 and understands the structure of that market. This implies that $E$ knows that $I$ will have a signal that is, with probability $\alpha$, distributed uniformly on an interval of size $2\alpha\varepsilon_I$ and with probability $1-\alpha$, $I$'s information is on an interval of size $2\left(1-\alpha\right)\varepsilon_I$. Firm $I$ can set a price $p_{1,I}^s$ and $p_{1,I}^n$, in market 1 that depends on whether it sold $(s)$ or not $(n)$ in market 2, respectively. Firm $E$'s strategy is a price $p_{1,E}$. For simplicity, suppose that $\varepsilon_I\leq\varepsilon_E$.[10]

**Proposition 8** *If $\varepsilon_I\leq\varepsilon_E$, there exists an equilibrium in which the single market firm $E$ sets $l_{1,E}=x_E$. The price $p_{1,E}$ and probability $\alpha$ are given by the solution to the two*

[10]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 $\alpha$ is endogenous. Overall, the benefit of learning would be reduced since $\frac{\partial\pi_B}{\partial\varepsilon_B}>0$ when $\varepsilon_B>\varepsilon_A$.

*equations*

$$p_{1,E} = \frac{1}{3}\varepsilon_E + \frac{1}{6}\varepsilon_I\left[\alpha^2 + (1-\alpha)^2\right] \tag{31}$$

$$\alpha = \min\left\{\frac{1}{2}\frac{v_2 + \frac{\varepsilon_I}{\varepsilon_E}\left(\frac{p_{1,E}+\varepsilon_E}{2} - \frac{3}{16}\varepsilon_I\right)}{\varepsilon_I + \frac{\varepsilon_I}{\varepsilon_E}\left(\frac{p_{1,E}+\varepsilon_E}{2} - \frac{3}{16}\varepsilon_I\right)}, 1\right\}. \tag{32}$$

*The multi-market firm $I$ chooses location and price in market 2 according to*

$$l_2 = x_I + (1-\alpha)\,\varepsilon_I,$$

$$p_2 = v_2 - \alpha\varepsilon_I,$$

*and in market 1*

$$l_{1,I} = \begin{cases} l_2 \quad \textit{if} \quad \textit{a sale occurred in market 2} \\ l_2 = x_I - \alpha\varepsilon_I \quad \textit{if no sale occurred in market 2} \end{cases},$$

$$p_{1,I} = \begin{cases} p_{1,I}^s = \frac{p_E+\varepsilon_E}{2} - \frac{1}{4}\alpha\varepsilon_I \quad \textit{if} \quad \textit{a sale occurred in market 2} \\ p_{1,I}^n = \frac{p_E+\varepsilon_E}{2} - \frac{1}{4}(1-\alpha)\,\varepsilon_I \quad \textit{if no sale occurred in market 2} \end{cases}.$$

Even without characterizing the explicit solution for $\alpha$ we can say a few things about it from equation (32). Take first the case where the good in market 2 is sufficiently valuable for the monopolist to cover the market fully, i.e., $v_2 > 2\varepsilon_I$ and thus $\alpha_2 = 1$. Then are cases where the desire to learn about market 1 would lead firm $I$ to cover market 2 only partially ($\alpha < 1$). In particular, this would be the case when $2\varepsilon_I < v_2 < 2\varepsilon_I + \frac{\varepsilon_I}{\varepsilon_E}\left(\frac{p_{1,E}+\varepsilon_E}{2} - \frac{3}{16}\varepsilon_I\right)$. When $v_2 \in (\varepsilon_I, 2\varepsilon_I)$ then $\alpha_2 > \frac{1}{2}$ and $\frac{1}{2} < \alpha < \alpha_2$, i.e., market coverage worsens, while when $v_2 \in (0, \varepsilon_I)$ 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".

## 6.2 Costly Choice of Information Quality and Firm Entry

In this section, we reconsider the endogenous choice of information quality already analyzed in Section 4 but now consider that an improvement in formation quality is *costly.* 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 $\varepsilon_i$.[11] We write for the profits net of information quality costs, $r(\varepsilon_i) = \pi(\varepsilon_i) - c(\varepsilon_i)$.

We explore two questions. In Subsection 6.2.1 we return to the expanded model where firms first choose their information precision before competing in prices and locations. We look in Subsection 6.2.2 at the strategic choice of information to gain advantage, deter entry or force exit.

### 6.2.1 Costly choice of information quality

Firms first make a sequential strategic choice on $\varepsilon_i$ (at cost $c(\varepsilon_i)$), with firm $I$ in the Stackelberg leadership position, and then simultaneously choose their product location and prices. For simplicity, we ignore the information cost $c(\varepsilon_E)$, assuming that it is negligible in our formal analysis. We add an informal discussion at the end about the case where cost $c(\varepsilon_E)$ are non-trivial.

**Proposition 9** *Firm $I$ chooses $\varepsilon_I = 0$ when cost $c(\varepsilon_I)$ is small, and $\varepsilon_I^{D*} \in \left(0, \frac{2}{3}v\right)$ when cost $c(\varepsilon_I)$ is large. In either case, firm $E$ chooses $\varepsilon_E = \frac{3}{2}v - \frac{5}{4}\varepsilon_I$, and hence $\varepsilon_I < \varepsilon_E$ in equilibrium.*

Proof: see Appendix.

[11] 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 $c(\varepsilon_i)$.

The result generalizes our finding in Section 4 that the Stackelberg leader $I$ always prefers a choice that leads to $\varepsilon_I < \varepsilon_E$, i.e. $I$ chooses $\varepsilon_I = 0$ or the closest level $\varepsilon_I > 0$ that maximizes its net profit $r_I(\varepsilon_I) = \pi_I(\varepsilon_I) - c(\varepsilon_I)$. Firm $E$ reacts by choosing a rather low level of information, $\varepsilon_E = \frac{3}{2}v - \frac{5}{4}\varepsilon_I$. This result confirms our earlier insight into the forces that lead firms to choose opposite positions of information precision to ease price competition, but adds that the asymmetry diminishes in the costs of information quality. As seen in Section 4, $E$ chooses a position as an opportunistic niche supplier, making chance sales at a low price when the consumer is accidentally sufficiently close to its product offer.

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

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

We also extend our analysis of entry and leap-frogging in Section 4 to the case of costly information choice. Assume that firm $E$ makes a strategic choice for an investment in

$\varepsilon_E$ at cost $c(\varepsilon_E)$, whereas the information quality of $I$ is given as $\varepsilon_I$; in the resulting equilibrium, Firm $E$ may leapfrog, i.e. obtain better information than $I$. We introduce the notation $\varepsilon_E^{D*} = \arg\max_{\varepsilon_I}(\pi_E(\varepsilon_E) - c(\varepsilon_E))$ to denote an internal solution for the choice of $\varepsilon_E$, i.e. a solution where the optimal $\varepsilon_E$ is not located at one of the points of $E$'s profit function $\pi_E$ where $\pi_E$ is not continuously differentiable (these boundary points are $\varepsilon_E \in \left\{0, \varepsilon_E, \frac{6}{5}v - \frac{4}{5}\varepsilon_E\right\}$). The most interesting case is when $E$ is sufficiently competitive so that its information $\varepsilon_E$ matters for prices. As shown in Proposition 1, in eqn. (7) and (8), this will be the case when $\frac{2}{3}\varepsilon_I + \frac{5}{6}\varepsilon_E < v$, and hence, since $\varepsilon_I$ is now endogenous, a necessary condition is that $\varepsilon_I \leq \frac{3}{2}v$.

We say that the cost $c(\varepsilon_E)$ is small when $\pi_E'(\varepsilon_E) - c'(\varepsilon_E) < 0$ for all $\varepsilon_I > 0$ (leading to a corner solution, $c(\varepsilon_I) = 0$).

**Proposition 10** *For sufficiently large values of $\varepsilon_I$, firm $E$ will leapfrog and choose $\varepsilon_E = 0$ when cost $c(\varepsilon_E)$ is small, and $\varepsilon_E^{D*} \in (0, \varepsilon_I)$ when cost $c(\varepsilon_E)$ is large. For small values of $\varepsilon_I$, firm $I$ will not leapfrog and choose $\varepsilon_E = \frac{6}{5}v - \frac{4}{5}\varepsilon_I$ for small cost $c(\varepsilon_E)$, and $\varepsilon_E^{D*} \geq \frac{6}{5}v - \frac{4}{5}\varepsilon_I$ for large $c(\varepsilon_E)$.*

Proof: see Appendix.

The intuition follows closely that of Lemma 3: The profit function of firm $E$ is V-shaped, with a trough when both firms have identical information precision, $\varepsilon_E = \varepsilon_I$. 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, $\varepsilon_E = \varepsilon_I$, it becomes the industry leader in information precision, then firms become more heterogeneous and less competitive. Such differentiation in information benefits the incumbent, but is detrimental for the less informed firm that now makes a sale less often has to compete in prices more aggressively. Thus, when $\varepsilon_I$ is small, firm $E$ benefits from increasing $\varepsilon_E$ and hence softening price competition, but only up to the point where the lessening of competition will induce it to raise its price. From that point onward, the loss of sales dominates, and $E$'s profit declines. When the information cost $c(\varepsilon_E)$ is not small,

$E$'s optimal equilibrium information quality changes gradually, and internal optima of $\varepsilon_E$ become likely. The structure of equilibrium, however, remains unchanged.

### 6.2.2 Information quality choice and entry deterrence

Finally, we consider strategic entry deterrence through information. This case can arise when entrant $E$ faces an entry cost $F_E > 0$, but $E$'s expected profit is in principle sufficient for entry to occur, $\pi_E - F_E > 0$. Firm $I$ may protect its monopoly when it is able to thwart entry of firm $E$. We add the superscript $D$ to denote the duopoly situation when $E$ enters the market, and we use the superscript $M$ for monopoly. The relevant consideration for firm $I$ is not just the reaction of duopoly profit $\pi_I^D(\varepsilon_E, \varepsilon_I)$ to a change in $\varepsilon_I$, but also the gain of a sufficient investment into $\varepsilon_I$ to deter entry and maintain monopoly profit $\pi_I^M(\varepsilon_I)$. We are particularly interested in the question whether the entry deterrence motive can lead to a distortion in the incentive to invest in $\varepsilon_I$. We again denote $I$'s optimal information choice in the duopoly case by $\varepsilon_I^{D*} = \arg\max_{\varepsilon_I} r_I^D(\varepsilon_I, \varepsilon_E)$, and introduce the notation $\varepsilon_I^{M*} = \arg\max_{\varepsilon_I} r_I^M(\varepsilon_I)$ for the monopoly case, where $r_I^M(\varepsilon_I) = \pi_I^M(\varepsilon_I) - c(\varepsilon_I)$ is $I$'s net profit.

We focus on a qualitative result on overinvestments. The interesting case arises when $c(\varepsilon_I)$ is sufficiently high so that an internal optimum with $\varepsilon_I > 0$ occurs, reflected in our assumption $\varepsilon_I^{M*} > 0$ below.[12] We find that the result is essentially the same in this case compared to the analysis when both firms make an information quality choice (Section 6.2.1). As in Section 6.2.1, we find that the parameter restrictions on $\varepsilon_E$ are less restrictive in the case of a two-sided information quality choice because $E$'s endogenous choice will rule out very imprecise information choices.

We consider that only firm $I$ optimally chooses its information quality $\varepsilon_I$, at cost $c(\varepsilon_I)$, whereas firm $E$'s information precision $\varepsilon_E$ is fixed. We assume $\varepsilon_E < \frac{2}{3}v$ in this case.[13] We

[12]If the cost $c(\varepsilon_B)$ is negligible, and provided that $\varepsilon_A$ is not too small, then firm $B$ will always choose maximal information precision, $\varepsilon_B = 0$, in a duopoly as shown in Propositions 10. It would do the same as a monopolist since its profit is then even more decreasing in $\varepsilon_B$, at least for small $\varepsilon_B$. This implies that overinvestment in information quality is not possible.

[13]This assumption guarantees that for all choices $\varepsilon_I \leq \varepsilon_E$, the market remains fully covered, i.e. Propo-

find:

**Proposition 11** *Suppose $c(\varepsilon_I)$ is sufficiently large so that the incumbent firm I would choose $\varepsilon_I^{M*} > 0$ (internal optimum in the monopoly case). If there is a threat of entry by E (or an option to force exit of E) then there are values of $F_E > 0$ so that I will choose $\hat{\varepsilon}_I < \varepsilon_I^{M*}$ if the choice of $\hat{\varepsilon}_I$ can deter entry (force exit) of E.*

Proof: see Appendix.

To see the intuition, consider $I$'s choice around the boundary value of $\varepsilon_I = 0$ is optimal for $I$ (this will be the case for small $c(\varepsilon_I)$). Consider $\varepsilon_E = \frac{3}{2}v$ which is the most profitable situation for $E$. With the choice $\varepsilon_I = 0$, the Stackelberg leader $I$ maximizes entry deterrence, but still cannot squeeze $E$'s profits upon entry below $\pi_E^D = \frac{v}{6}$. Hence $I$ can only keep the entrant $E$ out if $F_E > \frac{1}{6}v$. If this is the case, then $I$ enjoys a monopoly with $\pi_I^M = v$. By contrast, if $F_E \leq \frac{v}{6}$, then $E$ enters, and we have profits $\pi_I^D = \frac{2}{3}v$ and $\pi_E^D = \frac{1}{6}v$. So there is a substantial drop in profit for $I$ when accommodating entry, and hence $I$ is willing to invest in entry deterrence by overinvesting in information. This analysis carries over mutatis mutandis to the case of an internal optimum $\varepsilon_I^{M*} > 0$, and hence explains why $I$ prefers a choice $\hat{\varepsilon}_I < \varepsilon_I^{M*}$ in this case.

As Proposition 11 highlights, we can symmetrically consider the case of exit induced through overinvestment in information. Our analysis of entry of $E$ essentially carries over, with changes only in interpretation rather than formalities.

# 7 Conclusion

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

sition 1, cases (i) or (ii) apply.

We find that better information by the less well informed firm leads to lower prices by both firms, reflecting that firms are expected to offer products in closer proximity. The opposite is true when the incumbent's information quality improves. Still, in a duopoly, more information makes consumers better off in both cases as any price increase is limited by competitive pressure implying that the dominate effect on consumer surplus is the better product fit on average. This contrasts with the case when the consumer faces a monopoly. Then there is interior optimum to the information a consumer would like to firm to possess. While consumer surplus is always higher in a duopoly, compared to a monopoly, the same is not true for total surplus. Here the monopoly may dominate when the two duopoly firms have very different access to information. In that case, the duopoly features a strong allocation inefficiency, as the less well informed firm charges a significantly lower price, inducing the consumer to purchase a product some of the time even though a neter targeted product specification was on offer.

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 this mechanism jeopardizes the efficiency of regulatory data sharing policies and explains why data broker will increase rather than reduce data inequality between competitors and deteriorate competition.

We also study informational spillovers across markets and show that the implications of entry into a monopoly market on consumer surplus are ambiguous. While the market in which entry occurs ends up with higher consumer surplus, the informational spillover reduces consumer surplus in the monopolist's market. When a firm can experiment in one market to learn in another, we show that it will offer a biased product variety and may sometimes lower the price. This finding contrasts with standard results in the literature, whereby experimentation leads to a price increase so as to gauge maximum willingness to pay.

We show that an incumbent can use overinvestment in data quality as a strategic entry deterrent, or accelerator of exit.

# Appendix

**Proof of Lemma 1.** The monopolist sells one unit if

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

Hence, if $p \leq v - \varepsilon$, the monopolist always sells. When $p > v - \varepsilon$ the monopolist only sells if $|\tilde{e}| \leq v - p$, 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.

For $v \geq 2\varepsilon$ 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 $v < 2\varepsilon$, 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 $TS^M = CS^M + \pi^M$. ■

**Proof of Proposition 1.** Consider first location choices $l_i$. Suppose $E$ uses the candidate equilibrium location $l_E = x_E$. Denote by $\lambda_I \geq 0$ the bias relative to its signal that $I$ applies when choosing the location, i.e., $l_I = x_I + \lambda_I$. Thus $I$'s good is preferred over $E$'s, if

$$v - p_I - |\eta - (\lambda_I + \eta + \tilde{e}_I)| > v - p_E - |\eta - (\eta + \tilde{e}_E)| \, .$$

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

(a) Suppose $v - \varepsilon_E - 2\varepsilon_I + p_E \geq 0$. We can then calculate $I$'s probability of selling as:

(1) If $0 \leq \lambda_I \leq \varepsilon_E - \varepsilon_I - \Delta p$ then

$$\Pr(Sell_I) = 1 - \frac{\Delta p}{\varepsilon_E} - \frac{1}{2}\frac{\varepsilon_I}{\varepsilon_E} - \frac{1}{2}\frac{\lambda_I^2}{\varepsilon_E \varepsilon_I}.$$

(2) If $\varepsilon_E - \varepsilon_I - \Delta p < \lambda_I \leq \varepsilon_I$ then

$$\Pr(Sell_I) = \frac{\frac{1}{2}(\varepsilon_E - \Delta p)^2}{2\varepsilon_E \varepsilon_I} + \frac{\left(\varepsilon_E - \frac{1}{2}\varepsilon_I - \Delta p\right)\varepsilon_I}{2\varepsilon_E \varepsilon_I} - \frac{\lambda_I^2}{2\varepsilon_E \varepsilon_I} - \frac{\frac{1}{2}(\varepsilon_E - \varepsilon_I - \Delta p)\lambda_I}{2\varepsilon_E \varepsilon_I}.$$

(3) If $\varepsilon_I < \lambda_I \leq \varepsilon_E - \Delta p$ then

$$\Pr(Sell_I) = \frac{\frac{1}{2}(\varepsilon_E - \Delta p - \lambda_I)^2}{2\varepsilon_E \varepsilon_I} + \frac{(\varepsilon_E - \Delta p - \lambda_I)\varepsilon_I}{2\varepsilon_E \varepsilon_I} + \frac{\frac{1}{2}\varepsilon_I^2}{2\varepsilon_E \varepsilon_I}.$$

(4) If $\varepsilon_E - \Delta p < \lambda_I \leq \varepsilon_E + \varepsilon_I - \Delta p$ then

$$\Pr(Sell_I) = \frac{\frac{1}{2}(\varepsilon_E + \varepsilon_I - \Delta p - \lambda_I)^2}{2\varepsilon_E \varepsilon_I}$$

(5) If $\varepsilon_E + \varepsilon_I - \Delta p < \lambda_I$ then

$$\Pr(Sell_I) = 0$$

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

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

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

Proof of part (i). The probabilities of firm's $E$ or $I$ selling their product is given by

$$\Pr\left(Sell_E\right) = \begin{cases} \frac{p_I - p_E}{\varepsilon_E} + \frac{1}{2}\frac{\varepsilon_I}{\varepsilon_E} & \text{if} \quad p_I \leq v - \varepsilon_I \\ \frac{p_I - p_E}{\varepsilon_E} + \frac{1}{2}\frac{\varepsilon_I}{\varepsilon_E} - \frac{\frac{1}{2}(p_I - (v - \varepsilon_I))^2}{\varepsilon_E \varepsilon_I} & \text{if} \quad p_I > v - \varepsilon_I \end{cases} \tag{33}$$

and

$$\Pr\left(Sell_I\right) = \begin{cases} 1 - \frac{p_I - p_E}{\varepsilon_E} - \frac{1}{2}\frac{\varepsilon_I}{\varepsilon_E} & \text{if} \quad p_I \leq v - \varepsilon_I \\ \frac{\frac{1}{2}(v - p_I)^2}{\varepsilon_E \varepsilon_I} + \frac{(p_E - (v - \varepsilon_E))(v - p_I)}{\varepsilon_E \varepsilon_I} & \text{if} \quad p_I > v - \varepsilon_I \end{cases} \tag{34}$$

We can express the expected profits as

$$\pi_i = \Pr\left(Sell_i\right) p_i, \tag{35}$$

Conjecturing $p_I \leq v - \varepsilon_I$ we get the best-response functions

$$p_E\left(p_I\right) = \frac{p_I}{2} + \frac{1}{4}\varepsilon_I \tag{36}$$

$$p_I\left(p_E\right) = \frac{p_E + \varepsilon_E}{2} - \frac{1}{4}\varepsilon_I. \tag{37}$$

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. (36) shows, the less well informed firm $E$ does not alter its price when its own information changes. This is because an increase in its information precision $\frac{1}{\varepsilon_E}$ will lead to a linear increase in its profit, as eq. (35), and does not affect the optimal price. Solving (36) and (37) yields (7) and (8). Note that these prices satisfy $0 \leq p_I - p_E \leq \varepsilon_E - \varepsilon_I$. Moreover, we require that the solution $p_I < v - \varepsilon_I$ which yields $g(\varepsilon_E, \varepsilon_I) < v$.

Proof of part (ii):

For $p_I > v - \varepsilon_I$, no sale incurs with probability $\Pr(\mathit{No\ Sale}) = 1 - \Pr(\mathit{Sell}_E) - \Pr(\mathit{Sell}_I) = \frac{\frac{1}{2}(p_E-(v-\varepsilon_E))(p_I-(v-\varepsilon_I))^2}{\varepsilon_E\varepsilon_I} > 0$. Using (33), (34) and (35) we can take derivatives with respect to each firm's price. We find that $E$ wishes to increase the price $p_E$ as long as

$$\varepsilon_I(v-2p_E) - \frac{1}{2}(v-p_I)^2 \geq 0. \tag{38}$$

$I$ wishes to increase the price $p_I$ as long as

$$\frac{1}{2}(v-p_I)(v-3p_I) + [\varepsilon_E - (v-p_E)](v-2p_I) \geq 0. \tag{39}$$

Conjecture a price $p_I = v - \varepsilon_I$. $E's$ best response to this price is

$$p_E = \frac{v}{2} - \frac{\varepsilon_I}{4},$$

regardless of whether $E$ uses the profit function just above or just below the boundary (so there is no jump in $E$'s best response just around the price chosen by $I$). Using this price $p_E$ it can be shown that for $p_I < v - \varepsilon_I$ the derivative of $\pi_I$ is positive if

$$v < \frac{2}{3}\varepsilon_E + \frac{5}{6}\varepsilon_I.$$

For $p_I > v - \varepsilon_I$ the derivative of $\pi_I$ is negative if

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

Moreover, it can be shown that

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

Hence the binding constraints on $v$ are given by $v \in [f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I)]$. Note that for the symmetric case $\varepsilon_E = \varepsilon_I$, the lower and upper bounds take the same value of $\frac{3}{2}\varepsilon$. I.e., for symmetry this case disappears. Using the price (9) and (10) it can be shown that $v \leq g(\varepsilon_E, \varepsilon_I)$ implies $\Delta p \leq \varepsilon_E - \varepsilon_I$ and $f(\varepsilon_E, \varepsilon_I) \leq v$ implies $\Delta p \geq 0$.

Proof of part (iii): Denote $y(p_I)$ the function (11) and by $z(p_I)$ the function (12). We have the following properties: (a) For $p_I \in \left(v-\varepsilon_I, \frac{v}{2}\right)$ we get $z' > y' > 0$ . (b) $z\left(\frac{v}{2}\right) > y\left(\frac{v}{2}\right)$. (c) $z(v-\varepsilon_I) < y(v-\varepsilon_I)$ if either $v < \varepsilon_E + \frac{7}{4}\varepsilon_I - \sqrt{\left(\varepsilon_E - \frac{1}{4}\varepsilon_I\right)^2 + \varepsilon_I^2}$ or $v > \varepsilon_E + \frac{7}{4}\varepsilon_I + \sqrt{\left(\varepsilon_E - \frac{1}{4}\varepsilon_I\right)^2 + \varepsilon_I^2}$. For $\left(v-\varepsilon_I, \frac{v}{2}\right)$ to be an open interval, we require $v < 2\varepsilon_I$. Note that $v < f_1(\varepsilon_E, \varepsilon_I)$ implies $v < 2\varepsilon_I$. Moreover, $v < 2\varepsilon_I$ implies $y(0) > 0$, which guarantees that the intersection is in the positive quadrant. Moreover, $v > \varepsilon_E + \frac{7}{4}\varepsilon_I + \sqrt{\left(\varepsilon_E - \frac{1}{4}\varepsilon_I\right)^2 + \varepsilon_I^2}$ and $v < 2\varepsilon_I$ are incompatible. These properties imply that when $v < f(\varepsilon_E, \varepsilon_I)$, there is a unique point $p_I \in \left(v-\varepsilon_I, \frac{v}{2}\right)$ where $y$ and $z$ intersect. ■

**Proof of Lemma 2.** Since we are focusing on the case $f(\varepsilon_E, \varepsilon_I) \leq v$ 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 $E$ this loss is $|\tilde{e}_E|$ while it is $|\tilde{e}_I|$ when he buys from $I$. We denote the expected loss by $L$. Moreover, the consumer pays a price $p_E$ or $p_I$ for the product, depending on who he purchases from. Since marginal production costs are zero, the expected price paid also corresponds to the firm's expected profits. We can thus write

$$CS^D = v - \pi_E - \pi_I - L.$$

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

$$\begin{aligned} L = &\int_0^{\Delta p} \tilde{e}_E \frac{d\tilde{e}_E}{\varepsilon_E} + \int_{\Delta p}^{\Delta p+\varepsilon_I} \int_{\tilde{e}_E-\Delta p}^{\varepsilon_I} \tilde{e}_E \frac{d\tilde{e}_I}{\varepsilon_I}\frac{d\tilde{e}_E}{\varepsilon_E} \\ &+ \int_{\Delta p}^{\Delta p+\varepsilon_I} \int_0^{\tilde{e}_E-\Delta p} \tilde{e}_I \frac{d\tilde{e}_I}{\varepsilon_I}\frac{d\tilde{e}_E}{\varepsilon_E} + \int_{\Delta p+\varepsilon_I}^{\varepsilon_E} \int_0^{\varepsilon_I} \tilde{e}_I \frac{d\tilde{e}_I}{\varepsilon_I}\frac{d\tilde{e}_E}{\varepsilon_E} \\ &= \frac{1}{\varepsilon_E}\left\{\frac{(\Delta p)^2}{2} + \frac{\varepsilon_E\varepsilon_I}{2} - \frac{\varepsilon_I^2}{6}\right\}. \end{aligned} \tag{40}$$

Applying equilibrium prices from Proposition 1 and probabilties of selling in (33) and (34) to firm profits from (35) we get the result. ■

**Proof of Proposition 2.** We need to distinguish four cases, depending on whether the market is fully covered under monopoly ($v \lesseqgtr 2\varepsilon_I$) and whether we are in regions (i) or (ii) of duopolistic competition ($v \lesseqgtr g(\varepsilon_E, \varepsilon_I)$).

Proof of part (a): $v \geq 2\varepsilon_I$

**Case 1**: $v \geq 2\varepsilon_I, v \geq g(\varepsilon_E, \varepsilon_I)$ (market fully covered in monopoly and equilibrium in Region I). In eq. (3) set $\varepsilon = \varepsilon_I$ and compare. The monopoly generates higher total surplus if

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

$$\Leftrightarrow$$

$$\frac{1}{3}\varepsilon_I^2 < \frac{1}{9}(\varepsilon_E - \varepsilon_I)^2$$

$$\Longleftrightarrow$$

$$2\varepsilon_I^2 + 2\varepsilon_E\varepsilon_I - \varepsilon_E^2 < 0.$$

The latter can be simplified to

$$\varepsilon_I < \varepsilon_E\frac{\sqrt{3}-1}{2}. \tag{41}$$

Note that for $v$ large enough we can find values of $\varepsilon_E$ such that (41) as well as the conditions defining case 1, hold.

**Case 2**: $v \geq 2\varepsilon_I, v \in (f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I))$ (market fully covered under monopoly and equilibrium in Region II). We have $TS^M > TS^D$ if

$$v - \frac{\varepsilon_I}{2} > v - \frac{1}{2\varepsilon_E}\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 + \varepsilon_E\varepsilon_I - \frac{\varepsilon_I^2}{3}\right\},$$

which can, after some calculations can be re-written as

$$v > \varepsilon_I\left(\frac{2}{3}\sqrt{3} + \frac{3}{2}\right). \tag{42}$$

Since (42) implies $v \geq 2\varepsilon_I$, there can be high enough values of $v$ such that both conditions hold. Moreover, since $\varepsilon_E$ can be arbitrarily large, we can find values of $\varepsilon_E$ that are high enough such that for any $v$, $v \in (f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I))$.

Proof of part (b): $v \in (\frac{3}{2}\varepsilon_I, 2\varepsilon_I)$

**Case 3**: $v \in (\frac{3}{2}\varepsilon_I, 2\varepsilon_I), v \geq g(\varepsilon_E, \varepsilon_I)$ (partial coverage in monopoly and duopoly equilibrium in Region I). Total surplus is higher in duopoly if

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

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

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

$$\Leftrightarrow$$

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

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

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

$$\Leftrightarrow$$

$$4(\varepsilon_E - \varepsilon_I)^2 + \frac{3}{2}\varepsilon_E\varepsilon_I > 0,$$

which always holds. Hence, $TS^M$ is never higher than $TS^D$.

**Case 4**: $v \in (\frac{3}{2}\varepsilon_I, 2\varepsilon_I), v \in (f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I))$ (partial coverage under monopoly and equilibrium in Region II)

Total surplus in duopoly is higher than in monopoly in this region if

$$v - \frac{1}{2\varepsilon_E}\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 + \varepsilon_E\varepsilon_I - \frac{\varepsilon_I^2}{3}\right\} > \frac{3}{2}\frac{v^2}{4\varepsilon_I}.$$

This can be re-arranged as

$$\varepsilon_E\left(-\frac{3}{4}v^2 + 2v\varepsilon_I - \varepsilon_I^2\right) > \varepsilon_I\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 - \frac{\varepsilon_I^2}{3}\right\}.$$

It is easy to verify that on the interval $v \in \left(\frac{3}{2}\varepsilon_I, 2\varepsilon_I\right)$

$$-\frac{3}{4}v^2 + 2v\varepsilon_I - \varepsilon_I^2 > 0,$$
$$\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 - \frac{\varepsilon_I^2}{3} < 0.$$

Hence, $TS^D > TS^M$ in this region.

Proof of part (c). When $v \geq g(\varepsilon_E, \varepsilon_I)$, then consumer surplus under duopoly is higher than under monopoly if

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

This can be re-written as

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

Since $v > \frac{2}{3}\varepsilon_E + \frac{5}{6}\varepsilon_I$ by assumption (15) a sufficient condition is

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

After further simplification the inequality can be re-written as

$$\varepsilon_E^2 - 2\varepsilon_E\varepsilon_I + \varepsilon_I^2 > 0,$$

which is always true.

When $v \in (f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I))$ then consumer surplus under duopoly is higher than under monoply if

$$v - \left\{v - \varepsilon_I - \frac{1}{\varepsilon_E}\left(\frac{v}{2} - \frac{\varepsilon_I}{4}\right)\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)\right\} - \frac{1}{2\varepsilon_E}\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 + \varepsilon_E\varepsilon_I - \frac{\varepsilon_I^2}{3}\right\} > \frac{\varepsilon_I}{2}.$$

After some re-arranging this can be written as

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

which is true in this region ($v \geq 2\varepsilon_I$). ■

**Proof of Proposition 3.** Consider first Region I, i.e., where $v > g(\varepsilon_E, \varepsilon_I)$. Substituting (20) into (17) and taking the derivative yields

$$\frac{\partial TS^D_{(i)}}{\partial \varepsilon_E} < 0 \Longleftrightarrow -\frac{4}{9}\varepsilon_E^2 - \frac{2}{9}\varepsilon_I^2 < 0,$$

which is true. Moreover,

$$\frac{\partial TS^D_{(i)}}{\partial \varepsilon_I} < 0 \Longleftrightarrow \varepsilon_I < \frac{7}{4}\varepsilon_E,$$

which is true since $\varepsilon_I \leq \varepsilon_E$.

Substituting (18) and (19) into (16) and taking the derivative yields

$$\frac{\partial CS_{(i)}^D}{\partial \varepsilon_E} < 0 \iff -\frac{11}{18}\varepsilon_E^2 - \frac{1}{18}\varepsilon_I^2 < 0,$$

which always true. Similarly,

$$\frac{\partial CS_{(i)}^D}{\partial \varepsilon_I} < 0 \iff \varepsilon_I < \frac{5}{2}\varepsilon_E,$$

which holds by our assumption $\varepsilon_I \leq \varepsilon_E$.

Consider next Region II, i.e., $v \in [f(\varepsilon_E, \varepsilon_I), g(\varepsilon_E, \varepsilon_I)]$. Substituting (23) into (17), we can write

$$TS_{(ii)}^D = v - \frac{1}{2\varepsilon_E}\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 + \varepsilon_E\varepsilon_I - \frac{\varepsilon_I^2}{3}\right\}.$$

Taking the derivative with respect to $\varepsilon_E$ yields

$$\frac{\partial TS_{(ii)}^D}{\partial \varepsilon_E} < 0 \iff \varepsilon_E(\varepsilon_E - \varepsilon_I) > \left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)^2 - \frac{\varepsilon_I^2}{3}.$$

The right-hand side of the inequality is increasing in $v$ for any $v > \frac{3}{2}\varepsilon_I$, which holds by assumption. In Region II, the upper bound on $v$ is given by $v = g(\varepsilon_E, \varepsilon_I)$. Hence, the following is a sufficient condition for $\frac{\partial TS_{(i)}^D}{\partial \varepsilon_E} < 0$:

$$\varepsilon_E(\varepsilon_E - \varepsilon_I) > \left(\frac{g(\varepsilon_E, \varepsilon_I)}{2} - \frac{3}{4}\varepsilon_I\right)^2 - \frac{\varepsilon_I^2}{3}.$$

Substituting $g(\varepsilon_E, \varepsilon_I)$ into the above inequality gives us, after some re-arranging,

$$\varepsilon_E(\varepsilon_E - \varepsilon_I) > \frac{1}{9}(\varepsilon_E - \varepsilon_I)^2 - \frac{\varepsilon_I^2}{3}.$$

This inequality holds since $\varepsilon_E > \varepsilon_E - \varepsilon_I$ inside Region II.

Taking the derivative of $TS_{(ii)}^D$ with respect to $\varepsilon_I$ yields

$$\frac{\partial TS_{(ii)}^D}{\partial \varepsilon_I} < 0 \iff \varepsilon_E - \frac{2}{3}\varepsilon_I > \frac{3v}{4} - \frac{9}{8}\varepsilon_I.$$

A sufficient condition can be obtained by substituting $v$ by its upper bound $g(\varepsilon_E, \varepsilon_I)$. Doing so and re-arraning yields the sufficient condition

$$\varepsilon_E > \frac{1}{3}\varepsilon_I,$$

which always holds.

Moving to consumer surplus, note that $CS^D = TS^D - (\pi_E + \pi_I)$, where we can calculate

$$\pi_E + \pi_I = v - \varepsilon_I - \frac{1}{\varepsilon_E}\left(\frac{v}{2} - \frac{\varepsilon_I}{4}\right)\left(\frac{v}{2} - 3\frac{\varepsilon_I}{4}\right).$$

It is evident that

$$\frac{\partial(\pi_E + \pi_I)}{\partial \varepsilon_E} > 0,$$

and together with $\frac{\partial TS^D_{(i)}}{\partial \varepsilon_E} < 0$ it follows that $\frac{\partial CS^D_{(i)}}{\partial \varepsilon_E} < 0$.

Using Lemma 2 we can express consumer surplus in Region II, $CS^D_{(ii)}$, after some simplifications as follows

$$CS^D_{(ii)} = \frac{\varepsilon_I}{2} + \frac{1}{2\varepsilon_E}\left\{\left(\frac{v}{2} - \frac{3}{4}\varepsilon_I\right)\left(\frac{v}{2} + \frac{\varepsilon_I}{4}\right) + \frac{\varepsilon_I^2}{3}\right\}.$$

Therefore

$$\frac{\partial CS^D_{(ii)}}{\partial \varepsilon_I} > 0 \Longleftrightarrow v < 4\varepsilon_E + \frac{7}{6}\varepsilon_I,$$

which is true in Region II.

**Proof of Proposition ??** The proof is a special case with incormation cost $c(\varepsilon_i) = 0$ of the general proof of Proposition 9. ■

**Proof of Proposition 6.** Consider the case where the follower can leap-frog the leader ($\delta < \frac{\varepsilon_L}{\varepsilon_F}$).

$$\frac{1}{\varepsilon_F}\left(\frac{2}{3}\varepsilon_F - \frac{1}{6}\delta\varepsilon_L\right)^2 - \frac{1}{\varepsilon_L}\left(\frac{1}{3}\varepsilon_L + \frac{1}{6}\delta\varepsilon_F\right)^2 > \frac{1}{\varepsilon_L}\left(\frac{2}{3}\varepsilon_L - \frac{1}{6}\delta\varepsilon_F\right)^2 - \frac{1}{\varepsilon_F}\left(\frac{1}{3}\varepsilon_F + \frac{1}{6}\delta\varepsilon_L\right)^2 - .$$

After some simplications we can re-write the inequality as

$$\varepsilon_F\varepsilon_L\left(\varepsilon_F - \varepsilon_L\right)\left(\frac{5}{9} + \frac{1}{9}\delta\right) - \frac{1}{18}\delta^2\left(\varepsilon_F^3 - \varepsilon_L^3\right) > 0. \tag{43}$$

Denote by $x = \frac{\varepsilon_L}{\varepsilon_F}$. Inequality (43) holds if $z_1(x) > 0$ on the interval $x \in (\delta, 1)$ where

$$z_1(x) = x\left(1 - x\right)\left(10 + 2\delta\right) - \delta^2\left(1 - x^3\right).$$

We now show that this is the case using the following properties of $z_1(x)$: (i) $z_1(x = \delta) > 0$, (ii) $z_1(x = 1) = 0$, (iii) $z_1(x)$ is concave on $x \in (\delta, 1)$.

To prove (i), we write the inequality as

$$\begin{aligned} z_1(\delta) &= (1-\delta)(10+2\delta) - \delta\left(1-\delta^3\right) \\ &= 10 - 9\delta - 2\delta^2 + \delta^4 \\ &= 1 + 9(1-\delta) - \delta^2 - \delta^2 + \delta^4 \\ &= \left(1-\delta^2\right)\left(1-\delta^2\right) + 9(1-\delta) > 0 \end{aligned} \tag{44}$$

To prove (iii), take the second order derivative:

$$\frac{\partial^2 z_1}{\partial x^2} = -2\delta(2-3\delta x) - 20 < 0.$$

Consider next the case where the information follower cannot leap-frog the leader ($\delta > \frac{\varepsilon_L}{\varepsilon_F}$). Using (24) and (25), the leader's willingness to pay is higher than the follower's if

$$\frac{1}{\varepsilon_F}\left(\frac{2}{3}\varepsilon_F - \frac{1}{6}\delta\varepsilon_L\right)^2 - \frac{1}{\delta\varepsilon_F}\left(\frac{2}{3}\delta\varepsilon_F - \frac{1}{6}\varepsilon_L\right)^2 > \frac{1}{\delta\varepsilon_F}\left(\frac{1}{3}\delta\varepsilon_F + \frac{1}{6}\varepsilon_L\right)^2 - \frac{1}{\varepsilon_F}\left(\frac{1}{3}\varepsilon_F + \frac{1}{6}\delta\varepsilon_L\right)^2.$$

As before, the inequality can be re-written in the form $z_2(x) > 0$, with $x = \frac{\varepsilon_L}{\varepsilon_F}$ with $x \in (0, \delta]$ and

$$z_2(x) = \delta(1-\delta)(10+2x) - x^2\left(1-\delta^3\right).$$

We can again show that $z_2(x=0) > 0$ and that $z_2(x)$ is concave. Moreover, note that $z_2(x=\delta) = z_1(x=\delta) > 0$. Hence, $z_2(x) > 0$ on the relevant interval. ■

**Proof of Lemma 4.** When firm $I$ learns nothing from signal $\widetilde{x}_{I,a}$ it simply sets location and price as in Lemma 1 based on its signal $\widetilde{x}_{I,b}$. This covers the region $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} \in [-(\varepsilon_{I,a} - \varepsilon_{I,b}), \varepsilon_{I,a} - \varepsilon_{I,b}]$. When $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} < -(\varepsilon_{I,a} - \varepsilon_{I,b})$ or $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b}$ the firm changes its belief about the boundaries, but learns nothing further about where within that range $\eta$ lies (this follows from uniformity of the error terms). As in Lemma 1 it is therefore optimal to set the location at the mid-point of the admissible interval, i.e., $\frac{\widetilde{x}_{I,a}+\widetilde{x}_{I,b}+\varepsilon_{I,a}-\varepsilon_{I,b}}{2}$ if $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} < -(\varepsilon_{I,a} - \varepsilon_{I,b})$ and $\frac{\widetilde{x}_{I,a}+\widetilde{x}_{I,b}-(\varepsilon_{I,a}-\varepsilon_{I,b})}{2}$ if $\widetilde{x}_{I,a} - \widetilde{x}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b}$. The optimal price is given by $v_2$ minus the maximum distance between the true $\eta$ and the location choice, which is just half of the size of the admissible interval. ■

**Proof of Proposition 7.** Since the market is fully covered, the monopolist always sells. Expected profits are therefore given by the expected price. Using the price function (27) and calculating its expectation yields (28).

For given error terms $\tilde{e}_{I,b}$ and $\tilde{e}_{I,a}$ total surplus is given by

$$TS_b\left(\tilde{e}_{I,b}, \tilde{e}_{I,a}\right) = v_b - \left|\eta - l_b\left(\tilde{e}_{I,b}, \tilde{e}_{I,a}\right)\right|,$$

where $l_b\left(\tilde{e}_{I,b}, \tilde{e}_{I,a}\right)$ is given by substituting $\widetilde{x}_{I,b}$ and $\widetilde{x}_{I,a}$ into (26). We need to distinguish regions according to $\tilde{e}_{I,a} - \tilde{e}_{I,b} < -\left(\varepsilon_{I,a} - \varepsilon_{I,b}\right)$, $\tilde{e}_{I,a} - \tilde{e}_{I,b} \in \left[-\left(\varepsilon_{I,a} - \varepsilon_{I,b}\right), \varepsilon_{I,a} - \varepsilon_{I,b}\right]$ and $\tilde{e}_{I,a} - \tilde{e}_{I,b} > \varepsilon_{I,a} - \varepsilon_{I,b}$. Within those regions we also need to distinguish between the case where $\eta - l_b\left(\tilde{e}_{I,b}, \tilde{e}_{I,a}\right) \geq 0$ and $\eta - l_b\left(\tilde{e}_{I,b}, \tilde{e}_{I,a}\right) < 0$. This yields expected total surplus given by

$$\widehat{TS}_b = \frac{1}{\varepsilon_{I,b}\,\varepsilon_{I,a}} \int_{-\varepsilon_{I,a}}^{-\left(\varepsilon_{I,a}-\varepsilon_{I,b}\right)} \int_{\tilde{e}_{I,a}+\left(\varepsilon_{I,a}-\varepsilon_{I,b}\right)}^{-\tilde{e}_{I,a}-\left(\varepsilon_{I,a}-\varepsilon_{I,b}\right)} \left[v_b + \frac{\tilde{e}_{I,a} + \tilde{e}_{I,b} + \varepsilon_{I,a} - \varepsilon_{I,b}}{2}\right] d\tilde{e}_{I,b}\, d\tilde{e}_{I,a}$$
$$+ \frac{1}{2}\frac{1}{\varepsilon_{I,b}\,\varepsilon_{I,a}} \int_{0}^{\varepsilon_{I,a}} \int_{\tilde{e}_{I,b}-\left(\varepsilon_{I,a}-\varepsilon_{I,b}\right)}^{\tilde{e}_{I,b}+\left(\varepsilon_{I,a}-\varepsilon_{I,b}\right)} \left[v_b - \tilde{e}_{I,b}\right]\, d\tilde{e}_{I,a}\, d\tilde{e}_{I,b}.$$

Solving the integrals yields

$$\widehat{TS}_b = v_b - \frac{\varepsilon_{I,b}}{2}\left(1 - \frac{1}{3}\frac{\varepsilon_{I,b}}{\varepsilon_{I,a}}\right).$$

From risk neutrality it follows that we get $\widehat{CS}_b = \widehat{TS}_b - \widehat{\pi}_{I,b}$ which yields (29). ■

**Proof of Proposition 8.** Firm $E$ knows that $I$ has a uniformly distributed signal with a range $\alpha\varepsilon_I$ and uses a price $p_{1,I}^s$ with probability $\alpha$. With probability $1-\alpha$, it has a uniformly distributed signal with range $(1-\alpha)\,\varepsilon_I$ and sets a price $p_{1,I}^n$ with probability $1-\alpha$. If firm $E$ locates at $l_1 = x_E$ and sets price $p_E$ it sells with probability

$$\Pr(Sell_E) = \frac{\alpha p_{1,I}^s + (1-\alpha)\,p_{1,I}^n - p_E}{\varepsilon_E} + \frac{1}{2}\frac{\alpha^2 + (1-\alpha)^2}{\varepsilon_E}\varepsilon_I,$$

yielding profits $\pi_E = \Pr\left(Sell_E\right) p_E$. Note that $E$ cannot do better by changing its location as it has no information about where $I$ is likely to locate (the fact that $I$ will move to either side of its original signal is not useful to $E$ because of the uniform distribution).

Maximizing $\pi_E$ with respect to $p_E$ yields $E$'s best response

$$p_{1,E} = \frac{\alpha p_{1,I}^s + (1-\alpha)\,p_{1,I}^n}{2} + \frac{1}{4}\left(\alpha^2 + (1-\alpha)^2\right)\varepsilon_I.$$

Firm $I$ 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 $\eta$ may be located, i.e., at $l_2$ and sells with probability

$$\Pr\left(Sell_{1,I}\left|Sell_2\right.\right) = 1 - \frac{p_{1,I}^s - p_E}{\varepsilon_E} - \frac{1}{2}\frac{\alpha\varepsilon_I}{\varepsilon_E},$$

yielding a best response

$$p_{1,I}^s\left(\alpha\right) = \frac{p_{1,E} + \varepsilon_E}{2} - \frac{1}{4}\alpha\varepsilon_I.$$

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

$$p_{1,I}^n\left(\alpha\right) = \frac{p_{1,E} + \varepsilon_E}{2} - \frac{1}{4}\left(1-\alpha\right)\varepsilon_I.$$

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

$$\pi_{1,I}^s\left(\alpha\right) = \left(1 - \frac{p_{1,I}^s\left(\alpha\right) - p_E}{\varepsilon_E} - \frac{1}{2}\frac{\alpha\varepsilon_I}{\varepsilon_E}\right)p_{1,I}^s\left(\alpha\right),$$
$$\pi_{1,I}^n\left(\alpha\right) = \left(1 - \frac{p_{1,I}^n\left(\alpha\right) - p_E}{\varepsilon_E} - \frac{1}{2}\frac{\alpha\varepsilon_I}{\varepsilon_E}\right)p_{1,I}^n\left(\alpha\right).$$

Overall expected profits are given by

$$\begin{aligned}\pi_{1,I} &= \alpha\pi_{1,I}^s\left(\alpha\right) + \left(1-\alpha\right)\pi_{1,I}^n\left(\alpha\right)\\ &= \frac{1}{\varepsilon_E}\left[\left(\frac{p_E+\varepsilon_E}{2}\right)^2 - \frac{1}{2}\varepsilon_I\frac{p_E+\varepsilon_E}{2}\left(\alpha^2 + \left(1-\alpha\right)^2\right)\right]\\ &+ \frac{1}{16}\frac{\varepsilon_I^2}{\varepsilon_E}\left(1 - 3\alpha\left(1-\alpha\right)\right).\end{aligned}$$

Taking the first-order condition of $\pi_{1,I} + \pi_{2,I}$ and solving for $\alpha$ yields (32).

Note also that for our previous equilibrium analysis to hold, we require $p_{1,I}^s - p_{1,E} \leq \varepsilon_E - \alpha\varepsilon_I$ and $p_{1,I}^n - p_{1,E} \leq \varepsilon_E - \left(1-\alpha\right)\varepsilon_I$. It is sufficient to check one of the two conditions as they are symmetric in $\alpha$ around $\frac{1}{2}$. In order to check the validity of the first inequality we substitute the price $p_{1,I}^{s}$:

$$\frac{p_{1,E}+\varepsilon_E}{2}-\frac{1}{4}\alpha\varepsilon_I-p_{1,E}\leq\varepsilon_E-\alpha\varepsilon_I.$$

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

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

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

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

which holds.

Finally, we prove that (31) and (32) has a unique solution on $\alpha\in[0,1]$. Note that $p_{1,E}(\alpha)$ is decreasing (increasing) for $\alpha<\frac{1}{2}$ ($\alpha>\frac{1}{2}$). Moreover, from (32) it is clear that $\alpha<\frac{1}{2}$ if and only if $v_2<\varepsilon_I$. It is also the case that if $\alpha$ is interior, then $\frac{\partial\alpha}{\partial p_{1,E}}>0$ if and only if $v_2<\varepsilon_I$. Hence, if $v_2<\varepsilon_I$ then $\alpha<\frac{1}{2}$, $p_{1,E}(\alpha)$ is decreasing and $\alpha(p_{1,E})$ 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 $\alpha(p_{1,E})$ and denote it by $P(\alpha)$. It is straightforward to show that $P(\alpha=0)<0$ and $\lim\limits_{\alpha\to\frac{1}{2}}P(\alpha)=+\infty$. Since $p_{1,E}(\alpha)$ is positive and finite on $\alpha\in[0,1]$, there must be an intersection point. When $v_2>\varepsilon_I$ then $\alpha>\frac{1}{2}$, $p_{1,E}(\alpha)$ is increasing and $\alpha(p_{1,E})$ is decreasing. In this case we either get a unique interior solution, or the corner solution $\alpha=1$.

**Proof of Proposition 9.** We initially abstract from costs $c(\varepsilon_I)$ and $c(\varepsilon_E)$, and consider them to be arbitrarily small.

*Step 1:* Going backwards, we start by characterizing $E$'s best response (BR) function $\varepsilon_E(\varepsilon_I)$. We show that:

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

To derive this expression, we need to look at four cases, defined by two separating lines: the straight line $g(\varepsilon_E,\varepsilon_I)=\frac{2}{3}\varepsilon_E+\frac{5}{6}\varepsilon_I=v$ that separates Proposition 1(i) ($g(\varepsilon_E,\varepsilon_I)\leq v$) from

Proposition 1(ii), and $\varepsilon_E = \varepsilon_I$. We begin with the analysis of the simulateneous choice of locations and precisions where, as we recall, $(g(\varepsilon_E, \varepsilon_I) > v)$.

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

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

(iii) Next, consider $\varepsilon_I > \frac{2}{3}v$ , $g(.) > v$, and $\varepsilon_E < \varepsilon_I$. W are in the case of Proposition 1(ii) with inverted roles af $E$ and $I$, as long as $f < v$ (note that for small enough $\varepsilon_E$, this will hold since, with inverted roles af $E$ and $I$, $f(\varepsilon_I, 0) = 0$). Hence in this case $\pi_E = \frac{1}{4\varepsilon_I}(4\varepsilon_I + \varepsilon_E - 2v)(v - \varepsilon_E)$, with derivative: $\frac{\partial \pi_E}{\partial \varepsilon_E} < 0$, so $E$ chooses the smallest possible value $\varepsilon_E(\varepsilon_I)$. The smallest candidate value inside the parameter region, $\varepsilon_E(\varepsilon_I) = \frac{3}{2}v - \frac{5}{4}\varepsilon_I$, 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 $\varepsilon_E(\varepsilon_I) = 0$.

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

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

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

(ii) For $\varepsilon_I > \frac{2}{3}v$, using Proposition 1(i) by inverting the roles of $E$ and $I$, we have then: $\pi_I = \frac{1}{\varepsilon_I}\left(\frac{1}{3}\varepsilon_I + \frac{1}{6}\varepsilon_E\right)^2$, and substituting $E$'s BR function $\varepsilon_E(\varepsilon_I) = 0$, this becomes $\pi_I(\varepsilon_I, \varepsilon_E(\varepsilon_I)) = \frac{1}{9}\varepsilon_I$, with derivative:$\frac{\partial \pi_I}{\partial \varepsilon_I} = \frac{1}{9} > 0$. So $I$ wants to increase $\varepsilon_I$ at least up to the limit point where the profit function changes since the market is then not fully covered (boundary of Proposition 1(i) $E$ and $I$ are inverted). Since $\varepsilon_I > \varepsilon_E(\varepsilon_I)$, we need to invert $\varepsilon_E$ and $\varepsilon_I$ in determining this boundary, and hence $g(\varepsilon_E, \varepsilon_I) = \frac{5}{6}\varepsilon_E + \frac{2}{3}\varepsilon_I \leq v$ so $\varepsilon_I = \frac{3}{2}v$ is the limit point; after that point, $g(\varepsilon_E, \varepsilon_I) > v$ and the profit function is in the region of Proposition 1(ii). Consider whether $I$ wants to expand beyond $\varepsilon_I = \frac{3}{2}v$. Using Proposition 1(ii) by inverting the roles of $E$ and $I$, we know that $I$ maximizes: $\pi_I = \frac{1}{16\varepsilon_I}(2v-\varepsilon_E)^2$, with derivative: $\frac{\partial \pi_I}{\partial \varepsilon_I} = -\frac{1}{16\varepsilon_I^2}(2v-\varepsilon_E)^2 < 0$. Hence there is no equilibrium candidate of $I$ inside the parameter region of Proposition 1(ii), and $\varepsilon_I = \frac{3}{2}v$ is the candidate solution (local optimum for $\varepsilon_I > \frac{2}{3}v$). Note that this solution is identical to $E$'s best response function when $\varepsilon_I = 0$, $\varepsilon_E(\varepsilon_I) = \frac{3}{2}v$.

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

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

**Proof of Proposition 10.** We first consider first that $c(\varepsilon_E)$ is small, meaning that it can be neglected when determining equilibrium precision choices, and hence we can approximate

$r(\varepsilon_E) \approx \pi(\varepsilon_E)$. We consider first the case where $\varepsilon_I < \frac{3}{2}v$ which implies that $\varepsilon_I < \frac{6}{5}v - \frac{4}{5}\varepsilon_E$.

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

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

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

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

We consider then the case where $\varepsilon_I > \frac{2}{3}v$ which implies that $\varepsilon_E > \frac{6}{5}v - \frac{4}{5}\varepsilon_I$. For $\varepsilon_E < \frac{6}{5}v - \frac{4}{5}\varepsilon_I$, Proposition 1(i) applies and $\frac{\partial \pi_E}{\partial \varepsilon_E} < 0$ as above. For $\frac{6}{5}v - \frac{4}{5}\varepsilon_I < \varepsilon_E < \varepsilon_I$, Proposition 1(ii) applies and it can be shown that $\frac{\partial \pi_E}{\partial \varepsilon_E} < 0$. For $\varepsilon_E > \varepsilon_I$, Proposition 1(ii) applies, with the roles of $I$ and of $E$ inverted, and $\frac{\partial \pi_E}{\partial \varepsilon_E} < 0$ as shown above. Thus, when $\varepsilon_I > \frac{2}{3}v$, $E$'s profit decreases globally in $\varepsilon_E$, and only a choice of $\varepsilon_E = 0$ can be optimal.

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

**Proof of Proposition 11.** Recall $I$'s potential monopoly profit is (Section 2.1):

$$\pi_I^M(\varepsilon_I) = \begin{cases} v - \varepsilon_I & \text{if} \quad \varepsilon_I \leq \frac{v}{2} \\ \frac{v^2}{4\varepsilon_I} & \text{if} \quad \varepsilon_I > \frac{v}{2} \end{cases}$$

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

*One-sided information quality choice of firm B:* In this case, the duopoly profit $\pi_I^D(\varepsilon_I, \varepsilon_E)$ is: $\pi_I^D(\varepsilon_I < \varepsilon_E) = \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2$ (see Proposition 1(i) and Section 6.2.2), and hence $\frac{\partial \pi_I^D}{\partial \varepsilon_I} < 0$ over the range $\varepsilon_I \in (0, \varepsilon_E)$. As seen in Subsection 6.2.2, the duopoly profit of $I$ decreases in $\varepsilon_I$ for $\varepsilon_I < \varepsilon_E$, and increases in $\varepsilon_I$ for $\varepsilon_I \geq \varepsilon_E$. For the profit function of the potential entrant $E$, we get $\frac{\partial \pi_E^D}{\partial \varepsilon_I} > 0$ throughout, so the lower is $\varepsilon_I$, the lower is $E$'s profit and hence the more likely is entry deterrence. Moreover, with $c(\varepsilon_I)$ sufficiently small and $\varepsilon_E$ sufficiently large, $\varepsilon_I = 0$ is the optimal choice, with profit level $\pi_I^D(\varepsilon_I = 0) = \frac{4}{9}\varepsilon_E$ and $E$'s profit: $\pi_E^D(\varepsilon_I = 0) = \frac{1}{9}\varepsilon_E$, and $\pi_I^M(\varepsilon_I = 0) = v$. $\pi_E^D(\varepsilon_I = 0) = \frac{1}{9}\varepsilon_E$ is the lowest profit level of entrant $E$ that the incumbent $I$ can reach with its choice of $\varepsilon_I$.

Consider the difference between the monopoly and the duopoly profit, $\Delta\pi = \pi_I^M(\varepsilon_I) - \pi_I^D(\varepsilon_I, \varepsilon_E)$. We show that $\Delta\pi > 0$ for all values of $\varepsilon_E$ and optimal choices of $\varepsilon_I$ when $\varepsilon_E < \frac{2}{3}v$. We consider the points where $\pi_I^D(\varepsilon_I, \varepsilon_E)$ is not continuously differentiable, and then show that $\Delta\pi > 0$ also holds for all intermediate values. Consider $\varepsilon_E < \frac{2}{3}v$ which implies $\varepsilon_E < \frac{6}{5}v - \frac{4}{5}\varepsilon_E$. At $\varepsilon_I = 0$, $\Delta\pi = \pi_I^M(\varepsilon_I) - \pi_I^D(\varepsilon_I, \varepsilon_E) = v - \varepsilon_I - \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2 = v - \frac{4}{9}\varepsilon_E > \frac{1}{3}v$. At $\varepsilon_I = \varepsilon_E$, either $\varepsilon_E < \frac{v}{2}$ and $\Delta\pi = v - \varepsilon_I - \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2 = v - \frac{4}{9}\varepsilon_E > \frac{7}{9}v$ or $\varepsilon_E > \frac{v}{2}$ and $\Delta\pi = \frac{v^2}{4\varepsilon_I} - \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2 = \frac{1}{4\varepsilon_E}\left(v^2 - \varepsilon_E^2\right) > 0$ since we consider $\varepsilon_E < \frac{2}{3}v$.

At $\varepsilon_I = \frac{6}{5}v - \frac{4}{5}\varepsilon_E$, $\Delta\pi = \frac{v^2}{4\varepsilon_I} - \frac{1}{16\varepsilon_I}(2v-\varepsilon_E)^2 = \frac{v^2}{4\left(\frac{6}{5}v-\frac{4}{5}\varepsilon_E\right)} - \frac{1}{16\left(\frac{6}{5}v-\frac{4}{5}\varepsilon_E\right)}(2v-\varepsilon_E)^2 = \frac{5}{32}\frac{\varepsilon_E}{3v-2\varepsilon_E}(4v-\varepsilon_E) > 0$. Moreover, $\Delta\pi$ is continuous and piecewise differentiable between these points, showing that $\Delta\pi > 0$ everywhere. No choice $\varepsilon_I > \frac{6}{5}v - \frac{4}{5}\varepsilon_E$ can be optimal in a duopoly, as shown (in the inverted case, reaction function $\varepsilon_E(\varepsilon_I)$), in Proposition 9.

To show that there is a generic set of values $F_E$ such that $I$ prefers to decrease $\hat{\varepsilon}_I$ below $\varepsilon_I^{M*}$, consider a situation where $\varepsilon_I^{M*}$ is such that it does not deter entry, i.e. $\pi_E^D\left(\varepsilon_I^{M*}, \varepsilon_E\right) > F_E$. Then $I$ prefers $\hat{\varepsilon}_I < \varepsilon_I^{M*}$, at cost $c(\hat{\varepsilon}_I) - c\left(\varepsilon_I^{M*}\right) > 0$, provided that: (i) $\pi_E^D\left(\varepsilon_I^{M*}, \varepsilon_E\right) > F_E > \pi_E^D(\hat{\varepsilon}_I, \varepsilon_E)$, and (ii) $\pi_I^M(\hat{\varepsilon}_I) - c(\hat{\varepsilon}_I) > \pi_I^D\left(\varepsilon_I^{M*}, \varepsilon_E\right) - c\left(\varepsilon_I^{M*}\right)$.

Condition (i) is feasible for some $F_E > 0$ since $\pi_E^D(\varepsilon_I, \varepsilon_E)$ increases in $\varepsilon_E$. Condition (ii) is feasible for some $\hat{\varepsilon}_I < \varepsilon_I^{M*}$ since $\Delta\pi$ is bounded away from zero, and $\pi_I^M$ and $\pi_I^D$ are continuous functions in $\varepsilon_I$. Thus, there must exist a generic set of values $F_E$ satisfying both conditions.

*Two-sided information quality choice:* We next look at the case where firm $E$ can react to $I$'s entry and optimally adjust $\varepsilon_E$, under the same assumptions are the as in Subsection 6.2.1. Consider the difference between the monopoly and the duopoly profit, $\Delta\pi = \pi_I^M(\varepsilon_I) - \pi_I^D(\varepsilon_I, \varepsilon_E)$. When $E$ chooses $\varepsilon_E$, then we obtain $\pi_I^D(\varepsilon_I, \varepsilon_E)$ by substituting $E$'s best response function $\varepsilon_E(\varepsilon_I)$ (Proposition 9) into the expression for $\pi_I^D(\varepsilon_I, \varepsilon_E)$ following Proposition 1. This yields, for $\varepsilon_I \leq \frac{2}{3}v$, $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I)) = \frac{4}{6v-5\varepsilon_I}(v-\varepsilon_I)^2$, and for $\frac{2}{3}v < \varepsilon_I \leq \frac{3}{2}v$, $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I)) = \frac{1}{\varepsilon_I}\left(\frac{1}{3}\varepsilon_I + \frac{1}{6}\varepsilon_E\right)^2$. For $\frac{2}{3}v < \varepsilon_I \leq \frac{3}{2}v$, taking into account $E$'s best response $\varepsilon_E(\varepsilon_I) = 0$, $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I)) = \frac{1}{9}\varepsilon_I$.Since $\varepsilon_E(\varepsilon_I) = 0$ drops discontinuously at the point from $\varepsilon_E(\varepsilon_I) = \frac{2}{3}v$ to $\varepsilon_E(\varepsilon_I) = 0$, $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I))$ discontinuously falls at this point from $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I))\big|_{\varepsilon_I = \frac{2}{3}v^-} = \frac{1}{6}v$ to $\pi_I^D(\varepsilon_{I,}\varepsilon_E(\varepsilon_I))\big|_{\varepsilon_I = \frac{2}{3}v^+} = \frac{1}{9}v$.

We show that $\Delta\pi > 0$ for all choices of $\varepsilon_I$ and best responses of $\varepsilon_E$. We consider the points where $\pi_I^D(\varepsilon_I, \varepsilon_E)$ is not continuous or not continuously differentiable. At $\varepsilon_I = 0$, $\varepsilon_E(\varepsilon_I) = \frac{3}{2}v$, and hence $\Delta\pi = \pi_I^M(\varepsilon_I) - \pi_I^D(\varepsilon_I, \varepsilon_E) = v - \varepsilon_I - \frac{1}{\varepsilon_E}\left(\frac{2}{3}\varepsilon_E - \frac{1}{6}\varepsilon_I\right)^2 = v - \frac{4}{9}\varepsilon_E = \frac{1}{3}v$. At $\varepsilon_I = \frac{2}{3}v^-$, $\varepsilon_E(\varepsilon_I) = \frac{3}{2}v - \frac{5}{4}\varepsilon_I = \frac{5}{6}v$, hence $\Delta\pi = \frac{v^2}{4\varepsilon_I} - \frac{4}{6v-5\varepsilon_I}(v-\varepsilon_I)^2 = \frac{3}{8}v - \frac{1}{6}v > 0$. At $\varepsilon_I = \frac{2}{3}v^+$, $\varepsilon_E(\varepsilon_I) = 0$, hence $\Delta\pi = \frac{3}{8}v - \frac{1}{9}v > 0$; at $\varepsilon_I > \frac{3}{2}v$, $\varepsilon_E(\varepsilon_I) = 0$, hence $\Delta\pi = \frac{v^2}{4\varepsilon_I} - \frac{1}{9}\varepsilon_I > 0$ (and converging to $\Delta\pi = 0$ for $\varepsilon_I \to \frac{3}{2}v$). Since $\pi_I^M(\varepsilon_I)$ and $\pi_I^D(\varepsilon_I, \varepsilon_E)$ 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. ■

# References

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

[2] Acquisti, A., and H.R. Varian, 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] Armstrong, M. and J. Zhou, 2022, "Consumer Information and the Limits to Competition," *American Economic Review*, 112(2):534–577.

[5] Bedre-Defolie, O. and G. Biglaiser, 2022, "Competition for Exclusivity of a Superior Input and Quality Implications," unpublished manuscript.

[6] Belleflamme, P., W. M. Lam, and W. Vergote, 2020, "Competitive Imperfect Price Discrimination and Market Power," *Marketing Science*, 39(5):996–1015.

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

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

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

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

[11] Chen, Z., C. Choe and Matsushima, 2020, "Competitive Personalized Pricing," *Management Science*, 66(9):4003–4023.

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

[13] de Cornière, A. and R. Montes, 2017, "Consumer Privacy and the Incentives to Price-Discriminate in Online Markets," *Review of Network Economics*, 16(3):291–305.

[14] de Cornière, A. and G. Taylor, 2023, "Data and Competition: A Simple Framework," TSE working paper #23–1404.

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

[16] Gu, Y., L. Madio, and C. Reggiani, 2019, "Exclusive Data, Price Manipulation and Market Leadership," CESifo Working Paper #7853.

[17] Hagiu, A. and J. Wright, 2021, "Data-enabled learning, network effects and competitive advantage" RAND Journal of Economics, forthcoming.

[18] Hidir, S. and N. Vellodi, 2020, "Privacy, Personnalization and Price Discrimination," *Journal of the European Economic Association*, 19(2), 1342–1363.

[19] Ichihashi, S., 2020, "Online Privacy and Information Disclosure by Consumers," *American Economic Review*, 110(2):569–595.

[20] Iyer, G., D. Soberman, and J.M. Vilas-Boas, 2005 "The Targeting of Advertising," *Marketing Science*, 24:461–476.

[21] Jullien, B., Y. Lefouili, and M. Riordan, 2020, "Privacy Protection, Security, and Consumer Retention," TSE Working Paper # 18–947.

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

[23] Nageeb Ali, S., G. Lewis and S. Vasserman, 2022, "Voluntary Disclosure and Personalized Pricing," Research Papers 3890, Stanford University, Graduate School of Business.

[24] Prüfer, J. and C. Schottmüller, 2021, "Competing with Big Data," *Journal of Industrial Economics*, 69:967–1008.

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

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

[27] 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.