Market Segmentation, Information Sale, and Information Foreclosure*

Emanuele Tarantino† Stefan Terstiege‡ Adrien Vigier§

July 18, 2024

Abstract

We study the sale of consumer data allowing firms to engage in discriminatory pricing. When an independent data intermediary sells information, the consumer surplus is as if all firms were fully informed. We then consider the case where the data intermediary also competes in the final goods market. We show that the integrated data intermediary may choose not to sell full information to its rivals; this strategic behavior lowers consumer surplus. We study the industry's equilibrium structure and show that acquiring a firm by an independent data intermediary always reduces consumer surplus. We apply the insights of our analysis to the debate on self-preferencing in the digital advertising industry.

Keywords: data intermediaries, information sale, market segmentation.
JEL-Classification: D42, D83.

*We thank Alessandro Bonatti, Daniele Condorelli, Francesco Decarolis, Marco Pagnozzi, Alessandro Pavan, Patrick Rey, Bill Rogerson, Florian Schuett, Quitzé Valenzuela-Stookey, the participants at the Competition and Markets Authority and Maastricht University seminars, and at the Athens Industrial Organization Workshop, Annual Berkeley/Columbia/Duke/MIT/Northwestern IO Theory Conference, and Leuven Summer Event in Economic Theory for insightful comments. We also thank Qiyue Shen for his excellent research assistance.

†Luiss, EIEF, and CEPR. Email: etarantino@luiss.it.

‡Maastricht University. Email: s.terstiege@maastrichtuniversity.nl.

§University of Notthingam. Email: adrien.vigier@nottingham.ac.uk.

1 Introduction

The examination of the welfare effects of discriminatory pricing goes at least as far back as Pigou (1920). The advent of the digital era and, with it, the availability of huge amounts of data concerning consumers, has recently revived a strong interest in price discrimination (see, e.g., Rhodes and Zhou, 2024). The access to consumer data gives firms the possibility to segment the market, offering different prices in different market segments. The starting point of our paper is that, in practice, market segmentation often results from the sale of consumer data by intermediaries.[1] Furthermore, while consumer data may be sold by specialized intermediaries, in other cases it involves firms competing with those that they sell data to: Amazon, for instance, both displays sponsored search results and sells products of its own. Our contribution is twofold: (i) study market segmentation arising from information sale; (ii) examine how the sale of consumer data and the resulting welfare depend on the structure of the market considered and, in particular, whether the data intermediary also operates in the goods market.

We study a model with one consumer and multiple firms. The consumer has unit demand and attaches different valuations to the goods of the different firms. We first consider the case in which information is sold by an independent data intermediary through secret bilateral negotiations. We show that, in this case, every equilibrium is efficient and generates the same consumer surplus as when all firms are fully informed (Theorem 1). The broad logic is as follows. Since firms do not observe what information is sold to competitors, in equilibrium the data intermediary sells each firm all valuable information. This in turn implies that conditional on every single realization of the consumer's valuations, each firm must be setting a price that constitutes the best response to the prices of the other firms. Hence, for every realization of the valuations, the consumer buys from the same firm and at the same price as if all firms were fully informed.

We then examine the case in which one of the firms is better informed than the others,

[1]Consumer data may be sold "directly", by specialized intermediaries such as Nielsen and Oracle, or "indirectly", through access to targeted consumer segments, by companies like Google, Meta, Microsoft, and Amazon. Bergemann and Bonatti (2019) discuss the business models of data intermediaries.

and sells information to its competitors. As said firm operates both in the goods and data markets, we call it the "integrated" firm. We show that the equilibrium consumer surplus is bounded from above by the consumer surplus generated with an independent data intermediary (Theorem 2). If the integrated firm sells full information to every firm, the outcome is the same as with an independent data intermediary. If, instead, the integrated firm chooses to foreclose information (i.e., limit some competitors' access to information), it must do so to increase its profit in the goods market or some firm's willingness to pay for information. This enables us to show that, in equilibrium, industry profit—the sum of all firms' profits in the goods market—is at least as large as if all firms were fully informed. Since independent data intermediation maximizes social welfare, we conclude that "integrated" data intermediation can never be better for the consumer.

We go on to analyze the sale of information under integrated data intermediation within a variant of our setting embedding Varian and Narasimhan's model of captive and contested consumers (Varian, 1980; Narasimhan, 1988). Following Armstrong and Vickers (2019), we assume that there are three consumer types and two firms. The consumer is either "contested", or "captive": the contested type of the consumer considers buying from either firm, whereas a captive consumer only buys from one of the firms. We refer to the firms as "weak" and "strong", the value of the strong firm's good being greater than the value of the weak firm's good.

If the integrated firm is the strong firm, information is then fully foreclosed (Proposition 1). By contrast, if the integrated firm is weak, information may be sold; in the latter case, whether the integrated firm sells no information, partial information, or full information, depends on (i) the values that the consumer attaches to the different goods, and (ii) the likelihoods of the different consumer types (Proposition 2). Furthermore, if information is foreclosed (whether fully or partially) then industry profit increases while consumer surplus decreases. We thus obtain a characterization of the situations in which integrated data intermediation is strictly worse for the consumer compared to independent data intermediation.

Intuitively, when the integrated firm is strong, foreclosing information enables it to soften competition in the goods market. The reason is that being uncertain whether the consumer is captive or contested, the weak firm sets a high price. By contrast, when the integrated

firm is weak, *selling* information may soften competition. To illustrate this point, suppose that the likelihood of the contested type is high: in this case, the strong firm prices its good aggressively, and squeezes the weak firm out of the market for the contested type. Selling information indicating whether the consumer is captive or contested ensures that the strong firm sometimes sets a higher price.

To endogenize the market's structure, we augment the model with an initial stage in which an independent data intermediary can acquire one of the firms. We show that the data intermediary always finds the acquisition of one of the firms profitable. The data intermediary acquires the strong firm when either (a) the value of the strong firm's good is much larger than that of the weak firm's good, or (b) the probability that the consumer is captive to the strong firm is small relative to the corresponding probability for the weak firm. We also show that takeovers always reduce competition for the contested consumer type, thereby lowering consumer surplus (Theorem 3).

Our paper provides a framework for examining the ongoing policy discussion on the risks of anti-competitive behavior by data intermediaries with activities beyond the market for consumer data. Large data intermediaries like Google and Microsoft have actively pursued the acquisition of firms operating in vertically related markets (OECD, 2023).[2] A major concern of competition authorities, which our analysis substantiates, is that such business models could lead a data intermediary to favor its own or affiliated companies over competitors, a practice known as "self-preferencing".[3] In our setting, self-preferencing takes the form of information foreclosure: the presence of an integrated data intermediary results in privileged access to consumer information.

In the final part of the paper, we use our model to shed light on recent developments in the digital advertising industry. In 2020, Google announced plans to phase out third-party cookies in its Chrome browser by introducing the Google Privacy Sandbox. This move triggered concerns that Google would no longer transmit individual user information to its customer firms (see Motta, 2023). We first discuss how our model applies to the digital advertising industry. We then discuss how our analysis informs the decision taken by the

[2]Two prominent examples are Google's recent acquisition of Fitbit, and Microsoft acquisition of LinkedIn.

[3]In the European Union, the Digital Markets Act introduces measures against self-preferencing.

UK Antitrust authority.

The structure of our paper is as follows. The general setting is presented in Section 2. Section 3 studies independent data intermediation. Section 4 shows that integrated data intermediation results in lower consumer surplus. Section 5 studies integrated data intermediation and takeovers with captive and contested consumers. In Section 6, we describe how our model contributes to the recent policy debate concerning the digital advertising industry. Section 7 concludes. The Appendix contains the omitted proofs.

Related literature. Our paper contributes to several strands of the literature. It connects the literature on market segmentation and third-degree price discrimination in oligopolies with the literature on the sale of information in oligopolies.

The former literature includes Bergemann, Brooks, and Morris (2021), Shi and Zhang (2022), and Elliott, Galeotti, Koh, and Li (2022). This strand of research seeks to characterize the optimal segmentation of a market given exogenously fixed objectives, typically maximizing consumer surplus or producer surplus. In contrast to this literature, in this paper, market segmentation is endogenous, and results from the sale of information. Generally, information sold by a profit-maximizing data intermediary does not maximize consumer surplus or producer surplus. The reason is that the data intermediary seeks to maximize the value for information, i.e., the difference in profit with and without information.[4] Moreover, in the presence of an integrated data intermediary, the value of information changes. Whether the equilibrium information structure is closer to maximizing consumer or producer surplus depends on the bargaining protocol and the structure of the industry.

A nascent literature studies the sale of information in oligopolies.[5] Montes, Sand-Zantman, and Valletti (2019) consider two firms competing in a Hotelling setting and a data intermediary who chooses whether to sell full information about consumers either exclusively

[4]Two other related papers are Armstrong and Vickers (2019) and Yang (2022): the former explores how information about consumer types affects consumer and producer welfare in oligopolies with captive and contested consumers; the latter studies the sale of consumer data to a monopolistic firm with private information about production costs.

[5]A distinct literature studies the incentives of oligopolists to freely share information about costs or aggregate demand (see, e.g., Novshek and Sonnenschein, 1982; Clarke, 1983; Vives, 1984; Raith, 1996).

to one firm or to both firms. They show that the data intermediary's optimal strategy is to restrict the sale of information to a single firm.[6] Bounie, Dubus, and Waelbroeck (2021) and Bimpikis, Crapis, and Tahbaz-Salehi (2019) consider models in which a data intermediary can both sell information and choose the precision of information. Our work differs from theirs in three ways. First, in our setting the data intermediary is unrestricted in the design of information that it sells to the downstream firms. Second, building on the canonical approach followed by the literature on vertical relationships (Rey and Tirole, 2007), our data intermediary offers secret bilateral contracts to the downstream firms. Third, we determine the equilibrium structure of the industry and examine its impact on the sale of consumer data and welfare.[7]

Finally, our work is connected to the burgeoning literature on digital platforms. A strand of this literature studies the externalities that consumers exert on each other when sharing personal data with a platform (Ichihashi, 2021b; Acemoglu, Makhdoumi, Malekian, and Ozdaglar, 2022; Bergemann, Bonatti, and Gan, 2022; Bergemann and Bonatti, forthcoming). Other strands of this literature examine privacy regulation (see, among others, Taylor, 2004; Argenziano and Bonatti, 2023; Hidir and Vellodi, 2021), optimal pricing by two-sided platforms (e.g., Gomes and Pavan, 2016; Jullien and Pavan, 2018; Valenzuela-Stookey, 2020), or customer segmentation by competing platforms (Karle, Peitz, and Reisinger, 2020). Finally, Anderson and Bedre-Defolie (2023) model self-preferencing by allowing a platform to choose a hybrid structure, i.e., to display the products marketed by private labels. Our contribution to this literature consists in showing that self-preferencing can take the form of information foreclosure.

[6]Braulin and Valletti (2016) show that exclusivity also prevails in a model with vertical product differentiation.

[7]Bonatti, Dahleh, Horel, and Nouripour (2024) and Rodríguez Olivera (2024) study optimal mechanisms for selling information to privately informed agents who subsequently interact in a game, but do not analyze the optimal takeover decision of the data intermediary. All the aforementioned papers consider a monopolistic data intermediary. Competition among data intermediaries is examined in Ichihashi (2021a), Bounie, Dubus, and Waelbroeck (2022), and Terstiege and Vigier (2023).

2 General Environment

We consider an environment with $n \geq 2$ firms and one consumer.[8] The consumer's valuation for the good of firm $i = 1, \cdots, n$ is denoted by $\theta_i \in V$, where $V \subset \mathbb{R}_+$ is finite. The tuple

$$\boldsymbol{\theta} := (\theta_1, \cdots, \theta_n) \in V^n$$

is referred to as the consumer's *type*.

The consumer buys from at most one firm. His utility is $\theta_i - p_i$ if he buys from firm i at a price equal to $p_i \in \mathbb{R}_+$, and zero if he buys nothing. We break ties by assuming that the consumer buys if and only if $\max_i \theta_i - p_i \geq 0$, in which case he selects among the firms in $\arg\max_i \theta_i - p_i$ the firm $i^* = \min \arg\max_i \theta_i$.

The consumer's type $\boldsymbol{\theta}$ is initially unknown to all parties other than the consumer himself. The common prior is denoted by $\bar{\mu} \in \Delta V^n$: letting $\mathbf{v} = (v_1, \cdots, v_n)$ denote a generic element of V^n, $\bar{\mu}(\mathbf{v})$ is the prior probability that $\boldsymbol{\theta} = \mathbf{v}$.[9]

Data policies allow firms to receive additional information concerning the consumer's type. Formally, a *data policy* is a tuple (M, ψ), where M is a nonempty finite message set and

$$\psi : \text{supp } \bar{\mu} \ \rightarrow \ \Delta M.$$

The interpretation is as follows: for $\boldsymbol{\theta} = \mathbf{v}$, message $m \in M$ is drawn with probability $\psi(m \mid \mathbf{v})$. We refer to a tuple $\Big((M, \psi), t\Big)$ combining a data policy (M, ψ) and a fee $t \in \mathbb{R}$ as a *contract*. We will consider two models of information sale. In Section 3, firms buy data policies from an independent data intermediary. In Section 4, information is sold by one of the firms.

[8] Alternatively, we may interpret the model as having a continuum of consumers, and the prior $\bar{\mu}$ described below as representing the proportions of consumers with particular valuations.

[9] For any finite set X, ΔX denotes the set of all probability mass functions on X. $\Delta \mathbb{R}_+$ denotes the set of all Borel probability measures on $\mathbb{R}_+$.

2.1 Notation and Terminology

For $\mathbf{v} = (v_1, \cdots, v_n) \in V^n$, let v^1 and v^2 denote the highest and second highest valuations, that is,

$$v^1 := \max_i v_i,$$

and, fixing an arbitrary $i^* \in \arg\max_i v_i$,

$$v^2 := \max_{i \neq i^*} v_i.$$

Then define

$$\alpha_i(\mathbf{v}) := \begin{cases} v^1 - v^2 & \text{if } i \in \arg\max_{i' \in \{1,\ldots,n\}} v_{i'}, \\ 0 & \text{otherwise;} \end{cases}$$

we will refer to $\alpha_i(\mathbf{v})$ as *firm i's competitive advantage* at $\mathbf{v}$.

We say that a data policy (M, ψ) is *fully informative* if whenever $\mathbf{v} \neq \mathbf{v}'$, $\psi(m \mid \mathbf{v}) > 0$ implies $\psi(m \mid \mathbf{v}') = 0$. Thus, a fully informative data policy perfectly identifies the consumer's type. A data policy (M, ψ) is *uninformative* if $\psi(m \mid \mathbf{v}) = \psi(m \mid \mathbf{v}')$ for all $m \in M$ and all $\mathbf{v}$,$\mathbf{v}' \in \text{supp } \bar{\mu}$.

Finally, for every $\mathbf{v} \in \text{supp } \bar{\mu}$, the *pricing game* $G(\mathbf{v})$ is defined as follows: the consumer's type is commonly known to be $\boldsymbol{\theta} = \mathbf{v}$, and each firm i simultaneously sets a price $p_i \in \mathbb{R}_+$ and obtains a payoff equal to its revenue. The following basic result is well known (Wilson, 1967; Lee, 1984); we state it for future reference.[10]

Lemma 1. *Let $\mathbf{v} \in$ supp $\bar{\mu}$ such that $v^1 > 0$. In every Nash equilibrium of the pricing game $G(\mathbf{v})$, the consumer buys with probability one from a firm with valuation v^1 at a price equal to $v^1 - v^2$.*

3 Independent Data Intermediation

The model with independent data intermediation adds to the general environment described in Section 2 an independent data intermediary (DI). The DI offers each firm i a contract $\Big((M_i, \psi_i), t_i\Big)$; each firm either accepts or rejects, and pays the fee t_i in the former case. Firms

[10] For completeness, a proof is presented in the Appendix.

neither observe the contracts offered to their rivals, nor whether said contracts were accepted. The DI then learns the consumer's type, and independently draws messages $m_1, \cdots, m_n$ according to $(M_1, \psi_1), \cdots, (M_n, \psi_n)$, respectively. Finally, every firm i that accepted the contract that it was offered observes m_i and chooses its price p_i.

The payoff of firm i is its revenue, minus the fee t_i if the contract $\Big((M_i, \psi_i), t_i\Big)$ was accepted. The payoff of the DI is its revenue from selling data.

The solution concept is perfect Bayesian equilibrium, with "passive beliefs", that is, a firm's conjectures about the DI's contract offers to its rivals and about said rivals' actions are not influenced by an out-of-equilibrium offer.[11] We restrict attention to equilibria in which, on path, all offers are accepted; this streamlines the exposition and is inconsequential since the DI can always offer an uninformative data policy for free. We will refer to such equilibria as *equilibria with independent data intermediation.*

Throughout the paper, we use the terminology *industry profit* for the expected price paid by the consumer, and *consumer surplus* for the consumer's expected utility. We say that an equilibrium is *efficient* if, for every $\mathbf{v} \in \text{supp } \bar{\mu}$, the consumer buys with probability one from a firm with valuation v^1.

The following example illustrates the workings of the model with independent data intermediation.

Example 1. *Suppose there are $n = 2$ firms, and the possible consumer types are $(1, 0)$, $(0, 1)$, $(4/5, 1/5)$, and $(1/5, 4/5)$, all with equal probability. We first argue that an equilibrium with independent data intermediation must be such that some information is sold. If not, each firm would choose a price of $4/5$ and obtain a payoff of $4/5 * 1/2 = 0.4$. However, the DI could then profitably inform firm 1 of whether $\boldsymbol{\theta} = (1, 0)$ or $\boldsymbol{\theta} \neq (1, 0)$. Doing so would enable firm 1 to obtain an expected profit of $1/4 * 1 + 1/4 * 4/5 = 0.45$. This information could thus be sold for a fee $t_1 = 0.05$. We will see that an equilibrium with independent data intermediation exists in which the DI sells a fully informative data policy to both firms; the resulting industry profit is 0.8, and each firm pays a fee of 0.1 to the DI, corresponding to*

[11] The use of passive beliefs is common in the literature on multilateral contracting (see, e.g., Hart and Tirole, 1990; McAfee and Schwartz, 1994; Segal, 1999). Rey and Vergé (2004) argue that passive beliefs are particularly appropriate in settings like ours where information is used by firms to set prices.

the difference between a firm's expected revenues with and without information.

Applying Lemma 1 establishes that when the DI sells a fully informative data policy to all firms, the resulting industry profit and consumer surplus are

$$\Pi^f := \sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})(v^1 - v^2),$$

and

$$C^f := \sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})v^2.$$

We are now ready to state our first main result:

Theorem 1. *An equilibrium with independent data intermediation exists in which the DI sells a fully informative data policy to all firms, achieving industry profit Π^f and consumer surplus C^f. Moreover, all equilibria with independent data intermediation are efficient, generate industry profit Π^f, and consumer surplus C^f.*

Proof. We show that an equilibrium exists in which (a) the DI offers every firm i a contract $\Big((M_i, \psi_i), t_i\Big)$ where (M_i, ψ_i) is fully informative and

$$t_i = \sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})\alpha_i(\mathbf{v}) - \max_p \sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})\mathbf{1}_{[0,\alpha_i(\mathbf{v})]}(p)p,$$

(b) every firm i accepts its contract and sets the price $p_i = \alpha_i(\mathbf{v})$ when $\boldsymbol{\theta} = \mathbf{v}$.[12]

Suppose $\boldsymbol{\theta} = \mathbf{v}$ and every firm $j \neq i$ chooses $p_j = \alpha_j(\mathbf{v})$. If $\alpha_i(\mathbf{v}) = 0$, the consumer never buys from firm i for any $p_i > 0$; so choosing $p_i = 0$ is optimal for firm i. If instead $\alpha_i(\mathbf{v}) > 0$, then the tie-breaking assumptions imply that the consumer buys from firm i if and only if $p_i \leq \alpha_i(\mathbf{v})$; so choosing $p_i = \alpha_i(\mathbf{v})$ is optimal for firm i. Hence, in the putative equilibrium, every firm i sets prices optimally; it obtains an expected revenue equal to

$$\sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})\alpha_i(\mathbf{v}),$$

and an expected payoff of

$$\max_p \sum_{\mathbf{v}\in V^n} \bar{\mu}(\mathbf{v})\mathbf{1}_{[0,\alpha_i(\mathbf{v})]}(p)p.$$

[12] $\mathbf{1}_A$ denotes the indicator function of A.

This expected payoff is also firm i's maximum expected payoff in case firm i rejects $\Big((M_i, \psi_i), t_i\Big)$, whence firm i is indifferent between accepting and rejecting $\Big((M_i, \psi_i), t_i\Big)$; in particular, accepting the contract is optimal.

Finally, as the data policy (M_i, ψ_i) is fully informative and firm i is indifferent between accepting and rejecting the contract $\Big((M_i, \psi_i), t_i\Big)$, note that t_i is the largest fee that the DI can extract from firm i. It follows that the DI has no profitable deviation.

We now prove the second part of the theorem. Pick an equilibrium. Let $\Big((M_i, \psi_i), t_i\Big)$ be the contract that the DI offers to firm i, and $\sigma_i : M_i \to \Delta\mathbb{R}_+$ be firm i's pricing strategy in the equilibrium considered. For every $\mathbf{v} \in \text{supp } \bar{\mu}$, define $\zeta_i^{\mathbf{v}} \in \Delta\mathbb{R}_+$ by

$$\zeta_i^{\mathbf{v}}\Big([0, x]\Big) := \sum_{m_i \in M_i} \sigma_i\Big([0, x] \mid m_i\Big)\psi_i(m_i \mid \mathbf{v}),$$

for all $x \in \mathbb{R}$. In words, $\zeta_i^{\mathbf{v}}$ represents the distribution of prices induced by σ_i conditional on the consumer's type $\boldsymbol{\theta}$ being equal to $\mathbf{v}$.

We argue that for every $\mathbf{v} \in \text{supp } \bar{\mu}$, $(\zeta_1^{\mathbf{v}}, \cdots, \zeta_n^{\mathbf{v}})$ is a Nash equilibrium of the pricing game $G(\mathbf{v})$. Reason by contradiction, and suppose that we can find $\mathbf{v}^*$ where this is not the case. There exist then a firm i and a price p^* such that setting $p_i = p^*$ yields a strictly higher expected revenue conditional on $\boldsymbol{\theta} = \mathbf{v}^*$ than $\zeta_i^{\mathbf{v}^*}$. It follows that (M_i, ψ_i) is not fully informative, otherwise firm i could profitably deviate from σ_i. Next, suppose the DI deviated and offered firm i a contract comprising the fully informative data policy (M_i', ψ_i') with $M_i' = \text{supp } \bar{\mu}$ and $\psi_i'(\mathbf{v} \mid \mathbf{v}) = 1$ for all $\mathbf{v} \in \text{supp } \bar{\mu}$.

Now consider the pricing strategy $\sigma_i' : M_i' \to \Delta\mathbb{R}_+$ given by $\sigma_i'(\mathbf{v}) = \zeta_i^{\mathbf{v}}$ for all $\mathbf{v} \in M_i' \setminus \mathbf{v}^*$, and $\sigma_i'(\mathbf{v}^*) = \delta_{p^*}$, the Dirac measure with mass 1 on p^*. By passive beliefs, the pricing strategy σ_i' yields firm i a strictly higher expected revenue than the expected revenue which it obtains in the equilibrium considered; thus, the DI can sell (M_i', ψ_i') for a fee $t_i' > t_i$. This is impossible because in equilibrium the DI maximizes its expected revenue from selling data. Hence, $(\zeta_1^{\mathbf{v}}, \cdots, \zeta_n^{\mathbf{v}})$ is a Nash equilibrium of the pricing game $G(\mathbf{v})$.

We can now apply Lemma 1: for every $\mathbf{v} \in \text{supp } \bar{\mu}$ such that $v^1 > 0$, the consumer buys with probability one from a firm with valuation v^1 at a price equal to $v^1 - v^2$. The equilibrium is efficient. It generates industry profit Π^f, and consumer surplus C^f. □

The logic underlying Theorem 1 is as follows. Secret offers and passive beliefs imply that

any information that is valuable to some firm will be sold. Hence, in an equilibrium with independent data intermediation, any unsold information must have zero value. This, in turn, implies that each firm must best respond to the prices of its rivals for every possible realization of the consumer's type. Effectively, the DI floods the market with information, thus ensuring that no business-stealing opportunity exists for any of the firms.

Theorem 1 is related to the opportunism problem in Hart and Tirole (1990). However, our model of data intermediation is conceptually very different from the classic model studied by the industrial organization literature on vertical relationships (see, e.g., Rey and Tirole, 2007). In the classic formulation of the problem, a monopolistic manufacturer sells a physical input good to downstream retailers. Thus, the manufacturer controls the quantity produced by the downstream firms. In our model, the contract offered by the DI contains the tariff at which a firm gets access to information about the consumer's type. This means that the DI affects firms' price discrimination possibilities.

Under independent data intermediation, consumer surplus is not necessarily maximized.[13] However, as we will see in the section, the consumer is weakly better off under independent data intermediation than under integrated data intermediation.

4 Integrated Data Intermediation

In this section, we assume that one of the firms, say firm 1, observes the consumer's type and sells information to its rivals. Throughout this section, we therefore refer to firm 1 as the *integrated firm*.

The model with integrated data intermediation is as follows. The integrated firm offers each firm $i > 1$ a contract $\Big((M_i, \psi_i), t_i\Big)$, which a firm can either accept or reject. If it accepts, the firm pays the fee t_i. All firms $i > 1$ observe neither the contracts offered to their rivals, nor whether said contracts were accepted. The integrated firm, instead, perfectly learns the consumer's type, and independently draws messages $m_2, \cdots, m_n$ according to

[13]The consumer-surplus maximizing information structure is characterized by Elliott, Galeotti, Koh, and Li (2022). Each firm learns whether its product is the consumer's most preferred product, but not the exact willingness to pay of the consumer.

$(M_2, \psi_1), \cdots (M_n, \psi_n)$, respectively. Every firm i that accepted the contract that it was offered observes m_i. Finally, every firm i chooses its price p_i.

The payoff of firm $i > 1$ is as in Section 3. The payoff of the integrated firm is the sum of its revenues in the goods and data markets.

As in the previous section, the solution concept is perfect Bayesian equilibrium with passive beliefs. Thus, a firm's conjectures about the contracts offered to the other firms and about the actions of those firms are not influenced by an out-of-equilibrium contract offer. Here, passive beliefs further require that a firm's conjecture about the integrated firm's pricing strategy following rejection is independent of the content of the rejected contract. We will refer to such equilibria as *equilibria with integrated data intermediation*.

The following example illustrates the workings of the model with integrated data intermediation.

Example 2. *The setting is as in Example 1: there are $n = 2$ firms, and the possible consumer types are $(1, 0)$, $(0, 1)$, $(4/5, 1/5)$, and $(1/5, 4/5)$, all with equal probability. We will argue that the integrated firm can profitably foreclose information. A fully informative data policy would enable firm 2 to sell its product at a price $p_2 = 3/5$ to the consumer of type $(1/5, 4/5)$ and at a price $p_2 = 1$ to the consumer of type $(0, 1)$, thus achieving an expected profit of 0.4. In the absence of information, firm 2's expected profit would be 0.3. So the integrated firm can sell a fully informative data policy for a fee $t_2 = 0.1$. Now suppose that the integrated firm instead only informs firm 2 of whether $\boldsymbol{\theta} \in \{(0, 1), (4/5, 1/5)\}$ or $\boldsymbol{\theta} \notin \{(0, 1), (4/5, 1/5)\}$. In this case firm 2 chooses $p_2 = 1$ when $\boldsymbol{\theta} \in \{(0, 1), (4/5, 1/5)\}$ and $p_2 = 3/5$ when $\boldsymbol{\theta} \notin \{(0, 1), (4/5, 1/5)\}$. This information could still be sold for a fee $t_2 = 0.1$, but now the integrated firm can sell its good at a price $p_1 = 4/5$ when $\boldsymbol{\theta} = (4/5, 1/5)$ instead of $3/5$. In this example, foreclosing information enables the integrated firm to reduce consumer surplus and increase the sum of its expected revenues in the goods and data markets.*

The content of information sold by the integrated firm varies depending on the parameters of the model: we will see in the next section that the data policies sold by the integrated firm can range from fully informative to uninformative. In the rest of this section, our focus is on the consequences of integrated data intermediation for consumer welfare.

Theorem 2. *In every equilibrium with integrated data intermediation consumer surplus is at most C^f.*

Proof. To shorten notation, let

$$R_i^f := \sum_{\mathbf{v} \in V^n} \bar{\mu}(\mathbf{v}) \alpha_i(\mathbf{v}).$$

Then

$$\Pi^f = \sum_{i=1}^{n} R_i^f. \tag{1}$$

Furthermore, if firm $i = 2, \ldots, n$ has access to a fully informative data policy, then it can guarantee itself an expected revenue of at least R_i^f (by charging price $p_i = \alpha_i(\mathbf{v})$ when the consumer's type is $\mathbf{v}$). The integrated firm can always guarantee itself an expected revenue of R_1^f in the goods market.

Pick an arbitrary equilibrium. Let R_1^* denote the integrated firm's expected revenue in the goods market, and R_i^* the expected revenue of firm $i = 2, \ldots, n$. Let $\left((M_i^*, \psi_i^*), t_i^*\right)$ be the contract that the integrated firm offers to firm $i = 2 \ldots, n$. The integrated firm's payoff is then

$$R_1^* + \sum_{i \geq 2} t_i^*.$$

Next, observe that the consumer surplus in the equilibrium considered is bounded from above by the difference between the maximum social welfare and the firms' profits:

$$\sum_{\mathbf{v} \in V^n} \bar{\mu}(\mathbf{v}) v^1 - \sum_{i=1}^{n} R_i^* = \Pi^f + C^f - \sum_{i=1}^{n} R_i^*.$$

So to prove the proposition, it is enough to show that

$$\Pi^f \leq \sum_{i=1}^{n} R_i^*. \tag{2}$$

Reason by contradiction, and suppose that (2) is violated. By (1), we thus have

$$\sum_{i=1}^{n} R_i^f > \sum_{i=1}^{n} R_i^*. \tag{3}$$

As noted above, $R_1^* \geq R_1^f$, whence we can find a firm $j > 1$ whose data policy in the equilibrium considered is not fully informative. Now suppose that the integrated firm deviates and offers every firm i a contract $\left((M_i', \psi_i'), t_i'\right)$ comprising a fully informative data policy.

Let R_i' be firm i's expected revenue from accepting $\left((M_i', \psi_i'), t_i'\right)$. By passive beliefs, firm i must accept this contract whenever

$$t_i' \leq t_i^* + R_i' - R_i^* - \varepsilon,$$

for any $\varepsilon > 0$. Now pick an arbitrary $\varepsilon > 0$ and, for $i = 2, \cdots, n$, let t_i' be such that the previous inequality holds with equality. Letting R_1' denote the integrated firm's expected revenue in the goods market under the deviation, the integrated firm's gain from the deviation is then

$$\begin{aligned} R_1' + \sum_{i \geq 2}(t_i^* + R_i' - R_i^*) - (n-1)\varepsilon - \left(R_1^* + \sum_{i \geq 2} t_i^*\right) &= \sum_{i=1}^{n} R_i' - \sum_{i=1}^{n} R_i^* - (n-1)\varepsilon \\ &\geq \sum_{i=1}^{n} R_i^f - \sum_{i=1}^{n} R_i^* - (n-1)\varepsilon. \end{aligned}$$

It follows from inequality (3) that the deviation is profitable whenever ε is sufficiently small. This contradiction establishes inequality (2). □

Theorem 2 connects our work to the vertical integration literature studying settings in which an integrated supplier sells a physical input good to downstream retailers. Vertical integration allows the supplier to restrict the quantity produced on the product market by foreclosing rivals' access to the input good (see, among others, O'Brien and Shaffer, 1992; McAfee and Schwartz, 1994; Rey and Vergé, 2004; Rey and Tirole, 2007; Reisinger and Tarantino, 2015; Nocke and Rey, 2018). As we will show next, in our model foreclosure takes the form of self-preferencing: the data intermediary gives preferential access to information to the integrated firm, at the expense of other firms.

Combining Theorems 1 and 2 reveals that independent data intermediation can never be worse for consumers than integrated data intermediation. Whether independent data intermediation benefits consumers depends on the parameters of the model. In Section 5, we use an environment with captive and contested consumers to determine the conditions under which independent data intermediation raises consumer surplus, and the equilibrium structure of the industry.

5 Setting with Captive and Contested Consumers

In this section, we explore a special case of our general environment embedding Varian and Narasimhan's model of captive and contested consumers (Varian, 1980; Narasimhan, 1988). Recently, this model has been widely used by the literature studying how market segmentation affects product market competition (see, among others, Armstrong and Vickers, 2019; Bergemann, Brooks, and Morris, 2021; Shi and Zhang, 2022).

There are $n = 2$ firms, and three possible consumer types: (w_1, w_2), $(w_1, 0)$, and $(0, w_2)$. We shorten notation by letting

$$\bar{\mu}_0 := \bar{\mu}\Big((w_1, w_2)\Big), \quad \bar{\mu}_1 := \bar{\mu}\Big((w_1, 0)\Big), \quad \bar{\mu}_2 := \bar{\mu}\Big((0, w_2)\Big),$$

and assume $\bar{\mu}_0, \bar{\mu}_1, \bar{\mu}_2 > 0$.

The interpretation is as follows. The contested consumer, represented by the type (w_1, w_2), considers buying from both firms. Type $(w_1, 0)$ only considers buying from firm 1, and type $(0, w_2)$ only considers buying from firm 2: these consumer types represent captive consumers. For convenience, we henceforth refer to type 0, type 1, and type 2 of the consumer, respectively.

To streamline the exposition, we assume throughout that

$$w_1 \geq w_2 > 0,$$

and thus say that firm 1 is the *strong* firm and firm 2 is the *weak* firm. Accordingly, the types 0 and 1 of the consumer are in the *strong firm's reach*; and the the types 0 and 2 of the consumer are in the *weak firm's reach*. Since a firm can only extract revenue from the consumer if the latter is within its reach, a firm's only relevant information is the information that it possesses concerning its reach.

A data policy (M, ψ) is *fully informative concerning the strong firm's reach* if $\psi\Big(m \mid (w_1, 0)\Big) > 0$ implies $\psi\Big(m \mid (w_1, w_2)\Big) = 0$. A data policy (M, ψ) is *uninformative with respect to the strong firm's reach* if $\psi\Big(m \mid (w_1, 0)\Big) = \psi\Big(m \mid (w_1, w_2)\Big)$ for all $m \in M$. A data policy is *partially informative concerning the strong firm's reach* if it is neither uninformative nor fully informative concerning the strong firm's reach. The terminology is symmetric for the weak firm.

Section 3 examined the information sold by an independent DI, and showed that consumer surplus is as if the DI sells a fully informative data policy to every firm (Theorem 1). In what follows, our goal is threefold. First, we examine how information sold by the DI depends on whether it is integrated with the strong or the weak firm. Second, we study the equilibrium structure of the industry. Third, we determine the impact of integration on welfare, and draw the policy implications of our analysis.

5.1 Strong Integrated Firm

Throughout this subsection, the integrated firm is the strong firm (firm 1).

We first study a class of ancillary games. These games are useful because they capture the pricing "subgames" between the integrated firm and its rival. A generic distribution in $\Delta(\text{supp } \bar{\mu})$ will be denoted by μ. In line with previous notation, let

$$\mu_0 := \mu\Big((w_1, w_2)\Big), \quad \mu_1 := \mu\Big((w_1, 0)\Big), \quad \mu_2 := \mu\Big((0, w_2)\Big).$$

For $\mu \in \Delta(\text{supp } \bar{\mu})$, define the ancillary game $G(1, \mu)$ as follows. The consumer's type is drawn from μ and then observed only by firm 1. Subsequently, firms simultaneously choose prices. Each firm's payoff is its revenue.[14]

Lemma 2. *In all equilibria of the ancillary game $G(1, \mu)$:*

- *firm 2's expected revenue is $\mu_2 w_2$;*
- *if $\mu_1 < 1$, firm 1's expected revenue is*

$$\mu_0 \left(\frac{\mu_2 w_2}{\mu_0 + \mu_2} + w_1 - w_2 \right) + \mu_1 w_1; \tag{4}$$

  if $\mu_1 = 1$, firm 1's expected revenue is w_1.

Suppose $\mu_1 < 1$ (the other case is trivial). To gain intuition for the lemma, first recall that the consumer of type 1 never buys from firm 2. Next, firm 2 can guarantee itself an

[14] $G(1, \mu)$ and the ancillary game $G(2, \mu)$ in Subsection 5.2 are closely related to first-price auction models with asymmetrically informed bidders. See Engelbrecht-Wiggans, Milgrom, and Weber (1983) for an analysis of such auctions with common values.

expected revenue of $\mu_2 w_2$ by choosing $p_2 = w_2$. Hence, in equilibrium, firm 2 must pick a price in the interval

$$\left[\frac{\mu_2 w_2}{\mu_0 + \mu_2}, w_2\right].$$

This, in turn, enables firm 1 to sell to the type-0 consumer at a price equal to $\mu_2 w_2/(\mu_0 + \mu_2) + w_1 - w_2$. Since, in addition, firm 1 can sell to the type-1 consumer at a price equal to w_1, expression (4) gives firm 1's expected revenue.

For future use, let $\Pi_1(\mu)$ denote the sum of the expected revenues of firms 1 and 2 in the ancillary game $G(1, \mu)$. Thus, $\Pi_1(\mu) := w_1$ if $\mu_1 = 1$; instead, if $\mu_1 < 1$,

$$\Pi_1(\mu) := \mu_0 \left(\frac{\mu_2 w_2}{\mu_0 + \mu_2} + w_1 - w_2\right) + \mu_1 w_1 + \mu_2 w_2.$$

We are now ready to analyze the information sold. We start by making three basic remarks. First, each data policy (M, ψ) induces a distribution of posteriors $\tau \in \Delta\Delta(\text{supp } \bar{\mu})$ satisfying

$$\sum_{\mu} \tau(\mu)\mu(\mathbf{v}) = \bar{\mu}(\mathbf{v}), \quad \forall \mathbf{v} \in \text{supp } \bar{\mu}; \tag{5}$$

every $\mu \in \text{supp } \tau$ represents firm 2's posterior belief following the observation of a message $m \in M$.[15] Conversely, every distribution $\tau \in \Delta\Delta(\text{supp } \bar{\mu})$ satisfying (5) is the distribution of posterior beliefs induced by some data policy (see Kamenica and Gentzkow, 2011).

Second, if τ is the distribution of posteriors associated with (M, ψ), then applying Lemma 2 shows that this data policy induces an industry profit equal to

$$\sum_{\mu} \tau(\mu)\Pi_1(\mu).$$

Third, in equilibrium, the integrated firm must offer firm 2 a contract making it indifferent between accepting and rejecting.

Combining the previous remarks establishes that a data policy forms part of an equilibrium contract if and only if it induces a posterior belief distribution τ solving

$$\max_{\tau} \quad \sum_{\mu} \tau(\mu)\Pi_1(\mu) \quad \text{s.t. (5) holds.} \tag{P$_1$}$$

The solution to this maximization problem yields the following proposition.

[15] $\Delta\Delta(\text{supp } \bar{\mu})$ denotes the set of all Borel probability measures on $\Delta(\text{supp } \bar{\mu})$ with finite support. A data policy (M, ψ) can only generate finitely many possible posteriors since the message set M is finite.

Proposition 1. *Suppose that the integrated firm is the strong firm. Then, in equilibrium, the integrated firm offers a contract comprising a data policy that is uninformative with respect to the weak firm's reach.*

Proposition 1 tells us that a strong integrated firm sells no information to its rival. By foreclosing information, the integrated firm here engages in a form of self-preferencing. The logic is as follows. In equilibrium, choosing $p_2 = w_2$ is optimal for firm 2 regardless of its posterior belief about the consumer's type. So, for firm 2 information has no value. This in turn prevents the integrated firm from making any revenue by selling data. Hence, if the integrated firm offers an informative data policy, this has to be to increase its expected revenue in the goods market. Yet, sharing information with firm 2 lowers the integrated firm's expected revenue in the goods market. To see why, first notice that the integrated firm always sells to the type-1 consumer at a price equal to w_1, so the focus is on firm 2's reach. Given a posterior μ with $1 - \mu_1 > 0$, denote by $\rho := \mu_2/(1 - \mu_1)$ the conditional probability of type 2 given that the consumer belongs to firm 2's reach. We saw earlier that, in equilibrium, the lowest price charged by the integrated firm is $\rho w_2 + w_1 - w_2$; furthermore, by charging this price, the integrated firm sells with probability $1-\rho$. The resulting expected revenue is $(1-\rho)(\rho w_2 + w_1 - w_2)$; this revenue being strictly concave in ρ, sharing information with firm 2 lowers the integrated firm's expected revenue in the goods market (by Jensen's inequality).

5.2 Weak Integrated Firm

Throughout this subsection, the integrated firm is the weak firm (firm 2).

For $\mu \in \Delta(\text{supp } \bar{\mu})$, the ancillary game $G(2, \mu)$ is defined analogously to the ancillary game $G(1, \mu)$, only interchanging the roles of the two firms; thus, firm 2 is now the one observing the consumer's type.

Lemma 3. *In all equilibria of the ancillary game $G(2, \mu)$:*

- *firm 1's expected revenue is*

$$\max\Big\{\mu_1 w_1,\ (\mu_0 + \mu_1)(w_1 - w_2)\Big\};$$

- *if $\mu_2 < 1$, firm 2's expected revenue is*

$$\max\left\{\mu_0\left(\frac{\mu_1 w_1}{\mu_0+\mu_1}-(w_1-w_2)\right),\,0\right\}+\mu_2 w_2; \tag{6}$$

  if $\mu_2 = 1$, firm 2's expected revenue is w_2.

Suppose $\mu_2 < 1$ (the other case is trivial). To understand Lemma 3, first notice that firm 1 can guarantee itself an expected revenue of $\mu_1 w_1$ by choosing $p_1 = w_1$. On the other hand, as $w_1 \geq w_2$, setting $p_1 = w_1 - w_2$ yields firm 1 an expected revenue of $(\mu_0 + \mu_1)(w_1 - w_2)$. Two cases must be considered. If $\mu_1 w_1 > (\mu_0+\mu_1)(w_1-w_2)$, then noting that in equilibrium, firm 1 must pick a price in the interval

$$\left[\frac{\mu_1 w_1}{\mu_0+\mu_1}, w_1\right]$$

shows that firm 2 can effectively sell to the consumer of type 0 at a price of $\mu_1 w_1/(\mu_0+\mu_1) - (w_1 - w_2) > 0$. Since, in addition, firm 2 can sell to the consumer of type 2 at a price of w_2, expression (6) gives firm 2's expected revenue. By contrast, in the remaining case, firm 1 chooses $p_1 = w_1 - w_2$ with probability one in all equilibria, thereby preventing firm 2 from earning revenue off the consumer of type 0.

We are now ready to analyze the nature of information sold by the integrated firm. Let $\Pi_2(\mu)$ denote the sum of firm 1 and firm 2's expected revenues in the ancillary game $G(2, \mu)$. Thus $\Pi_2(\mu) := w_2$ if $\mu_2 = 1$, while, if $\mu_2 < 1$,

$$\Pi_2(\mu) := \max\left\{\mu_0\left(\frac{\mu_1 w_1}{\mu_0+\mu_1}-(w_1-w_2)\right)+\mu_1 w_1,\,(\mu_0+\mu_1)(w_1-w_2)\right\}+\mu_2 w_2.$$

The same reasoning as in the previous subsection implies that the data policies forming part of an equilibrium contract are those that induce a posterior belief distribution τ, solving

$$\max_{\tau} \quad \sum_{\mu} \tau(\mu)\Pi_2(\mu) \quad \text{s.t. (5) holds.} \tag{P$_2$}$$

The solution of this maximization problem yields the next proposition.

Proposition 2. *Suppose that the integrated firm is the weak firm. In equilibrium:*

- *if $w_1 \geq 2w_2$, the integrated firm sells a data policy that is fully informative with respect to the strong firm's reach;*

- *if* $w_1 < 2w_2$ *and* $\bar{\mu}_1 < (1-\bar{\mu}_2)\sqrt{2(w_1-w_2)/(w_1)}$, *the integrated firm sells a data policy that is partially informative with respect to the strong firm's reach;*

- *otherwise, the integrated firm offers a contract comprising a data policy that is uninformative with respect to the strong firm's reach.*

In sharp contrast to Proposition 1, Proposition 2 tells us that a weak integrated firm sells information to its rival when either (a) the value of the rival's good is much larger than that of the integrated firm's good, or (b) the probability that the consumer is of type 1, conditional on being within the rival's reach, is sufficiently small.

The broad logic is as follows. In equilibrium, depending on firm 1's posterior belief about the consumer's type, either $p_1 = w_1$ is optimal, or $p_1 = w_1 - w_2$ is optimal, not both. So information is valuable to the strong firm. This allows the integrated firm to sell data in exchange for a positive fee. However, the integrated firm may face a trade-off. To illustrate this trade-off, given a posterior μ with $1-\mu_2 > 0$, denote by $\rho := \mu_1/(1-\mu_2)$ the conditional probability of type 1 given that the consumer belongs to the strong firm's reach. We saw above that, if

$$\rho > \frac{w_1 - w_2}{w_1},$$

in equilibrium the lowest price charged by the integrated firm is $\rho w_1 - (w_1 - w_2)$; furthermore, by charging this price, the integrated firm sells with probability $1-\rho$. The resulting expected revenue is $(1-\rho)\Big(\rho w_1 - (w_1 - w_2)\Big)$; this revenue being strictly concave in ρ, sharing information with firm 1 can lower the integrated firm's expected revenue in the goods market. Instead, if

$$\rho \leq \frac{w_1 - w_2}{w_1},$$

firm 1 charges $p_1 = w_1 - w_2$, squeezing the integrated firm out of the market for the consumer of type 0. Thus, when the probability that the consumer is of type 0 is large relative to the probability that the consumer is of type 1, sharing information increases the integrated firm's expected revenue in the goods market.

5.3 Takeovers

In this subsection, we endogenize the market's structure by allowing an independent DI to acquire one of the firms.

The model with endogenous market structure is as follows. At the outset, the DI makes a takeover bid $b_i \in \mathbb{R}$ to a firm i of its choice. If firm i accepts the bid, the DI and firm i form an integrated firm, in which case the game continues as in Subsections 5.1 and 5.2. If firm i rejects, the process is repeated with the other firm; if that firm rejects as well, the DI stays independent, and the game continues as in Section 3. Takeover bids, as well as a firm's decision whether to accept or reject, are publicly observable.

The payoff of firm i is equal to b_i if it accepted this bid; otherwise the payoff of firm i consists of its revenue, minus the fee t_i if the contract $\Big((M_i, \psi_i), t_i\Big)$ was accepted. If b_i was accepted, the DI's payoff is the sum of its revenues in the goods and data markets, minus b_i. If both bids were rejected, the payoff of the DI is its revenue from selling data. The solution concept is perfect Bayesian equilibrium; we here use the terminology *equilibrium with endogenous market structure.*

We saw in the previous subsections that, if the integrated firm is firm i, then the resulting industry profit is

$$\Pi_i^* := \max_{\tau} \sum_{\mu} \tau(\mu)\Pi_i(\mu) \quad \text{s.t. (5) holds.}$$

The following lemma establishes that the DI always acquires one of the firms, and that the DI's choice of the takeover target is determined by comparing Π_1^* and Π_2^*.

Lemma 4. *In every equilibrium with endogenous market structure, the DI acquires one of the firms. If $\Pi_1^* > \Pi_2^*$, the DI acquires the strong firm; if $\Pi_2^* > \Pi_1^*$, the DI acquires the weak firm.*

Let c_i denote firm i's expected payoff when firm $j \neq i$ is integrated. The key step of the proof of Lemma 4 shows that if the DI first makes a takeover bid to firm i, then firm i's outside option is equal to c_i. Thus, the DI can acquire firm i by bidding c_i, generate industry profit Π_i^*, and leave firm j with an expected payoff of c_j. This implies that the DI's choice of the takeover target is determined by whether Π_1^* is greater than Π_2^*. Finally, we show,

that for the DI, acquiring dominates staying independent, so the DI must acquire one of the firms.

We next build on Lemma 4 to determine which firm the DI chooses to acquire as a function of the parameters: w_1, w_2, $\bar{\mu}_1$, and $\bar{\mu}_2$.

Proposition 3. *Fix $\bar{\mu}_0$. If $w_1 \geq 2w_2$ then, in every equilibrium with endogenous market structure, the DI acquires the strong firm. Otherwise, there exists a cutoff κ such that the DI acquires the strong firm when $\bar{\mu}_1/\bar{\mu}_2 < \kappa$, and the weak firm when $\bar{\mu}_1/\bar{\mu}_2 > \kappa$.*

If the value of the strong firm's good is much larger than that of the weak firm's good, acquiring the weak firm cannot be optimal for the DI. The reason is as follows. If the integrated firm is weak and $w_1 \geq 2w_2$, the integrated firm can do no better than selling a data policy that is fully informative with respect to the strong firm's reach (Proposition 2). The price for the consumer of type 0 is then competed down to the strong firm's competitive advantage, $w_1 - w_2$. When instead the integrated firm is the strong firm, the integrated firm refrains from selling any information to its rival (Proposition 1); as a consequence, the rival competes less aggressively. Thus, acquiring the strong firm enables the DI to raise the price paid by the consumer of type 0.

However, if $w_1 < 2w_2$, acquiring the weak firm can be optimal for the DI. If the integrated firm is weak and $w_1 < 2w_2$, the integrated firm refrains from selling a fully informative data policy with respect to the strong firm's reach (Proposition 2). This softens competition for the consumer of type 0. Which firm the DI chooses to acquire is then determined by $\bar{\mu}_1$ and $\bar{\mu}_2$. Intuitively, if $\bar{\mu}_1$ is large and the DI acquires firm 2, then firm 1 will not compete very aggressively for the consumer of type 0; so in this case acquiring firm 2 may be preferable to acquiring firm 1, as illustrated in Figure 1.

5.4 Welfare Analysis and Policy Formulation

In this subsection, we draw the welfare implications of the results in the previous subsections. We study the effects of integration on consumer surplus and social welfare, and examine how

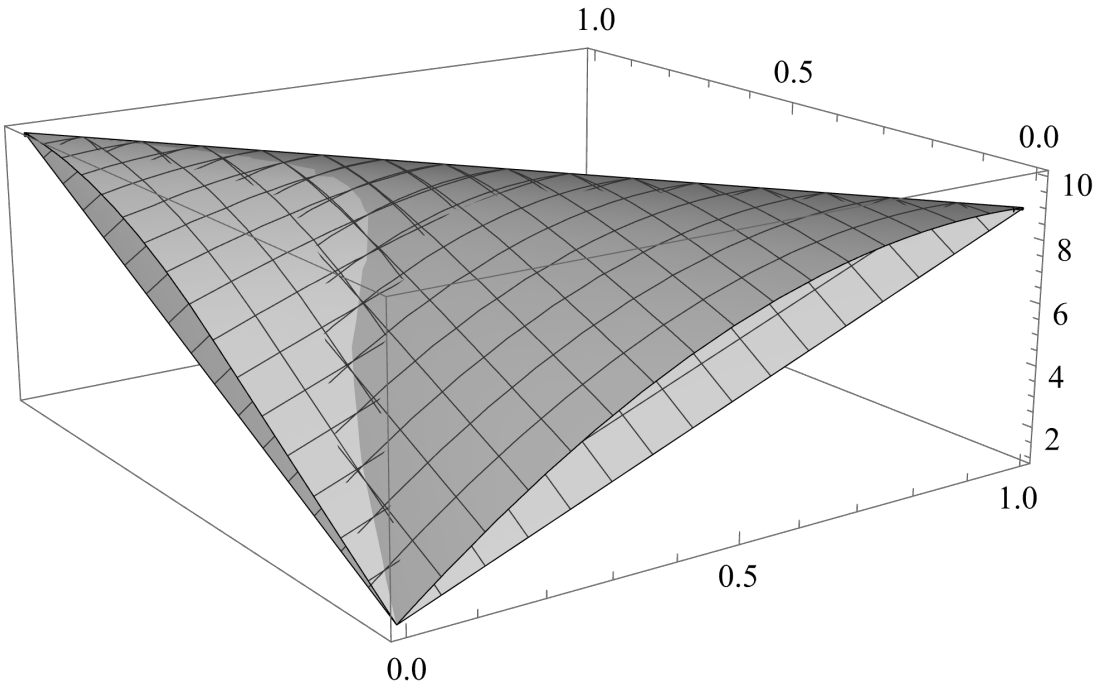


Figure 1: Industry profits Π_1^* (dark) and Π_2^* (light) depending on $\bar{\mu}_1$ (from front to back) and $\bar{\mu}_2$ (from left to right) for $w_1 = 10$ and $w_2 = 9$.

the revenue of the non-integrated firm changes with integration.[16]

This subsection's first result is as follows:

Proposition 4. *Suppose that the integrated firm is the strong firm. Then, every equilibrium is inefficient and generates a consumer surplus that is strictly smaller than C^f.*

Combining Theorem 1 and Proposition 4 shows that, relative to situations in which information is sold by an independent DI, data intermediation by a strong integrated firm lowers both social welfare and consumer surplus. The reason is that, as shown in Proposition 1, the integrated firm then fully forecloses information.

We now consider the case in which the integrated firm is the weak firm.

Proposition 5. *Suppose that the integrated firm is the weak firm. If $w_1 \geq 2w_2$, then every equilibrium is efficient and generates consumer surplus C^f; otherwise, every equilibrium is*

[16]In this subsection, we exclude throughout the knife-edge case $w_1 = w_2$. If $w_1 = w_2$, all equilibria are trivially efficient regardless of which firm is the integrated firm, and the expected revenue of the non-integrated firm is always the same as with independent data intermediation.

inefficient and generates consumer surplus that is strictly smaller than C^f.

If the integrated firm is the weak firm and the value of the strong firm's good is much larger than that of the weak firm's good ($w_1 \geq 2w_2$), independent and integrated data intermediation generate the same consumer surplus and social welfare. Instead, consumer surplus and social welfare fall if $w_1 < 2w_2$. Taken together, Propositions 4 and 5 show that self-preferencing (in the form of data policies that are either partially or fully uninformative) always reduces welfare.

We now examine the impact of integration on the revenue of the non-integrated firm. We show that the expected revenue of a non-integrated firm never increases compared to the benchmark with an independent DI.

Proposition 6. *a) If the integrated firm is the strong firm, then the weak firm's expected revenue is the same as with an independent DI.*

b) Let the integrated firm be the weak firm. If the equilibrium data policy is fully informative with respect to the strong firm's reach, then the strong firm's exptected revenue is the same as with an independent DI; otherwise, the expected revenue is smaller.

The proposition establishes that integration gives rise to consumer steering. In effect, with integrated data intermediation the consumer is directed towards the integrated firm.

We now analyze the welfare consequences of the takeover decision made by an independent DI.

Theorem 3. *Every equilibrium with endogenous market structure is inefficient and generates a consumer surplus that is strictly smaller than C^f.*

Combining Theorems 1 and 3 shows that takeovers are unambiguously harmful for consumers and social welfare. For policy makers, the implications are as follows. In order to restore efficiency, it is necessary that the DI does not internalize the impact that the sale of data to the non-integrated firm has on the integrated unit. As shown in Section 3, secrecy has this effect. In the next section, we discuss how secrecy can be implemented within an integrated company, taking as example a recent antitrust case dealing with the impact of data sale and self-preferencing within the digital advertising industry.

6 The Digital Advertising Industry

Our model provides a framework to examine recent developments that occurred in the digital advertising industry. First, we discuss how our model relates to this industry.

The good on the product market is a slot of advertising inventory—a space on a digital property, be it a webpage or mobile app, that can be utilized for displaying advertisements. As of 2020, the sale of advertising inventory stood as the primary revenue source for major companies such as Google, Meta, and Twitter.

Digital advertising encompasses visually oriented ads, including text, images, or videos, showcased on a publisher's website. For instance, Expedia and Kayak might be interested in purchasing ad inventory on the travel section of the New York Times. Transactions in this market occur through online advertising exchange platforms owned by companies like Google or Microsoft. Upon a user loading a participating publisher's webpage, a "request-for-content" is dispatched to the online advertising exchange platform, specifying the type, size, and other details of the intended advertisement (Bourreau, de Streel, and Graef, 2017).

In this industry, the equivalent of the downstream firms acquiring data from the data intermediary in our model are either the publishers, like the New York Times, or the online advertising exchange platforms that publishers use to manage their advertising inventory. In turn, the model's consumer is an advertiser that seeks to post digital advertisement on publishers' properties, like Expedia and Kayak in the example of the travel agencies.

The data intermediary of the model corresponds to a company that owns detailed information on consumers' valuation of downstream firms' goods and has significant market power. Google holds a (close to) monopolistic position in the market for search data by holding market shares in the search engine market exceeding 90% in most European countries.[17] It owns individual-level information that is often incidentally collected through the plethora of services it offers on its platforms, including popular services like maps and general search and various shopping search engines. Google is also dominant in the digital advertising market (EC, 2019), where it operates through its integrated companies. These can be pub-

[17]See the decision of the European Commission in the Google Shopping case (`https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1784`).

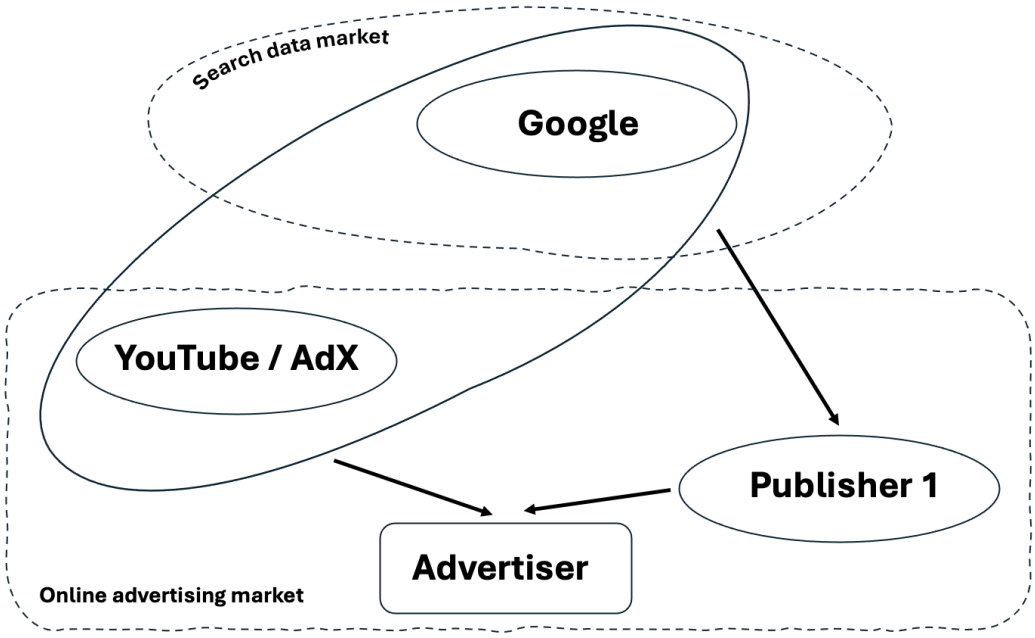


Figure 2: The figure plots a configuration of the digital advertising market in which an integrated DI operates next to an independent publisher.

lishers, like YouTube, or online advertising exchange platforms, like AdX, as documented by the decision of the European Commission in Google/Fitbit (EC, 2020).

In the digital advertising market, publishers price discriminate against advertisers (Celis, Lewis, Mobius, and Nazerzadeh, 2014). Whether personalized or group pricing occurs depends on the granularity of the information publishers have about the value advertisers attach to their slots. Ultimately, the latter is determined by the profile of the users who are potentially interested in the products the advertisers offer.

Publishers can obtain such information from data intermediaries, like Google (Alcobendas, Kobayashi, Shi, and Shum, 2023). A significant source of information is found in third-party cookies, which are commonly utilized to track individual user behavior and browsing patterns. Publishers' access to such information can happen via direct negotiation with the data intermediary, or indirectly through the online advertising exchange platforms that manage their advertising inventories (Bourreau, de Streel, and Graef, 2017).

Figure 2 illustrates the structure of the industry. Google is the dominant firm in the search data market. It also operates in the digital advertising market through its integrated

online exchange platform (AdX) or publisher (YouTube). We assume that Publisher 1 is independent (or collaborates with an online advertising exchange platform not integrated with Google). YouTube or AdX, instead, have privileged access to information that Google collects on the search data market.

The Google Privacy Sandbox Our model allows us to shed light on the competitive effects of practices leading to the aggregation of information sold by a data intermediary on the digital advertising industry.

A recent example is the introduction of the Google Privacy Sandbox. In January 2020, Google announced that it would no longer allow websites to access the third-party cookies stored in its web browser Chrome. Under the new policy, known as the Google Privacy Sandbox, Google aggregates user information by creating "cohorts", groups of users sharing certain features (Motta, 2023). With the introduction of the privacy sandbox, Google reduced the granularity of the information available to third-party publishers. According to the Competition and Markets Authority (CMA), which investigated the case, the Privacy Sandbox had the potential to give rise to a form of self-preferencing, by limiting the information available to Google's rivals in the digital advertising market, while leaving Google's ability to use this information relatively unaffected (CMA, 2022). This conclusion aligns with our Propositions 1 and 2.

Remark 1. *The aggregation of individual third-party cookies in cohorts of users via the Google Privacy Sandbox reduces the granularity of the information available to third-party publishers. This is equivalent to a form of information foreclosure.*

Alcobendas, Kobayashi, Shi, and Shum (2023) estimate that the information disadvantage caused by the elimination of third-party cookies reduces the advertising revenues of publishers that are not integrated with Google (or collaborate with online advertising exchange platforms that are not integrated with Google). Consistent with this empirical analysis, Proposition 6 shows that, in our model with captive and contested consumers, as a consequence of self-preferencing, the product market revenue of the non-integrated firm never increases compared to the benchmark without integration. In other words, self-preferencing gives rise to a form of steering of consumers towards the integrated firm.

Remark 2. *The Privacy Sandbox allows Google to soften product market competition by limiting publishers' opportunities to engage in third-degree price discrimination against advertisers in the digital advertising market.*

In its decision, the UK Competition and Markets Authority requested that Google simultaneously restrict data sharing to its integrated companies. The authority took this decision intending to avoid any distortion to competition between Google and other market participants in the digital advertising industry, which would limit the ability of publishers to generate revenue from ad inventory (CMA, 2022, p. 6). Our theoretical analysis supports this approach. With the introduction of a Chinese Wall between the data unit and its integrated (downstream) companies, Google's incentives in selling information should be aligned with those of an independent data intermediary.

Remark 3. *The imposition of a firewall restricting information sharing with downstream affiliates reduces Google's incentives to engage in information foreclosure.*

As discussed at the end of Section 5.4, secrecy has the effect of restoring efficiency. With secrecy, the data unit does not internalize the impact that the sale of data to the non-integrated firm has on the integrated unit. To obtain the intended effects, two conditions must apply. The first is that the integrated downstream unit does not observe the contract offer made to the non-integrated firms. The second is that the downstream unit of the integrated firm does not have access to the consumer information held by the data intermediary. In other words, the firewall must not only apply to data, but also to the contractual information related to the sale of data to the downstream competitors. Recently, the European Commission adopted a decision imposing a limit on data sharing within the newly integrated firm in Google/Fitbit (EC, 2020). There, the concern was about Google's use of health data collected through Fitbit. The Commission required that this information be contained into a "strictly permissioned virtual data storage environment within Google" (EC, 2020, p. 183).

7 Conclusion

Discriminatory pricing is a widespread phenomenon in digital markets. The welfare consequences of price discrimination depend on retailers' information about consumers. In practice, this kind of information is often sold by data intermediaries. Therefore, studying the sale of information by such intermediaries is crucial to evaluating the welfare consequences of price discrimination.

Data intermediaries such as Acxiom and Nielsen specialize in selling consumer data. By contrast, Google, Meta, Microsoft, and Amazon sell not only information but also goods and services; this dual role affects their incentives to sell information. In this paper, we have shown that intermediaries of the second kind often sell less information than intermediaries of the first kind, resulting in a lower consumer surplus. Furthermore, the acquisition of a downstream firm by an independent data intermediary always reduces consumer surplus.

These results speak directly to the ongoing policy debate surrounding the wave of acquisitions by Google, Microsoft, and other large digital companies. They suggest that such acquisitions should not be allowed. Alternatively, regulators could impose remedies that restore integrated firms' incentives to share information with third parties. The policy implications of our research support a recent decision by the UK Antitrust authority in a high-profile case in the digital advertising market.

This paper is a first step towards understanding the interaction between market structure and information design. Further research could consider alternative bargaining protocols. An open question concerns how information design changes when downstream firms collect consumer data. Another question relates to the optimal boundaries of digital conglomerates and how regulation can curb their market power.

Appendix

Proof of Lemma 1. Let $\mathbf{v} \in \text{supp } \bar{\mu}$ with $v^1 > 0$. Without loss of generality, suppose $v_1 \geq ... \geq v_n$.

We first state the Nash equilibrium conditions of the game $G(\mathbf{v})$. A mixed strategy of a firm i is a distribution of prices $\zeta_i \in \Delta\mathbb{R}_+$. For a given strategy profile $(\zeta_1, \ldots, \zeta_n) \in (\Delta\mathbb{R}_+)^n$, let $\zeta_{-i} \in (\Delta\mathbb{R}_+)^{n-1}$ be the profile of the other firms' strategies. If firm i charges price p_i, then its expected revenue is

$$R_i(p_i, \zeta_{-i}) := \prod_{j<i} \zeta_j\Big((v_j - v_i + p_i, \infty)\Big) \prod_{j>i} \zeta_j\Big([\max\{0, v_j - v_i + p_i\}, \infty)\Big)\mathbf{1}_{[0,v_i]}(p_i)p_i.$$

A strategy profile $(\zeta_1, \ldots, \zeta_n)$ is a Nash equilibrium of $G(\mathbf{v})$ if

$$R_i(p_i, \zeta_{-i}) = \sup_p R_i(p, \zeta_{-i}), \quad \zeta_i\text{-a.e. } p_i.$$

Let now $(\zeta_1, \ldots, \zeta_n)$ be a Nash equilibrium. We show first that

$$\inf \text{supp } \zeta_{i'} \geq v^1 - v^2, \quad \forall i' \in \arg\max_i v_i. \tag{A.1}$$

Indeed, if $v_1 = v_2$ then (A.1) holds because prices are nonnegative. Suppose $v_1 > v_2$, that is, $\arg\max_i v_i = \{1\}$, and assume by contradiction that $\inf \text{supp } \zeta_1 < v^1 - v^2$. Then, ζ_1 is not a best response against ζ_{-1} because the consumer would buy from firm 1 with probability one when $p_1 = v_1 - v_2$. This shows (A.1).

We show next that

$$\text{there exists a firm } i' \in \arg\max_i v_i \text{ such that } \zeta_{i'}\Big(\{v^1 - v^2\}\Big) = 1. \tag{A.2}$$

By contradiction, suppose that $\sup \text{supp } \zeta_{i'} > v^1 - v^2$ for all firms $i' \in \arg\max_i v_i$. Then, $R_1(\varepsilon, \zeta_{-1}) > 0$ for small $\varepsilon > 0$, whence

$$\int R_1(p_1, \zeta_{-1}) \mathrm{d}\zeta_1 > 0 \tag{A.3}$$

and thus

$$\prod_{i>1} \zeta_i\Big([\max\{0, v_i - v^1 + \sup \text{supp } \zeta_1\}, \infty)\Big) > 0.$$

For $0 < p < \min\Big\{\sup\operatorname{supp}\zeta_1 - v^1 + v^2, v^2\Big\}$, we have

$$\zeta_1\Big((v^1 - v^2 + p, \infty)\Big)\prod_{i>2}\zeta_i\Big([\max\{0, v_i - v^2 + p\}, \infty\Big) > 0,$$

whence $R_2(p, \zeta_{-2}) > 0$ and thus

$$\int R_2(p_2, \zeta_{-2})\mathrm{d}\zeta_2 > 0.$$

It follows that $\zeta_2\Big([v^2 - v^1 + \sup\operatorname{supp}\zeta_1, \infty)\Big) = 0$. Since $p_1 \mapsto \zeta_2\Big([\max\{0, v^2 - v^1 + p_1\}, \infty)\Big)$ is left-continuous, we thus have

$$\lim_{p_1 \uparrow \sup\operatorname{supp}\zeta_1} R_1(p_1, \zeta_{-1}) = 0.$$

But then $R_1(p_1, \zeta_{-1}) = 0$ for ζ_1-a.e. p_1, contradicting (A.3). Hence, (A.2) holds.

Together, (A.1) and (A.2) imply that Lemma 1 holds true: by (A.2), the consumer buys from a firm with valuation v^1; taking (A.1) into account, the price at which he buys is $v^1 - v^2$. □

Proof of Lemma 2. Let $\mu \in \Delta(\operatorname{supp}\bar{\mu})$. The case that some type occurs with probability one is covered in Lemma 1. If $\mu_1 > 0$, firm 1 sets price w_1 with probability one to the type-1 consumer in every equilibrium. The case that $\mu_0 = 0$ is trivial. If $\mu_2 = 0$, the proof is analogous to the one of Lemma 1. Therefore, suppose that $\mu_0 > 0$ and $\mu_2 > 0$.

Let $\zeta \in \Delta\mathbb{R}_+$ be firm 1's strategy for type 0. Let $\xi \in \Delta\mathbb{R}_+$ be the strategy of firm 2 for all types. If firm 1 charges price p_1 to type 0, then its expected revenue is

$$R_1(p_1, \xi) := \mu_0\xi\Big([\max\{0, w_2 - w_1 + p_1\}, \infty)\Big)\mathbf{1}_{[0,w_1]}(p_1)p_1 + \mu_1 w_1.$$

If firm 2 charges price p_2, then its expected revenue is

$$R_2(p_2, \zeta) := \Big(\mu_0\zeta\Big((w_1 - w_2 + p_2, \infty)\Big) + \mu_2\Big)\mathbf{1}_{[0,w_2]}(p_2)p_2.$$

The tuple (ζ, ξ) is an equilibrium if

$$R_1(p_1, \xi) = \sup_p R_1(p, \xi), \quad \zeta\text{-a.e. } p_1,$$

$$R_2(p_2, \zeta) = \sup_p R_2(p, \zeta), \quad \xi\text{-a.e. } p_2.$$

Firm 2's expected revenue. We show: *if* (ζ, ξ) *is an equilibrium, then*

$$\int R_2(p_2, \zeta) \mathrm{d}\xi(p_2) = \mu_2 w_2.$$

Because $R_2(w_2, \zeta) \geq \mu_2 w_2$, we have

$$\int R_2(p_2, \zeta) \mathrm{d}\xi(p_2) > 0, \tag{A.4}$$

which implies $\inf \operatorname{supp} \xi > 0$. Then, $R_1(\inf \operatorname{supp} \xi, \xi) > \mu_1 w_1$, whence

$$\int R_1(p_1, \xi) \mathrm{d}\zeta(p_1) > \mu_1 w_1. \tag{A.5}$$

By (A.4), it follows that $\zeta\Big([w_1 - w_2 + \sup \operatorname{supp} \xi, \infty)\Big) = 0$ and hence

$$\begin{aligned} \lim_{p_2 \uparrow \sup \operatorname{supp} \xi} R_2(p_2, \zeta) &\leq \lim_{p_2 \uparrow \sup \operatorname{supp} \xi} \Big(\mu_0 \zeta\Big([w_1 - w_2 + p_2, \infty)\Big) + \mu_2\Big) \mathbf{1}_{[0, w_2]}(p_2) p_2 \\ &= \mu_2 \mathbf{1}_{[0, w_2]}(\sup \operatorname{supp} \xi) \sup \operatorname{supp} \xi \\ &\leq \mu_2 w_2. \end{aligned}$$

By $R_2(w_2, \zeta) \geq \mu_2 w_2$, we thus have $R_2(p_2, \zeta) = \mu_2 w_2$ for ξ-a.e. p_2.

Firm 1's expected revenue. We show: *if* (ζ, ξ) *is an equilibrium, then*

$$\int R_1(p_1, \xi) \mathrm{d}\zeta(p_1) = \mu_0 \left(\frac{\mu_2 w_2}{\mu_0 + \mu_2} + w_1 - w_2 \right) + \mu_1 w_1.$$

In equilibrium, (A.4) and (A.5) imply $\inf \operatorname{supp} \zeta = w_1 - w_2 + \inf \operatorname{supp} \xi$ and

$$\zeta\Big(\{w_1 - w_2 + \inf \operatorname{supp} \xi\}\Big) = \xi\Big(\{\inf \operatorname{supp} \xi\}\Big) = 0.$$

Then,

$$\lim_{p_2 \downarrow \inf \operatorname{supp} \xi} R_2(p_2, \zeta) = R_2(\inf \operatorname{supp} \xi, \zeta) = (\mu_0 + \mu_2) \inf \operatorname{supp} \xi,$$

whence $\inf \operatorname{supp} \xi = \mu_2 w_2 / (\mu_0 + \mu_2)$ and thus $\inf \operatorname{supp} \zeta = \mu_2 w_2 / (\mu_0 + \mu_2) + w_1 - w_2$. Hence,

$$\lim_{p_1 \downarrow \inf \operatorname{supp} \zeta} R_1(p_1, \xi) = R_1(\inf \operatorname{supp} \zeta, \xi) = \mu_0 \left(\frac{\mu_2}{\mu_0 + \mu_2} w_2 + w_1 - w_2 \right) + \mu_1 w_1,$$

which shows that

$$R_1(p_1, \xi) = \mu_0 \left(\frac{\mu_2}{\mu_0 + \mu_2} w_2 + w_1 - w_2 \right) + \mu_1 w_1, \quad \zeta\text{-a.e. } p_1.$$

□

Proof of Proposition 1. Let τ be an arbitrary distribution of posterior beliefs that satisfies (5). For $\rho \in [0,1]$, define

$$f(\rho) := (1-\rho)\left(\rho w_2 + w_1 - w_2\right) + \rho w_2.$$

Then,

$$\begin{aligned}\sum_{\mu} \tau(\mu)\Pi_1(\mu) &= \sum_{\mu:\mu_1=1} \tau(\mu)\mu_1 w_1 + \sum_{\mu:\mu_1<1} \tau(\mu)\left(\mu_0\left(\frac{\mu_2 w_2}{\mu_0+\mu_2} + w_1 - w_2\right) + \mu_1 w_1 + \mu_2 w_2\right) \\ &= \bar{\mu}_1 w_1 + (1-\bar{\mu}_1)\sum_{\mu:\mu_1<1} \frac{\tau(\mu)(1-\mu_1)}{1-\bar{\mu}_1} f\left(\frac{\mu_2}{1-\mu_1}\right).\end{aligned}$$

Now note that

$$\sum_{\mu:\mu_1<1} \frac{\tau(\mu)(1-\mu_1)}{1-\bar{\mu}_1} = 1 \quad \text{and} \quad \sum_{\mu:\mu_1<1} \frac{\tau(\mu)(1-\mu_1)}{1-\bar{\mu}_1}\frac{\mu_2}{1-\mu_1} = \frac{\bar{\mu}_2}{1-\bar{\mu}_1}.$$

The function f being strictly concave, we thus obtain

$$\sum_{\mu} \tau(\mu)\Pi_1(\mu) \leq \bar{\mu}_1 w_1 + (1-\bar{\mu}_1) f\left(\frac{\bar{\mu}_2}{1-\bar{\mu}_1}\right) = \Pi_1(\bar{\mu}),$$

where the inequality is strict whenever $\tau(\mu) > 0$ for some μ such that $\mu_1 < 1$ and $\mu_2/(1-\mu_1) \neq \bar{\mu}_2/(1-\bar{\mu}_1)$, that is, whenever the underlying data policy is not uninformative with respect to firm 2's reach. □

Proof of Lemma 3. Let $\mu \in \Delta(\operatorname{supp} \bar{\mu})$. The case that some type occurs with probability one is covered in Lemma 1. If $\mu_2 > 0$, firm 2 sets price w_2 with probability one to the type-2 consumer in every equilibrium. The case that $\mu_0 = 0$ is trivial. If $\mu_1 = 0$, the proof is analogous to the one of Lemma 1. Therefore, suppose that $\mu_0 > 0$ and $\mu_1 > 0$.

Let $\xi \in \Delta\mathbb{R}_+$ be firm 2's strategy for type 0. Let $\zeta \in \Delta\mathbb{R}_+$ be the strategy of firm 1 for all types. If firm 1 charges price p_1, then its expected revenue is

$$R_1(p_1, \xi) := \Big(\mu_0 \xi\Big([\max\{0, w_2 - w_1 + p_1\}, \infty)\Big) + \mu_1\Big)\mathbf{1}_{([0,w_1])}(p_1)p_1.$$

If firm 2 charges price p_2 to type 0, then its expected revenue is

$$R_2(\zeta, p_2) := \mu_0 \zeta\Big((w_1 - w_2 + p_2, \infty)\Big)\mathbf{1}_{[0,w_2]}(p_2)p_2 + \mu_2 w_2.$$

The tuple (ζ, ξ) is an equilibrium if

$$R_1(p_1, \xi) = \sup_p R_1(p, \xi), \quad \zeta\text{-a.e. } p_1,$$
$$R_2(p_2, \zeta) = \sup_p R_2(p, \zeta), \quad \xi\text{-a.e. } p_2.$$

Firm 1's expected revenue. We show: *if* (ζ, ξ) *is an equilibrium, then*

$$\int R_1(p_1, \xi) \mathrm{d}\zeta(p_1) = \max\Big\{\mu_1 w_1, (\mu_0 + \mu_1)(w_1 - w_2)\Big\}. \tag{A.6}$$

We have $R_1(w_1, \xi) \geq \mu_1 w_1$. Moreover, $R_1(w_1 - w_2, \xi) = (\mu_0 + \mu_1)(w_1 - w_2)$, whence $\inf \operatorname{supp} \zeta \geq w_1 - w_2$. If

$$\int R_1(p_1, \xi) \mathrm{d}\zeta(p_1) > (\mu_0 + \mu_1)(w_1 - w_2),$$

then $\sup \operatorname{supp} \zeta > w_1 - w_2$. Suppose this is the case. For $0 < p < \min\Big\{\sup \operatorname{supp} \zeta - w_1 + w_2, w_2\Big\}$, we then have $R_2(p, \zeta) > \mu_2 w_2$, whence

$$\int R_2(p_2, \zeta) \mathrm{d}\xi(p_2) > \mu_2 w_2. \tag{A.7}$$

It follows that $\xi\Big((w_2 - w_1 + \sup \operatorname{supp} \zeta, \infty)\Big) = 0$ and hence

$$\begin{aligned} \lim_{p_1 \uparrow \sup \operatorname{supp} \zeta} R_1(p_1, \xi) &= R_1(\sup \operatorname{supp} \zeta, \xi) \\ &= \mu_1 \mathbf{1}_{[0, w_1]}(\sup \operatorname{supp} \zeta) \sup \operatorname{supp} \zeta \\ &\leq \mu_1 w_1. \end{aligned}$$

By $R_1(w_1, \xi) \geq \mu_1 w_1$, we thus have $R_1(p_1, \xi) = \mu_1 w_1$ for ζ-a.e. p_1. This shows (A.6).

Firm 2's expected revenue: We show: *if* (ζ, ξ) *is an equilibrium, then*

$$\int R_2(\zeta, p_2) \mathrm{d}\xi(p_2) = \max\left\{\mu_0 \left(\frac{\mu_1 w_1}{\mu_0 + \mu_1} - (w_1 - w_2)\right), 0\right\} + \mu_2 w_2. \tag{A.8}$$

The proof of (A.6) showed that if $(\mu_0 + \mu_1)(w_1 - w_2) \geq \mu_1 w_1$, then $\zeta\Big(\{w_1 - w_2\}\Big) = 1$. Because $R_2(w_2, \zeta) \geq \mu_2 w_2$, it follows that then

$$\int R_2(\zeta, p_2) \mathrm{d}\xi(p_2) = \mu_2 w_2.$$

If $(\mu_0+\mu_1)(w_1-w_2)<\mu_1 w_1$, (A.6) and (A.7) imply $\inf\operatorname{supp}\zeta = w_1-w_2+\inf\operatorname{supp}\xi$ and

$$\zeta\Big(\{w_1-w_2+\inf\operatorname{supp}\xi\}\Big)=\xi\Big(\{\inf\operatorname{supp}\xi\}\Big)=0.$$

Then,

$$\lim_{p_1\downarrow\inf\operatorname{supp}\zeta} R_1(p_1,\xi)=R_1(\inf\operatorname{supp}\zeta,\xi)=(\mu_0+\mu_1)\inf\operatorname{supp}\zeta,$$

whence $\inf\operatorname{supp}\zeta=\mu_1 w_1/(\mu_0+\mu_1)$ and thus $\inf\operatorname{supp}\xi=\mu_1 w_1/(\mu_0+\mu_1)-(w_1-w_2)$. Hence,

$$\lim_{p_2\downarrow\inf\operatorname{supp}\zeta} R_2(p_2,\zeta)=R_2(\inf\operatorname{supp}\xi,\zeta)=\mu_0\left(\frac{\mu_1}{\mu_0+\mu_1}w_1-(w_1-w_2)\right)+\mu_2 w_2,$$

which shows that

$$R_2(p_2,\zeta)=\mu_0\left(\frac{\mu_1}{\mu_0+\mu_1}w_1-(w_1-w_2)\right)+\mu_2 w_2,\quad \xi\text{-a.e. } p_2.$$

□

Proof of Proposition 2. Let τ be an arbitrary distribution of posterior beliefs that satisfies (5). For $\rho\in[0,1]$, define

$$f(\rho):=\max\Big\{(1-\rho)\Big(\rho w_1-(w_1-w_2)\Big)+\rho w_1,\ w_1-w_2\Big\}$$

As in the proof of Proposition 1,

$$\sum_\mu \tau(\mu)\Pi_2(\mu)=\bar\mu_2 w_2+(1-\bar\mu_2)\sum_{\mu:\mu_2<1}\frac{\tau(\mu)(1-\mu_2)}{1-\bar\mu_2}f\left(\frac{\mu_1}{1-\mu_2}\right),$$

where

$$\sum_{\mu:\mu_2<1}\frac{\tau(\mu)(1-\mu_2)}{1-\bar\mu_2}=1\quad\text{and}\quad\sum_{\mu:\mu_2<1}\frac{\tau(\mu)(1-\mu_2)}{1-\bar\mu_2}\frac{\mu_1}{1-\mu_2}=\frac{\bar\mu_1}{1-\bar\mu_2}.$$

Thus, using the notation $\bar\rho:=\bar\mu_1/(1-\bar\mu_2)$,

$$\sum_\mu \tau(\mu)\Pi_2(\mu)\le\bar\mu_2 w_2+(1-\bar\mu_2)\left(\max_{\hat\tau\in\Delta[0,1]}\sum_\rho\hat\tau(\rho)f(\rho)\ \text{ s.t }\ \sum_\rho\hat\tau(\rho)\rho=\bar\rho\right).$$

By standard arguments (see Kamenica and Gentzkow, 2011), there exists $\lambda,\rho',\rho''\in[0,1]$ such that $\lambda\rho'+(1-\lambda)\rho''=\bar\rho$ and

$$\sum_\rho\hat\tau(\rho)f(\rho)\le\lambda f(\rho')+(1-\lambda)f(\rho'')\quad\text{for all }\hat\tau\in\Delta[0,1]\text{ s.t. }\sum_\rho\hat\tau(\rho)\rho=\bar\rho;$$

we may set $\rho'' = 0$, and ρ' such that $f(\rho') = (1-\rho')\Big(\rho' w_1 - (w_1 - w_2)\Big) + \rho' w_1$, implying

$$\lambda f(\rho') + (1-\lambda) f(\rho'') = \lambda \left(\left(1 - \frac{\bar{\rho}}{\lambda}\right) \left(\frac{\bar{\rho}}{\lambda} w_1 - (w_1 - w_2) \right) + \frac{\bar{\rho}}{\lambda} w_1 \right) + (1-\lambda)(w_1 - w_2).$$

Optimal data policies can thus be derived from the solutions to

$$\max_{\lambda \in [\bar{\rho}, 1]} \left(2 - \frac{\bar{\rho}}{\lambda}\right) \bar{\rho} w_1 + (1 + \bar{\rho} - 2\lambda)(w_1 - w_2).$$

The derivative of the objective function with respect to λ is

$$\frac{\bar{\rho}^2}{\lambda^2} w_1 - 2(w_1 - w_2).$$

The objective function is strictly concave. Hence, the optimal λ is unique; denote it by λ^*. If $w_1 \leq 2(w_1 - w_2) \iff w_1 \geq 2w_2$, then $\lambda^* = \bar{\rho}$ (fully informative). If $\bar{\rho} \geq \sqrt{2(w_1 - w_2)/w_1}$, then $\lambda^* = 1$ (uninformative). Finally, if $w_1 < 2w_2$ and $\bar{\rho} < \sqrt{2(w_1 - w_2)/w_1}$, then $\lambda^* = \bar{\rho}\sqrt{w_1/(2(w_1 - w_2))} \in (0,1)$ (partially informative). □

Proof of Lemma 4. Working by backward induction, suppose that both firms rejected the respective takeover bid that the DI has made to them. Thus, the DI stays independent. In the equilibrium of this subgame, if firm i accepts contract $\Big((M_i, \psi_i), t_i\Big)$, then the fee t_i must leave the firm indifferent between accepting the contract and rejecting it. If the strong firm rejects its contract, it can secure itself an expected revenue of at least

$$c_1 := \max\Big\{\bar{\mu}_1 w_1, (\bar{\mu}_0 + \bar{\mu}_1)(w_1 - w_2)\Big\}$$

by charging price $p_1 = w_1$ or $p_1 = w_1 - w_2$, respectively. Similarly, if the weak firm rejects, it can secure itself an expected revenue of at least

$$c_2 := \bar{\mu}_2 w_2$$

by charging price $p_2 = w_2$. Thus, c_i is a lower bound on firm i's expected payoff in any equilibrium of the subgame.

Consider next a subgame that starts after one firm rejected the takeover bid that the DI has made. Let $i \in \{1, 2\}$ be the firm that receives the next bid. If it rejects and the DI stays independent, let x_1 and x_2 be the resulting expected payoffs of the strong and the weak firm, respectively, where $x_1 \geq c_1$ and $x_2 \geq c_2$ as shown above. Acquiring firm i results

in industry profit Π_i^*. In Subsections 5.1 and 5.2, we saw that the integrated firm firm i sells firm $j \neq i$ a data policy such that firm j's expected payoff is equal to its expected revenue in the ancillary game $G(i, \bar{\mu})$. By Lemmas 2 and 3, this expected revenue is equal to c_j. Hence, forming an integrated firm with firm i results for the DI in an expected payoff of $\Pi_i^* - c_j$. By Theorem 1, staying independent results in industry profit Π^f, and hence for the DI in an expected payoff of $\Pi^f - x_1 - x_2$. We have

$$\Pi_i^* \geq \Pi^f$$

because the integrated firm can sell firm j a fully informative data policy. Moreover

$$c_j \leq x_j.$$

If any of these two inequalities is strict, then there exists $b > x_i$ such that

$$\Pi_i^* - c_j - b > \Pi^f - x_1 - x_2,$$

whence in any equilibrium of the subgame, the DI acquires firm i.

Consider now the beginning of the game, where the DI makes the first takeover bid. Let $i \in \{1, 2\}$ be the firm that receives the bid. If it rejects, then the DI will subsequently either acquire firm $j \neq i$, in which case firm i's expected payoff is c_i, or the DI will remain independent, in which case firm i's expected payoff is also $c_i = x_i$ as shown above.

We conclude that if the DI first makes a takeover bid to firm $i \in \{1, 2\}$ and this bid is strictly smaller than c_i, then the bid will be rejected; if the bid is equal to $c_i + \varepsilon$ for any $\varepsilon > 0$, by contrast, then the bid will be accepted, resulting for the DI in a payoff of

$$\Pi_i^* - c_1 - c_2 - \varepsilon.$$

This shows that if the DI acquires the strong firm in equilibrium, then $\Pi_1^* \geq \Pi_2^*$, and if the DI acquires the weak firm, then $\Pi_1^* \leq \Pi_2^*$. Finally, the DI always acquires a firm in equilibrium because

$$\begin{aligned} \Pi_1^* &= \bar{\mu}_0 \left(\frac{\bar{\mu}_2 w_2}{\bar{\mu}_0 + \bar{\mu}_2} + w_1 - w_2 \right) + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2 \\ &> \bar{\mu}_0 (w_1 - w_2) + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2 \\ &= \Pi^f. \end{aligned}$$

□

Proof of Proposition 3. Fix $w_1 < 2w_2$ and $\bar{\mu}_0 \in (0,1)$. By Lemma 2 and Proposition 1,

$$\begin{aligned}\Pi_1^* &= \bar{\mu}_0 \left(\frac{\bar{\mu}_2 w_2}{\bar{\mu}_0 + \bar{\mu}_2} + w_1 - w_2 \right) + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2 \\ &= (1-\bar{\mu}_2) w_2 - \frac{\bar{\mu}_0^2 w_2}{1-\bar{\mu}_1} + \bar{\mu}_2 w_2 + (1-\bar{\mu}_2)(w_1 - w_2).\end{aligned}$$

According to the proof of Proposition 2,

$$\begin{aligned}\Pi_2^* &= \bar{\mu}_2 w_2 + (1-\bar{\mu}_2) \left(\left(2 - \frac{\bar{\rho}}{\lambda^*} \right) \bar{\rho} w_1 + (1 + \bar{\rho} - 2\lambda^*)(w_1 - w_2) \right) \\ &= (1-\bar{\mu}_2) \left(\left(2 - \frac{\bar{\rho}}{\lambda^*} \right) \bar{\rho} w_1 + (\bar{\rho} - 2\lambda^*)(w_1 - w_2) \right) + \bar{\mu}_2 w_2 + (1-\bar{\mu}_2)(w_1 - w_2),\end{aligned}$$

where $\bar{\rho} := \bar{\mu}_1/(1-\bar{\mu}_2)$ and

$$\lambda^* := \begin{cases} 1 & \text{if } \bar{\rho} \geq \sqrt{2(w_1 - w_2)/w_1}, \\ \bar{\rho}\sqrt{\frac{w_1}{2(w_1 - w_2)}} & \text{if } \bar{\rho} < \sqrt{2(w_1 - w_2)/w_1}. \end{cases}$$

Thus, if $\bar{\rho} \geq \sqrt{2(w_1 - w_2)/w_1}$, then

$$\Pi_2 = (1-\bar{\mu}_2) w_2 - \mu_0 (w_1 - w_2) - \frac{\bar{\mu}_0^2 w_1}{1-\bar{\mu}_2} + \bar{\mu}_2 w_2 + (1-\bar{\mu}_2)(w_1 - w_2),$$

and if $\bar{\rho} < \sqrt{2(w_1 - w_2)/w_1}$, then

$$\Pi_2^* = \bar{\mu}_1 \left(3w_1 - w_2 - 2\sqrt{2w_1(w_1 - w_2)} \right) + \bar{\mu}_2 w_2 + (1-\bar{\mu}_2)(w_1 - w_2).$$

Define the functions f and g on $(0, 1-\bar{\mu}_0)$ as follows:

$$f(s) := (1-s) w_2 - \frac{\bar{\mu}_0^2 w_2}{\bar{\mu}_0 + s}, \quad \forall s \in (0, 1-\bar{\mu}_0);$$

if $(1 - \bar{\mu}_0 - s)/(1-s) \geq \sqrt{2(w_1 - w_2)/w_1}$, then

$$g(s) := (1-s) w_2 - \mu_0 (w_1 - w_2) - \frac{\bar{\mu}_0^2 w_1}{1-s} := g_1(s),$$

and if $(1 - \bar{\mu}_0 - s)/(1-s) < \sqrt{2(w_1 - w_2)/w_1}$, then

$$g(s) := (1 - \bar{\mu}_0 - s) \left(3w_1 - w_2 - 2\sqrt{2w_1(w_1 - w_2)} \right) := g_2(s).$$

Fix an arbitrary $s \in (0, 1-\bar{\mu}_0)$ and an arbitrary $s' \in (s, 1-\bar{\mu}_0)$, and suppose that $f(s) \geq g(s)$. To prove the proposition, we must show that $f(s') > g(s')$. There are three possible

cases. **Case 1:** $g(s) = g_1(s)$ **and** $g(s') = g_1(s)$. Then $f(s') > g(s')$, as is routine to verify. **Case 2:** $g(s) = g_2(s)$ **and** $g(s') = g_2(s')$. Then, $f(s') > g(s')$ because $3w_1 - w_2 - 2\sqrt{2w_1(w_1 - w_2)} > w_2$; to see the latter, observe that for every $w_2 > 0$, the function $w_1 \mapsto 3w_1 - w_2 - 2\sqrt{2w_1(w_1 - w_2)}$ on (w_2, ∞) is strictly convex and minimized at $w_1 = 2w_2$, where the value is equal to w_2. **Case 3:** $g(s) = g_1(s)$ **and** $g(s') = g_2(s')$. Then, there exists $s'' \in (s, s')$ such that $g_1(s'') = g_2(s'')$. As in Case 1, $f(s'') > g_1(s'')$; applying the logic from Case 2, it follows that $f(s') > g_2(s')$. □

Proof of Proposition 4. Consider any equilibrium, and let C be the consumer surplus. By Propositions 2 and 1, the industry profit is equal to

$$\Pi_1(\bar{\mu}) = \bar{\mu}_0 \left(\frac{\bar{\mu}_2 w_2}{\bar{\mu}_0 + \bar{\mu}_2} + w_1 - w_2 \right) + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2.$$

Because

$$\Pi^f = \bar{\mu}_0(w_1 - w_2) + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2 < \Pi_1(\bar{\mu})$$

whereas

$$C + \Pi_1(\bar{\mu}) \leq \bar{\mu}_0 w_1 + \bar{\mu}_1 w_1 + \bar{\mu}_2 w_2 = C^f + \Pi^f,$$

we have $C < C^f$. Next, let $w_1 > w_2$, and suppose by contradiction that the equilibrium is efficient. Then, the type-0 consumer buys from the integrated firm with probability one. Hence, firm 2 sets price $p_2 = w_2$ with probability one, implying that the integrated firm charges $p_1 = w_1$ to the type-0 consumer with probability one. But then firm 2's pricing is not a best response; a contradiction. Thus, every equilibrium is inefficient. □

Proof of Proposition 5. If $w_1 \geq 2w_2$, then in equilibrium the integrated firm sells a data policy that is fully informative with respect to the strong firm's reach (Proposition 2). By Lemma 1, this implies that every equilibrium is efficient and generates consumer surplus C^f.

If $w_1 < 2w_2$, then in equilibrium the integrated firm does not sell a data policy that is fully informative with respect to the strong firm's reach. Because the data policy maximizes industry profit, it follows that every equilibrium generates industry profit that is strictly larger than Π^f. The same argument as in the proof of Proposition 4 shows that the consumer surplus must then be strictly smaller than C^f. The inefficiency of equilibrium follows also as in that proof. □

Proof of Proposition 6. In what follows, we prove the two claims in the proposition:

a) With independent data intermediation, the weak firm's expected revenue is $\bar{\mu}_2 w_2$ by Lemma 1 and Theorem 1. With integrated data intermediation, when the integrated firm is the strong firm, the weak firm's expected revenue is $\bar{\mu}_2 w_2$ as well by Lemma 2 and Proposition 1.

b) With independent data intermediation, the strong firm's expected revenue is $\bar{\mu}_1 w_1 + \bar{\mu}_0(w_1 - w_2)$ by Lemma 1 and Theorem 1. With integrated data intermediation, when the integrated firm is the weak firm, the strong firm's expected revenue is

$$\sum_{\mu} \tau(\mu) \max\Big\{\mu_1 w_1,\ (\mu_0 + \mu_1)(w_1 - w_2)\Big\},$$

where the distribution of posteriors $\tau \in \Delta\Delta(\text{supp } \bar{\mu})$ solves (P_2). Now, if the data policy that induces τ is fully informative concerning the strong firm's reach, then

$$\sum_{\mu} \tau(\mu) \max\Big\{\mu_1 w_1,\ (\mu_0 + \mu_1)(w_1 - w_2)\Big\} = \bar{\mu}_1 w_1 + \bar{\mu}_0(w_1 - w_2).$$

Otherwise, it holds that

$$\sum_{\mu} \tau(\mu) \max\Big\{\mu_1 w_1,\ (\mu_0 + \mu_1)(w_1 - w_2)\Big\} < \bar{\mu}_1 w_1 + \bar{\mu}_0(w_1 - w_2),$$

since for every $\lambda \in (0,1)$ and $\mu \in \Delta(\text{supp } \bar{\mu})$ with $\mu_2 < 1$ and $(\mu_0, \mu_1) = \lambda(1 - \mu_2, 0) + (1 - \lambda)(0, 1 - \mu_2)$, we have

$$\begin{aligned}(1 - \lambda)(1 - \mu_2) \max\Big\{w_1,\ (w_1 - w_2)\Big\} + \lambda(1 - \mu_2) \max\Big\{0,\ (w_1 - w_2)\Big\} \\ > \max\Big\{(1 - \lambda)(1 - \mu_2) w_1,\ (1 - \mu_2)(w_1 - w_2)\Big\}.\end{aligned}$$

□

Proof of Theorem 3. Fix an arbitrary equilibrium. By Lemma 4, the industry profit is at least as large as $\Pi_1^* = \Pi_1(\bar{\mu})$, the industry profit when the DI is integrated with firm 1. As in the proof of Proposition 4, it follows that the consumer surplus is strictly smaller than C^f. By Propositions 4 and 5, this implies that the equilibrium is furthermore inefficient. □

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022): "Too Much Data: Prices and Inefficiencies in Data Markets," *American Economic Journal: Microeconomics*, 14(4), 218–56.

Alcobendas, M., S. J. Kobayashi, K. Shi, and M. Shum (2023): "The Impact of Privacy Protection on Online Advertising Markets," mimeo.

Anderson, S., and Bedre-Defolie (2023): "Hybrid Platform Model: Monopolistic Competition and a Dominant Firm," *RAND Journal of Economics*.

Argenziano, R., and A. Bonatti (2023): "Data Linkages and Privacy Regulation," mimeo.

Armstrong, M., and J. Vickers (2019): "Discriminating Against Captive Consumers," *American Economic Review: Insights*, 1(3), 257–272.

Bergemann, D., and A. Bonatti (2019): "Markets for Information: An Introduction," *Annual Review of Economics*, 11, 85–107.

——— (forthcoming): "Data, Competition, and Digital Platforms," *American Economic Review*.

Bergemann, D., A. Bonatti, and T. Gan (2022): "The Economics of Social Data," *RAND Journal of Economics*, 53(2), 263–296.

Bergemann, D., B. Brooks, and S. Morris (2021): "Search, Information, and Prices," *Journal of Political Economy*, 129(8), 2275–2319.

Bimpikis, K., D. Crapis, and A. Tahbaz-Salehi (2019): "Information Sale and Competition," *Management Science*, 65(6), 2646–2664.

Bonatti, A., M. Dahleh, T. Horel, and A. Nouripour (2024): "Selling Information in Competitive Environments," *Journal of Economic Theory*, 216, 105779.

Bounie, D., A. Dubus, and P. Waelbroeck (2021): "Selling Strategic Information in Digital Competitive Markets," *RAND Journal of Economics*, 52(2), 283–313.

——— (2022): "Competition and Mergers with Strategic Data Intermediaries," *mimeo*.

Bourreau, M., A. de Streel, and I. Graef (2017): "Big Data and Competition Policy: Market Power, Personalised Pricing and Advertising," .

Braulin, F. C., and T. Valletti (2016): "Selling Customer Information to Competing Firms," *Economics Letters*, 149, 10–14.

Celis, L. E., G. Lewis, M. Mobius, and H. Nazerzadeh (2014): "Buy-It-Now or Take-a-Chance: Price Discrimination Through Randomized Auctions," *Management Science*, 60(12), 2927–2948.

Clarke, R. (1983): "Collusion and the Incentives for Information Sharing," *Bell Journal of Economics*, 14(2), 383–394.

CMA (2022): "Decision to accept commitments offered by Google in relation to its Privacy Sandbox Proposalss," Case number 50972, Competition and Markets Authority (CMA).

EC (2019): "Google Search (AdSense)," Case AT.40411, European Commission (EC).

——— (2020): "Google/Fitbit," Case M.9660, European Commission (EC).

Elliott, M., A. Galeotti, A. Koh, and W. Li (2022): "Market Segmentation Through Information," mimeo.

Engelbrecht-Wiggans, R., P. R. Milgrom, and R. J. Weber (1983): "Competitive Bidding and Proprietary Information," *Journal of Mathematical Economics*, 11, 161–169.

Gomes, R., and A. Pavan (2016): "Many-to-many matching and price discrimination," *Theoretical Economics*, 11(3), 1005–1052.

Hart, O., and J. Tirole (1990): "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity*, 21(1990 Microeconomics), 205–286.

Hidir, S., and N. Vellodi (2021): "Privacy, Personalization, and Price Discrimination," *Journal of the European Economic Association*, 19(2), 1342–1363.

Ichihashi, S. (2021a): "Competing Data Intermediaries," *RAND Journal of Economics*, 52(3), 515–537.

——— (2021b): "The economics of data externalities," *Journal of Economic Theory*, 196(C).

Jullien, B., and A. Pavan (2018): "Information Management and Pricing in Platform Markets," *The Review of Economic Studies*, 86(4), 1666–1703.

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Karle, H., M. Peitz, and M. Reisinger (2020): "Segmentation versus Agglomeration: Competition between Platforms with Competitive Sellers," *Journal of Political Economy*, 128(6), 2329–2374.

Lee, T. K. (1984): "Incomplete Information, High-Low Bidding and Public Information in First Price Auctions," *Management Science*, 30(12), 1490–1496.

McAfee, R. P., and M. Schwartz (1994): "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity," *American Economic Review*, 84(1), 210–230.

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.

Motta, M. (2023): "Self-Preferencing and Foreclosure in Digital Markets: Theories of Harm for Abuse Cases," *International Journal of Industrial Organization*, 90, 102974.

Narasimhan, C. (1988): "Competitive Promotional Strategies," *Journal of Business*, 61(4), 427–449.

Nocke, V., and P. Rey (2018): "Exclusive Dealing and Vertical Integration in Interlocking Relationships," *Journal of Economic Theory*, 177, 183–221.

Novshek, W., and H. Sonnenschein (1982): "Fulfilled Expectations Cournot Duopoly with Information Acquisition and Release," *Bell Journal of Economics*, 13(1), 214–218.

O'Brien, D. P., and G. Shaffer (1992): "Vertical Control with Bilateral Contracts," *RAND Journal of Economics*, 23(3), 299–308.

OECD (2023): "Theories of Harm for Digital Mergers," OECD Competition Policy Roundtable.

Pigou, A. (1920): *The Economics of Welfare*. Macmillan, New York, NY.

Raith, M. (1996): "A General Model of Information Sharing in Oligopoly," *Journal of Economic Theory*, 71(1), 260–288.

Reisinger, M., and E. Tarantino (2015): "Vertical Integration, Foreclosure, and Productive Efficiency," *RAND Journal of Economics*, 46(3), 461–479.

Rey, P., and J. Tirole (2007): "A Primer on Foreclosure," in *Handbook of Industrial Organization*, ed. by M. Armstrong, and R. Porter, vol. 3, chap. 33, pp. 2145–2220. Elsevier.

Rey, P., and T. Vergé (2004): "Bilateral Control with Vertical Contracts," *RAND Journal of Economics*, 35(4), 728–746.

Rhodes, A., and J. Zhou (2024): "Personalized Pricing and Competition," *American Economic Review*, 114(7), 2141–70.

Rodríguez Olivera, R. (2024): "Strategic Incentives and the Optimal Sale of Information," *American Economic Journal: Microeconomics*, 16(2), 296–353.

Segal, I. (1999): "Contracting with Externalities," *Quartely Journal of Economics*, 114(2), 337–388.

Shi, X., and J. Zhang (2022): "Welfare of Price Discrimination and Market Segmentation in Duopoly," mimeo.

Taylor, C. R. (2004): "Consumer Privacy and the Market for Customer Information," *The RAND Journal of Economics*, 35(4), 631–650.

Terstiege, S., and A. Vigier (2023): "Market Segmentation and Product Steering," mimeo.

Valenzuela-Stookey, Q. (2020): "Platform-mediated Competition," mimeo.

Varian, H. R. (1980): "A Model of Sales," *American Economic Review*, 70(4), 651–659.

Vives, X. (1984): "Duopoly Information Equilibrium: Cournot and Bertrand," *Journal of Economic Theory*, 34(1), 71–94.

Wilson, R. B. (1967): "Competitive Bidding with Asymmetric Information," *Management Science*, 13(11), 816–820.

Yang, K. H. (2022): "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and Its Consequences," *American Economic Review*, 112(4), 1364–1394.