

Selling Data to a Competitor

Extended Abstract

Ronen Gradwohl

Department of Economics and Business Administration
Ariel University
roneng@ariel.ac.il

Moshe Tennenholtz

Faculty of Data and Decision Sciences
The Technion – Israel Institute of Technology
moshet@ie.technion.ac.il

We study the costs and benefits of selling data to a competitor. Although selling all consumers' data may decrease total firm profits, there exist other selling mechanisms—in which only some consumers' data is sold—that render both firms better off. We identify the profit-maximizing mechanism, and show that the benefit to firms comes at a cost to consumers. We then construct Pareto-improving mechanisms, in which each consumers' welfare, as well as both firms' profits, increase. Finally, we show that consumer opt-in can serve as an instrument to induce firms to choose a Pareto-improving mechanism over a profit-maximizing one.

1 Introduction

In recent years, it has become common wisdom that data is a dominant source of power. This power is perhaps most clearly illustrated in markets where an incumbent with access to consumer data competes with an entrant who does not have such data. As stated in a crisp manner by [25], common wisdom holds that the incumbent's key advantage is data superiority: "If you run a market-leading company, you should never be blindsided by an invader. Locked within your own records is a huge, largely untapped asset that no attacker can hope to match: what we call the incumbent's advantage." The situation is not uncommon: In our data-driven economy, competing firms often find themselves in asymmetric situations where one of them has superior or even exclusive access to relevant data.

Such data asymmetry has become a major issue for debate. For example, in a June 2021 press release, the EU declared that it has opened an antitrust investigation that will "examine whether Google is distorting competition by restricting access by third parties to user data for advertising purposes on websites and apps, while reserving such data for its own use," [13]. One of the issues in such debates is the question of data sharing: Should the incumbent share its data in order to increase market competition and consumer welfare? And can the incumbent profitably sell its data to the entrant?

Strategic decisions about the sale of data to competitors appear in the online economy frequently, although they are not always stated explicitly. For example, when an advertiser buys a sponsored-search campaign through an ad exchange, the advertiser obtains useful information about a segment of consumers as part of the ad exchange reports. The advertiser might then use this information when bidding directly for display ads on other platforms, including platforms on which the ad exchange is also a competitor. The data-holder (e.g., ad exchange) thus faces a strategic decision about which consumer segments to sell to a competitor, and at what prices. The data-buyer (e.g., advertiser), in turn, must decide whether to pay the price and obtain data about these consumer segments, or whether to enter into competition without the data on offer.

These strategic considerations raise numerous questions: What are the data-holder's costs and benefits from selling data to a data-buyer? What are the effects of data sale on consumer welfare? And,

in a regulated market, can data sale be regulated in a way that leads to Pareto-improving transactions, benefitting consumers as well as firms?

In this paper we study these questions in a paradigmatic model of imperfect competition between two firms who have asymmetric access to data. We consider the classic Hotelling model of imperfect competition: There are two firms, each located at a different endpoint of a unit interval, with a unit mass of consumers distributed across this interval. We model data about a consumer as information about the consumer's location on the interval. In the classic model, neither firm has data about any consumer, and so firms engage in competition via uniform prices that each offers to all consumers. In our variation of this model, in contrast, one firm is a data-holder who knows the locations of all consumers, whereas the other firm is a data-buyer who has no such data. The data-holder can use its data advantage in order to personalize prices to consumers, and can thus sometimes undercut the data-buyer's uniform price.

In order to study the costs and benefits of data sale to a competitor, we suppose the firms engage in a data-sharing mechanism. Such a mechanism consists of a segment of consumers whose data the data-holder shares with the data-buyer, as well as a price the data-buyer pays the data-holder. After engaging in such a mechanism, the data-buyer will hold location data about all consumers in the shared segment, allowing that firm to also personalize prices to them.

Within this model, we first show that full data-sharing, in which the data-holder shares all its data with the data-buyer, is harmful to the firms. We then show that there exist other data-sharing mechanisms—in which only some consumers' data is shared—that increase both firms' profits. In fact, we identify the mechanism that maximizes total firm profits. This last mechanism, however, increases firm profits at the expense of consumers. We thus proceed to show that there exist Pareto-improving mechanisms, in which each consumers' welfare, as well as both firms' profits, increase. Finally, we consider the question of how a regulator can induce firms to utilize a Pareto-improving mechanism rather than a profit-maximizing one that may harm consumers. We show that consumer opt-in may serve as such an instrument: If consumers are given the opportunity to opt-in to having their data sold, and if the data-holder is only permitted to share data about consumers who have opted-in, then in equilibrium the firms will choose a Pareto-improving mechanism.

Our results are driven by two forces, which we identify as the direct effect and the indirect effect of data sharing. The direct effect is the following: if the data-holder shares data about a particular consumer, then the data-buyer can now offer that consumer a personalized rather than the uniform price. This affects both firms' equilibrium personalized prices to that consumer, and may thus impact profits and welfare. The indirect effect of data sharing, on the other hand, is the following: by sharing data about a segment of consumers, the data-holder changes the set of consumers to whom the data-buyer's uniform price applies (since additional consumers will now be offered personalized prices). And since the uniform price is determined in equilibrium in part by the locations of consumers to whom that price will apply, a change in the set of consumers may effect a change in the equilibrium uniform price, thereby affecting profits and welfare. Our results highlight how the interplay between the direct and indirect effects of data sharing lead to changes in firms' profits and consumers' welfare.

In addition to identifying the two effects of data sharing, our analysis generates several general insights. First, and perhaps surprisingly, selling data to a competitor can be strictly beneficial to both firms.¹ Second, data can be sold in a way that is Pareto improving. And finally, such Pareto-improving data-sale can be induced by consumer opt-in regulation.

We note that the idea of selling data to a competitor has been advocated in financial markets (see, for example, [2]). In that context, the possibility of data sale allows a decision maker to choose between

¹We note that this holds even if data is sold at no cost—see Proposition 2.

taking investment risks or obtaining direct monetary rewards. The incredibly fast-growing data-economy, in which some firms hold massive amounts of data about consumers, raises calls to consider such data sale in a broader context: Can it lead to increased profits to both data-holders and data-buyers? And can it benefit all of society, including consumers whose data is exchanged?

Related literature This paper is part of a large and growing literature on data markets (see, e.g., the survey of [5]). Work in this area focuses on related but orthogonal questions, such as the effects of data-sale by a third-party data-provider and of information sharing between competitors. Our paper bridges these strands by considering data sale to a competitor. To the best of our knowledge, the work of [2] is the only other paper to study such a scenario, and ours is the first to focus on the effects of such data sale on firm profits as well as consumer welfare.

The literature on the sale of data by a data provider (e.g., [1, 6, 26, 29, 34] and others) studies how a third-party data-provider can maximize profits by selling data to a monopolist or to competing firms who use this data to price discriminate. Within this literature, one paper that is closely related to ours is that of [12]. [12] consider an information designer who provides consumer information to oligopolists, and characterize the different market outcomes that can be achieved by the designer. Our paper differs from this research in that we suppose data is not held by a third-party, but rather by one of the competing firms. This firm may sell data to its direct competitor, affecting both firms' respective market positions.

A different but related setting is that of [3], where the consumers are holders of information who may share it with one or both firms so as to intensify competition. The model and results of [3] bear some similarity to ours. For example, they also consider a Hotelling model, and show that consumers are better off whenever those sufficiently closer to one firm than another share their location with that firm, and those closer to the middle share their location with both firms. Despite the similarities, our paper studies an orthogonal question, as we assume one of the firms already has data about consumers, and focus on whether that firm will sell data to its competitor. In contrast to the model of [3], in which each consumer chooses which firm has access to that consumer's location, in our model the informed firm strategically chooses whether or not to share this information. Under consumer opt-in the role of consumers is in determining whether such sharing could potentially take place, but not in whether it actually takes place. Finally, while [3] show that, consumers are always better off when they share some of their information, we show that when firms choose what information to share this may no longer be the case.

Because our work considers the sale of data from one firm to another, it is related to the literature on information sharing. Although information sharing between firms has been studied in a variety of settings,² our paper is most-closely related to that of competitive price discrimination—see, for example, the surveys of [31] and [14]. One of the main insights from this literature is that when firms have more data about consumers, competition between them is more intense, leading to lower prices and profits. In our paper, in contrast, data is sold by one firm to another in such a way as to increase profits.

Two papers that specifically analyze the effects of data sharing within a Hotelling model are [24] and [7]. [24] study a model in which each of two firms may have data both about consumers' locations and about their transportation costs, and consider the eight permutations in which each firm may have either a dataset about locations, a dataset about transportation costs, both datasets, or neither datasets. They then analyze the market effects of firms sharing one or both of their (full) datasets with each other, and provide conditions under which sharing is beneficial to the firms. [7] studies a Hotelling model in which locations are two-dimensional, and firms hold all data about one dimension, both dimensions, or neither dimension. He analyzes the various scenarios in terms of firm profits and consumer welfare, with

²These include oligopolistic competition [11, 28], financial intermediation [27, 23, 15], supply chain management [21, 30], competition between data brokers [20, 22], and advertising [18].

a particular emphasis on the comparison to the regimes of full privacy (neither firm has any data) and no privacy (both firms have full data). Interestingly, [7] shows that total firm profits are hump-shaped in the amount of information they hold; for example, the scenario in which each firm holds data about a different dimension yields higher profits than both full privacy and no privacy. Our work differs from both of these papers in that we study the sale of partial data from one firm to another, with an emphasis on mutually increasing profits.

In terms of modeling, our paper is most closely related to [26] and [17]. [26] consider a one-dimensional Hotelling model in which consumers' locations may be known to one, both, or neither firm. Their concern is not the sale of data from one firm to another, but rather the optimal strategy of a data broker who sells the data to the firms. They also consider the effects of a consumer-side technology that allows consumers the ability to protect their privacy. [17] also study a Hotelling model, but suppose that both firms have some data about consumers. Their main focus is on various forms of mutual data sharing between the firms.

2 The Model

We focus on a standard Hotelling model, in which a unit mass of consumers is spread over the unit interval according to an atomless distribution F with continuous, strictly positive density f that has full support. There are two firms: firm A is located at $\theta_A = 0$, and firm B is located at $\theta_B = 1$. Each consumer chooses at most one firm from which to purchase a good. Consumers derive value v from the good, but pay two costs: the price, and a linear transportation cost that scales with the distance between the consumer and the firm providing the good. Thus, a consumer located at θ who buys from firm i at price p_i obtains utility $v - p_i - t|\theta - \theta_i|$, where t is the marginal transportation cost. We assume throughout that the market is covered—namely, that $v > 2t$ —so that all consumers purchase a good even when there is a monopolist firm. Finally, we also assume for simplicity that firms' marginal costs are 0, and so their profit from the sale of a good is equal to the price. These are all standard assumptions in Hotelling games.

The standard setup consists of a two-stage game: First, firms simultaneously set prices; second, consumers choose a firm and make a purchase. In the simple case where the distribution F of consumers is uniform the game has a unique subgame perfect equilibrium: firms' prices are $p_A = p_B = t$, consumers in $[0, 0.5]$ buy from A , and consumers in $(0.5, 1]$ buy from B (see, e.g., [4]).³

In this paper we will consider a variant of the standard model by supposing that firms may have additional information about some of the consumers. In particular, we will suppose that, for each consumer, one or both firms know the location of that consumer on the unit interval. For such consumers, firms will be able to offer a *personalized price*—a special offer specifically tailored to that consumer. If a firm does not know a consumer's location, however, then it cannot distinguish between that consumer and all other consumers whose location it does not know. All such consumers are offered the same *uniform price*.

In our model, firm B is the data-holder and firm A is the data-buyer. Thus, initially, we assume that firm B knows the locations of all consumers, whereas firm A does not know any consumer's location. Given this informational environment, a data-sharing mechanism $M = (M_B, r)$ between firms specifies a subset $M_B \subseteq [0, 1]$ and a number $r \in \mathbb{R}$, with the interpretation that firm B shares with firm A the locations of consumers in M_B , and firm A transfers to firm B a payment r . Two simple examples of data-sharing mechanisms are one that involves *no sharing*, $M = (\emptyset, r)$, and one that involves *full sharing*, $M = ([0, 1], r)$. Alternatively, firm B may share data about a subset of consumers. For example, under mechanism $([x, y], r)$, if consumer $\theta \in [x, y]$ arrives, both firms will know that consumer's location. On

³The equilibrium is unique up to the choice of the indifferent consumer located at $\theta = 0.5$.

the other hand, if consumer $\theta \in [0, 1] \setminus [x, y]$ arrives, firm B will know that consumer's location, and firm A will only be able to deduce that the consumer is not located within $[x, y]$.

In our analysis, we consider the following order of events:

1. Firms engage in a data-sharing mechanism $M = (M_B, r)$.
2. Firm A announces uniform price p_A .⁴
3. A consumer arrives, and all firms who know the consumer's location θ simultaneously offer that consumer a personalized price, $p_A(\theta)$ and $p_B(\theta)$.
4. The consumer chooses a firm from which to buy, and payoffs are realized.

Note that firms share data, and firm A announces its uniform prices, before consumers arrive. After a consumer arrives to the market, the firms who know the consumer's specific location simultaneously offer personalized prices. If firm A offers a consumer a personalized price, this offer subsumes the firm's original uniform price. Thus, the uniform price p_A will apply only to those consumers who will not subsequently be offered a personalized price by firm A .

Importantly, when firms set personalized prices, they know the uniform price set by firm A in the previous stage. This is the standard timing considered in the literature (see, e.g., [33, 10, 9, 26, 8]).⁵

For any fixed mechanism M , we will consider the pure subgame perfect equilibria of the game that starts with data-sharing mechanism M . Such equilibria always exists, and consist of a uniform price for firm A followed by personalized prices for both firms. Once the uniform price is fixed, the equilibrium personalized prices for each consumer θ are uniquely fixed. We will be interested in designing mechanisms M that lead to equilibria with high firm-profits and high consumer-welfare.

One important desideratum of data-sharing mechanisms (with corresponding equilibria) is that they be *individually rational (IR)*: That the expected utility of each firm with data sharing be at least as high as without data sharing. A data-sharing mechanism should be IR if we expect firms to participate.

Our main focus will be on mechanisms that are not only IR, but also *Pareto-improving*: when sharing takes place, (i) the expected utility of each firm and *every* consumer be at least as high as without data sharing, and that (ii) either firm A 's profits, firm B 's profits, or total consumer welfare be strictly higher.

We note that many of our results make no assumptions about the distribution of consumers. In such a general setting there may be multiple equilibria, even with no data-sharing, each with different uniform prices. Hence, we will often describe mechanisms as being IR or Pareto-improving *relative to* a particular no-sharing equilibrium.

3 No Data-Sharing

We begin by analyzing equilibria under no data-sharing. To this end, define $\mu(p_A) = \frac{1}{2} - \frac{p_A}{2t}$. If firm A charges uniform price p_A , then the consumer located at $\mu(p_A)$ is indifferent between purchasing from firm A at that price and purchasing from firm B at price 0. All consumers located to the left of $\mu(p_A)$ will thus strictly prefer purchasing from firm A at price p_A than from firm B at any nonnegative price. In contrast, for every consumer located to the right of $\mu(p_A)$ there exists a nonnegative price of firm B such that that consumer will prefer to purchase from B than from A . This is formalized in the following proposition:

⁴Note that firm B knows all consumers' locations, and so personalizes prices to each. It therefore need not post a uniform price.

⁵An alternative model that we do not analyze is one in which firms set uniform and personalized prices simultaneously, for each consumer. [26] show that, in this case, a (pure) equilibrium may fail to exist.

Proposition 1 Let $P_A = \arg \max_p p \cdot F(\mu(p_A))$. Without data sharing, the set of equilibria consist of any uniform price $p_A \in P_A$ for firm A and corresponding personalized prices $p_B(\theta) = \max\{0, p_A + t(2\theta - 1)\}$ for firm B. In the equilibrium with uniform price $p_A \in P_A$, consumers in $[0, \mu(p_A))$ purchase from firm A, whereas consumers in $[\mu(p_A), 1]$ purchase from B. The equilibrium with $p_A = \max\{P_A\}$ is strictly dominant for the firms.

The proof of Proposition 1, and all other propositions, appear in the full version of this paper [19].

Throughout the paper we will illustrate our results with the simple case in which consumers are uniformly distributed on $[0, 1]$. We note that this is the standard setup in Hotelling games.

Example 1 When consumers are uniformly distributed on $[0, 1]$, the set $P_A = \{t/2\}$. In the unique equilibrium, then, consumers between 0 and $\mu(t/2) = 1/4$ purchase from A at uniform price $p_A = t/2$, whereas the rest purchase from B at personalized prices $p_B(\theta) = \max\{t(2\theta - 1/2), 0\}$. Total firm profits are $\pi_A = t/8$ and

$$\pi_B = \int_{1/4}^1 t(2\theta - 1/2)d\theta = \frac{9t}{16},$$

whereas consumer welfare is

$$CW = \int_0^1 \max\{v - \theta t - p_A, v - t(1 - \theta) - p_B(\theta)\}d\theta = \int_0^1 (v - t/2 - \theta t)d\theta = v - t.$$

4 The Direct Effect and Full Data-Sharing

In this section we begin our analysis of how data-sharing impacts profits and welfare. Data sharing has a direct effect and an indirect effect. The direct effect is that if firm A obtains information about a consumer's locations via the sharing mechanism, it can now offer that consumer a personalized price. This affects firm B's equilibrium personalized price to that consumer, and hence also profits and welfare. The indirect effect of data sharing is that it may change the set of consumers to whom firm A's uniform price applies, since additional consumers will now be offered personalized prices. And since the uniform price is determined in equilibrium in part by the locations of consumers to whom that price will apply, a change in the set of consumers may effect a change in the equilibrium uniform price. In this section we explore the direct effect, and then in Section 5 we explore the indirect effect.

Suppose that, absent data-sharing, firm A's uniform price is p_A . If firm B shares the location θ of some consumer with A, then the firms compete in personalized prices over that consumer, yielding equilibrium prices $p_A(\theta) = \max\{t(1 - 2\theta), 0\}$ and $p_B(\theta) = \max\{t(2\theta - 1), 0\}$. The direct effect of firm B sharing the location of a consumer is summarized in Lemma 1:

Lemma 1 Consider mechanism $M = (\{\theta\}, 0)$ relative to no sharing, and suppose that consumer θ shows up.

1. If $\theta \in (1/2, 1]$, consumer θ still buys from B, but now at price $t(2\theta - 1)$. This is a net loss of p_A to firm B and a net gain of p_A to the consumer.
2. If $\theta \in [\mu(p_A), 1/2)$, consumer θ switches to purchasing from A, at price $t(1 - 2\theta)$. This is a loss of $p_A + t(2\theta - 1)$ to firm B, a gain of $t(1 - 2\theta)$ to firm A, and a gain of $p_A - t(1 - 2\theta) \geq 0$ to the consumer. Also, the gain to A is greater than the loss to B if and only if

$$\theta < \frac{1}{2} \left(\mu(p_A) + \frac{1}{2} \right),$$

the midpoint of the interval of θ -s in the case under consideration.

3. If $\theta \in [0, \mu(p_A))$, consumer θ still buys from A, now at personalized price $p_A(\theta) = t(1 - 2\theta) > p_A$.

Given these direct effects, we now consider full data-sharing, namely, $M = ([0, 1], r)$ for some r . Under this mechanism, both firms know the location of every consumer, and so both engage in personalized pricing. Firm A's uniform price thus applies to no consumer, and so only the direct effect has any bite. By Lemma 1, relative to the no-sharing mechanism with price p_A , consumers $\theta \in [\mu(p_A), 1]$ are better off, whereas consumers $\theta \in [0, \mu(p_A))$ are worse off, under full data-sharing. For the firms, naturally firm B is better off with no sharing and firm A with full sharing. The effect on total profits, however, depends on the distribution F . For the case of uniformly distributed consumers, full data-sharing harms firms:

Example 2 When consumers are uniformly distributed, [32] show that profits are $\pi_A = \pi_B = t/4$ (see also [33]). Note that total profits $\pi_A + \pi_B$ are higher under no sharing ($t/8 + 9t/16 = 11t/16$, by Example 1) than under full sharing ($t/4 + t/4 = t/2$). This implies that no mechanism $([0, 1], r)$ is IR, regardless of r .

Although full data-sharing decreases total firm profits when consumers are uniformly distributed, there exist distributions of consumers under which full data-sharing increases profits—for example, this is the case when a $(1 - \varepsilon)$ -fraction of consumers are uniformly distributed on the sub-interval $[0, 1/4]$, and the remaining ε on $(1/4, 1]$, for some small enough $\varepsilon > 0$.⁶ However, even then full sharing does not lead to maximal profits. We now turn to mechanisms that do.

5 The Indirect Effect and Firm-Optimal Data-Sharing

In this section we describe firm-optimal mechanisms, which exploit the *indirect* effect of data sharing. By Lemma 1, firm B's profit from a consumer $\theta \in (1/2, 1]$ is $p_A + t(2\theta - 1)$. If A's uniform price were to increase, this would likewise increase B's profit from consumer θ . Now, recall that, when there is no sharing, firm A sets its uniform price by choosing $p_A \in P_A = \arg \max_p p \cdot F(\mu(p))$. If B were to share data about consumers in some interval $[\underline{\theta}, \mu(p_A)]$, however, then A would offer consumers on this interval a personalized price. The uniform price would no longer apply to them, but would instead apply only to consumers $[0, \underline{\theta}) \cup (\mu(p_A), 1]$. Firm A may then benefit from increasing (decreasing) the uniform price above (below) p_A , at the same time increasing (decreasing) the profits of firm B from consumers $\theta \in (1/2, 1]$. This is the indirect effect of data sharing.

Firm B can exploit both the indirect and direct effects of data sharing by sharing data both about consumers in $[\underline{\theta}, \mu(p_A)]$ and about consumers in $(\mu(p_A), 1/2]$. Note, however, that sharing data about consumers in $(1/2, 1]$ is never beneficial, since it only results in a net loss to firms and has no indirect effect (by Lemma 1, above).

In general, the firm-optimal mechanism may depend on the distribution of consumers and other primitives of the model. In Proposition 2, however, we show that when v (the consumers' value for the good) is sufficiently high, then there is an essentially unique mechanism, with a corresponding equilibrium, that yield the firms maximal joint profits. The mechanism makes extreme use of the indirect effect of data sharing: Firm B shares data about consumers $[0, 1/2]$, implying that firm A's uniform price no longer applies to these consumers, and hence that this price can be almost arbitrarily high. A's uniform price does apply to consumers in $(1/2, 1]$, for whom it serves as an outside option. However, because these

⁶Such a consumer distribution does not satisfy our continuity assumption. However, the same result holds also when the kink at $1/4$ is smoothed out.

consumers will always purchase from B in equilibrium (by Lemma 1, above), the high outside option allows that firm to extract these consumers' entire surplus.

Proposition 2 Fix $v > \frac{5t}{2(1-F(1/2))}$. Mechanism $M = ([0, 1/2], 0)$ with equilibrium uniform price $p_A = v - t/2$ maximizes joint firm profits and is IR relative to any no-sharing equilibrium. Every other firm-optimal mechanism is of the form $M' = ([0, 1/2], r)$.

Example 3 When consumers are uniformly distributed, the mechanism described in Proposition 2 is actually firm-optimal for all $v > 2t$, as we now show. This mechanism leads to profits $\pi_A = t/4$ and

$$\pi_B = \int_{1/2}^1 (v - t(1 - \theta)) d\theta = \frac{v}{2} - \frac{t}{8} > \frac{7t}{8},$$

where the inequality follows since $v > 2t$. Thus, total profits are at least $9t/8$. In contrast, consider any mechanism $M' = (M_B, r)$ in which firm A 's uniform price applies to a consumer in $[0, 1/2]$, and fix some uniform price $p'_A \leq t$. Consumers $\theta \in (1/2, 1]$ buy from B , leading to total profits at most $3t/4$ from these consumers. Consumers $\theta \in [0, 1/2]$ either buy from A at uniform price p'_A or at personalized price $t(1 - 2\theta)$, or from B at personalized price $t(2\theta - 1)$ or $p'_A + t(2\theta - 1)$ (depending on whether $\theta \in M_B$). Total profits are maximized when $p'_A = t$, consumers $\theta \in [0, 1/4]$ buy at A 's personalized price, and the rest buy from B at price $t + t(2\theta - 1)$. Profits to A from $[0, 1/4]$ and to B from $(1/4, 1]$ are each equal to $3t/16$. Total profits from M' are thus bounded above by $3t/4 + 2(3t/16) = 9t/8$, which is equal to the lower bound on profits from $([0, 1/2], 0)$.

Remark 1 Proposition 2 provides a sufficient condition under which mechanism $M = ([0, 1/2], 0)$ is firm-optimal for *some* equilibrium (namely, the one with uniform price $p_A = v - t/2$). However, under this mechanism there are other equilibria, which involve lower uniform prices, and that yield lower firm profits. In Proposition 6 in the full version of this paper [19] we describe a different mechanism with $r = 0$ that, while not firm-optimal, yields both firms strictly higher profits than under no sharing in *every* equilibrium.

6 Pareto-Improving Data-Sharing

In Section 5 above we show that firm B can sell data in a way that maximizes joint firm profits, and hence allows that firm to charge a high price for the data. Such sharing, however, comes at the expense of consumers. In particular, under the equilibrium of Proposition 2, firms extract the entire surplus of consumers located in $[1/2, 1]$. In this section we show that there exist other data-sharing mechanisms that increase firm profits relative to no sharing, while at the same time also increasing consumers' utilities.

Recall that $P_A = \arg \max_p p \cdot F(\mu(p_A))$, and that, by Proposition 1, the set of equilibria under no data-sharing consist of uniform prices $p_A \in P_A$ by firm A and respective personalized prices $p_B(\theta) = \max\{0, p_A + t(2\theta - 1)\}$ by firm B . Denote by $E(p_A)$ the no-sharing equilibrium with uniform price p_A . In the following proposition we show that for each such no-sharing equilibrium there exists a Pareto-improving mechanism.

Proposition 3 For every $p_A \in P_A$ there exists r such that mechanism $M = \left(\left[\mu(p_A), \frac{1}{4} + \frac{\mu(p_A)}{2} \right], r \right)$ with uniform price p_A is IR and weakly beneficial to every consumer, relative to $E(p_A)$. Furthermore, M yields higher total firm profits than any other mechanism that is weakly beneficial to every consumer relative to $E(p_A)$.

The mechanism M described in Proposition 3 does not decrease the utility of any consumer. Moreover, by bullet 2 of Lemma 1, that mechanism *strictly increases* the utilities of a subset of consumers—namely, those located in $\left(\mu(p_A), \frac{1}{4} + \frac{\mu(p_A)}{2}\right]$.

The main idea underlying the construction for Proposition 3 is that firm B shares data about every consumer θ that satisfies two conditions: (i) with no sharing, consumer θ prefers to pay B 's personalized price than A 's uniform price; (ii) sharing consumer θ 's location leads to a net increase in firm profits. Note that these consumers are all closer to A than to B , so that the welfare and profit gain is obtained due to an increase in efficiency. Finally, the construction is such that A 's uniform price under M remains the same as with no sharing, which guarantees that consumers close to A do not pay a higher price than under no sharing, but also that firms maximize their joint profits subject to this constraint.

7 Consumer Opt-In

Proposition 3 above shows that there exist mechanisms that are strictly Pareto-improving, increasing firm profits as well as consumer welfare. However, these mechanisms are not optimal for firms—Proposition 2 identifies a different mechanism as maximizing firm profits, a mechanism that does so at the expense of consumers. How can a policymaker induce firms to share data in a Pareto-improving manner, rather than in a profit-maximizing manner? In this section we identify one way in which a policymaker can do this: by asking each consumer whether or not they agree to have their data shared, and then permitting firms to share data only about consumers who have agreed.

In order to analyze such consumer opt-in regulation, we first extend the model to include a preliminary opt-in stage. After setting up the model, we present two results. The first, Proposition 4 in Section 7.2, states that, under consumer opt-in, there is an equilibrium of the extended model wherein firms choose the Pareto-improving mechanism of Proposition 3. Now, although consumer opt-in can lead to the choice of the Pareto-improving mechanism, there are other equilibria that do not. However, in our second result here—Proposition 5 in Section 7.3—we show that the equilibrium of Section 7.2, where the Pareto-improving mechanism is chosen, is, in a sense, optimal for the consumers.

7.1 The Extended Model

We begin by extending the model of Section 2 with a preliminary stage, in which each consumer simultaneously chooses whether or not to opt in to having location data shared. Denote the set of consumers who opted in as C . Only then do firms engage in a data-sharing mechanism $M = (M_B, r)$; however, firms are restricted to choosing a mechanism for which $M_B \subseteq C$. Such mechanisms are *feasible for C* .

We assume that firms bargain over the choice of mechanism efficiently—that is, they choose a mechanism M that maximizes total firm profits, subject to the opt-in constraint. One way to implement such efficient bargaining is when one of the firms makes the other a take-it-or-leave-it offer by suggesting a mechanism (M_B, r) that is feasible for C . Depending on which firm makes the offer, the chosen price transfer r will vary to favor the offering firm. Either way, however, firms will choose to offer a mechanism that maximizes joint firm-profits. This assumption is stated formally in Definition 1 below as part of the solution concept. In addition, as in the previous sections, we assume that, absent data-sharing, firms play the no-sharing equilibrium $E(p_A)$ for some $p_A \in P_A$.

In this extended model there is an additional, technical complication. We are assuming that firms choose a mechanism that is feasible for some C . However, since C is generated by the set of consumers who choose to opt in, it may not be a measurable set. Thus, the firms' optimization problem may not be

well-defined at every C . One way to get around this problem is to consider the Nash equilibria of this extensive-form game (rather than the subgame perfect equilibria). However, this is somewhat unsatisfying, as such equilibria may be sustained by strange off-equilibrium behavior—namely, the presence of empty threats. Instead, we will use an equilibrium notion that is weaker than subgame perfect equilibrium but nonetheless suffices to eliminate empty threats. The general definition is due to [16]; here we give a specialized version that applies to our specific game.

For the definition, let L denote the set of subsets of $[0, 1]$, and let $\mathcal{M}(C)$ denote the set of all mechanisms feasible for C .

Definition 1 *A set $C^* \in L$ and functions $m : L \rightarrow \mathcal{M}(L)$ and $p : L \rightarrow \mathbb{R}_+$ form a threat-free Nash equilibrium (TFNE) if*

1. *For every C , mechanism $m(C)$ is feasible for C , and $p(C)$ is an equilibrium uniform price for firm A under $m(C)$.*
2. *For every $\theta \in C^*$, consumer θ is weakly better off under $m(C^*)$ than under $m(C^* \setminus \{\theta\})$ (with respective uniform prices $p(C^*)$ and $p(C^* \setminus \{\theta\})$).*
3. *For every $\theta \in [0, 1] \setminus C^*$, consumer θ is weakly better under $m(C^*)$ than under $m(C^* \cup \{\theta\})$ (with respective uniform prices $p(C^*)$ and $p(C^* \cup \{\theta\})$).*
4. *For every $\theta \in [0, 1]$ and $C \in \{C^* \setminus \{\theta\}, C^* \cup \{\theta\}\}$, mechanism $m(C)$ with uniform price $p(C)$ is IR and jointly firm-optimal relative to all mechanisms that are feasible for C (with corresponding uniform prices).*

For comparison, in a Nash equilibrium bullet 4 would be replaced by requiring IR and joint firm-optimality only for C^* . In a subgame perfect equilibrium, in contrast, bullet 4 would require these for all sets C . A TFNE is a compromise between the two, requiring IR and joint firm-optimality for C^* and for all sets C that differ from C^* by a single consumer's unilateral deviation.

7.2 Pareto-Improving Equilibrium

Given the extended model above, we can now state our proposition on the benefit of consumer opt-in.

Proposition 4 *For every $p_A \in P_A$ there exists a TFNE (C^*, m, p) of the extended model in which $m(C^*)$ is the Pareto-improving mechanism $M = \left(\left[\mu(p_A), \frac{1}{4} + \frac{\mu(p_A)}{2} \right], r \right)$, for some r .*

Proposition 4 shows that, when consumers can choose whether or not to opt in to having their data shared, and firms are allowed to only share the data of consumers who have opted in, then the equilibrium mechanism is Pareto improving. There are other equilibria that lead to a Pareto-improving mechanism. In fact, as long as consumers $\theta \in [0, \mu(p_A))$ do *not* opt in to having their data shared, the mechanism that maximizes firms' profits will be Pareto improving.

However, there are also other equilibria in which the chosen mechanism is not Pareto improving. Consider the following strategies: Consumers $[0, 1/2]$ opt in to having their data shared, and firms choose the firm-optimal mechanism $M = ([0, 1/2], 0)$ from Proposition 2. If some consumer $\theta \in [0, 1/2]$ does not opt in, then firms use the mechanism $M_\theta = ([0, 1/2] \setminus \{\theta\}, 0)$. This mechanism is identical to M , except that consumer θ faces firm A 's uniform price $p_A = v - t/2$ rather than the personalized price $p_A(\theta) = t(1 - 2\theta)$. This is no better for consumer θ , and so these strategies form an equilibrium. Why, then, would consumers choose to collectively opt in as in Proposition 4?

7.3 Consumer-Optimal Equilibrium

We now show that the equilibrium of Proposition 4 is focal for the consumers. In particular, we show that it maximizes consumer welfare, relative to all other equilibria that leave no consumer worse off.

Fix some $p_A \in P_A$, and observe that there always exists a TFNE of the extended game in which no consumer opts in, and that this leads to consumer utilities as derived from equilibrium $E(p_A)$ in mechanism $M_\emptyset = (\emptyset, 0)$. Next, let us consider other opt-in choices for consumers. For a set C and mechanism M feasible for C , say that M is *Pareto-improving for the consumers* if the resulting utility of every consumer is weakly higher than under M_\emptyset . We now show that consumers' utilities in the equilibrium of Proposition 4 are optimal:

Proposition 5 *Fix $C \subseteq [0, 1]$ and a mechanism M with uniform price q_A that is feasible for C and that is Pareto-improving for the consumers. If M yields strictly higher total utility to the consumers than M^* , then M will not be chosen by the firms in any TFNE in which consumers C opt in.*

That is, if we assume consumers make their opt-in decisions in a way that leads to a weak improvement for each, then they can do no better than the opt-in strategy of Proposition 4.

8 Conclusion

In this paper we analyzed the benefits to a data-holder of selling consumer data to a data-buyer in a Hotelling model of imperfect competition. We identified the two effects of data sharing, and showed that the interplay of these effects can lead to Pareto-improving mechanisms that benefit consumers as well as firms. Finally, we showed that consumer opt-in can induce firms to choose such a Pareto-improving mechanism.

Acknowledgements

The work of Tennenholtz was supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant number 740435).

References

- [1] Anat R Admati & Paul Pfleiderer (1986): *A monopolistic market for information*. *Journal of Economic Theory* 39(2), pp. 400–438, doi:10.1016/0022-0531(86)90052-9.
- [2] Anat R Admati & Paul Pfleiderer (1988): *Selling and trading on information in financial markets*. *The American Economic Review* 78(2), pp. 96–103.
- [3] S Nageeb Ali, Greg Lewis & Shoshana Vasserman (forthcoming): *Voluntary disclosure and personalized pricing*. *Review of Economic Studies*, doi:10.1093/restud/rdac033.
- [4] Paul Belleflamme & Martin Peitz (2015): *Industrial organization: markets and strategies*. Cambridge University Press, doi:10.1017/CB09781107707139.
- [5] Dirk Bergemann & Alessandro Bonatti (2019): *Markets for information: An introduction*. *Annual Review of Economics* 11, pp. 85–107, doi:10.2307/1912702.
- [6] Dirk Bergemann, Alessandro Bonatti & Alex Smolin (2018): *The design and price of information*. *American Economic Review* 108(1), pp. 1–48, doi:10.1016/j.jet.2011.06.003.

- [7] Francesco Clavorà Braulin (2023): *The effects of personal information on competition: Consumer privacy and partial price discrimination*. *International Journal of Industrial Organization* 87, p. 102923, doi:10.1016/j.ijindorg.2023.102923.
- [8] Zhijun Chen, Chongwoo Choe & Noriaki Matsushima (2020): *Competitive personalized pricing*. *Management Science* 66(9), pp. 4003–4023, doi:10.1287/mksc.1100.0607.
- [9] Chongwoo Choe, Stephen King & Noriaki Matsushima (2018): *Pricing with cookies: Behavior-based price discrimination and spatial competition*. *Management Science* 64(12), pp. 5669–5687, doi:10.1287/mksc.1100.0607.
- [10] Vidyanand Choudhary, Anindya Ghose, Tridas Mukhopadhyay & Uday Rajan (2005): *Personalized pricing and quality differentiation*. *Management Science* 51(7), pp. 1120–1130, doi:10.1509/jmkr.37.3.292.18777.
- [11] Richard N Clarke (1983): *Collusion and the incentives for information sharing*. *The Bell Journal of Economics*, pp. 383–394, doi:10.2307/3003640.
- [12] Matthew Elliott, Andrea Galeotti, Andrew Koh & Wenhao Li (2021): *Market segmentation through information*. Available at SSRN 3432315, doi:10.2139/ssrn.3432315.
- [13] European Commission (2021): *Antitrust: Commission opens investigation into possible anticompetitive conduct by Google in the online advertising technology sector*. Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_21_3143.
- [14] Drew Fudenberg & J Miguel Villas-Boas (2012): *Price discrimination in the digital economy*. *The Oxford handbook of the digital economy*, pp. 254–272.
- [15] Thomas Gehrig & Rune Stenbacka (2007): *Information sharing and lending market competition with switching costs and poaching*. *European Economic Review* 51(1), pp. 77–99, doi:10.1016/j.euroecorev.2006.01.009.
- [16] Ronen Gradwohl, Noam Livne & Alon Rosen (2013): *Sequential rationality in cryptographic protocols*. *ACM Transactions on Economics and Computation (TEAC)* 1(1), pp. 1–38, doi:10.1007/978-3-642-00457-5_3.
- [17] Ronen Gradwohl & Moshe Tennenholtz (2022): *Pareto-Improving Data-Sharing*. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, doi:10.1145/3298981.
- [18] Ronen Gradwohl & Moshe Tennenholtz (2023): *Coopetition against an Amazon*. *Journal of Artificial Intelligence Research* 76, pp. 1077–1116, doi:10.1613/jair.1.14074.
- [19] Ronen Gradwohl & Moshe Tennenholtz (2023): *Selling Data to a Competitor*. SSRN 4343686, doi:10.2139/ssrn.4343686. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4343686.
- [20] Yiquan Gu, Leonardo Madio & Carlo Reggiani (2022): *Data brokers co-opetition*. *Oxford Economic Papers* 74(3), pp. 820–839, doi:10.1093/oep/gpab042.
- [21] Albert Y Ha & Shilu Tong (2008): *Contracting and information sharing under supply chain competition*. *Management Science* 54(4), pp. 701–715, doi:10.1111/j.1937-5956.2002.tb00476.x.
- [22] Shota Ichihashi (2021): *Competing data intermediaries*. *The RAND Journal of Economics* 52(3), pp. 515–537, doi:10.1111/1756-2171.12382.
- [23] Tullio Jappelli & Marco Pagano (2002): *Information sharing, lending and defaults: Cross-country evidence*. *Journal of Banking & Finance* 26(10), pp. 2017–2045, doi:10.1016/S0378-4266(01)00185-6.
- [24] Nicola Jentzsch, Geza Sapi & Irina Suleymanova (2013): *Targeted pricing and customer data sharing among rivals*. *International Journal of Industrial Organization* 31(2), pp. 131–144, doi:10.1016/j.ijindorg.2012.11.004.
- [25] Ian Macmillan & Larry Selden (2008): *The incumbent's advantage*. *Harvard Business Review* 86(10), pp. 111–121.

- [26] Rodrigo Montes, Wilfried Sand-Zantman & Tommaso Valletti (2019): *The value of personal information in online markets with endogenous privacy*. *Management Science* 65(3), pp. 1342–1362, doi:10.1016/0022-0531(84)90162-5.
- [27] Marco Pagano & Tullio Jappelli (1993): *Information sharing in credit markets*. *The Journal of Finance* 48(5), pp. 1693–1718, doi:10.2307/2329064.
- [28] Michael Raith (1996): *A general model of information sharing in oligopoly*. *Journal of economic theory* 71(1), pp. 260–288, doi:10.1006/jeth.1996.0117.
- [29] Carlos Segura-Rodriguez (2021): *Selling Data*. *PIER Working Paper No. 19-006*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3385500.
- [30] Noam Shamir & Hyoduk Shin (2016): *Public forecast information sharing in a market with competing supply chains*. *Management Science* 62(10), pp. 2994–3022, doi:10.1287/mnsc.1040.0226.
- [31] Lars A Stole (2007): *Price discrimination and competition*. *Handbook of industrial organization* 3, pp. 2221–2299, doi:10.1016/S1573-448X(06)03034-2.
- [32] Curtis Taylor & Liad Wagman (2014): *Consumer privacy in oligopolistic markets: Winners, losers, and welfare*. *International Journal of Industrial Organization* 34, pp. 80–84, doi:10.1016/j.ijindorg.2014.02.010.
- [33] Jacques-Francois Thisse & Xavier Vives (1988): *On the strategic choice of spatial price policy*. *The American Economic Review*, pp. 122–137.
- [34] Kai Hao Yang (2022): *Selling consumer data for profit: Optimal market-segmentation design and its consequences*. *American Economic Review* 112(4), pp. 1364–93, doi:10.1016/0304-4068(87)90007-3.