Natural Monopoly for Data Intermediaries

Shota Ichihashi*

January 30, 2019

Abstract

I study how the market structure of data intermediaries—technology companies and data brokers that collect and distribute personal data—affects the welfare of consumers and other economic agents. I consider a model in which intermediaries obtain data from consumers and sell them to a downstream firm. Intermediaries have to compensate consumers, who are negatively affected by the firm's use of personal data. I show that competition among intermediaries does not occur: If intermediaries offer high compensations to consumers, consumers would share their data with multiple intermediaries because data are non-rivalrous; however, this lowers the price of data in the downstream market and hurts intermediaries. The lack of competition among intermediaries leads to an equilibrium where a single intermediary acquires personal data at the monopoly level of compensation. I also consider a number of extensions. For example, if the downstream firm's use of personal data positively affects consumers, competition among intermediaries benefits consumers.

Keywords: information markets, intermediaries, personal data, privacy

*Bank of Canada, 234 Wellington Street West, Ottawa, ON K1A 0G9, Canada. Email: shotaichihashi@gmail.com.
I am grateful to Sitian Liu for carefully reading the first draft of this paper. The opinions expressed in this article are the author's own and do not reflect the view of Bank of Canada.

1 Introduction

Why are consumers not paid for their personal data? Technology companies (e.g., Google and Facebook) and data brokers (e.g., LiveRamp and Nielsen) collect and monetize a vast amount of personal information, such as consumers' demographic characteristics and browsing activities; however, consumers are usually not paid for their data contributions. It is argued that the reason could be historical: Consumers are accustomed to surrendering their data in exchange for free services, or they are unaware of the collection and productive value of their data (Arrieta-Ibarra et al., 2018; Carrascal et al., 2013; Federal Trade Commission, 2014). I attempt to answer the question using a model of data intermediaries in markets for personal data.

Specifically, this paper studies how the market structure of data intermediaries—technology companies and data brokers that collect and distribute personal data—affects consumers and other economic agents. In particular, can competition among intermediaries be sustained to ensure that consumers get a fair fraction of surplus created by their personal data? Does the answer depend on how downstream firms (i.e., firms that obtain data from intermediaries) use personal data?

To answer these questions, I build a model in which data intermediaries obtain personal data from consumers and sell them to a downstream firm. The firm's use of personal data imposes consumers negative externalities, such as intensive price discrimination, intrusive advertising, and potential data leakage. Therefore, intermediaries have to compensate consumers for their data provision. The level of compensation determines how much personal data each intermediary can acquire, which in turn affects price competition among intermediaries in the downstream market.

The following examples illustrate the main idea of this paper by contrasting intermediaries for physical goods (cars) with intermediaries for personal data.

Example 1 (Intermediares in Used Car Markets). Suppose that a car owner values her car at \$5,000, and a downstream buyer values it at \$7,000. An intermediary buys the car from the owner and resells it to the buyer by posting prices. First, a monopoly intermediary can extract full surplus by posting prices of \$5,000 and \$7,000 to the owner and the buyer, respectively. Second, suppose that there are two intermediaries. Since the intermediary that obtains the car from the owner can charge \$7,000 to the buyer as a monopoly, intermediaries will raise prices up to \$7,000 to compete for the car in the upstream market. Therefore, competition yields a net gain of \$2,000 to the owner.

Example 2 (Intermediaries in Markets for Personal Data). Suppose that a firm (e.g., retailer or advertiser) values a consumer's personal data at \$70, but the consumer incurs a loss of \$50 if the firm uses her data. Moreover, suppose that data intermediaries post prices (compensations) to obtain the data from the consumer and sell them to the firm. First, as in Example 1, a monopoly intermediary can extract full surplus by posting prices \$50 and \$70 to the consumer and the firm, respectively. Second, if there are two intermediaries (1 and 2), there is no longer an equilibrium where both intermediaries offer the consumer to pay \$70. This is because, if they do so, the consumer will share her data with both intermediaries—*data are non-rivalrous*. Moreover, once both intermediaries obtain the data, in the downstream market, Bertrand competition forces them to set a price of zero. This deters the ex ante competition for data. Indeed, there is an equilibrium in which intermediaries 1 and 2 offer compensations of \$50 and \$0, respectively. The consumer will accept the offer of \$50, but this does not make her any better off than the monopoly outcome. Note that if intermediary 2 deviates and offers a positive compensation, the consumer will share her data with both intermediaries, which deters such a deviation. Thus, competition between data intermediaries may not occur.[1]

To generalize these examples, in Section 3, I consider a model of a continuum of consumers, K data intermediaries, and a single downstream firm. Intermediaries offer consumers compensations to obtain their data, and sell the data to a downstream firm. The firm's use of personal data imposes negative externalities on consumers. The loss incurred by each consumer depends on whether the firm acquires her data and the mass of consumers sharing their data. The firm's revenue increases in the amount of data but the marginal revenue is decreasing.

There are two main findings. First, I show that a monopoly outcome can arise regardless of the number of intermediaries in the market. The set of equilibria consists of a continuum of equilibria that differ in the allocation of data across intermediaries and the amount of compensation each consumer earns. Importantly, for any number of intermediaries, there is a "monopoly equilibrium," where the largest mass of consumers give up their data at the minimum level of compensation. The intuition is similar to Example 2: If intermediaries compete for data by offering high compensations, consumers will share their data with multiple intermediaries, which will lower the price of

[1]This is not a unique equilibrium. Indeed, there is also an equilibrium where the consumer earns more than \$50 (Theorem 1); however, only the compensation of \$50 is "robust" with respect to a small transaction cost that each intermediary incurs to make an offer to the consumer (Proposition 3).

the data in the downstream market. This relaxes competition among intermediaries and enables them to secure a monopoly profit in (some) equilibrium. I also show that the industry profit of intermediaries is greater in equilibrium where data are concentrated to a smaller number of intermediaries, because they can extract the infra-marginal value of data from the downstream firm.

The second main finding is that only the monopoly outcome is robust with respect to small transaction costs. Specifically, I show that if it is costly for intermediaries to make offers to consumers, in any equilibrium, the largest mass of consumers share their data with the monopoly level of compensation; moreover, the result holds no matter how small the cost is. For example, in the absence of transaction costs, there are equilibria where intermediaries offer greater compensations than the monopoly level. However, those equilibria rely on a particular off-path behavior of consumers: If an intermediary deviates and lowers compensations, a consumer "punishes" the deviating intermediary by sharing her data with other intermediaries, which offer zero compensations. This requires that intermediaries make offers of zero compensations that are rejected on the equilibrium path. Such an equilibrium is eliminated if each intermediary incurs small costs to make offers to consumers.

I consider several extensions in Section 7 and Section 8. For instance, I consider that the downstream firm's use of personal data brings consumers net benefits, such as personalized services and customized products. In this case, competition among intermediaries is sustainable and benefits consumers. For another instance, I consider that each consumer has multiple pieces of personal data. I show that in equilibrium, different intermediaries could specialize in acquiring different pieces of data, and each intermediary compensates consumers for the marginal disutiltiy from sharing the piece of data. If the marginal disutility is increasing, the concentration of data among a smaller number of intermediaries leads to a lower consumer surplus.

In conclusion, the model illustrates that price competition among data intermediaries does not in general yield the competitive outcome. The non-rivalry of data deters competition among intermediaries in the upstream market, which in turn enables them to secure the monopoly outcome. In particular, consumer surplus could be no greater than the monopoly level regardless of the number of intermediaries in the market.

The rest of the paper is organized as follows. Section 2 provides a literature review and Section 3 describes the model. Section 4 considers the case of monopoly intermediary. Section 5 considers the case of multiple intermediaries and shows that the monopoly outcome could arise in equilibrium (e.g., Theorem 2). Section 6 shows the robustness of the monopoly equilibrium, and Section 7 generalizes the model by allowing consumers to have multiple pieces of personal data. Section 8 provides extensions including the case where firms' use of data benefits consumers. Section 9 concludes.

2 Literature Review

This paper is related to two strands of literature. First, it is related to the literature of platform competition in two-sided markets. The literature often focuses on the case where each side of agents benefit from interacting with the other side of agents (e.g., Armstrong (2006); Caillaud and Jullien (2003); Rochet and Tirole (2003)). This is natural in many applications such as video games (consumers and game developers) and credit cards (cardholders and merchants). In this case, platform competition involves undercutting prices for at least one side, which is sustainable even if multi-homing is possible.[2] In contrast, I consider the case where one side (i.e., firms) benefits and the other side (i.e., consumers) loses from the interactions (i.e., transfer of personal data). In this case, competition involves raising compensations for consumers, which cannot be sustained whenever data are non-rivalrous. Caillaud and Jullien (2003) show that intermediaries have an incentive to make their services non-exclusive to relax price competition. This might look analogous to my result, but it is logically distinct. In their model, the profits of intermediaries offering non-exclusive services still go to zero as matching frictions disappear. Negative cross-side externalities also appear in models of advertising platforms, such as Anderson and Coate (2005) and Reisinger (2012), where the presence of advertisers impose negative externalities on users (due to nuisance costs).

The paper also relates to the more classical literature of competing intermediaries. Stahl (1988) studies a model of market markers who obtain stock from suppliers and resell it to buyers. My model resembles his model if we replace physical (rivalrous) goods with personal data. In Stahl (1988), competition among market makers drives down their profits to zero, while in my model,

[2]The reason price competition is sustainable is similar to the one in Section 8, which shows that competition benefits consumers if the downstream firm's use of data benefits consumers.

data intermediaries can obtain the monopoly profit regardless of the market structure. Yanelle (1997) studies a model of competing financial intermediaries and shows that an equilibrium may not be competitive. The result is driven by increasing returns to scale of monitoring technologies and coordination problem in each side of the market (i.e., debtors or creditors), both of which are absent in my model.

Second, this paper is related to the recent literature on markets for data. The model in this paper is closely related to the one presented in Bergemann and Bonatti (2019). They consider a model of a monopoly data intermediary, where a downstream firm uses personal data to price discriminate consumers with quadratic utility and Gaussian signals. My model allows multiple intermediaries and more general ways in which a downstream firm can use personal data (in addition to price discrimination). This enables me to address new questions such as how the market structure of intermediaries affect economic outcomes and how the concentration of data affects consumers. In contrast to Bergemann and Bonatti (2019), my model is not amenable to intermediaries' information design problem.

More broadly, this paper relates to work on markets for data beyond the context of price discrimination. For instance, Choi et al. (2018) consider consumers' privacy choices in the presence of information externality. For another instance, Bergemann et al. (2018) consider a model of data provision and data pricing. I contribute to this literature by illustrating how the non-rivalry of data affects the nature of competition.[3]

3 Model

I consider a model in which data intermediaries buy personal data from consumers and sell them to a downstream firm. For now, I focus on the case where the firm's use of personal data negatively affects consumers. Thus, intermediaries have to compensate consumers for sharing their data.

The formal description is as follows. The market consists of a unit mass of consumers, $K \geq 1$ data intermediaries, and a single downstream firm.[4] If it is clear from context, I use K also to mean

[3] Jones et al. (2018) study a macroeconomic model that explicitly takes into account the non-rivalry of data and data property rights.

[4] Section 8 considers the case of multiple downstream firms. If firms do not interact with each other, it is without loss of generality to consider a single firm.

the set of intermediaries. The game consists of two parts—intermediaries buy data in the upstream market and sell them in the downstream market.

Upstream Market

Each consumer $i \in N := [0, 1]$ is endowed with some personal data, such as her demographic characteristics and browsing histories. At the beginning of the game, each intermediary $k \in K$ simultaneously makes an *offer* $(\tau_i^k)_{i \in N} \in \mathbb{R}^N$, where τ_i^k is the compensation that intermediary k pays consumer i in exchange for her personal data.[5] Assume that τ_i^k is measurable in i. I assume that consumer i does not observe offers made to other consumers, i.e., $(\tau_m^k)_{k \in K, m \in N \setminus \{i\}}$.

After observing offers $(\tau_i^1, \ldots, \tau_i^K)$, each consumer i simultaneously chooses $K_i \subset K$, the set of intermediaries with which i shares her personal data. Hereafter, I say that consumer i *shares her data* whenever $K_i \neq \emptyset$. Each consumer's decision determines $N_k := \{i \in N : k \in K_i\}$, the set of consumers who share their data with intermediary k. Hereafter, I call any subset of $\cup_k N_k$ as *data*. All intermediaries and the firm publicly observe $(N_1, \ldots, N_K)$, which I call an *allocation of data* across intermediaries.

Downstream Market

Given the allocation of data $(N_1, \ldots, N_K)$, intermediaries sell their data to the firm. Namely, each intermediary k simultaneously posts price $p_k \in \mathbb{R}$ as a take-it-or-leave-it offer. Then, the firm chooses the set $K' \subset K$ of intermediaries from which it buys data. The firm's decision determines $D := \cup_{k \in K'} N_k \subset N$, the set of consumers whose personal data are acquired by the firm. Finally, payoffs are realized and the game ends.

Preferences

All players are risk-neutral, and their ex post payoffs are as follows. The payoff of an intermediary is revenue minus compensation: If intermediary k obtains data $N_k \subset N$ and the firm buys data from a set K' of intermediaries, k obtains a payoff of $\mathbf{1}_{\{k \in K'\}} p_k - \int_{N_k} \tau_i^k di$, where $\mathbf{1}_{\{k \in K'\}}$ is the indicator function that is equal to 1 and 0 if $k \in K'$ and $k \notin K'$, respectively.

The payoff of a consumer depends on whether the firm acquires her data and the fraction of consumers who give up their data. Suppose that consumer i sells her data to intermediaries in K_i

[5] I assume that each intermediary makes offers to all consumers, because making no offer is payoff equivalent to offering a negative compensation $\tau_i^k < 0$. This is not without loss of generality if an intermediary incurs a cost to send an offer, as I study in Section 6.

and the firm acquires the personal data of consumers in $D \subset N$. Let $|D|$ denote the Lebesgue measure of D. If the firm acquires the data of i, her payoff is $u(1, |D|) + \sum_{k \in K_i} \tau_i^k$; if the firm does not, i's payoff is $u(0, |D|) + \sum_{k \in K_i} \tau_i^k$. The first term of each expression captures how the firm's use of personal data affects consumer i, and the second term captures the total compensation from intermediaries. I impose the following assumption on $u : \{0,1\} \times [0,1] \to \mathbb{R}$. Intuitively, it requires that the firm's use of personal data negatively affects the data subject. The assumption does not restrict how the sharing of data by other consumers affects the welfare of each consumer.

Assumption 1. $\forall n \in [0,1]$, $u(1,n) < u(0,n)$.

Next, I describe the payoffs of the firm. Suppose that the firm obtains data from the set K' of intermediaries. Let $D = \cup_{k \in K'} N_k$ denote the resulting data that the firm acquires. Then, the firm's payoff is $\Pi(|D|) - \sum_{k \in K'} p_k$. The first term, which I call revenue, depends only on the size of data D.[6] The second term is the total payments to intermediaries in K'. I assume that Π satisfies the following.

Assumption 2. $\Pi : [0,1] \to \mathbb{R}$ is differentiable on $(0,1)$ and satisfies the following.

1. $\Pi(n)$ is strictly increasing in $n \in [0,1]$.
2. $\Pi'(n)$ is strictly decreasing in $n \in (0,1)$.
3. $\Pi(0) = 0$.

Point 1 means that a new piece of data always increases revenue. Point 2 says that the marginal contribution of data to revenue is decreasing. This is motivated by the idea that "data typically exhibits decreasing returns to scale like any other factor of production" (Varian, 2018).[7] Point 2 also simplifies the derivation of the equilibrium prices of data in the downstream market. Point 3 is a normalization.

Finally, I impose the following assumption on (u, Π).

[6] This excludes the case where the firm values the data of some consumers more highly than the data of other consumers. The extension in Section 7 could take into account such a situation, because I assume that the firm's revenue is a submodular set function of data.

[7] Nonetheless, Point 2 is not a mathematical property of the value of information in decision-making. See, for example, Radner and Stiglitz (1984).

Assumption 3. There is a unique $n^* \in (0,1)$ such that $\Pi'(n) - |u(1,n) - u(0,n)|$ crosses zero at n^* and from above.

The assumption ensures the existence of pure-strategy equilibrium.[8] It holds, for example, if a consumer incurs a constant privacy cost, i.e., $\exists c > 0$, $\forall n$, $|u(1,n) - u(0,n)| = c \in (\Pi'(1), \Pi'(0))$. n^* can be different from an efficient level, because this "first-order condition" ignores how acquiring the data of one consumer affects the welfare of other consumers.[9]

Timing

The timing of the game is summarized as follows. First, each intermediary simultaneously makes an offer to each consumer. After privately observing the offers made to her, each consumer simultaneously decides the set of offers to accept. The decision of each consumer determines the allocation of data. Then, each intermediary simultaneously posts a price to the firm. Finally, the firm chooses the set of intermediaries from which it buys data.

Solution Concept

The solution concept is pure-strategy perfect Bayesian equilibrium (PBE) with the following property: For each $k \in K$ and $i \in N$, even if intermediary k makes a deviating offer to consumer i, the deviation does not change i's beliefs over the offers that each intermediary makes to other consumers. To simplify exposition, I consider equilibrium such that the set of consumers who share their data takes the form of $[0, n]$ for some $n \in [0,1]$.[10] From now on, I abbreviate PBE satisfying these conditions as "equilibrium."

3.1 Interpretations of Data Intermediaries

I provide three potential interpretations of data intermediaries in the model. First, we can interpret an intermediary as a company that directly buys data from consumers and sell them to firms. For example, a mobile application Killi offers users gift cards or monetary payments in exchange for

[8]All the results continue to hold under a weaker assumption that $\Pi'(n) - |u(1,n) - u(0,n)|$ crosses zero at most once and from above.

[9]If mass n of consumers share their data, the total surplus, which is defined as the sum of the payoffs of all players, is $nu(1,n) + (1-n)u(0,n) + \Pi(n)$.

[10]Any pure-strategy equilibrium is "equivalent" to an equilibrium where the set of consumers who share their data is written as $[0, n]$. The equivalence is up to the indices of consumers sharing their data and the outcomes associated with a measure-zero set of consumers.

providing personal data such as demographic characteristics, location, and email address. Users can select which pieces of data to share, and when downstream firms (e.g., retailers and advertisers) purchase their data, users are paid.[11] The model helps us understand how competition in this new but growing market can differ from competition in traditional markets.[12]

Second, we can interpret intermediaries as large data brokers such as LiveRamp (formerly known as Acxiom), Nielsen, and Oracle. The modeling assumption that intermediaries compensate consumers seems to contradict the observation that data brokers usually do not interact with consumers. However, I argue that the model could still be useful for understanding how the incentives of data brokers would look like if they had to source data directly from consumers. The question is of growing importance, as awareness of data sharing practices increase and policymakers try to ensure that consumers have control over their data (e.g., The EU General Data Protection Regulation and California Consumer Privacy Act).

Finally, one attractive but controversial interpretation would be to see data intermediaries as online platforms such as Google and Facebook. Arguably, the model abstracts from many institutional features of these platforms. For example, Google and Facebook distribute personal data indirectly through sponsored search or targeted display advertising. In this case, these platforms could be competing for not only consumers' data but also their attention, which is a scarce resource due to consumers' time and cognitive constraints. Nonetheless, the model seems to capture an economic force relevant to competition between these platforms: The non-rivalry of data allows two platforms to obtain the same data. However, then, downstream advertisers will perceive advertising spaces of two platforms (which are bundled with the same data) as more substitutable. The model suggests that this in turn deters competition for data in the upstream market.

4 Benchmark: Monopoly Data Intermediary

Consider a monopoly data intermediary ($K = 1$). Suppose that the intermediary holds the data of consumers in $[0, n]$. To (additionally) acquire data $(n, n + \Delta]$, the intermediary has to pay the

[11] https://www.killi.io/killi-app/

[12] Another example is Nielsen Mobile App, which tracks users' browsing activities in exchange for rewards. Startup companies, such as CitizenMe, Datacoup, Datawallet, and Digi.me, aim for offering more sophisticated marketplaces for personal data.

total compensation of (at least) $\Delta \cdot |u(1,n) - u(0,n)|$ to consumers in $(n, n+\Delta]$. By doing so, it can increase the revenue by $\Pi(n+\Delta) - \Pi(n)$ in the downstream market. Thus, the monopoly intermediary chooses the amount of personal data to acquire by balancing the marginal revenue $\Pi'(n)$ and cost $|u(1,n) - u(0,n)|$. Assumption 3 implies that the intermediary acquires data from consumers in $[0, n^*]$.[13] The proof of the following result is relegated to Appendix A.

Proposition 1. *If the market consists of a monopoly intermediary, in any equilibrium, consumers in* $[0, n^*]$ *provide their data where* $n^* \in (0,1)$ *satisfies* $\Pi'(n^*) = |u(1,n^*) - u(0,n^*)|$. *Each consumer obtains a payoff of* $u(0,n^*)$*, and the firm obtains a payoff of zero.*

A data intermediary has to compensate each consumer for the disutility of sharing her data with the firm, but not for externalities among consumers. Thus, the intermediary can make a profit even if transferring personal data lowers total surplus due to negative externalities among consumers.[14] In an extreme case, if $u(1,n) = u(0,n)$ for all $n \in [0,1]$ but $u(1,n)$ and $u(0,n)$ are decreasing in n, the intermediary can acquire personal data from all consumers for free, whereas total and consumer surplus could be maximized when the firm acquires no data. Conversely, the equilibrium level of data sharing n^* could be inefficiently low if there are positive externalities among consumers.

5 Competing Data Intermediaries

Now, I assume that the market consists of multiple intermediaries ($K \geq 2$). I solve the game backwards. First, given the allocation $(N_1, \ldots, N_K)$ of data among intermediaries, I derive the equilibrium prices of data in the downstream market. Second, given the equilibrium pricing, I derive the equilibrium compensation in the upstream market. Hereafter, for any set $N' \subset N$, I write $\Pi(N')$ to mean $\Pi(|N'|)$.

[13] Note that this argument relies on the assumption that consumer i does not observe offers $\tau_{-i} := (\tau_j^k)_{(j,k)\in(N\setminus\{i\})\times K}$ made to other consumers and that she does not change her beliefs about τ_{-i} after detecting a deviation that affects i. This implies that when the intermediary deviates and acquires data from consumers in $(n, n+\Delta]$, it does not need to change compensations to consumers in $[0,n]$, and in order to calculate the compensation to consumers in $(n, n+\Delta]$, it can use $(u(1,n), u(0,n))$ instead of $(u(1,n'), u(0,n'))$ for some $n' \neq n$.

[14] Bergemann and Bonatti (2019) point this out in the context of price discrimination with quadratic utility and Gaussian information structure.

5.1 Price of Data

The following lemma shows that the equilibrium price of data is equal to their marginal contribution to the firm's revenue.

Lemma 1. *Suppose that each intermediary $k \in K$ holds data $N_k \subset N$. This subgame has a subgame perfect equilibrium in which intermediary k posts a price of*

$$\Pi_k := \Pi\left(\bigcup_{j \in K} N_j\right) - \Pi\left(\bigcup_{j \in K \setminus \{k\}} N_j\right) \tag{1}$$

for its data, and the firm purchases data from all intermediaries.

Proof. Take any $(N_1, \ldots, N_K) \in (2^N)^K$. Consider a strategy profile in which each intermediary $k \in K$ sets a price of Π_k in (1) and the firm buys data from all intermediaries. Also, if an intermediary deviates and sets a higher price, the firm buys data from all but the deviating intermediary. I show that this is an equilibrium (of the subgame). First, if no intermediary deviates, it is optimal for the firm to buy all data: Point 2 of Assumption 2 implies that $\Pi(\cup_{j \in K' \cup \{k\}} N_j) - \Pi(\cup_{j \in K'} N_j) - \Pi_k \geq 0$ for any $K' \subset K$. Thus, the firm is willing to buy N_k at price Π_k regardless of the prices posted by other firms. Second, if intermediary k unilaterally deviates and sets a price of $p_k > \Pi_k$, the firm prefers to buy data from intermediaries in $K \setminus \{k\}$, and thus k cannot benefit by raising a price. Finally, any price $p_k < \Pi_k$ strictly lowers the payoff of intermediary k. □

Hereafter, I focus on equilibrium where the prices of data satisfy the condition in Lemma 1. The lemma implies that two intermediaries obtain zero revenue in the downstream market if they hold the same data, which is similar to the Bertrand competition with homogeneous products. More generally, the revenue of an intermediary depends only on the part of the data that other intermediaries do not hold.

Corollary 1. *Suppose that each intermediary $j \neq k$ holds data N_j. Take any $N_k \subset N$ and any $D \subset \cup_{j \neq k} N_j$. Then, the equilibrium revenue of intermediary k in the downstream market is identical between when it holds N_k and $N_k \cup D$.*

5.2 Natural Monopoly for Data Intermediaries

Given the pricing rule in Lemma 1, how do the equilibrium compensation and the allocation of data look like? First, the following result shows that the non-rivalry of personal data deters intermediaries from making competing offers.

Lemma 2. *Take any equilibrium. Let D denote the set of all consumers who share their personal data on the equilibrium path. Let $\hat{N} \subset D$ denote the set of all consumers in D to whom two or more intermediaries offer positive compensations. Then, $|\hat{N}| = 0$.*

Proof. Lemma 1 implies that the downstream firm acquires the data of consumers in D. For each $(k, j) \in K^2$ with $k \neq j$, let $\hat{N}_{kj}$ denote the set of all consumers in D to whom both intermediaries k and j offer positive compensations. Suppose to the contrary that there is $(k, j) \in K^2$ such that $|\hat{N}_{kj}| > 0$. Then, consumer $i \in \hat{N}_{kj}$ accepts the offers from both k and j, because conditional on accepting one offer, accepting other offers (with positive compensations) does not affect her payoff from the firm's use of data (i.e., $u(1, n)$ where n is the mass of consumers sharing data) but strictly increases her payoff from compensations ($\sum_{k \in K_i} \tau_i^k$). However, intermediary k can profitably deviate by offering consumers in $\hat{N}_{kj}$ zero compensation, because the deviation does not change the revenue in the downstream market (Corollary 1) but decreases the total compensation k has to pay. This implies that $|\hat{N}| \leq \cup_{(k,j) \in K^2} |\hat{N}_{kj}| = 0$. □

Lemma 2 indicates the lack of competition among data intermediaries in the upstream market; however, it does not necessarily imply that the equilibrium coincides with the monopoly outcome in Proposition 1, as the following example illustrates.

Example 3 (Equilibrium with Complete Privacy). Consider the following strategy profile: All intermediaries offer zero compensation to each consumer, who rejects all offers. If intermediary k unilaterally deviates and offers positive compensation $\tau_i^k \geq |u(1, 0) - u(0, 0)|$ to consumer i, she accepts the offers of *all* intermediaries.[15] This strategy profile is an equilibrium. In particular, no intermediary has an incentive to deviate and acquire data by offering high compensations, because consumers will also share their data with other intermediaries. In this equilibrium, the firm and

[15]Each player's action on other off-path information sets are naturally defined. For example, if an intermediary unilaterally deviates and offers $\tau_i^k < |u(1, 0) - u(0, 0)|$, i rejects all offers.

intermediaries obtain zero profits, and each consumer obtains $u(0,0)$, which can be higher or lower than the monopoly outcome.

Example 3 indicates the multiplicity of equilibria. Indeed, if there is an equilibrium where consumers in $[0,n]$ share their data, we can also construct an equilibrium where consumers in $[0,n'] \subsetneq [0,n]$ share their data, by applying the argument in Example 3 to consumers in $(n',n]$.

Before characterizing the set of equilibria, I introduce two definitions about allocation of data. Recall that an allocation of data, denoted by $(N_k)_k \in (2^N)^K$, specifies whose data each intermediary holds.

Definition 1. An allocation of data $(N_k)_k$ is *disjoint* if $|N_k \cap N_j| = 0$ for all $k, j \in K$ with $k \neq j$. The *size* of the allocation refers to $| \cup_k N_k|$.

Note that in equilibrium, the size of an allocation of data is equal to the mass of consumers sharing their data. The following result characterizes the set of all equilibria. The proof is in Appendix B.

Theorem 1. *Let $n^* \in (0,1)$ be the unique value satisfying $\Pi'(n^*) = |u(1,n^*) - u(0,n^*)|$. The following two conditions are equivalent.*

1. *There exists an equilibrium such that the allocation of data is $(N_k)_k$ with size n, and each consumer $i \in N$ earns compensation τ_i.*

2. *Allocation $(N_k)_k$ is disjoint and $\cup_{k\in K} N_k = [0,n]$ where $n \leq n^*$. Moreover, compensation τ_i satisfies*

$$\begin{aligned} |u(1,n) - u(0,n)| \leq \tau_i \leq \Pi'(n) \quad & \textit{if} \quad i \leq n; \\ \tau_i = 0 \quad & \textit{if} \quad i > n. \end{aligned}$$

for almost every $i \in [0,1]$.

Theorem 1 implies that regardless of the number of intermediaries, the monopoly outcome (Proposition 1) can arise. For example, it is an equilibrium that a single intermediary (say intermediary 1) obtains data $[0,n^*]$ at the monopoly level of compensation, because we can set

$N_1 = [0, n^*]$, $N_k = \emptyset$ for $k \geq 2$, and $\tau_i = u(0, n^*) - u(1, n^*)$ for $i \in [0, n^*]$. In this equilibrium, any intermediary $k \geq 2$ has no incentive to obtain data from consumer $i \leq n^*$ because i will share her data not only with intermediary k but also with 1. Also, intermediaries have no incentive to obtain data from consumer $i > n^*$, because the compensation i asks for is greater than the price of her data in the downstream market.

Note that in equilibrium where the mass of consumers sharing their data is strictly less than n^*, there is a gap between the marginal value of data ($\Pi'(n)$) and the disutility of sharing data ($|u(1, n) - u(0, n)|$). In this case, Theorem 1 states that a compensation strictly above $|u(1, n) - u(0, n)|$ is sustained. In such an equilibrium, intermediaries have no incentive to *lower* compensations because consumers will punish such deviations by sharing data with other intermediaries.

Despite the non-rivalry of data, Theorem 1 states that no two intermediaries obtain the same piece of data, because such data have a price of zero in the downstream market. Conversely, any disjoint allocation of data can arise in some equilibrium, because given an allocation, no intermediary prefers to acquire data that other intermediaries will acquire.

Next, I study how the allocation of data affects the industry profit of data intermediaries. To simplify exposition, I introduce the following notion.

Definition 2. Let $(N_k)_k$ and $(N'_k)_k$ denote two disjoint allocations of data with the same size. $(N'_k)_k$ is *more concentrated* than $(N_k)_k$ if, for any k, there is ℓ such that $N_k \subset N'_\ell$.

Roughly, one allocation is more concentrated than the other if the former is coarser than the latter, when we regard these allocations as partitions of the set of consumers sharing their data. The following result states that an equilibrium where the allocation of data is more concentrated leads to higher industry profits of intermediaries.

Proposition 2. *Let $(N_k)_k$ and $(N'_k)_k$ denote two disjoint allocations of data with the same size. If $(N'_k)_k$ is more concentrated than $(N_k)_k$, the total revenue of intermediaries in the downstream market is higher under $(N'_k)_k$ than under $(N_k)_k$.*

Proof. In this proof, for $X, Y \subset N$, $X - Y$ stands for $X \setminus Y$. Let $(N'_k)_{k \in K}$ and $(N_k)_{k \in K}$ denote two disjoint allocations of data such that the former is more concentrated than the latter. Without loss of generality, assume that they have size 1 (i.e., $\cup_k N'_k = \cup_k N_k = [0, 1]$). Note that in general, for any $N_0 \subset N$ and a partition $(M_1, \ldots, M_K)$ of N_0, we have

$$
\begin{aligned}
&\Pi(N) - \Pi(N - N_0) \\
=&\Pi(N) - \Pi(N - M_1) + \Pi(N - M_1) - \Pi(N - M_1 - M_2) + \cdots \\
&+ \Pi(N - M_1 - M_2 - \cdots - M_{K-1}) - \Pi(N - M_1 - M_2 - \cdots - M_K) \\
\geq& \sum_{k \in K} \left[\Pi(N) - \Pi(N - M_k)\right],
\end{aligned}
$$

where the last inequality follows from Point 2 of Assumption 2. Now, for any $\ell \in K$, let $K(\ell) \subset K$ satisfy $N'_\ell = \sum_{k \in K(\ell)} N_k$. The above inequality implies

$$
\begin{aligned}
\Pi(N) - \Pi(N - N'_\ell) &\geq \sum_{k \in K(\ell)} \left[\Pi(N) - \Pi(N - N_k)\right], \forall \ell \in K \\
\Rightarrow \sum_{\ell \in K} \left[\Pi(N) - \Pi(N - N'_\ell)\right] &\geq \sum_{\ell \in K} \sum_{k \in K(\ell)} \left[\Pi(N) - \Pi(N - N_k)\right].
\end{aligned}
$$

In the last inequality, the left and the right hand sides are the total revenue for intermediaries in the downstream market under (N'_k) and (N_k), respectively. $\square$

To gain some intuition, suppose $K = 2$ and $\Pi(n) = \sqrt{n}$. If a single intermediary holds all data $[0, 1]$, it can set a price of $\Pi(1) = 1$ for its data. In contrast, if intermediaries 1 and 2 hold data $[0, 1/2]$ and $[1/2, 1]$, respectively, each intermediary sets a price of $\Pi(1) - \Pi(1/2)$. This yields the industry profit of $2[\Pi(1) - \Pi(1/2)] = 2 - \sqrt{2} < 1$. The concentration of data allows intermediaries to extract the infra-marginal value of data from downstream firms.

To state the next result, I define monopoly equilibrium:

Definition 3. A *monopoly equilibrium* is an equilibrium in which a single intermediary acquires personal data $[0, n^*]$.

Combining the cases of monopoly (Proposition 1) and multiple intermediaries (Theorem 1) and the comparative statics in the allocation of data (Proposition 2), I obtain the following.

Theorem 2. *For any number of intermediaries in the market, there exists a monopoly equilibrium. Among all equilibria, a monopoly equilibrium maximizes the total payoffs of intermediaries and minimizes the firm's payoff.*

It would be illustrative to contrast Theorem 2 with the case of physical (rivalrous) goods. To do so, the next result assumes that each consumer can share her data with *at most one* intermediary, which makes data rivalrous. The proof of the following result is in Appendix C.

Claim 1. *Suppose that each consumer can share her data with at most one intermediary. If there is a single intermediary ($K = 1$), the equilibrium is identical with the monopoly equilibrium in Proposition 1. If there are multiple intermediaries ($K \geq 2$), in any equilibrium, all intermediaries obtain zero payoffs.*

If goods are rivalrous, competition dissipates intermediaries' profits for a similar reason to the standard Bertrand competition with homogeneous products. Namely, if there is an intermediary making a positive profit, another intermediary can offer a slightly higher compensation to exclusively obtain the goods from consumers. Thus, in equilibrium, all intermediaries obtain zero profits. This contrasts with intermediaries for (non-rivalrous) data, where intermediaries can secure a monopoly profit.

6 Robustness of Monopoly Outcome

This section shows that if intermediaries incur small transaction costs to interact with each consumer, any equilibrium has the monopoly level of consumer surplus. Formally, I modify the action space and the preferences of each intermediary as follows. First, each intermediary can choose to not make an offer, which is now distinguished from offering nonpositive compensation. Second, there is $c > 0$ such that if an intermediary makes an offer to mass n of consumers, it incurs a cost of $n \cdot c$, which enters additively into its payoff function. I assume that c is small so that

$$0 < c < \min \left\{ \Pi'(n^*), \inf_{0 \leq n \leq 1} |u(1, n) - u(0, n)| \right\}.$$

The following result states that small transaction costs for intermediaries single out the monopoly level of consumer surplus. The proof is in Appendix D.

Proposition 3. *The set of equilibria is nonempty, and in any equilibrium, consumers in $[0, n^*]$ share their data and earn monopoly compensation $\tau^* = |u(1, n^*) - u(0, n^*)|$, whereas consumers in $(n^*, 1]$ do not share data.*

The intuition is as follows. If there are no transaction costs, compensations greater than the monopoly level are sustainable in equilibrium. In such an equilibrium, if an intermediary deviates and lowers a compensation, consumers will punish it by sharing their data with multiple intermediaries. On the equilibrium path, these intermediaries make offers of zero compensations, which are rejected by consumers. This cannot consist of an equilibrium if it is costly to make an offer.

Similarly, if there are no transaction costs, there are equilibria where the mass of consumers sharing their data is $n < n^*$. In such equilibria, intermediaries offer zero compensations to consumers in $(n, 1]$, which are rejected on the equilibrium path. Consumer $i \in (n, 1]$ accepts these offers only in off-path information sets where an intermediary deviates to acquire data from i by offering high compensations. The transaction cost eliminates offers with zero compensations and singles out the monopoly level of data sharing and compensation.

7 Extension: Multidimensional Personal Data

I extend the model in Section 3 by allowing a richer structure of personal data. This formulation allows me to find a condition under which the concentration of data to a small number of intermediaries is detrimental to consumers.

To simplify exposition, I assume that there are finitely many consumers $1, \ldots, N$. (Abusing notation, in this section, I use N to mean both the number and the set of consumers.) Formally, each consumer $i \in N$ is now endowed with a finite set $\mathcal{D}_i$ of characteristics, such as her name, location, income, and browsing histories. *Personal data* of consumer i refer to any set $D_i \subset \mathcal{D}_i$, and a *dataset* refers to any set $D \subset \mathcal{D} := \cup_{i \in N} \mathcal{D}_i$.

The firm obtains a revenue of $\Pi(D)$ from dataset D. The firm's revenue may depend on whose and what pieces of personal data it holds. Formally, Π satisfies the following, which is a natural generalization of Assumption 2:

Assumption 4. $\Pi : 2^{\mathcal{D}} \to \mathbb{R}$ satisfies the following.

1. Π is increasing: $\Pi(X) < \Pi(Y)$ for any $X, Y \subset \mathcal{D}$ such that $X \subset Y$ and $X \neq Y$.

2. Π is submodular: For all $X, Y \subset \mathcal{D}$ with $X \subset Y$ and $d \in \mathcal{D} \setminus Y$, it holds that

$$\Pi(X \cup \{d\}) - \Pi(X) \geq \Pi(Y \cup \{d\}) - \Pi(Y).$$

3. $\Pi(\emptyset) = 0$.

The payoff of each consumer $i \in N$ is the sum of $u_i(D_i, D_{-i})$ and total compensation from intermediaries, where $D_i \subset \mathcal{D}_i$ is the data i shares with the firm and $D_{-i} \subset \mathcal{D}_{-i} := \cup_{j \neq i} \mathcal{D}_j$ is the data shared by other consumers. For simplicity, I drop the subscript i from u_i and write it as u. u satisfies the following.

Assumption 5. $u : 2^{\mathcal{D}_i} \times 2^{\mathcal{D}_{-i}} \to \mathbb{R}$ satisfies the following.

1. u is decreasing in D_i: For any $Z \subset \mathcal{D}_{-i}$ and for any $X, Y \subset \mathcal{D}_i$ such that $X \subset Y$ and $X \neq Y$, $u(X, Z) > u(Y, Z)$.

2. u is submodular in D_i: For any $Z \subset \mathcal{D}_{-i}$ and for all $X, Y \subset \mathcal{D}_i$ with $X \subset Y$ and $d \in \mathcal{D}_i \setminus Y$, it holds that

$$u(X \cup \{d\}, Z) - u(X, Z) \geq u(Y \cup \{d\}, Z) - u(Y, Z). \tag{2}$$

Point 1 is a natural extension of Point 1 in Assumption 1. Point 2 is increasing marginal disutility from sharing personal data. This is a key assumption to later derive the result that the concentration of data hurts consumers. Also, it simplifies the derivation of equilibrium compensations. Nonetheless, the validity of Point 2 would depend on a specific application.

The action space of each intermediary is extended as follows. In the upstream market, an offer from intermediary k to consumer i is $(\tau_i^k, D_i^k) \in \mathbb{R} \times 2^{\mathcal{D}_i}$, which is a pair of compensation τ_i^k and personal data $D_i^k \subset \mathcal{D}_i$. Each consumer i decides whether to share D_i^k with intermediary k in exchange for τ_i^k.[16] Because there are finitely many consumers, the total compensation k pays is given by $\sum_{i \in N_k} \tau_i^k$ where N_k is the set of consumers accepting k's offer.

The interaction in the downstream market remains the same. Each intermediary simultaneously posts a price for its data, and the firm decides a set of intermediaries from which it buys data.

[16] I assume that a consumer cannot choose to share a strict subset of D_i^k.

Finally, to simplify the analysis, I assume that the marginal contribution of data to the firm's revenue is sufficiently high relative to the marginal disutility of sharing data:

Assumption 6. For any $i \in N$, $D_i \subset \mathcal{D}_i$, and $D_{-i} \subset \mathcal{D}_{-i}$,

$$\Pi(\mathcal{D}_i \cup D_{-i}) - \Pi\left((\mathcal{D}_i \setminus D_i) \cup D_{-i}\right) > |u(\mathcal{D}_i, D_{-i}) - u(\mathcal{D}_i \setminus D_i, D_{-i})|.$$

7.1 Equilibrium Analysis

Monopoly Intermediary

Under Assumption 6, the case of monopoly is straightforward.

Claim 2. *If $K = 1$, in equilibrium, each consumer i shares all data $\mathcal{D}_i$ and earns compensation $|u(\mathcal{D}_i, \mathcal{D}_{-i}) - u(\emptyset, \mathcal{D}_{-i})|$.*

Proof. Suppose that the monopoly intermediary obtains data $\hat{D}_j$ from each consumer j and $\hat{D}_i \subsetneq \mathcal{D}_i$ for some i. Then, it can strictly increase its payoff by additionally obtaining data $\mathcal{D}_i \setminus \hat{D}_i$ by raising compensation to i by $\Delta\hat{\tau}_i = |u(\mathcal{D}_i, D_{-i}) - u(\hat{D}_i, D_{-i})|$. This is because $\Delta\hat{\tau}_i < \Pi(\cup_{j \in N} \hat{D}_j \cup \mathcal{D}_i) - \Pi(\cup_{j \in N} \hat{D}_j)$ by Assumption 6. Thus, in any equilibrium, all consumers share all data. The lowest compensation the intermediary has to pay is $\tau_i^* = |u(\mathcal{D}_i, \mathcal{D}_{-i}) - u(\emptyset, \mathcal{D}_{-i})|$. This completes the proof. $\square$

Competing Intermediaries

Now, suppose $K \geq 2$. The following lemma generalizes Lemma 1. The proof is in Appendix E.

Lemma 3. *Suppose that each intermediary $k \in K$ holds data $D^k \subset \mathcal{D}$. This subgame has a subgame perfect equilibrium in which intermediary k posts a price of*

$$\Pi_k := \Pi\left(\bigcup_{j \in K} D^j\right) - \Pi\left(\bigcup_{j \in K \setminus \{k\}} D^j\right)$$

for its data, and the firm buys data from all intermediaries.

Before proceeding to the next result, I extend the various notions about allocation of data introduced in Section 5 in a straightforward manner:

Definition 4. An *allocation of data* is any element $(D^1, \ldots, D^K)$ of $(2^{\mathcal{D}})^K$, and its *size* refers to $|\cup_{k\in K} D^k|$. An allocation of data is *disjoint* if $D^j \cap D^k = \emptyset$ for any $j, k \in K$ with $j \neq k$. Let $(D^k)_{k\in K}$ and $(\hat{D}^k)_{k\in K}$ denote two disjoint allocations of data with the same size. $(\hat{D}^k)_k$ is *more concentrated* than $(D^k)_k$ if for any k, there is ℓ such that $D^k \subset \hat{D}^\ell$.

The following result, which generalizes a part of Theorem 1, states that the lowest equilibrium compensation from intermediary k to consumer i is equal to the marginal disutility that k's acquisition of i's data imposes on her.

Lemma 4. *Take any disjoint allocation of data $(D^1, \ldots, D^K)$ that can arise in some equilibrium. Define $D_i^k := \mathcal{D}_i \cap D^k$, $D_i = \cup_{k\in K} D_i^k$, and $D_{-i} = \cup_{j\in N\setminus\{i\}} D_j$, and $D = \cup_{k\in K} D^k$. The lowest equilibrium compensation consistent with $(D^k)_k$ is such that intermediary k pays consumer i a compensation of*

$$\tau_i^k = \left| u(D_i, D_{-i}) - u(\cup_{j\in K\setminus\{k\}} D_i^j, D_{-i}) \right|. \tag{3}$$

Proof. Take any $i \in N$. First, consumer i is willing to share D_i^k with each intermediary k at compensation τ_i^k. This is because for any (possibly empty) set K' of intermediaries with which i shares her data, it holds that $\tau_i^k \geq u(\cup_{j\in K'\setminus\{k\}} D_i^j, D_{-i}) - u(\cup_{j\in K'} D_i^j, D_{-i})$ due to the submodularity of u. Thus, i can maximize her payoff by sharing her data with all intermediaries.

Second, each intermediary has no incentive to unilaterally deviate from (τ_i^k). In particular, if intermediary k lowers compensation to i, she rejects the offer of k. Intermediary k can save at most τ_i^k with this, but it loses revenue $\Pi(\mathcal{D}) - \Pi(\mathcal{D} \setminus D_i^k) > \tau_i^k$, where the inequality is by Assumption 6. Therefore, no intermediary prefers to deviate from (τ_i^k).

Third, take any profile of compensations $(\hat{\tau}_i^k)$ such that $\hat{\tau}_i^k < \tau_i^k$ for some (i, k) with $D_i^k \neq \emptyset$. Then, i does not share her data with at least one intermediary k such that $D_i^k \neq \emptyset$. This is because i obtains a strictly greater payoff by sharing data with $K \setminus \{k\}$ than with K, by construction of τ_i^k. Thus, equation 3 is the minimum compensation sustaining $(D^1, \ldots, D^K)$. $\square$

The following result shows that under Assumption 6, any disjoint allocation of data can arise in equilibrium with the compensation in Lemma 4.

Lemma 5. *For any disjoint allocation of data* $(D^k)_k$ *with size* $|\mathcal{D}|$ *and compensation* $(\tau_i^k)_{(i,k)\in N\times K}$ *satisfying equality (3), there is an equilibrium sustaining them.*

Proof. Take any allocation and compensations described in the lemma. Consider the following strategy profile. Intermediary k offers (D_i^k, τ_i^k) to consumer i whenever τ_i^k in equality (3) is positive. Let $N_k := \{i \in N : \tau_i^k > 0\}$. On the equilibrium path, each consumer i accepts (D_i^k, τ_i^k) if and only if $i \in N_k$. The equilibrium price is given by Lemma 3. The off-path behavior of each player is naturally defined.

I show that this is an equilibrium. Suppose that intermediary k deviates in the first stage and offers $(\hat{D}_i^k, \hat{\tau}_i^k)$ to $i \in N_k$. There can be multiple consumers affected by this deviation, but I consider a particular i and prove that the deviation cannot be profitable for k. Any deviation with $\hat{\tau}_i^k > \tau_i^k$ cannot be optimal, because it increases the total compensation without strictly increasing the price of data in the downstream market.

Suppose $\hat{\tau}_i^k < \tau_i^k$. The deviating offer is rejected if $D_i^k \subset \hat{D}_i^k$: Consumer i prefers to accept τ_i^j for $j \neq k$, but conditional on that she accepts the offers of all intermediaries but k, i prefers to reject the deviating offer of k, because compensation $\hat{\tau}_i^k$ is strictly lower than the additional disutility of sharing $\hat{D}_i^k$, which is equal to τ_i^k. Thus, assume $D_i^k \not\subset \hat{D}_i^k$. Because obtaining the data held by other intermediaries does not increase k's payoff, without loss of generality, assume $\hat{D}_i^k \subset D_i^k$ and $\hat{D}_i^k \neq D_i^k$.[17]

Define $X = D_i^k \setminus \hat{D}_i^k$. With this deviation, intermediary k can save the compensation by $|u(\mathcal{D}_i \setminus X, \mathcal{D}_{-i}) - u(\mathcal{D}_i, \mathcal{D}_{-i})|$. However, the price of k's data in the downstream market decreases by at least $\Pi(\mathcal{D}) - \Pi(\mathcal{D} \setminus X)$. The latter is greater than the former by Assumption 6. Thus, an intermediary k has no profitable deviation to $i \in N_k$. Finally, intermediary k has no profitable

[17] To see this, suppose that intermediary k offers $\hat{D}_i^k$ such that $\hat{D}_i^k \cap (\cup_{j\neq k} D_i^j) \neq \emptyset$. Compare this with the case where intermediary k offers $\tilde{D}_i^k := \hat{D}_i^k \setminus \cup_{j\neq k} D_i^j$ without changing compensation. First, if i accepts $\tilde{D}_i^k \subset D_i^k$, the submodularity of u implies that she also accepts all D_i^j with $j \neq i$. This implies that i facing $\hat{D}_i^k$ can maximize her payoff by accepting $\hat{D}_i^k$ and D_i^j $(j \neq i)$, because i facing $\hat{D}_i^k$ never obtains a strictly greater payoff than i facing $\tilde{D}_i^k$. Second, if i rejects $\tilde{D}_i^k$, the submodularity of u implies that i accepts all D_i^j with $j \neq i$. This again implies that i facing $\hat{D}_i^k$ can maximize her payoff by accepting D_i^j $(j \neq i)$. Thus, consumer i's response is the same between offering $\hat{D}_i^k$ and $\tilde{D}_i^k$.

deviation that affects $i \notin N_k$ alone, because it merely increases the total compensation that k has to pay. $\square$

Recall that in Theorem 1 where each consumer has a single piece of personal data, there is no systematic relationship between the equilibrium allocation of data and consumer surplus. In contrast, the increasing marginal disutility of sharing data implies that concentration of data lowers consumer surplus:

Proposition 4. *In equilibrium described in Lemma 5, consumer surplus is lower in equilibrium where the allocation of data is more concentrated.*

Proof. If each intermediary k pays the compensation in equality (3) to each consumer i, the total compensations paid to consumers is

$$\sum_{i \in N, k \in K} \left| u(D_i, \mathcal{D}_{-i}) - u(\cup_{j \in K \setminus \{k\}} D_i^j, \mathcal{D}_{-i}) \right|. \tag{4}$$

Because $u(X, \mathcal{D}_{-i})$ is submodular in $X \subset \mathcal{D}_i$, (4) is greater under more fragmented allocation of data by the same calculation as Proposition 2 combined with Lemma 3 and Lemma 4. $\square$

8 Other Extensions

This section maintains the original formulation of Section 3 that there is a continuum of consumers.

8.1 When the Use of Personal Data Benefits Consumers

So far, I have assumed that the use of personal data by the downstream firm negatively affects consumers. In reality, sharing personal data may also bring consumers benefits, such as personalized services and products. To capture such a situation, I replace Assumption 1 with the following.

Assumption 7. $\forall n \in [0, 1]$, $u(1, n) > u(0, n)$.

Assumption 7 does not change the equilibrium pricing in the downstream market (Lemma 1), but it changes the interaction in the upstream market: An intermediary now may offer a negative compensation, which is interpreted as a fee to transfer personal data. Thus, the relevant question is whether competing intermediaries have an incentive to lower fees. The following result shows that the answer is yes. The proof is relegated to Appendix F.

Proposition 5. *When the firm's use of data benefits consumers, competition among intermediaries can benefit consumers. Formally, under Assumption 7, the following holds.*

1. *If there is a monopoly intermediary, in any equilibrium, almost every consumer shares her data, pays a positive fee, and obtains a payoff of* $u(0,1)$*;*

2. *If there are at least two intermediaries, there is an equilibrium in which every consumer shares her data. In any such equilibrium, each consumer receives a non-negative compensation and obtains a payoff of at least* $u(1,1) > u(0,1)$.

The intuition is as follows. For simplicity, suppose that intermediaries are restricted to offering nonnegative fees. Because consumers never benefit from accepting multiple offers with nonnegative fees, consumers can credibly share their data with at most one intermediary. Then, intermediaries compete for data by lowering fees as if data are rivalrous. As a result, the existence of multiple intermediaries benefits consumers. The result contrasts with the baseline model, where intermediaries have no incentive to raise compensations due to the non-rivalry of data. This illustrates that the way in which (downstream) firms use personal data could be a key parameter to understand how competition in markets for data works.

Finally, the result suggests that there are multiple equilbria if $K \geq 2$. Indeed, by the same logic as Theorem 1, I can show that any compensation between 0 and $\Pi'(1)$ can be sustained in equilibrium. However, as in Proposition 3, only the equilibrium with zero fees remains if it is costly for intermediaries to make offers.

8.2 Multiple Downstream Firms

The model can readily take into account multiple downstream firms if they do not interact with each other. To see this, suppose that there are L firms, where firm $\ell \in L$ has revenue function Π_ℓ that depends only on the amount of data available to ℓ. Each consumer i's payoff (without transfer) is $\sum_{\ell \in L} u_\ell$, where u_ℓ depends on whether firm ℓ has i's personal data and how many of other consumers share their data with ℓ.

This setting is equivalent to the case of a single firm. To see this, first, Lemma 1 implies that each intermediary k posts a price of $\Pi_\ell(\cup_k N_k) - \Pi_\ell(\cup_{j \neq k} N_k)$ to firm ℓ in the downstream market. Note that I implicitly assume that intermediaries can price discriminate firms.

Given this pricing rule, the revenue of intermediary k given the allocation of data $(N_k)_k$ is $\sum_{\ell \in L}[\Pi_\ell(\cup_k N_k) - \Pi_\ell(\cup_{j \neq k} N_k)]$. By setting $\Pi := \sum_{\ell \in L} \Pi_\ell$, we can calculate the equilibrium revenue of each intermediary in the downstream market as in Lemma 1.

Second, intermediaries in this paper cannot commit to not sell personal data to downstream firms. Thus, once a consumer shares her data with one intermediary, the data is sold to all firms. This means that in equilibrium, each consumer i decides whether to accept an offer by comparing $\sum_{\ell \in L} u_\ell(1, n') - \sum_{\ell \in L} u_\ell(0, n')$ with the compensation.[18] Therefore, by setting $u := \sum_{\ell \in L} u_\ell$, we can apply the same analysis as before. Note that this extension can accommodate the case where some firms impose negative externalities and some impose positive externalities on consumers, because the analysis only requires that $u(1, n) > u(0, n)$ for all $n \in [0, 1]$.

It could be interesting to consider the case where an intermediary can commit to not sell data to some of the downstream firms (say, in the form of a "privacy policy") when it acquires data from consumers. In this case, an intermediary may profit from committing to not sell data to firms whose value of data exceeds the negative externality it imposes on consumers.

9 Conclusion

One important question in the data economy would be whether consumers are properly compensated in return for sharing their personal data with firms. I study a simple model of markets for personal data in which transfer of data between consumers and firms are facilitated by data intermediaries. The model gives a negative answer to the question. It shows that data intermediaries may have no incentive to compete over personal data, because competition in the upstream market lowers the price of data in the downstream markets.

Arguably, the model is highly stylized and abstracts from various important elements of personal data sharing (such as consumer unawareness). Nonetheless, I believe that the model highlights an economically important difference between competition for personal data and that for

[18] n' is the mass of consumers who share their data in a candidate strategy profile.

physical goods.

References

Anderson, Simon P and Stephen Coate (2005), "Market provision of broadcasting: A welfare analysis." *The Review of Economic studies*, 72, 947–972.

Armstrong, Mark (2006), "Competition in two-sided markets." *The RAND Journal of Economics*, 37, 668–691.

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E Glen Weyl (2018), "Should we treat data as labor? Moving beyond "Free"." In *AEA Papers and Proceedings*, volume 108, 38–42.

Bergemann, Dirk and Alessandro Bonatti (2019), "Markets for information: An introduction." *Annual Review of Economics*, 11, 1–23.

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin (2018), "The design and price of information." *American Economic Review*, 108, 1–48.

Caillaud, Bernard and Bruno Jullien (2003), "Chicken & egg: Competition among intermediation service providers." *RAND journal of Economics*, 309–328.

Carrascal, Juan Pablo, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira (2013), "Your browsing behavior for a big mac: Economics of personal information online." In *Proceedings of the 22nd international conference on World Wide Web*, 189–200, ACM.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim (2018), "Privacy and personal data collection with information externalities."

Federal Trade Commission (2014), "Data brokers: A call for transparency and accountability." *Washington, DC*.

Jones, Charles, Christopher Tonetti, et al. (2018), "Nonrivalry and the economics of data." In *2018 Meeting Papers*, 477, Society for Economic Dynamics.

Radner, Roy and Joseph Stiglitz (1984), "A nonconcavity in the value of information." *Bayesian models in economic theory*, 5, 33–52.

Reisinger, Markus (2012), "Platform competition for advertisers and users in media markets." *International Journal of Industrial Organization*, 30, 243–252.

Rochet, Jean-Charles and Jean Tirole (2003), "Platform competition in two-sided markets." *Journal of the european economic association*, 1, 990–1029.

Stahl, Dale O (1988), "Bertrand competition for inputs and walrasian outcomes." *The American Economic Review*, 189–201.

Varian, Hal (2018), "Artificial intelligence, economics, and industrial organization." In *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

Yanelle, Marie-Odile (1997), "Banking competition and market efficiency." *The Review of Economic Studies*, 64, 215–239.

Appendix

A Proof of Proposition 1

Proof. Consider the following strategy profile: The intermediary offers $|u(1,n^*) - u(0,n^*)|$ to consumer $i \leq n^*$, and offers zero to other consumer $i > n^*$. Also, it sets a price of $\Pi(n)$ at a node in which the intermediary has the data of size n. First, no consumer $i \leq n^*$ strictly benefits from unilaterally rejecting the offer. Also, no consumer $i > n^*$ benefits from accepting the offer of zero compensation.

Suppose that the intermediary deviates and buys data from consumers in $[0, n]$. If $n > n^*$, the total compensation the intermediary has to pay increases by at least $\Delta \cdot |u(1,n^*) - u(0,n^*)|$ where $\Delta = n - n^*$; in contrast, the revenue of the intermediary in the downstream market increases by

$\Pi(n^* + \Delta) - \Pi(n^*)$. By the definition of n^* in Assumption 3, such a deviation cannot be strictly profitable. (Recall that the deviation does not alter the belief of the consumer over the offers made to other consumers.) If $n < n^*$, the total compensation the intermediary has to pay decreases by at most $\Delta \cdot |u(1, n^*) - u(0, n^*)|$ where $\Delta = n^* - n$; in contrast, the revenue of the intermediary in the downstream market decreases by $\Pi(n^*) - \Pi(n^* - \Delta)$. Again, such a deviation cannot be strictly profitable. Thus, the proposed strategy profile is an equilibrium.

By the same logic, we can show that if the intermediary buys data from $n' > n^*$ mass of consumers under some strategy profile, it can strictly increase its payoff by not purchasing data from consumers in $(n' - \Delta, n')$ (say, by offering a negative compensation) because $\Pi'(n') < |u(1, n') - u(0, n')|$ implies that $\Pi'(n') - \Pi'(n' - \Delta) < \Delta \cdot |u(1, n') - u(0, n')|$ for a small $\Delta > 0$. Note that the left hand side is the revenue loss from not buying data from the consumer, and the right hand side is the compensation that the intermediary can save by not purchasing the data.

Similarly, if the intermediary buys data only from $n' < n^*$ consumers under some strategy profile, it can strictly increase revenue by purchasing data from consumers in $(n', n' + \Delta]$. Therefore, there is no equilibrium in which the intermediary buys data from mass $n' \neq n^*$ of consumers, which also uniquely pins down equilibrium compensations. $\square$

B Proof of Theorem 1

Proof. First, I prove that Point 2 implies Point 1. Take any allocation and compensations described in Point 2. Consider the following strategy profile. Intermediary k offers τ_i to consumer $i \in N_k$ and offers zero compensation to consumers in $N_{-k} := N \setminus N_k$. On the equilibrium path, consumers $i \in [0, n]$ accept the offer of only intermediary k with $i \in N_k$. The equilibrium prices in the downstream market are given by Lemma 1.

The off-path behaviors of consumers are as follows. Suppose that a consumer detects a deviation by any intermediary in the first stage. Then, the consumer accepts a set of offers to maximize her payoff, but here, the consumer accepts an offer if she is indifferent between accepting and rejecting it. Note that this is different from the on-path behavior, where each consumer rejects the offer of zero compensation by intermediary j such that $i \notin N_j$, although she is indifferent between rejecting and accepting it given that she accepts the offer of k with $i \in N_k$.

I show that this strategy profile is an equilibrium. First of all, it is optimal for (almost) every consumer to accept an offer because $u(1,n)+\tau_i \geq u(0,n)$. Second, suppose that intermediary k unilaterally deviates in the first stage and offers $\hat{\tau}_i^k$ to each consumer i. Consider a deviation such that $\hat{\tau}_i^k > 0$ for a positive mass of consumers $i \in N' \subset N \setminus N_k$. There are two cases to consider. First, if consumer $i \in N'$ rejects $\hat{\tau}_i^k$, intermediary k can obtain the same payoff by offering $\hat{\tau}_i^k = 0$. Second, if a positive mass of consumers accept $\hat{\tau}_i^k$, intermediary k has to pay a positive compensation. However, because these consumers accept the offers of other intermediaries as well (by construction of consumers' off-path behavior), the deviation does not increase k's revenue in the downstream market (Corollary 1). Thus, setting $\hat{\tau}_i^k = 0$ for all $i \in N'$ weakly increases intermediary k's payoff relative to the original deviation.

Thus, it is sufficient to consider deviations by intermediary k that only affect consumers $i \in N_k$, i.e., $\hat{\tau}_j^k = 0$ for any $j \notin N_k$. Consider any such deviation, and let $N_D := \left\{i \in N_k : \hat{\tau}_i^k \neq \tau_i\right\}$. N_D is the set of consumers in N_k who receive different offers from τ_i, where τ_i is the on-path level of compensation to i. Let N_{D_1} denote the set of consumers in N_D who accept no offer as a result of the deviation, and let N_{D_2} denote the set of consumers in N_D who accept at least one offer. Because all intermediaries other than k offer zero compensations to consumers in N_D, consumers in N_{D_2} accept the offer of intermediary k; moreover, by the way I define the off-path behavior, these consumers also accept offers of other intermediaries. Note that $|N_D| = |N_{D_1}| + |N_{D_2}|$, and the mass of consumers whose personal data are bought only by intermediary k (i.e., consumers who are not affected by the deviation of k) is $|N_k| - |N_{D_1}| - |N_{D_2}|$. I show that the revenue of intermediary k in the downstream market decreases by more than the total compensation that k can save. First, the total compensation that k pays in the upstream market decreases by at most $N_D \cdot \Pi'(n)$, because $\tau_i \leq \Pi'(n)$ on the equilibrium path. Next, consider the equilibrium price of data held by k. After the deviation, the firm's revenue from (aggregate) data becomes $\Pi(N^* - N_{D_1})$, where $N^* := [0,n]$. Without intermediary k's data, it would be $\Pi(N^* - N_{D_1} - (N_k - N_{D_1} - N_{D_2}))$, because now only $N_k - N_{D_1} - N_{D_2}$ is exclusive to k's data. By Lemma 1, the deviation decreases

k's revenue in the downstream market by

$$
\begin{aligned}
&\underbrace{\Pi(N^*)-\Pi(N^*-N_k)}_{\text{revenue without deviation}}-\underbrace{[\Pi(N^*-N_{D_1})-\Pi(N^*-N_{D_1}-(N_k-N_{D_1}-N_{D_2})]}_{\text{revenue with deviation}}\\
=&\Pi(N^*)-\Pi(N^*-N_{D_1})+\Pi(N^*-N_k+N_{D_2})-\Pi(N^*-N_k)\\
\geq&\Pi(N^*)-\Pi(N^*-N_{D_1})+\Pi(N^*)-\Pi(N^*-N_{D_2})\\
\geq&(N_{D_1}+N_{D_2})\cdot\Pi'(n)\\
=&N_D\cdot\Pi'(n)
\end{aligned}
$$

Here, the first and the second inequalities are by Point 2 of Assumption 2. Therefore, the deviation cannot be strictly profitable for k. We can also verify that a unilateral deviation by a consumer, an intermediary in the downstream market, and the firm cannot be profitable.

Next, I show that Point 1 implies Point 2. First, suppose to the contrary that there is an equilibrium in which consumers in $[0,n]$ with $n>n^*$ share their data. Then, by Assumption 1, the minimum (total) compensation that any consumer $i\in[0,n]$ has to receive is $|u(1,n)-u(0,n)|$. Lemma 2 implies that there is a measure zero set D such that for each $i\in[0,n]\setminus D$, exactly one intermediary gives such an offer to i. Without loss of generality, suppose that intermediary 1 buys data from consumers in $[0,\Delta]$ with $\Delta>0$. Then, intermediary 1 can strictly increase its payoff by giving an offer of zero to consumers in $[0,\delta]$ with a small $\delta\in(0,\Delta)$. Indeed, this deviation increases its payoff by $\delta|u(1,n)-u(0,n)|-[\Pi(n)-\Pi(n-\delta)]>0$. This establishes that there is no equilibrium in which the mass of consumers sharing their data is strictly more than n^*. Thus, it holds that $\cup_k N_k=[0,n]$ for $n\leq n^*$.

Next, I show that the compensation that each consumer i earns satisfies the inequalities (described in the theorem) for almost every $i\in[0,1]$. Note that for almost every $i\in[0,n]$, Lemma 2 implies that only one intermediary, say k, offers i a positive compensation $\tau_i^k>0$. Suppose that intermediary k offers $\tau_i^k>\Pi'(n)$ to a positive mass of consumers. Then, intermediary k can profitably deviate by offering the compensation to zero to a small but positive mass (say ε) of those consumers. Indeed, this increases the payoff of intermediary k by $\varepsilon\tau_i^k-[\Pi(n)-\Pi(n-\varepsilon)]>0$. Also, if $\tau_i^k<|u(1,n)-u(0,n)|$, then consumer i would reject the offer. Thus, $\tau_i^k\in[u(0,n)-u(1,n),\Pi'(n)]$ for almost every $i\in[0,n]$ in equilibrium. Almost every consumer in $(n, 1]$ earns no compensation because she reject any offer. Finally, if the allocation is not disjoint, some intermediary k such that $|N_k \cap N_j| > 0$ for some $j \neq k$ can strictly increase its payoff by offering zero compensation to consumer $i \in N_k \cap N_j$. Thus, any equilibrium allocation must be disjoint. $\square$

C Proof of Claim 1

In this appendix, I prove Claim 1 and the existence of an equilibrium.

Proof of Claim 1. Take any $K \geq 2$ and suppose to the contrary that there is an equilibrium in which one intermediary, say 1, obtains a positive payoff. Let N_1 denote the set of consumers from whom intermediary 1 buys data. (As before, we consider pure-strategy equilibrium. Thus, N_1 is deterministic.) Suppose that intermediary 2 deviates and offers each consumer $i \in N_1$ a compensation of $\tau_i^1 + \varepsilon$, where τ_i^1 is the compensation by intermediary 1. Then, all consumers in N_1 accept the offer of only intermediary 2. In the downstream market, the revenue of intermediary 2 increases from $\Pi(N^*) - \Pi(N^* - N_2)$ to $\Pi(N^*) - \Pi(N^* - N_1 - N_2)$, which yields a net gain of $\Pi(N^* - N_2) - \Pi(N^* - N_1 - N_2)$. By Assumption 2, $\Pi(N^* - N_2) - \Pi(N^* - N_1 - N_2) \geq \Pi(N^*) - \Pi(N^* - N_1)$. As intermediary 1 obtains a positive payoff without 2's deviation, $\Pi(N^*) - \Pi(N^* - N_1) - \int_{N_1} \tau_i^1 di > 0$, which implies $\Pi(N^* - N_2) - \Pi(N^* - N_1 - N_2) - \int_{N_1} (\tau_i^1 + \varepsilon) di > 0$ for a small $\varepsilon > 0$. Thus, intermediary 2 has a profitable deviation, which is a contradiction. $\square$

Existence of a Pure-Strategy Equilibrium

To see that a pure-strategy equilibrium exists when data are rivalrous, consider the following strategy profile: All intermediaries offer compensation $\Pi'(i)$ to consumer $i \in [0, n^*]$ and zero compensation to consumer $i \in (n^*, 1]$. Each consumer $i \leq n^*$ accepts the offer of intermediary 1. Each player's action in other information sets are naturally defined. In particular, the equilibrium pricing in the downstream market is given by Lemma 1.

Clearly no consumer has a profitable deviation. Thus, I show that no intermediary has a profitable deviation. Suppose that intermediary $k \geq 2$ deviates and obtains data $D_1 \subset [0, n^*]$ and $D_2 \subset (n^*, 1]$. The deviation increases the revenue by $\Pi(n^* + D_2) - \Pi(n^* - D_1)$ in the downstream market, but k has to pay the compensation of weakly greater than $\int_{n^*-D_1}^{n^*+D_2} \Pi'(i) di \geq$

$\Pi(n^* + D_2) - \Pi(n^* - D_1)$. This is because any consumer $i > n^*$ requires a compensation of at least $|u(1, n^*) - u(0, n^*)| \geq \Pi'(i)$.

Finally, suppose that intermediary 1 deviates, so that it does not obtain data $D_1 \subset [0, n^*]$ and obtains data $D_2 \subset (n^*, 1]$. Following this deviation, $i \in D_1$ provides her data exclusively to another intermediary. Thus, the deviation reduces the total compensation 1 has to pay by at most $\int_{D_1} \Pi'(i)di - |D_2| \cdot |u(1, n^*) - u(0, n^*)|$ but decreases the revenue in the downstream market by $\Pi(n^*) - [\Pi(n^* + D_2) - \Pi(D_1)]$. Thus, to show that the deviation is not profitable, it is enough to show that

$$\Pi(n^*) - [\Pi(n^* + D_2) - \Pi(D_1)] \geq \int_{D_1} \Pi'(i)di - |D_2| \cdot |u(1, n^*) - u(0, n^*)|. \tag{5}$$

First, the concavity of Π implies that $\Pi(D_1) = \int_0^{|D_1|} \Pi'(i)di \geq \int_{D_1} \Pi'(i)di$. Second, $|D_2| \cdot |u(1, n^*) - u(0, n^*)| = |D_2| \cdot \Pi'(n^*) \geq \Pi(n^* + D_2) - \Pi(n^*)$. These inequalities imply inequality (5). Thus, the deviation cannot be profitable.

D Proof of Proposition 3

Proof. Let $\tau_i^k = \emptyset$ denote intermediary k's action of making no offer to consumer i. To see that an equilibrium exists, consider the following strategy profile: For each $i \leq n^*$, $\tau_i^1 = \tau^*$; for each $i > n^*$, $\tau_i^1 = \emptyset$; for each $k \neq 1$ and $i \in N$, $\tau_i^k = \emptyset$. On the equilibrium path, consumers $i \in N^* := [0, n^*]$ accept offers from intermediary 1. The equilibrium price in the downstream market is given by Lemma 1. If an intermediary deviates in the first stage, each consumer accepts a set of offers to maximize her payoff. By the same argument as Theorem 1 where I show that Point 2 implies Point 1, we can confirm that this consists of an equilibrium.

Now, take any equilibrium where consumers $[0, n^*]$ share their data. I show that consumers $i \leq n^*$ and $i > n^*$ earn compensations τ^* and zero, respectively. First, any consumer $i > n^*$ does not receive compensation because she does not share her data. Second, suppose to the contrary that there is an equilibrium in which a positive mass of consumers in $D \subset [0, n^*]$ receives a compensation $\tau > \tau^*$ from, say, intermediary 1. Suppose that intermediary 1 unilaterally deviates and offers all consumers $i \in D$ a compensation of $\tau' \in (\tau, \tau^*)$. Suppose that for a positive mass of consumers in D, τ' is the only offer they receive. Let D' denote the set of those consumers. Then they continue to accept the offer. However, this means that intermediary 1 can profitably deviate by lowering offers to consumers in D', which is a contradiction. Thus, $|D'| = 0$. In other words, (almost) every consumer in D provides her data to another intermediary $k \neq 1$ following the deviation. By the same logic as Lemma 2, we can show that such intermediary k must be offering a non-positive compensation. However, this is a contradiction: On the equilibrium path where intermediary 1 offers τ, intermediary k can strictly increase its payoff by not sending an offer to consumers in D. Thus, intermediary 1 can profitably deviate by lowering the compensation, which is a contradiction.

Next, I show that there is no equilibrium where mass $n \neq n^*$ of consumers share their data. First, take any equilibrium, and suppose that mass $n < n^*$ of consumers share their data. Suppose that intermediary 1 deviates and offers consumers $(n, n + \Delta]$ a compensation strictly greater than but close to $|u(1, n) - u(0, n)|$. Consumer $i \in (n, n + \Delta]$ accepts this offer, and importantly, intermediary 1 is the only one that acquires the data of consumer i. Indeed, given the transaction costs, there cannot be other intermediaries making offers, which are rejected for sure in the proposed equilibrium. Thus, intermediary 1's deviation increases its payoff by (arbitrarily close to) $\Pi(n + \Delta) - \Pi(n) - \Delta|u(1, n) - u(0, n)| > 0$, which is a contradiction. We can also show that there is no equilibrium in which mass $n > n^*$ of consumers provide their data, in the same way as the proof of Theorem 1. $\square$

E Proof of Lemma 3

Proof. Take any $(D^1, \ldots, D^K)$. Consider a strategy profile in which each intermediary $k \in K$ sets a price of Π_k and the firm buys data from all intermediaries. First, it is optimal for the firm to buy all data: Point 2 of Assumption 4 implies that $\Pi(\cup_{j \in K' \cup \{k\}} D^j) - \Pi(\cup_{j \in K'} D^j) - \Pi_k \geq 0$ for any $K' \subset K$. Thus, the firm is willing to buy D^k at price Π_k regardless of the prices posted by other firms. Second, if intermediary k unilaterally deviates and sets a price of $p_k > \Pi_k$, the firm strictly prefers to buy data from intermediaries in $K \setminus \{k\}$, and thus k cannot benefit by raising a price. Finally, any price $p_k < \Pi_k$ strictly lowers the payoff of intermediary k. $\square$

F Proof of Proposition 5

Proof. To show Point 1, suppose that the market consists of a monopoly intermediary. Take any equilibrium, and suppose to the contrary that the mass of consumers sharing their data is $n < 1$. Then, the intermediary can strictly increase its payoff by offering a compensation of 0 to consumers in $(n, 1]$. This increases the net payoff of the intermediary by $\Pi(1) - \Pi(n) > 0$. Note that all consumers in $(n, 1]$ accept this offer. This shows that in any equilibrium, almost every consumer shares her data. Now, if a positive mass of consumers (in the set D) obtain a payoff of strictly greater than $u(0, 1)$, it means that the intermediary charges a fee of strictly less than $u(1, 1) - u(0, 1) > 0$. Then, the intermediary can strictly increase its payoff by slightly increasing the fees offered to consumers in D, which is a contradiction. Thus, in any equilibrium, almost every consumer shares her data and pays a fee of $u(1, 1) - u(0, 1)$. Finally, it is straightforward to show that it is indeed an equilibrium that the intermediary offers a fee of $u(1, 1) - u(0, 1)$ to all consumers, all of whom accept the offer.

Second, I show Point 2. First, I show that there is an equilibrium where all consumers share their data. Consider the strategy profile where all intermediaries offer zero fee to all consumers, who accept all offers; if intermediary k unilaterally deviates, consumers who are affected by the deviation share their data with all intermediaries $j \neq k$, and they share their data with k if and only if k offers a non-positive fee. First, the strategy of each consumer is optimal both on and off the equilibrium paths because accepting zero fee increases their payoffs by $u(1, 1) - u(0, 1) > 0$. Second, no intermediary has an incentive to deviate, because it either obtains no data or obtains data that other intermediaries hold. Therefore, the proposed strategy profile is an equilibrium.

Next, suppose to the contrary that there is an equilibrium in which all consumers share data but a positive mass of consumers in the set D obtain payoffs strictly lower than $u(1, 1)$. This means that consumers in D pay positive fees. Without loss of generality, suppose that intermediary 2 charges positive fees to all consumers in D. Then, if intermediary 1 can deviate and offers them a fee of zero, consumers in D share their data *only* with 1. This strictly benefits intermediary 1 because its payoff in the downstream market increases by at least $\Pi(1) - \Pi(1 - |D|) > 0$. Therefore, all consumers share their data with zero or lower fees and obtain payoffs of at least $u(1, 1)$. $\square$