Natural Monopoly for Data Intermediaries

Shota Ichihashi*

January 4, 2019

Abstract

I study how the market structure of data intermediaries—technology companies and data brokers that collect and distribute personal data—affects consumer surplus and the welfare of 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 competing intermediaries have no incentive to raise compensations: 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 firm's use of personal data positively affects consumers, competition among intermediaries improves consumer welfare.

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 data; 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.

This paper studies how the market structure of data intermediaries—technology companies and data brokers that collect and distribute personal data—affects consumers and the welfare of 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 tackle 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 the price competition among intermediaries in the downstream market.

The following examples illustrate the main idea of this paper by contrasting intermediaries for 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 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 (compensation) 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 over data, and 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 may not benefit the consumer.[1]

To generalize these examples, in Section 3, I consider a model of N consumers, K data intermediaries, and a single downstream firm. Intermediaries offer compensations to consumers to obtain their personal 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 a consumer depends on whether the firm acquires her data and the number of other 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 that each consumer earns. Importantly, for any number of intermediaries, there is a "monopoly equilibrium" where the maximum number of consumers give up their data at the minimum level of compensation. An intuition is similar to Example 2: If intermediaries compete over data by offering high compensations, consumers will share their data with multiple intermediaries, because data are non-rivalrous. However, this will lower the price of data in the downstream market. Thus, inter-

[1] This is not a unique equilibrium: Theorem 1 implies that any compensation between \$50 and \$70 can be sustained in equilibrium; 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 conusmer.

mediaries have no incentive to raise compensations beyond the minimum acceptable level. I also show that the industry profit of intermediaries is lower in equilibrium where data are fragmented across intermediaries, because each of them has small pricing power against 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 maximum number 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 a small cost to make an offer to a consumer.

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 multiple intermediaries is sustainable, and competition 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. I also show that consumer surplus is low if data are concentrated to a small number of intermediaries.

In conclusion, the model suggests that competition among data intermediaries may not occur: Competing over personal data by offering high compensations decreases the price of the data in downstream markets, because high compensations encourage consumers to share their data with multiple intermediaries due to the non-rivalry of data. Therefore, competition among intermediaries is unattainable, and this leads to the monopoly level of compensation in equilibrium.

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 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). In contrast to Bergemann and Bonatti (2019), I do not consider 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. (2015) consider a model of data provision and data pricing. My paper adds to this literature by providing a model where the non-rivalry of personal data interacts with the market structure.[2]

Second, the paper is also related to the literature on intermediaries for physical goods. For example, Stahl (1988) studies a model of market markers who obtain stock from suppliers and resell it to buyers. My model resembles the model of Stahl (1988) 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, data intermediaries obtain the monopoly profit regardless of the market structure, due to the non-rivalry of data.

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

3 Model

I consider a model where data intermediaries buy personal data from consumers and sell them to a downstream firm. For now, I focus on the case in which 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. There are $N \geq 1$ consumers, $K \geq 1$ intermediaries, and a single downstream firm.[3] If it is clear from context, I use N or K also to mean the set of players. The game consists of two parts—intermediaries buy data in the upstream market and sell them in the downstream market.

Upstream Market

Each consumer $i \in N$ is endowed with some personal data, which can be, for example, her demographic characteristics or browsing histories. At the beginning of the game, each intermediary $k \in K$ simultaneously makes an *offer* $(\tau_1^k, \dots, \tau_N^k) \in \mathbb{R}^N$, where τ_i^k is the compensation that intermediary k pays consumer i in exchange for her personal data.[4] 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 $(\tau_i^1, \dots, \tau_i^K)$, 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 personal data with intermediary k. Hereafter, I call any subset of $\cup_k N_k$ as *data*. All intermediaries and the firm publicly observe $(N_1, \dots, N_K)$, which I call an *allocation of data* across intermediaries.

Downstream Market

Given the allocation of data $(N_1, \dots, N_K)$, 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.

All players are risk-netural, and their ex post payoffs are as follows. The payoff of an interme-

[3] 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.

[4] I assume that each intermediary makes an offer to all consumers, because making no offer is 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 Subsection 6.

diary 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 $\mathbf{1}_{\{k \in K'\}} p_k - \sum_{i \in N_k} \tau_i^k$, 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 how many consumers give up their data. Suppose that consumer i sells her data to intermediaries in K_i and the firm acquires the personal data of consumers in $D_{-i} \subset N \setminus \{i\}$. If the firm acquires the data of i, her payoff is $u(1, |D_{-i}|) + \sum_{k \in K_i} \tau_i^k$; if the firm does not, i's payoff is $u(0, |D_{-i}|) + \sum_{k \in K_i} \tau_i^k$. 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 $u : \{0, 1\} \times \mathbb{Z}_+ \to \mathbb{R}$. 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 one consumer affects the welfare of other consumers.

Assumption 1. For each $n \in \mathbb{Z}_+$, $u(1, n) < u(0, n)$.

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.[5] The second term is the total payments to intermediaries in K'. I assume that Π satisfies the following.

Assumption 2. $\Pi : \mathbb{Z}_+ \to \mathbb{R}$ satisfies the following properties.

1. $\Pi(n)$ is strictly increasing in n.

2. For any $z \in \mathbb{Z}_+$, $\Pi(n + z) - \Pi(n)$ is decreasing[6] in $n \in \mathbb{Z}_+$.

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. Point 2 is motivated by the observation that "data

[5] This excludes the case where the firm values the data of some consumers more than the data of other consumers. In Section 7, I consider a more general setting where the firm's revenue is a submodular set function of data.

[6] To simplify exposition, I use "decreasing" to mean "non-increasing." The same thing applies to other words such as "increasing" and "greater."

typically exhibits decreasing returns to scale like any other factor of production" (Varian, 2018).[7] Moreover, it enables me to characterize the equilibrium prices of data in the downstream market in a simple way. Point 3 is a normalization.

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

Assumption 3. There is a unique $N^* \in \{1, \ldots, N\}$ such that $\Pi(n+1) - \Pi(n) > |u(1,n) - u(0,n)|$ for all $n < N^*$ and $\Pi(n+1) - \Pi(n) < |u(1,n) - u(0,n)|$ for all $n \geq N^*$, where $\Pi(N+1) = -\infty$.

The assumption ensures the existence of pure-strategy equilibrium. It holds, for example, if a consumer incurs constant privacy cost and the firm's revenue exhibits constant returns to scale, i.e., $\exists(c, \rho)$, $\forall n$, $|u(1,n) - u(0,n)| = c < \rho = \Pi(n+1) - \Pi(n)$. N^* can be different from an efficient level, because this "first-order condition" ignores how acquiring the personal data of one consumer affects the welfare of other consumers.

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.

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 deviates and offers a different compensation to consumer i, the deviation does not change i's beliefs over the set of offers that intermediaries make to other consumers. To simplify exposition, I consider equilibrium such that if consumer $i \in N$ shares her personal data, any consumer $j \leq i$ also shares her data. From now on, I abbreviate PBE satisfying these conditions as "equilibrium."

4 Benchmark: Monopoly Data Intermediary

First, I consider a monopoly data intermediary ($K = 1$).

Proposition 1. *If the market consists of a monopoly intermediary, in any equilibrium, consumers $i \leq N^*$ provide their data where N^* is defined in Assumption 3. Any consumer $i \leq N^*$ obtains*

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

compensation $|u(1,N^*-1)-u(0,N^*-1)|$ *and payoff* $u(0,N^*-1)$*, while any consumer* $i>N^*$ *obtains zero compensation and payoff* $u(0,N^*)$*. The firm obtains a payoff of zero.*

Proof. Consider the following strategy profile: The intermediary offers $|u(1,N^*-1)-u(0,N^*-1)|$ to consumers $1,\ldots,N^*$, and offers zero to other consumers; Also, it sets a price of $\Pi(n)$ at a node in which the intermediary has the data of size n. First, no consumer in N^* strictly benefits from unilaterally rejecting the offer. Also, no consumer outside of N^* benefits from accepting the offer of zero compensation. If the intermediary deviates and buys data from a consumers outside of N^*, it has to pay at least $|u(1,N^*)-u(0,N^*)|$; in contrast, the revenue of the intermediary in the downstream market increases by at most $\Pi(N^*+1)-\Pi(N^*)$. 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.) Thus, the proposed strategy profile is an equilibrium.

By the same logic, we can show that if the intermediary buys data from $N'>N^*$ consumers, it can strictly increase its payoff by not purchasing data from one of N^* consumers (say, by offering a negative compensation) because $\Pi(N')-\Pi(N'-1)<|u(1,N'-1)-u(0,N'-1)|$. 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 has to pay her.

Similarly, if the intermediary buys data only from $N'<N^*$ consumers, it can strictly increase revenue by purchasing data from one of consumers in $N\setminus N'$ by offering $|u(1,N')-u(0,N')|+\varepsilon$ with a small $\varepsilon>0$, because $\Pi(N'+1)-\Pi(N')>|u(1,N')-u(0,N')|$. Therefore, there is no equilibrium in which the intermediary buys data from $N'\neq N^*$ consumers. □

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.[8] In an extreme case, if $u(1,n)=u(0,n)$ for all n but $u(1,n+1)-u(1,n)$ and $u(0,n+1)-u(0,n)$ are negative and large in magnitude, the intermediary acquires 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.

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

5 Competing Data Intermediaries

Now, I assume that the market consists of multiple intermediaries ($K \geq 2$). I solve the game in two steps. 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 $\Pi(|N'|)$.

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 Nash 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. If an intermediary deviates and sets a higher price, the firm buys data from all but the deviating intermediary. 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 are determined by 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 $i \in \cup_{j \neq k} N_j$. Then, the equilibrium revenue of intermediary k in the downstream market is identical between when it holds N_k and $N_k \cup \{i\}$.*

5.2 Natural Monopoly for Data Intermediaries

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

Lemma 2. *Take any equilibrium and any consumer $i \in N$ who shares her personal data.*[9] *Then, only one intermediary offers a positive compensation to consumer i.*

Proof. Lemma 1 implies that the downstream firm acquires consumer i's data. Suppose to the contrary that multiple intermediaries, say k and j, offer positive compensations to consumer i. Then, consumer i 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 number of other 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 consumer i 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 is a contradiction. $\square$

Lemma 2 generalizes the point made by Example 2 in the introduction: If intermediaries compete for data by offering high compensations, consumers will share their data with multiple intermediaries; however, this will lower the price of data in the downstream market. Thus, it cannot be an equilibrium that multiple intermediaries offer positive compensations to the same consumer.

[9] Recall that this means that consumer i shares her data with at least one intermediary. By Lemma 1, this is equivalent to the condition that the firm buys i's data.

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 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.[10] This strategy profile is an equilibrium. In particular, an intermediary has no incentive to deviate and acquire data by offering a high compensation, because consumers will share their data with other intermediaries. In this equilibrium, the firm and intermediaries obtain zero profits, and each consumer obtains $u(0,0)$, which can be higher or lower than the monopoly outcome.

Example 3 motivates the following equilibrium.

Definition 1. A *maximum revelation equilibrium (MRE)* is an equilibrium where the number of consumers who share their data[11] is maximized among all equilibria.

I focus on maximum revelation equilibrium for two reasons. First, it fits the casual observation that a large number of consumers seem to share their personal data with firms. Second, Section 6 shows that other equilibria (such as the one in Example 3) are not robust with respect to small transaction costs incurred by intermediaries.

Before characterizing maximum revelation equilibrium, I introduce two more definitions about allocation of data. Recall that an allocation of data, denoted by $(N_k)_k \in (2^N)^K$, specifies what data each intermediary holds.

Definition 2. An allocation of data $(N_k)_k$ is *disjoint* if $N_k \cap N_j = \emptyset$ for all $k, j \in K$ with $k \neq j$. The *size* of the allocation refers to $| \cup_k N_k|$.

The following result characterizes the set of maximum revelation equilibria. The proof is relegated to Appendix A.

[10] 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.

[11] By Lemma 1, the definition does not depend on whether we focus on the number of consumers who share their data with at least one intermediary or on the number of consumers whose data is acquired by the downstream firm.

Theorem 1. *The following two conditions are equivalent.*

1. *There exists a maximum revelation equilibrium such that the allocation of data is $(N_k)_k$ and each consumer $i \in N$ earns compensation τ_i.*

2. *Allocation $(N_k)_k$ is disjoint, $\cup_k N_k = \{1, \ldots, N^*\}$, and compensation τ_i satisfies*

$$u(0, N^* - 1) - u(1, N^* - 1) \leq \tau_i \leq \Pi(N^*) - \Pi(N^* - 1), \forall i \leq N^*;$$
$$\tau_i = 0, \forall i > N^*.$$

Theorem 1 implies that regardless of the number of intermediaries, the monopoly outcome (Proposition 1) can arise. For example, it is an equilibrium that intermediary 1 obtains data from N^* consumers with the monopoly level of compensation, because we can set $N_1 = \{1, .., N^*\}$, $N_k = \emptyset$ for $k \geq 2$, and $\tau_i = u(0, N^* - 1) - u(1, N^* - 1)$ for $i \leq N^*$. In this equilibrium, any intermediary $k \geq 2$ has no incentive to raise compensations because consumers will share their data with multiple intermediaries due to the non-rivalry of data.

The non-rivalry of data can also lead to equilibria where consumer surplus is strictly greater than under the monopoly. In these equilibria, intermediaries have no incentive to *lower* compensations because consumers will punish such deviations by sharing data with other intermediaries. However, I show in Section 6 that these equilibria disappear if it is costly for an intermediary to make an offer, no matter how small the cost is.

Theorem 1 also states that any disjoint allocation of data can arise in equilibrium, because given an allocation, no intermediary prefers to acquire data that other intermediaries will acquire. Also, the allocation of data does not restrict possible equilibrium consumer surplus. In contrast, allocation affects how intermediaries and the firm divide surplus. To see this, I introduce the following notion.

Definition 3. Let $(N_k)_k$ and $(N'_k)_k$ denote two disjoint allocations of data with the same size. $(N_k)_k$ is *more fragmented* than $(N'_k)_k$ if, for any k, there is ℓ such that $N_k \subset N'_\ell$.

Proposition 2. *If allocation $(N_k)_k$ is more fragmented than $(N'_k)_k$, the total revenue of intermediaries in the downstream market is lower 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 fragmented than the latter. Without loss of generality, assume that they have size N. 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

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

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

Intuitively, if the allocation of data is more fragmented, each intermediary has lower pricing power because the marginal contribution of its data to the firm's revenue is small. Thus, more fragmented allocations, though consistent with equilibrium, lead to lower industry profits for intermediaries.

To state the next result, I define monopoly equilibrium:

Definition 4. A *monopoly equilibrium* is an equilibrium in which a single intermediary acquires personal data from N^* consumers at the monopoly level of compensation $|u(1, N^*-1) - u(0, N^*-1)|$ and extracts full surplus $\Pi(N^*)$ from the firm.

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 $K \geq 1$, there exists a monopoly equilibrium. Among all maximum revelation equilibria, a monopoly equilibrium maximizes the total payoffs of intermediaries and minimizes consumer surplus and 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 B.

Claim 1. *Suppose that each consumer can share her data with at most one intermediary. If there is a monopoly 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 starkly contrasts with intermediaries for (non-rivalrous) data, where intermediaries can secure 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 with the maximum number (i.e., N^*) of consumers sharing their data.

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 distinguished from offering zero compensation. Second, if an intermediary makes an offer to N' consumers, it incurs a cost of $N'c$, which enters additively into its payoff function. I assume that c is positive but small so that

$$0 < c < \min\left\{\Pi(N) - \Pi(N-1), \min_{n} |u(1,n) - u(0,n)|\right\}.$$

The following result states that small transaction costs for intermediaries lead to a maximum revelation equilibrium with the monopoly level of compensation. The proof is in Appendix C.

Proposition 3. *The set of equilibria is nonempty, and in any equilibrium, consumers $i \leq N^*$ share their data and earn monopoly compensation $\tau^* = |u(1, N^* - 1) - u(0, N^* - 1)|$, whereas consumers $i > N^*$ 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, in any equilibrium where $N' < N^*$ consumers[12] share their data, intermediaries offer zero compensations to consumers in $N^* \setminus N'$, which are rejected on the equilibrium path. Consumers accept these offers only at off-path information sets where an intermediary deviates to acquire data from a consumer in $N \setminus N'$ by raising compensations. Therefore, transaction costs to make offers single out maximum revelation equilibria with the monopoly compensation.

7 Generalization: Multidimensional Personal Data

I generalize the model in Section 3 by allowing a richer structure of personal data. 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 personal data it holds. Formally, Π satisfies the following, which is a natural generalization of Assumption 2:

[12] N' stands for both the number and the set of consumers from whom intermediaries obtain data.

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 is the sum of $u(D_i, D_{-i})$ and total compensation from intermediaries, where $D_{-i} \subset \mathcal{D}_{-i} := \cup_{j \neq i} \mathcal{D}_j$. 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, which simplifies the derivation of equilibrium compensations. The validity of Point 2 would depend on a specific application. For example, the assumption could be natural if $\mathcal{D}_i$ consists of i's email address and browsing history, and the firm can engage in targeted marketing campaign (which i dislikes) only by knowing both pieces of information.

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.[13]

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.

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

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 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 Intermediary

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

Lemma 3. *Suppose that each intermediary $k \in K$ holds data $D^k \subset \mathcal{D}$. This subgame has a Nash 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.

To state the main result, I extend the various notions introduced in Section 5 in a straightforward manner:

Definition 5. An *allocation of data* is any element $(D^1, \dots, D^K)$ 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. $(D^k)_k$ is *more fragmented* than $(\hat{D}^k)_k$ 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 $(D^1, \dots, D^K)$ 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 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. $\square$

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

Proposition 4. *In any maximum revelation equilibrium (MRE), each consumer shares all personal data $\mathcal{D}_i$. Moreover, 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 a MRE sustaining them.*

Proof. Take any allocation and compensations described in the proposition. 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 $\hat{D}_i^k \neq D_i^k$.[14]

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

[14] 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, the equilibrium allocation of data has no implication on consumer surplus. In contrast, the submodularity of consumer's payoff u implies that concentration of data lowers consumer surplus:

Proposition 5. *Consumer surplus is lower if the data are more concentrated to one intermediary: Formally, in MREs with the lowest compensation (i.e., equilibrium in Proposition 4), consumer surplus is greater if the allocation of data is more fragmented.*

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 Extensions

8.1 When 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. For all $n \in \mathbb{Z}_+$, $u(1, n) > u(0, n)$.

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.

Proposition 6. *When the firm's use of data benefits consumers, competition benefits consumers. Formally, under Assumption 7, in any equilibrium, all consumers share their personal data. If $K = 1$, each consumer pays a positive fee and obtains a payoff of $u(0, N-1)$; if $K \geq 2$, each consumer receives non-negative compensation and obtains a payoff of at least $u(1, N-1)$.*

The intuition is as follows. When downstream firms use personal data to benefit consumers, intermediaries compete over data by lowering fees to transfer data. Because consumers never benefit from accepting multiple offers with non-negative fees, consumers can credibly share their data with at most one intermediary. This gives intermediaries an incentive to compete for data as if data are rivalrous. The result contrasts with the baseline model, where intermediaries have no incentive to raise compensations due to the non-rivalry of data.

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 $\Pi(N) - \Pi(N-1)$ can be sustained in equilibrium. However, as in Proposition 3, only the equilibrium with zero compensations 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.[15] 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.

It could be interesting to consider the case where an intermediary can commit to not sell data to some of the firms when it acquires data from consumers. In this case, an intermediary can make its offer attractive not only by raising compensations but also by committing to not sell data (say, in the form of "privacy policy") to firms whose data use negatively affects 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: Consumers are not compensated for their data, because data are not scarce. More precisely, the model 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 downstream markets.

References

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." *Annu. Rev. Econ*, 11, 1–23.

[15] N' is the number of consumers in $N \setminus \{i\}$ who share their data with at least one intermediary in a candidate strategy profile.

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin (2015), "Selling experiments." Technical report, Mimeo, MIT and Yale.

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.

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.

Appendix

A 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

$i \leq N^* := \{1, \ldots, N^*\}$ 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, she 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. 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 some $i \notin N_k$. There are two cases to consider. First, if consumer i rejects $\hat{\tau}_i^k$, intermediary k can obtain the same payoff by offering $\hat{\tau}_i^k = 0$. Second, if consumer i accepts $\hat{\tau}_i^k$, intermediary k has to pay a positive compensation. However, as consumer i accepts 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 $\hat{\tau}_i^k = 0$ 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 number 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 has to pay. First, the total compensation that k pays in the upstream market decreases by at most $N_D \cdot [\Pi(N^*) - \Pi(N^* - 1)]$, because $\tau_i \leq \Pi(N^*) - \Pi(N^* - 1)$ 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 $\Pi(N^* - N_{D_1})$.

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

$$
\begin{aligned}
&\underbrace{\Pi(N^*) - \Pi(N^* - N_k)}_{\text{revenue without deviation}} - \underbrace{[\Pi(N^* - N_{D_1}) - \Pi(N^* - N_{D_1} - (N_k - N_{D_1} - N_{D_2})]}_{\text{revenue with deviation}} \\
=&\Pi(N^*) - \Pi(N^* - N_{D_1}) + \Pi(N^* - N_k + N_{D_2}) - \Pi(N^* - N_k) \\
\geq&\Pi(N^*) - \Pi(N^* - N_{D_1}) + \Pi(N^*) - \Pi(N^* - N_{D_2}) \\
\geq&(N_{D_1} + N_{D_2}) \cdot [\Pi(N^*) - \Pi(N^* - 1)] \\
=&N_D \cdot [\Pi(N^*) - \Pi(N^* - 1)].
\end{aligned}
$$

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.

Second, suppose to the contrary that there is an equilibrium in which consumers in $\{1, \dots, M\}$ with $M > N^*$ provide their data. Then, by Point 2 of Assumption 1, the minimum (total) compensation that any consumer $i \in M$ has to receive is $|u(1, M-1) - u(0, M-1)|$. Lemma 2 implies that exactly one intermediary gives such an offer to i. Without loss of generality, suppose that intermediary 1 buys data from consumer $i \in M$. Then, intermediary 1 can strictly increase its payoff by giving an offer of zero to consumer i. Indeed, this deviation increases its payoff by $|u(1, M-1) - u(0, M-1)| - [\Pi(M) - \Pi(M-1)] > 0$. This establishes that any MRE has N^* consumers giving up their personal data. Thus, Point 2 implies Point 1.

Next, I show that Point 1 implies Point 2. The previous step establishes that any equilibrium has at most N^* consumers sharing their data, and I construct an equilibrium with N^* consumers sharing data. Thus, any MRE has consumers $1, \dots, N^*$ sharing data. Take any MRE and consumer $i \leq N^*$. Lemma 2 implies that only one intermediary, say k, offers i a positive compensation $\tau_i^k > 0$. If $\tau_i^k > \tau^* := \Pi(N^*) - \Pi(N^* - 1)$, then intermediary k can profitably deviate by lowering the compensation to zero. Also, if $\tau_i^k < |u(1, N^* - 1) - u(0, N^* - 1)|$, then consumer i would reject the offer. Thus, $\tau_i^k \in [u(0, N^* - 1) - u(N^* - 1), \Pi(N^*) - \Pi(N^* - 1)]$ for $i \leq N^*$ in any MREs that we consider. Any consumer $i > N^*$ earns no compensation because she reject any offer. Finally, if the allocation is not disjoint, some intermediary k such that $N_k \cap N_j \neq \emptyset$ 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. □

B Proof of Claim 1

Proof. 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, $\Pi(N^*) - \Pi(N^* - N_1) - \sum_{i \in N_1} \tau_i^1 > 0$, which implies $\Pi(N^* - N_2) - \Pi(N^* - N_1 - N_2) - \sum_{i \in N_1} (\tau_i^1 + \varepsilon) > 0$ for a small $\varepsilon > 0$. Thus, intermediary 2 has a profitable deviation, which is a contradiction. □

C 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 $i \leq N^*$ 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 $1, \ldots, N^*$ 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 consumer $i \leq N^*$ receives a compensation $\tau > \tau^*$ from, say, intermediary 1. Suppose that intermediary 1 unilaterally deviates and offers i a compensation of $\tau' \in (\tau, \tau^*)$. If τ' is the only offer that i receives, then she accepts the offer, which makes the deviation profitable. Thus, for such a deviation to be unprofitable for 1, it must be the case that consumer i 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 increase its payoff by $c > 0$ by not sending an offer to consumer i. Thus, intermediary 1 can profitably deviate by lowering the compensation, which is a contradiction.

Next, I show that there is no equilibrium where $M \neq N^*$ consumers share their data. First, take any equilibrium, and suppose that consumers $1, \ldots, M < N^*$ share their data. Suppose that intermediary 1 deviates and offers consumer $i := M + 1$ a compensation strictly greater than but close to $|u(1, M) - u(0, M)|$. Consumer i 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 $\Pi(M + 1) - \Pi(M) - |u(1, M) - u(0, M)| > 0$ as $M < N^*$, which is a contradiction. We can show that there is no equilibrium in which $M > N^*$ consumers provide their data, in the same way as the proof of Theorem 1. □

D 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. □

E Proof of Proposition 6

Proof. First, suppose to the contrary that there is an equilibrium in which consumer i does not share her personal data. This means that i either receives no offer or receives an offer that charges a fee greater than $\min_n[u(1,n) - u(0,n)] > 0$. Then, one intermediary (say 1) can deviate and offers her a fee (negative compensation) of $\delta > 0$. For a sufficiently small $\delta > 0$, consumer i shares her data only with 1. It strictly benefits intermediary 1 because the payoff in the downstream market increases by at least $\Pi(N) - \Pi(N-1) > 0$.

Second, suppose $K = 1$. Given that all of N consumers share their data, each consumer is willing to pay at most $u(1, N-1) - u(0, N-1)$. Thus, the intermediary charges a fee of $u(1, N-1) - u(0, N-1) > 0$, which yields each consumer a payoff of $u(0, N-1)$.

Next, suppose $K \geq 2$. Suppose that consumer i provides her data and pays a positive fee. This means that no intermediaries offer i a nonnegative compensation, because otherwise, i would never pay the fee. Also, this means that consumer i provides her data to only one intermediary, say 1. However, then intermediary 2, which does not buy data from i, can profitably deviate by offering i a fee of zero. Following this deviation, i accepts the offer of intermediary 2 and rejects that of 1. This deviation is strictly profitable for 2, because it can increase the price of the data by at least $\Pi(N) - \Pi(N-1) > 0$. Therefore, no consumers pay a positive fee in equilibrium. $\square$