Information Revelation and Privacy Protection*

Rossella Argenziano[†] Alessandro Bonatti[‡]

August 18, 2020

Abstract

We propose a microfoundation for consumers' privacy preferences and examine how it shapes the outcome of regulation. A single consumer interacts sequentially with two heterogeneous firms: the first firm collects data on consumer behavior, which the second firm uses to set a quality level and a price. Thus, the consumer manipulates her behavior to influence the future terms of trade. In equilibrium, manipulation is beneficial to the consumer when the recipient firm is sufficiently similar to the collecting firm (as measured by the relative salience of quality and price of their two products). We then evaluate the impact of privacy regulation, including mandatory transparency, explicit consent requirements, and limits to discriminatory offers. We show that transparency has an ambiguous effect on consumer welfare and that consent requirements are unambiguously beneficial to consumers but that limits to discrimination are harmful to consumers in equilibrium.

KEYWORDS: consumer privacy; consumer consent; personal information; data linkages; data rights; price discrimination; signaling; ratchet effect.

JEL CLASSIFICATION: D44, D82, D83.

*We thank Charles Angelucci, Gonzalo Cisternas, Jacques Crémer, Martin Peitz, and seminar participants at Mannheim and Toulouse for helpful discussions. We also thank Roi Orzach for stellar research assistance.

[†]Economics Department, University of Essex, Colchester CO4 3SQ, UK, `rargenz@essex.ac.uk`.

[‡]MIT Sloan School of Management, Cambridge, MA 02142, `bonatti@mit.edu`.

1 Introduction

Motivation The widespread collection, analysis, and distribution of individual data creates linkages across many seemingly unrelated transactions. From browsing and search histories to geolocation data to social media activity, large online platforms gather vast amounts of data about their consumers. The information gained from one transaction can then enable targeted advertising, tailored product offers, and even personalized prices in a different transaction. Indeed, a key and often overlooked feature of such data linkages is that consumers do not reveal personal information directly. Instead, data about *behavior* are acquired and transmitted, and information about individual *preferences* is inferred.[1]

In response to current market practices, several proposed and enacted privacy-protection regulatory interventions promise to augment consumers' control over their own data. This is the case, for example, of the European Union's General Data Protection Regulation (GDPR) and of the California Consumer Privacy Act (CCPA). The rationale for such interventions is that empowering consumers to control who can track them will enable the socially efficient use of information, or at least ensure that consumers are compensated for the use of their personal information.

However, because most data collected online is about behavior, consumers who know they are being tracked have an incentive to distort their actions to misrepresent their preferences. In particular, if consumers anticipate their data will be used for price discrimination, they have an incentive to understate their willingness to pay—the canonical *ratchet* effect of Laffont and Tirole (1988). Conversely, if a firm uses data to target the quality of the products and services it offers, consumers have an incentive to overstate their willingness to pay—the *niche envy* effect of Turow (2008). Both kinds of distortions impact on profits of firms that collect data, and they affect the terms of trade offered to the consumer. To properly assess the impact of existing policies and to clarify the need for further regulation, we must then examine the *equilibrium effects* of data linkages across heterogeneous transactions.[2]

[1] Customer Lifetime Value (CLV) scores—aggregate measures of profitability that merchants use to determine the level of service, prices, and perks to offer individual consumers—are a concrete channel through which behavior data links different transactions. See Bonatti and Cisternas (2020) for an in-depth treatment. Other prominent examples include travel and health insurance contracts that condition on third-party information about insurees' behavior and the Chinese Social Credit system that determines conditions and access to credit, housing, and travel; see the discussion in Tirole (2020).

[2] The incentive implications of linked transactions are not limited to consumer markets. In the B2B market for targeted advertising, the prevalence of personalized reserve prices is documented by Paes Leme, Pal, and Vassilvitskii (2016). Likewise, bidders may want to overstate their willingness to pay for campaigns if that means accessing higher-quality content (i.e., more profitable eyeballs).

Model To tease out these effects, we develop a simple dynamic model in which a consumer interacts sequentially with two heterogeneous firms. In each period, a firm sets a quality level and a price (the *terms of trade*), and the consumer chooses how much to consume. Firms differ in the salience of their products' quality relative to its price. At the onset of the game, the consumer has private information about her willingness to pay. However, if a *data linkage* exists between the two firms, the second firm can observe the consumer's first-period behavior and use this information to match both the quality level and the price to the consumer's perceived willingness to pay. We focus on equilibria in which the consumer's strategy is linear in her willingness to pay and the second firm's quality and price levels are linear in the observed first-period data.

Results We find that there exists a unique equilibrium in linear strategies. In this fully separating equilibrium, the second firm learns the consumer's type perfectly. Thus, a data linkage enables both price and quality discrimination by the second firm. We then show that price and quality discrimination increase the consumer's second-period welfare when the second firm's quality is more salient than its price.

A data linkage has two additional strategic effects in the first period. First, the consumer distorts her first-period behavior away from the static optimum to manipulate the second firm's beliefs. Specifically, if the second firm's quality is very salient, the consumer has an incentive to distort consumption upwards to receive a higher-quality (but more expensive) product. Conversely, if the second firm's quality is not very salient, the consumer distorts consumption downwards to receive a less expensive (but lower quality) product. Regardless of its direction, this distortion in behavior is costly for the consumer in the first period. Second, a data linkage affects the first period equilibrium terms of trade. If the consumer distorts her demand downward, firm 1 lowers its price and quality level, which increases consumer surplus when firm 1's quality is less salient than its price. If instead the consumer distorts her demand upwards, firm 1 increases both quality and price, which benefits the consumer when quality is more salient.[3]

Next, we characterize the set of data linkages that benefit consumers and producers from an *ex ante* perspective, relative to the benchmark of fully anonymous trading. The terms of trade effect above suggests that consumers prefer data linkages to form when the recipient firm is sufficiently similar to the collecting firm. When taking all three effects of data linkages into account, we obtain an additional necessary condition for a consumer to benefit from a data linkage between two firms: the salience of the second-period firm's quality must not

[3]In Section 6, we show that these strategic forces are robust to several extensions, including multiple or uncertain uses of period-1 data, and competition among period-2 firms.

be too extreme so as to not induce large upward or downward distortions in first-period consumption. Finally, the consumer is more likely to benefit from linkages to high-quality period-2 firms when the prior uncertainty about her willingness to pay is large. In that case, the effect of information in the second period is larger, and the direct impact of data linkages is stronger.

We also characterize the data linkages that improve the two firms' total profits. While most regulatory agencies use consumer welfare as their objective, producer surplus is a key determinant of linkage formation when consumers have no control rights over their data. We find that all linkages create value in period 2 by allowing price and quality discrimination. Conversely, in period 1, only linkages to period-2 firms with high quality-salience increase profits, stimulating demand, while linkages to firms with low quality-salience depress period-1 demand. Therefore, the two firms are collectively better off creating a linkage only when the period-2 firm has quality-salience above a critical level.

Policy Evaluation We then leverage the tractability of our model to evaluate the impact of recently enacted regulatory policies that transfer partial control rights from firms to consumers on consumer surplus. Specifically, we consider mandatory transparency (consumers need to be informed about the existence of a linkage), explicit consent requirements (a linkage can be formed only with the explicit consent of the consumer) and limits to discrimination that forbid conditioning terms of trade on the consumer's consent choice. We extend the model to endogenize linkage formation, assuming efficient bargaining ex ante (i.e., before the consumer learns her type). We can therefore ask whether a linkage will form in equilibrium as a function of the two firms' types under each given policy rule and whether each policy does indeed increase consumer surplus.

In the absence of regulation or commitment power by the first-period firm, transaction data are always shared—all linkages form. The effects of any policy on consumer welfare are twofold: first, privacy regulation directly constrains the formation of data linkages across different transactions; second, for the data linkages that do form, the consumer's privacy concerns lead to behavior distortions and modified terms of trade.

We show that requiring transparency by the firms allows them to commit to sharing data only when a linkage increases producer surplus. Because the sets of firm-optimal and consumer-optimal linkages are different, transparency policies have an ambiguous effect on consumer welfare. However, requiring consumer consent in addition to transparency is equivalent to imposing mutual veto rights on linkage formation. Because consent requirements lead to data sharing only when it is Pareto improving, they are unambiguously beneficial to consumers but provide insufficient data linkages from a social-welfare perspective. Finally, we show that forbidding firms to condition their terms of trade on the consumer's consent decision actually hurts consumers—it induces firms not to propose beneficial data linkages, and it induces consumers to consent to harmful linkages, which they would have vetoed under the equilibrium discriminatory offers.

Related Literature The diffusion of markets for individual data has prompted growing interest in the welfare effects of consumer information online. From a modeling standpoint, our paper is part of the behavior-based price discrimination literature, with Taylor (2004), Acquisti and Varian (2005), Calzolari and Pavan (2006), and most recently Baye and Sappington (2020) as the most closely related papers. In particular, Taylor (2004) introduces an explicit market for consumer information, with the goal of quantifying the differential welfare effects of price discrimination for naive and sophisticated consumers. In contrast to these papers, our model allows for heterogeneous sources and heterogeneous uses of data, thus clarifying the terms of trade effect. Moreover, we introduce a mechanism for endogenous linkage formation and show how it operates under different regulatory regimes.

A growing literature studies how correlated preferences facilitate the collection of information from multiple consumers. Specifically, Choi, Jeon, and Kim (2019), Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2019), Bergemann, Bonatti, and Gan (2020), and Ichihashi (2020) study the externalities associated with individuals selling their own data. In all these papers, consumers directly sell, control, report, or disclose their preferences. In other words, the work on data externalities has largely abstracted from distortions in behavior arising from the collection of behavior data. One notable exception is Liang and Madsen (2020), who analyze the incentive effects of correlated types in a model of career concerns, distinguishing between correlation in fundamentals (quality linkages) and in observation errors (circumstance linkages).

In all the above papers, consumers cannot affect the quality of information available to the firms. A number of recent contributions, including Cummings, Ligett, Pai, and Roth (2016), Frankel and Kartik (2019), Ball (2020), Bonatti and Cisternas (2020), and Jann and Schottmüller (2020), study how the consumer's manipulation incentives reduce the amount of information transmitted in equilibrium, and they suggest mechanisms to mitigate this loss. Another set of related contributions, such as Conitzer, Taylor, and Wagman (2012), Belleflamme and Vergote (2016), and Ali, Lewis, and Vasserman (2019), focus on the amount of information available in equilibrium when consumers can actively protect their privacy by remaining (partially) anonymous. Our model is simpler in this respect—all information is revealed in equilibrium, which allows us to focus on the welfare cost of behavior distortions.

Finally, a few recent papers—for example, Fainmesser, Galeotti, and Momot (2020) and Jullien, Lefouili, and Riordan (2020)—study privacy protection from data *leakages.* These leakages are modeled as reduced-form negative consequences of information diffusion. The survey by Acquisti, Taylor, and Wagman (2016) covers other such models in greater detail. In our paper, we instead focus on a specific microfoundation for privacy preferences, and we examine how it shapes the outcome of regulation.

2 Model

We consider a single consumer who lives for two periods and interacts with a different firm in each period $t = 1, 2$. The firm active at time t offers a unit price p_t and a quality level y_t to the consumer who, in turn, chooses a quantity q_t. The consumer's per-period utility is given by

$$U\left(p_t, y_t, q_t\right) = \left(\theta + b_t y_t - p_t\right) q_t - \frac{q_t^2}{2}. \tag{1}$$

We can interpret θ as the consumer's baseline willingness to pay for the good (i.e., before adjusting for quality) and henceforth refer to θ as the consumer's type. The parameter b_t is a firm-level characteristic that represents the marginal value of the quality of the good produced by firm t. Thus, b_t captures the nature of the interaction between the consumer and firm t. In particular, the case $b_t = 0$ corresponds to pure price discrimination. Finally, we refer to the quantity $b_t y_t - p_t$ as the *terms of trade* that firm t offers to the consumer.

Each firm t has a constant marginal cost of producing quantity q_t that we normalize to zero and a fixed per-consumer cost of producing quality y_t. Firm t's profit function is then given by

$$\Pi\left(p_t, y_t, q_t\right) = p_t q_t - \frac{y_t^2}{2}. \tag{2}$$

To keep the analysis well-behaved, we impose an assumption on the magnitude of the returns from producing quality y_t.

Assumption 1 *The marginal value of quality satisfies $b_t \in [0, \sqrt{2})$ in each period $t = 1, 2$.*

The firm-level parameters b_1 and b_2 are commonly known at the onset of the game. The two firms share a common prior on the consumer's type θ, which is distributed on an interval $\Theta \subset \mathbb{R}_+$ according to a distribution G. We assume that $G(\theta)$ admits an everywhere positive density $g(\theta)$ with finite mean $\mu \triangleq \mathbb{E}[\theta]$ and variance $\sigma^2 \triangleq \text{var}[\theta]$. Firm 1 sets (p_1, y_1) on the basis of the prior distribution only. Firm 2 can also observe the first-period outcome (p_1, y_1, q_1) before interacting with the consumer, if a data linkage is active.

We summarize the timing of the game as follows, and Figure 1 provides an illustration.

1. Firm 1 offers a price and a quality level (p_1, y_1) to the consumer.

2. The consumer learns her type θ and selects a quantity q_1.

3. If the data linkage is active, firm 2 observes the first-period outcome (p_1, y_1, q_1) before setting its price and quality level (p_2, y_2).

4. The consumer selects a quantity q_2.

We focus on *linear equilibria*—Bayesian Nash equilibria in which the consumer's strategy is linear in her type and the second-period firm's strategy is linear in any variable it observes.[4]

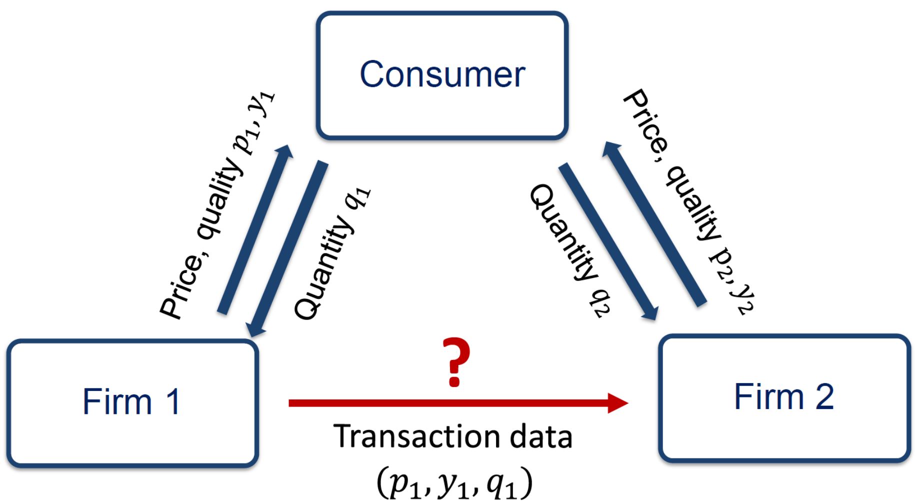

Figure 1: Model Sketch

In what follows, we first analyze the linear equilibria of our game under different information structures. We then evaluate which data linkages benefit consumers and firms, respectively (i.e., which pairs of firms b_1 and b_2 should trade consumer information). Finally, we endogenize the formation of data linkages as a function of the decision rights allocated to firms and consumers, and we map the outcome to existing consumer protection regulation.

[4]Linear equilibria are fully separating. However, because we define the consumer's type θ on a compact support, the consumer can choose actions that are off the equilibrium path. The linearity requirement disciplines firm 2's prices and qualities if this occurs, which has the effect of discouraging jumps in the consumer choice of q_1. See also the discussion in Ball (2020). Alternatively, we could have assumed that θ is distributed on $\mathbb{R}$ with full support, in which case all separating equilibria are linear.

3 Equilibrium Analysis

We begin our analysis with a static (one-period) benchmark, where we illustrate the welfare effects of information about the consumer's preferences when this is exogenously given to the firm. Therefore, Section 3.1 also serves as the naive-consumer benchmark, as well as the analysis of period $t = 2$. Section 3.2 builds upon these results, proceeding by backward induction to uncover the value of information in a dynamic model where firm 2 can infer the consumer type by observing the action she took in the first period.

3.1 Exogenous Information

Consider a game between a consumer and a single firm that sells a product with quality of value b. The firm is endowed with an arbitrary information structure $\mathcal{I}$ consistent with the prior G. The consumer observes the firm's offer (p, y) and simply maximizes her current-period utility (1). This yields consumer demand

$$q(\theta, p, y) = \theta + by - p. \tag{3}$$

The firm maximizes its expected profits (2) given the available information $\mathcal{I}$ and the demand function (3). This yields the following quality and price level

$$y^*(m, b) = \frac{bm}{2 - b^2}, \tag{4}$$

$$p^*(m, b) = \frac{m}{2 - b^2}, \tag{5}$$

where m denotes the firm's posterior mean

$$m \triangleq \mathbb{E}[\theta \mid \mathcal{I}].$$

The resulting *terms of trade* for the consumer are given by

$$by^*(m, b) - p^*(m, b) = \lambda(b) \cdot m, \tag{6}$$

where

$$\lambda(b) \triangleq \frac{b^2 - 1}{2 - b^2}.$$

Under Assumption 1, the parameter λ takes values in $[-1/2, \infty)$. This parameter summarizes the (static) *equilibrium* effect of the firm's beliefs on the terms of trade. We henceforth refer to λ_t as firm t's *type*. Intuitively, when the value of a firm's quality b is high, consumers with a higher type θ buy considerably more units when they are offered a higher quality level, which in turn justifies the firm's investment in y. In equilibrium, firms with $\lambda > 0$ (i.e., $b > 1$) then offer better terms of trade to higher-θ consumers.

Substituting (6) into the demand function (3), we obtain the realized consumer utility

$$U(\theta, m, b) = \frac{1}{2} q^* (\theta, y^* (m; b), p^* (m; b))^2 = \frac{1}{2} (\theta + \lambda (b) m)^2 . \tag{7}$$

We may then ask how the availability of information affects the consumer and the firm *ex ante*. For this purpose, we consider two information structures: $\mathcal{I} = \varnothing$, in which the firm has no information and hence $m \equiv \mu$; and $\mathcal{I} = \mathcal{I}^* \triangleq \Theta$, i.e., the complete information structure. We obtain the following result.

Proposition 1 (Value of Exogenous Information)

1. *Firm profits are higher under $\mathcal{I}^*$ for all λ.*
2. *Consumer surplus is higher under $\mathcal{I}^*$ for all $\lambda > 0$.*
3. *There exists a unique $\lambda^* < 0$ such that total surplus is higher under $\mathcal{I}^*$ for all $\lambda > \lambda^*$.*

Proposition 1 shows that the firm always prefers having (ideally, complete) information about the consumer's type to tailor its price and quality offers. The consumer, on the other hand, benefits in expectation from discriminatory offers if and only if $\lambda > 0$. Intuitively, information creates positive correlation between the firm's beliefs m and the consumer's type θ. When $\lambda > 0$, the consumer benefits through better terms of trade when her true willingness to pay is in fact high. Finally, total surplus combines these two effects and, hence, increases with information for some moderately negative and all positive values of λ.

How do the welfare effects of exogenous information depend on the heterogeneity in consumer type? One can show that all three effects in Proposition 1 are proportional to the prior variance σ^2. Specifically, consider the consumer's surplus. Substituting $m = \mu$ (for $\mathcal{I} = \varnothing$) and $\mu = \theta$ (for $\mathcal{I} = \mathcal{I}^*$) into 7, we can write the ex ante value of information for a consumer as

$$\mathbb{E}_\theta [U(\theta, \theta, b) - U(\theta, \mu, b)] = \frac{1}{2} \sigma^2 (2 + \lambda (b)) \lambda (b) , \tag{8}$$

which has the same sign of $\lambda (b)$ as shown in Proposition 1. A similar calculation yields the result for firm profits and social welfare.

3.2 Endogenous Information

We now turn to the dynamic game played by the consumer with the two firms $t = 1, 2$ with types λ_t. We assume the data linkage is active, i.e., firm 2 observes the terms of trade offered and the quantity purchased in the first period. Based on the previous analysis, we know the consumer benefits from the data linkage at t_2 if and only if $\lambda_2 > 0$. We now ask under which conditions a forward-looking consumer benefits from the data linkage $\lambda_1 \to \lambda_2$. In Section 6, we introduce competition at t_2 among firms with differentiated products but identical λ_2. We also allow for uncertainty over the type of the t_2 firm and consider which realizations of λ_2 should observe the first-period outcome.

We begin our analysis of linear equilibria by illustrating the consumer's manipulation incentives. In any linear equilibrium, the first period quantity q_1 signals the consumer's type θ to firm 2. In particular, firm 2 holds degenerate posterior beliefs over θ, which are captured by an increasing, linear function $m(q_1)$. Since the consumer knows λ_2, she can compute her continuation payoff (7). She then solves the following problem

$$\max_{q_1} \left[U(\theta, q_1, p_1, y_1) + \frac{1}{2} (\theta + \lambda_2 m(q_1))^2 \right]. \tag{9}$$

The consumer thus faces a trade-off between maximizing her t_1 utility and manipulating the t_2 terms of trade through a different choice of q_1. Critically, the direction of the manipulation incentives depends on the *sign* of λ_2, while the strength of such incentives depends on both the magnitude of λ_2 and on the sensitivity of firm 2's posterior $m(q_1)$.

Proposition 2 establishes the existence and uniqueness of linear equilibria in our dynamic game and characterizes the equilibrium strategies at t_1.

Proposition 2 (Linear Equilibrium)
There exists a unique linear equilibrium of the game. In the linear equilibrium:

1. *The consumer's first-period demand function is given by*

$$q_1^*(\theta, p_1, y_1) = \theta(1 + \lambda_2) + b_1 y_1 - p_1. \tag{10}$$

2. *Firm 1 offers terms of trade (p_1^*, y_1^*) that satisfy*

$$b_1 y_1^* - p_1^* = (1 + \lambda_2)\lambda_1 \mu. \tag{11}$$

3. *All players follow the second-period strategies (3)-(5), with $m = \theta$.*

Since a linear equilibrium is separating, the consumer's type is fully revealed; hence, the second-period outcome is the same as that under complete information (i.e., with $m = \theta$). In the first period, the consumer distorts her equilibrium behavior away from the naive benchmark in (3). This distortion affects only the demand intercept, and it does so through the weight placed on θ. In particular, the consumer places more weight on her type if $\lambda_2 > 0$ and less weight if $\lambda_2 < 0$.

To compute the magnitude of the equilibrium distortion in behavior, consider the second-period equilibrium utility of a type θ consumer as a function of the firm's beliefs, which is given in (7). Specifically, compute its derivative with respect to the firm's belief m, evaluated at the equilibrium (correct) beliefs $m = \theta$,

$$\frac{\partial U(\theta, \theta, \lambda_2)}{\partial m} = \lambda_2 (1 + \lambda_2) \theta. \tag{12}$$

From the first-order condition for the consumer's problem (9), we obtain the following equilibrium condition:

$$q_1^* (\theta, p_1, y_1) = \theta + b_1 y_1 - p_1 + \lambda_2 (1 + \lambda_2) \frac{\theta}{\alpha^*}, \tag{13}$$

where

$$\alpha^* \triangleq \frac{\partial q_1^* (\theta, p_1, y_1)}{\partial \theta} = \frac{1}{m'(q_1)}$$

is firm 2's equilibrium conjecture of the weight placed on θ by the consumer's strategy.

The first-order condition (13) helps clarify why statically optimal behavior cannot be part of an equilibrium if $\lambda_2 \neq 0$. Indeed, if firm 2 expected the consumer to maximize her t_1 utility (i.e., $\alpha^* = 1$), the weight on θ on the right-hand side of (13) would be given by $1 + \lambda_2 (1 + \lambda_2)$. Recalling that $\lambda_2 \geq -1/2$, we immediately obtain that the consumer would buy more (less) quantity than optimal at t_1 depending on whether $\lambda_2 > 0$ (to raise firm 2's beliefs) or $\lambda_2 < 0$ (to depress them). Intuitively, a small deviation from static optimization has no first-order impact on t_1 utility but strictly improves the terms of trade at t_2.

Finally, matching the coefficients on (θ, p_1, y_1) in (13) then yields the equilibrium demand function (10) and, in particular, $\alpha^* = 1 + \lambda_2$. From the perspective of firm 1 (because the expected type μ is positive), the individual consumer's behavior translates into an upward shift of the demand curve if $\lambda_2 > 0$ and a downward distortion if λ_2. This shift causes the equilibrium terms of trade (11) to shift by a factor of $1 + \lambda_2$, relative to the static benchmark in (6) with $\lambda = \lambda_1$. Combining the two parts of Proposition 2, it is immediate to verify that the quantity q_1^* traded in equilibrium also shifts by the same factor $1 + \lambda_2$.

4 Welfare Effects

In this section, we analyze the welfare implications of a data linkage between firms λ_1 and λ_2. Data linkages impact both periods of our game. The second-period welfare implications are described in Proposition 1. In the first period, a data linkage introduces the equilibrium distortions and the terms-of-trade effect described in the previous section. We begin with the latter, i.e., with the welfare effects of the consumer's first-period actions.

While the distortion in consumer behavior (Proposition 2) is a function of λ_2 only, the effect of the resulting demand shift on the t_1 terms of trade depends critically on both firms' types. We begin by illustrating two examples in Figure 2. In both examples, firm 1 is a pure price-setting firm (i.e., $b_1 = 0$; hence, $\lambda_1 = -1/2$).

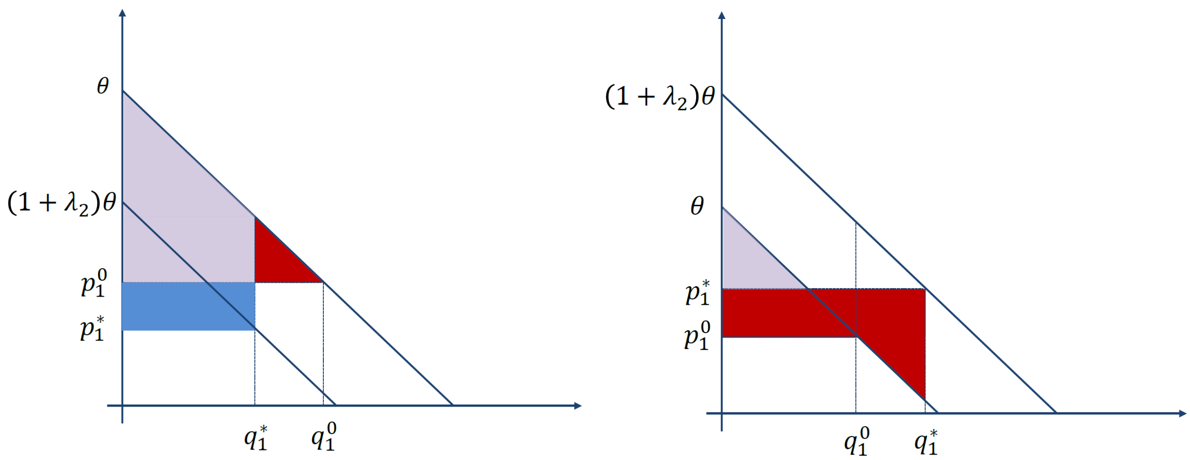

Figure 2: Demand and Price Shifts for $\lambda_1 = -1/2$, with $\lambda_2 < 0$ (left) and $\lambda_2 > 0$ (right)

The left panel captures the *ratchet* effect: firm 2's type is $\lambda_2 < 0$, i.e., the consumer knows that a higher posterior belief by firm 2 will lead to worse terms of trade. Firm 1 anticipates the consumer's concern over the second period price, expects a lower demand curve, and charges a lower monopoly price p_1. In equilibrium, the consumer buys smaller quantity q_1^* at a lower price p_1^* relative to the (q_1^0, p_1^0) outcome of a static game. Importantly, the consumer may benefit from this outcome, as the inframarginal discount $p_1^0 - p_1^*$ on q_1^* units (i.e., the blue rectangle) can compensate the loss from foregoing consumption of the marginal units (i.e., the red triangle). In addition to these effects, the consumer faces a certain loss at t_2, where firm 2 will know her type perfectly.

The right panel captures the *niche envy* effect introduced in Turow (2008): firm 2's type is $\lambda_2 > 0$, which means the consumer wishes to manipulate the firm's beliefs upward to obtain better terms of trade. Thus, firm 1 expects a higher demand curve than that in a static game and charges a higher price, $p_1^* > p_1^0$. This price nonetheless leads the consumer to buy more units than statically optimal, $q_1^* > q_1^0$. Therefore, in the first period, the consumer buys "too many" units (the red trapezoid) at a higher price (the red rectangle). However, the consumer also enjoys better terms of trade at t_2.

More cases than those depicted in Figure 2are possible. For example, if $\lambda_1 > 0$ and $\lambda_2 > 0$, then the consumer's upward demand shift in the first period would lead to more generous terms of trade at t_1. For any (λ_1, λ_2), however, the welfare effects of a data linkage at t_1 operate through two channels only.

First, the terms of trade offered by firm 1 change to reflect the shifts in consumer demand. Comparing Propositions 1 and 2, the difference in terms of trade $b_1 y_1^* - p_1^*$ between the static and dynamic cases is related to the sign of λ_1 and λ_2, i.e.,

$$(b_1 y_1^* - p_1^*) - \left(b_1 y_1^0 - p_1^0\right) = \lambda_1 \cdot \lambda_2 \cdot \mu.$$

Thus, the consumer obtains better t_1 terms of trade if the two firms are similar in a very specific sense: the terms of trade improve when both firms produce quality that is of high *or* low value relative to money (i.e., λ_1 and λ_2 have the same sign). Specifically, if both firms are low quality, then the distortion causes a helpful reduction in price; and if they are both high quality, the consumer's distortion causes a helpful increase in price-adjusted quality.

Secondly, the consumer's manipulation concerns introduce losses at t_1 due to the ensuing costly signaling that, despite being fully anticipated by firm 1, distorts the consumer's t_1 quantity away from the best reply to (p_1, y_1). Since demand is distorted up or down by an amount $\lambda_2 \theta$, the magnitude of this loss is proportional to $(\lambda_2)^2$, as can also be seen from the red triangles in Figure 2. Thus, the consumer's cost of signaling is related to the strength of her manipulation incentives, regardless of their direction.

We now summarize the combination of these two effects. In Proposition 3, we let $\hat{\sigma} \triangleq \sigma/\mu$ denote the coefficient of variation of the distribution $G(\theta)$. We then compare the expected consumer and producer surplus at t_1 when the $\lambda_1 \to \lambda_2$ data linkage is active to the expected surplus levels in the static benchmark with firm type λ_1.

Proposition 3 (First-Period Welfare Effects)

1. *A data linkage increases t_1-consumer surplus if and only if the following hold:*

$$\lambda_1 \cdot \lambda_2 > 0; \textit{ and}$$
$$|\lambda_2| < |\lambda_1| \frac{2(1+\lambda_1)}{\hat{\sigma}^2 + 1 - \lambda_1^2} \textit{ for all } \lambda_1 < \sqrt{1+\hat{\sigma}^2}.$$

2. *A data linkage increases firm 1's profits if and only if $\lambda_2 > 0$.*

Intuitively, firm 1 benefits from a data linkage if and only if the resulting change in consumer behavior increases the demand for its product. The effect on consumer surplus is slightly more involved. Figure 3 illustrates the set of pairs (λ_1, λ_2) whose linkage is beneficial to consumers *in the first period.*

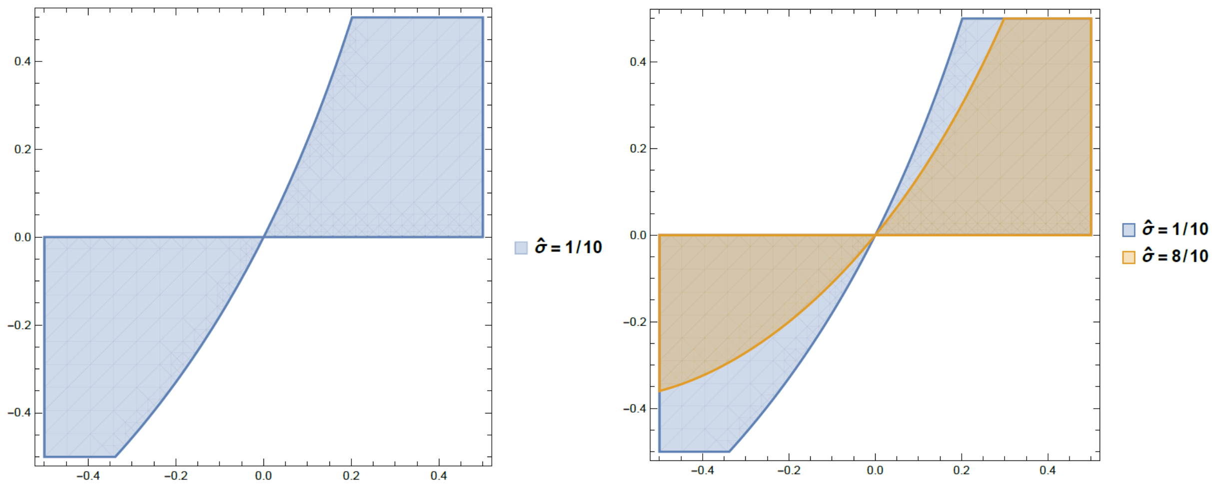

Figure 3: First-Period Consumer Surplus Improving Linkages ($\hat{\sigma} \in \{1/10, 8/10\}$)

Consistent with the trade-offs highlighted above, consumers can benefit in the first period only if two conditions are met: first, the firms' types must have the same sign; and second, λ_2 needs to be sufficiently small in magnitude such that the distortion in consumption does not trump the value of improved terms of trade. However, the latter condition applies only if λ_1 is smaller than the threshold—if the value of firm 1's quality is sufficiently large, then any $\lambda_2 > 0$ improves consumer surplus because the terms of trade effect dominates.

Finally, larger prior uncertainty $\hat{\sigma}$ does not affect the amount of distortion in behavior, but it unambiguously worsens its impact on expected consumer surplus. In particular, the region $\lambda_1 > \sqrt{1 + \hat{\sigma}^2}$, where all $\lambda_2 > 0$ benefit the consumer, shrinks as $\hat{\sigma}$ increases. Because the distortion in behavior relative to the best reply at t_1 impacts the weight the consumer places on her type, the consumer suffers a convex loss equal to $(\lambda_2\theta)^2/2$. Thus, the variance of θ increases the expected loss to the consumer.[5]

In Proposition 4 below, we calculate the intertemporal welfare impact of a data linkage across both periods.

[5] The social cost of data linkages also increases with $\hat{\sigma}$ because the impact on profits is constant in $\hat{\sigma}$.

Proposition 4 (Intertemporal Welfare Effects)

1. *A data linkage increases consumer welfare if and only if the following hold:*

$$\begin{aligned} &\left(\hat{\sigma}^2 + \lambda_1 (1+\lambda_1)\right) \cdot \lambda_2 > 0; \text{ and} \\ &|\lambda_2| < \left|\hat{\sigma}^2 + \lambda_1 (1+\lambda_1)\right| \frac{2}{1-\lambda_1^2} \text{ for all } \lambda_1 < 1. \end{aligned} \tag{14}$$

2. *A data linkage increases total firm profits if and only if*

$$\lambda_2 > \sqrt{\left(\frac{\hat{\sigma}^2}{2(\lambda_1+1)}\right)^2 + 1} - \frac{\hat{\sigma}^2}{2(\lambda_1+1)} - 1.$$

Figures 4 and 5 illustrate the sets of consumer- and firm-beneficial linkages (λ_1, λ_2) for different values of prior uncertainty $\hat{\sigma}$. We label them Λ^{CS} and Λ^{PS}, respectively.

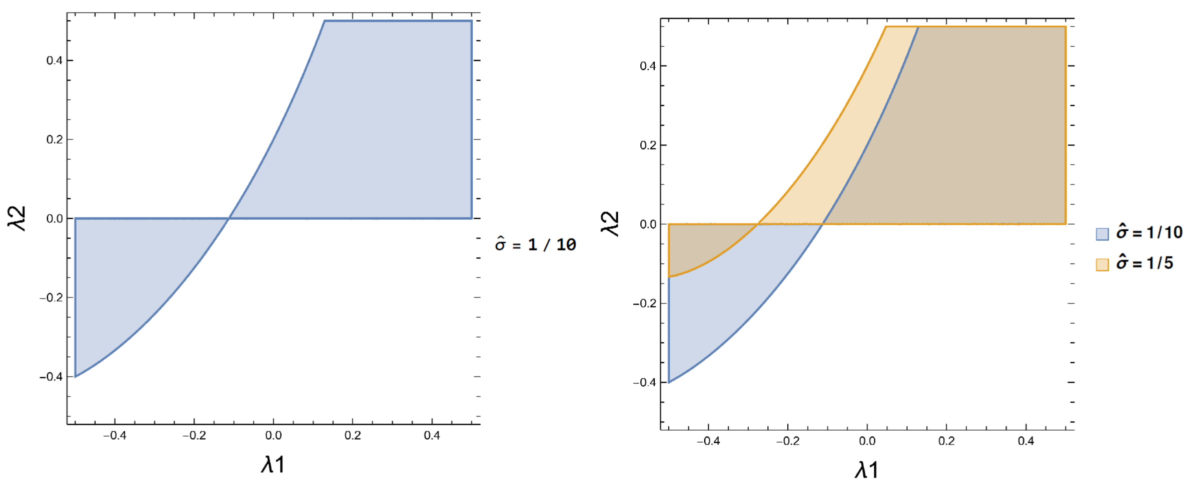

Figure 4: Λ^{CS}: Consumer Surplus Improving Linkages, $(\hat{\sigma} \in \{1/10, 1/5\})$

The total welfare impact of a data linkage combines the effect of exogenous information at t_2 with the first-period equilibrium forces. The overall effects of data linkages are thus best understood by comparing Propositions 3 and 4. In particular, two properties of Propositions 3 carry over to the intertemporal welfare effects: first, for any given λ_1, all λ_2 that benefit consumers have the same sign (see Figure 4); second, all linkages with $\lambda_2 > 0$ benefit the firms. There are, however, important differences that we discuss below, beginning with the consumer's perspective.

Consider the case of $\lambda_2 < 0$. Any linkage to such a firm 2 reduces consumer surplus at t_2 and introduces costly distortions in behavior at t_1. Thus, a linkage with $\lambda_2 < 0$ can be

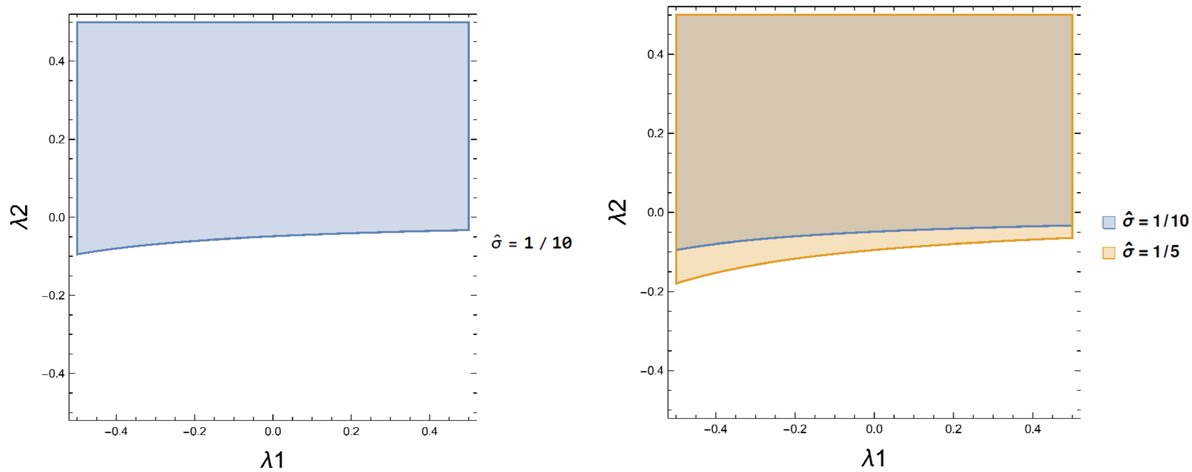

Figure 5: Λ^{PS}: Producer Surplus Improving Linkages, ($\hat{\sigma} \in \{1/10, 1/5\}$)

beneficial only if $\lambda_1 < 0$ (so that t_1 terms of trade improve) and $|\lambda_2|$ is sufficiently small such that distortions are not excessively costly. Conversely, all linkages with $\lambda_2 > 0$ have a positive effect on consumer surplus at t_2. Therefore, if a (λ_1, λ_2) linkage leads to a sufficiently small worsening of terms of trade and a sufficiently small behavior distortion, then such a linkage can be beneficial to consumers even if $\lambda_1 < 0 < \lambda_2$. This is exactly what occurs in Figure 4 (but not in Figure 3): there exists a threshold $\lambda_1 < 0$ above which all beneficial linkages have $\lambda_2 > 0$. This threshold is implicitly given on the left-hand side of (14).

From the firms' perspective (Figure 5), the problem is easier: all firms benefit from linkages with $\lambda_2 > 0$. These linkages increase profits at t_2 due to price discrimination and raise demand at t_1 by means of the consumer's manipulation incentives. By continuity, for any λ_1, there exists a threshold $\lambda_2 < 0$ above which linkages also increase total profits. Furthermore, recall that the distortion in a consumer's behavior is proportional to her average demand, which is an increasing function of λ_1. Therefore, high λ_1 firms are less willing to link to negative λ_2 firms (i.e., their threshold λ_2 is higher) because the resulting downward distortion in consumer behavior is more costly for them.

The effects of data linkages depend *quantitatively* on the distribution of consumer types G. In particular, recall that the welfare effect of exogenous information (Proposition 1) is increasing in σ^2, while the terms-of-trade effect is proportional to μ^2. Finally, as $\hat{\sigma}$ increases, the cost of distortions increases and the relative importance of the t_2 effects grows relative to the terms of trade effect. This shifts the consumer-beneficial set of linkages to the left in Figure 4. Likewise, as $\hat{\sigma}$ increases, the value of information for the firms at t_2 grows, as does the set of firm-beneficial linkages in Figure 5 (i.e., more $\lambda_2 < 0$ linkages become profitable).

5 The Impact of Privacy Regulation

In this section, we leverage our understanding of the welfare consequences of data linkages to examine the impact of policies that regulate data sharing. Our discussion follows the General Data Protection Regulation (GDPR) adopted by the EU in 2016 and the California Consumer Privacy Act (CCPA) of 2018 quite closely. The basic principle underlying these policies is that consumers' interests are best protected by transferring the control rights over the use of personal data from firms to consumers.[6] We now investigate the validity and limitations of this principle by explicitly modeling the negotiations between firm 1 and firm 2 over the formation of a data linkage, taking into account the constraints imposed by each policy and examining the welfare implications. Our exposition assumes that firm 1 is a large online platform that holds all the bargaining power vis-à-vis firm 2, but our results do not rely on a specific bargaining protocol.

We begin with the benchmark of a fully unregulated market, where linkages can be freely established by the two firms. We then look at the impact of introducing mandatory *transparency*. As prescribed by the GDPR, transparency requires that firm 1 must inform the consumer of whether a linkage with firm 2 is in place before the consumer makes a purchase.[7] Next, we consider the additional requirement of explicit consumer *consent* (also a provision of GDPR and CCPA). Mandatory consent means that firm 1 must inform the consumer of its intention to form a linkage with firm 2 and that the linkage can be activated only with the explicit permission of the consumer. Finally, we consider the additional impact of forbidding discriminatory offers by requiring that firms offer the same terms of trade to consumers whether they grant or deny their consent to the formation of a linkage.

Throughout this section, we maintain several assumptions. First, we assume that the consumer must complete a transaction at time 1, or equivalently, that the mean of her willingness to pay is sufficiently high. Second, we assume that the firms cannot commit to terms of trade before a data linkage is formed, i.e., they cannot induce the consumer to share her data with the promise of lower prices or better products. Finally, we consider ex ante consumer surplus as our welfare criterion.

5.1 No Regulation

We first consider a fully unregulated environment, in which firms have no commitment power (and consumers have no veto rights) over linkage formation.

In this scenario, the timing of the game is as follows:

[6] See, for example, the European Strategy for Data outlined by European Commission (2020).

[7] Importantly, transparency policies also let firm 1 commit to not share data with firm 2.

1. Firm 1 offers a price and a quality level (p_1, y_1) to the consumer.

2. The consumer learns her type θ and selects a quantity q_1.

3. Firm 1 decides whether to activate a linkage with firm 2.

4. The game described in Section 2 is played under the resulting information structure.

While we do not specify an explicit bargaining protocol between the two firms for the formation of a linkage, we assume that they bargain efficiently under complete information. Therefore, the links that form are those that increase the sum of the two firms' profits after the first transaction with the consumer has occurred. From Proposition 1, we know that the value of information for firm 2 is strictly positive for any value of λ_2. By sequential rationality, the consumer anticipates that the outcome of efficient bargaining is that every possible linkage is formed and, hence, chooses her first-period quantity as in the baseline model with a data linkage. Proposition 5 summarizes the above discussion.

Proposition 5 (No Regulation)
In the absence of privacy regulation, data linkage $\lambda_1 \to \lambda_2$ forms for every pair (λ_1, λ_2)

In terms of consumer surplus, the total absence of privacy is clearly problematic. Consumers would like a linkage to form if and only if $(\lambda_1, \lambda_2) \in \Lambda^{CS}$, which we characterized in Proposition 4. In terms of total producer surplus, some privacy would also be desirable: data sharing benefits firm 2 for all values of λ_2, but it decreases firm 1's overall profits if λ_2 is sufficiently negative, even if firm 1 can extract the entire value of information from firm 2. The overall effect of a linkage on total profits is positive only if $(\lambda_1, \lambda_2) \in \Lambda^{PS}$, which we also characterized in Proposition 4. Clearly, firm 1 suffers from the lack of commitment power. In particular, firm 1 would like to maintain the consumer's privacy to avoid strong ratchet forces, but it cannot commit to doing so.

5.2 Transparency

A first principle of the recent privacy legislation is that of *transparency*, according to which the consumer should be informed of how her personal data will be used and shared. For example, the GDPR establishes that "*Where personal data relating to a data subject are collected (...) the controller shall provide the data subject with (...) the recipients or categories of recipients of the personal data.*" Similarly, the California Consumer Privacy Act (CCPA), which was passed in 2018, establishes that "*A consumer shall have the right to request that* *a business that collects personal information about the consumer disclose (...) the categories of third parties with whom the business shares personal information."*

In the context of our model, mandatory transparency requires firm 1 to announce whether a linkage with firm 2 will be formed *before* any interaction with consumer 1 occurs. The timing of the game therefore becomes as follows:

1. Firm 1 commits to whether a linkage with firm 2 will be formed.

2. The game described in Section 2 is played under the resulting information structure.

At the beginning of the game, the two firms bargain efficiently. Therefore, they agree to form the linkage $\lambda_1 \to \lambda_2$ if and only if it increases total producer surplus, as in Figure 5.

Proposition 6 (Mandatory Transparency)
Under mandatory transparency of data sharing, the linkage $\lambda_1 \to \lambda_2$ forms if and only if $(\lambda_1, \lambda_2) \in \Lambda^{PS}$.

This regulation clearly increases total producer surplus because it allows firm 1 to commit to privacy whenever the profit loss for firm 1 exceeds the gains from trade of information. The effect on consumer surplus depends on the two firms' types. On the one hand, requiring transparency improves consumer surplus compared to the absence of regulation by preventing the formation of some harmful linkages, i.e., all those $(\lambda_1, \lambda_2) \notin \Lambda^{PS} \cup \Lambda^{CS}$. On the other hand, transparency also prevents the formation of a beneficial linkage whenever $(\lambda_1, \lambda_2) \in \Lambda^{CS} \setminus \Lambda^{PS}$.

5.3 Consent

A second principle that is currently well-established in the privacy protection legislation is that consumer data processing and sharing requires the explicit consent of the consumer. In particular, the GDPR requires that *"Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data."* The CCPA requires that *"A consumer shall have the right, at any time, to direct a business that sells personal information about the consumer to third parties not to sell the consumer's personal information."* We now study the consequences of combining transparency and consent requirements.

In our model, requiring the consumer's consent for the sale of transaction data grants her "veto power" over the formation of a linkage. Because transparency already offers the firms an opportunity to prevent unprofitable linkages, the addition of consent gives both parties *de facto* veto rights.

In this environment, the timing of the game is as follows:

1. Firm 1 decides whether to propose forming a linkage with firm 2 to the consumer.
2. The consumer decides whether to grant or deny consent.
3. The consumer learns her type θ.
4. The game described in Section 2 is played under the resulting information structure.

In this game, the consumer makes her consent decision before learning her realized willingness to pay.[8] Therefore, the consumer will grant consent to data sharing if and only if the linkage increases her expected surplus, i.e., if $(\lambda_1, \lambda_2) \in \Lambda^{CS}$. In turn, firm 1 will only ask for the consumer's permission if it expects her to grant consent and if a data linkage improves producer surplus. Therefore, as a result of mutual veto rights, the linkages that will form are only those that constitute a Pareto improvement over anonymous trading. We formalize this intuition in Proposition 7.

Proposition 7 (Mandatory Transparency and Consent)
If transparency and consumer consent are mandatory requirements for data sharing, the linkage $\lambda_1 \to \lambda_2$ is formed if and only if $(\lambda_1, \lambda_2) \in \Lambda^{PS} \cap \Lambda^{CS}$.

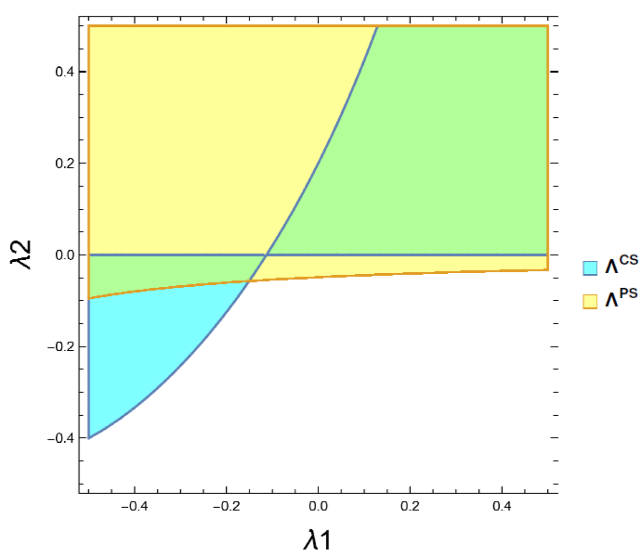

Figure 6: Linkages Formes Under Mandatory Transparency and Consent $(\hat{\sigma} = 1/10)$

We then immediately obtain the following corollary.

[8] With this assumption, we capture the idea that each consumer visits a specific firm's website repeatedly and agrees or disagrees with its "terms of use" independently of her current-day inclination to shop.

Corollary 1 (Consent vs. Transparency)
Requiring consumer consent for the formation of a linkage benefits consumers relative to a transparency requirement only.

Despite the improvement in consumer surplus, mandating consent imposes stringent conditions (i.e., Pareto improvements) on the formation of a link. As such, it prevents the formation of many *socially beneficial* linkages, which we characterize in Section 5.5 below.

5.4 No Discrimination

An even stronger form of privacy regulation allows firms to share transaction data if the consumer is made aware of the data sharing and explicitly consents to it but forbids the firms to offer different terms of trade to the consumer based on her consent choice. The GDPR introduces the no-discrimination principle through the requirement that *"Consent should not be regarded as freely given if the data subject has no genuine or free choice."* The CCPA makes this principle more explicit by requiring that "*A business shall not discriminate against a consumer because the consumer exercised any of the consumer's rights (...) by: (1) Denying goods or services to the consumer, (2) Charging different prices or rates, (3) Providing a different level or quality.*"

The conventional motivation for this rule is fairness. In the current policy debate, there is growing concern that allowing firms to offer different terms of trade to consumers who deny their consent to data transfer can turn privacy into a luxury good, *"worsening unequal access to privacy and further enabling predatory and discriminatory behavior"* (Elvy, 2017).

In our model, a no-discrimination clause modifies the timing of the game as follows:

1. Firm 1 decides whether to propose forming a linkage with firm 2 to the consumer.

2. The consumer decides whether to grant or deny consent.

3. Firm 1 sets the terms of trade (p_1, y_1) *without observing the consumer's decision.*

4. The consumer learns her type θ.

5. The game described in Section 2 is played under the resulting information structure.

Our choice of making the consumer's decision unobservable to firm 1 rules out discriminatory behavior without allowing the firm to obtain the consumer's consent by committing to the terms of trade.[9]

[9]It also has the realistic feature that most "consent boxes" appear on a webpage before the consumer can see the price for any product. However, the characterization of the equilibrium set of linkages in Proposition 8 below does not rely on this no-commitment assumption.

The main difference from a simple consent requirement is that firm 1 must set the terms of trade (p_1, y_1) anticipating whether the consumer will decide to grant or deny consent. In particular, firm 1 offers the equilibrium terms of trade under privacy (Proposition 1) if it anticipates that the consumer will deny her consent. If instead the firm expects that the consumer will consent to the linkage, the firm optimally offers the terms of trade in the separating equilibrium of Proposition 2

In turn, the consumer chooses whether to consent to the formation of a data linkage $\lambda_1 \to \lambda_2$ taking firm 1's choice of terms of trade (p_1, y_1) as given. However, we show in Proposition 8 that the consumer prefers to grant consent if and only if

$$\lambda_2(2\widehat{\sigma}^2 - \lambda_2) \geq 0, \tag{15}$$

for any first-period terms of trade. Furthermore, because condition (15) can hold only if $\lambda_2 \geq 0$, firm 1 will profitably propose all linkages that satisfy this condition. We can then characterize and illustrate the equilibrium set of data linkages as follows.

Proposition 8 (No Discrimination)
If transparency and consumer consent are required for data sharing and discrimination is forbidden, the linkage $\lambda_1 \to \lambda_2$ is formed if and only if $\lambda_2 \in [0, 2\widehat{\sigma}^2]$, for all λ_1.

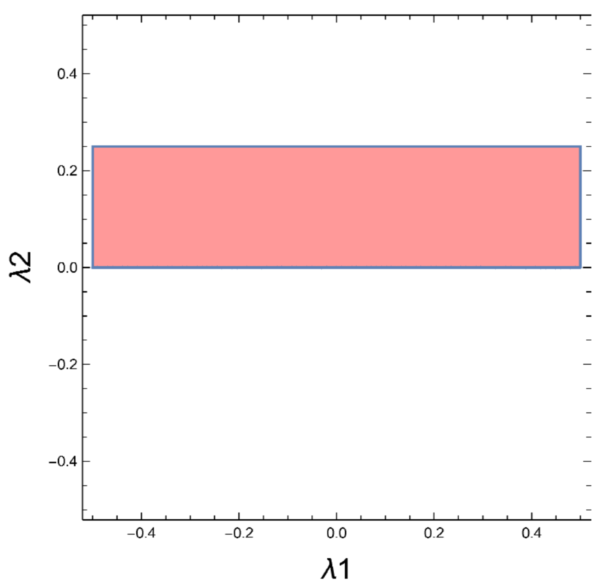

Figure 7: Linkages Formes Under No Discrimination $(\hat{\sigma} = 1/8)$

In the presence of this type of regulation, granting consent is always detrimental to consumer surplus in the first period. The reason is simple: when a data linkage is formed, the consumer distorts her demand away from the myopic optimum, but firm 1 cannot react by adjusting the terms of trade. Therefore, no compensating terms of trade effect can occur in the first period.[10]

Why, then, would the consumer want to consent to the transmission of her data? If $\lambda_2 < 0$, information revelation would worsen her terms of trade at time 2, thus reducing her time 2 surplus as well. Therefore, the consumer refuses consent for any $\lambda_2 < 0$, which explains the bottom half of Figure 7. If instead $\lambda_2 > 0$, the consumer obtains a higher surplus at time 2 by granting consent, which can potentially offset the first-period loss. However, the consumer surplus "triangle" lost by distorting behavior is proportional to $(\lambda_2)^2$, while the value of information at time 2 is proportional to λ_2. Therefore, the consumer grants consent if λ_2 is positive but small. Finally, because the value of information is increasing in the prior uncertainty about her type, the threshold λ_2 for granting consent is also increasing in $\hat{\sigma}$.

How does consumer welfare under this policy compare to the outcome of the previous, less restrictive policies? Contrary to the common wisdom that discrimination allows predatory behavior, the comparison of the equilibrium set of linkages in Propositions 7 and 8 suggests the opposite. In particular, for $\lambda_2 < 0$, no linkages form if discrimination is not allowed. However, for sufficiently negative λ_1, the consumer would allow data sharing in exchange for better terms of trade, which is the equilibrium outcome under a simple consent policy. The same is true for $\lambda_2 > 2\widehat{\sigma}^2$ and positive and sufficiently large λ_1. Finally, for $0 < \lambda_2 < 2\widehat{\sigma}^2$, firm 1 successfully proposes forming a linkage under a no-discrimination policy. When λ_1 is sufficiently negative, however, the consumer pays a higher price in the first period than she would under anonymity, if discrimination were allowed. In other words, she would be better off denying consent if, by doing so, she induced the equilibrium terms of trade under privacy, but that cannot happen under this policy.

We then draw a stark conclusion about adding the no-discrimination requirement to a policy that already requires the consumer's explicit consent for data sharing.

Corollary 2 (Banning Discrimination)

If transparency and consumer consent are mandatory requirements for data sharing, banning price and quality discrimination unambiguously damages the consumer.

[10] This observation also explains why the terms of trade (p_1, y_1) do not impact the consumer's consent decision: their effect on the quantity purchased is independent of θ; hence, they do not affect the distortion in quantity relative to the static optimum (which is given by $\lambda_2\theta$). An implication of this property is that commitment to the terms of trade would have no value for firm 1.

5.5 Direct Payments for Consent

We conclude this section by sketching a complete and efficient market for consumer information. We assume that transparency and consumer consent are mandatory and that firm 1 is allowed to offer a direct (positive or negative) payment to the consumer in exchange for her consent to forming a linkage with firm 2. For ease of exposition, assume further that firm 1 has all the bargaining power vis-à-vis firm 2 and the consumer, i.e., that it extracts all the surplus from the formation of a link. Because bargaining is assumed to be efficient, firm 1 proposes the linkage $\lambda_1 \to \lambda_2$ if and only if this linkage increases *social* surplus. In Figure 8, we characterize the set of welfare-improving linkages (λ_1, λ_2).

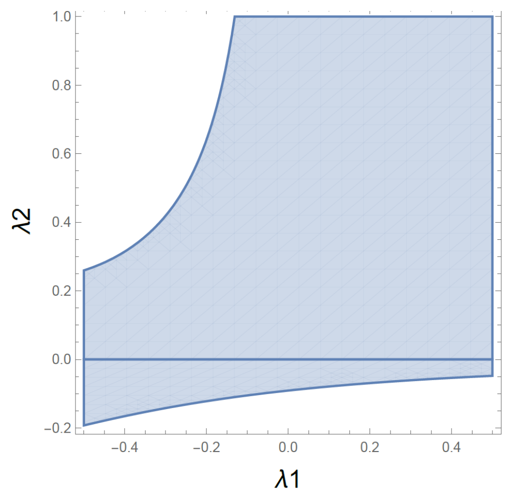

Figure 8: Socially Efficient Linkages $(\hat{\sigma} = 1/2)$

In a static version of our model, the social value of information is positive for all λ larger than a threshold $\lambda^* < 0$. In a dynamic model with a data linkage, the consumer has an incentive to distort her demand, and the situation becomes more complex. Specifically, suppose the second-period firm has a large $\lambda_2 > 0$: if the first period firm has $\lambda_1 < 0$, any linkage between these two firms causes a considerable loss in consumer surplus due to higher monopoly prices and upward quantity distortions in the first period. Likewise, for large $\lambda_1 > 0$, any linkage with $\lambda_2 < 0$ causes an inefficient reduction in consumer demand and underinvestment in product quality. The resulting loss is more severe for larger values of λ_1, for which the consumer's average consumption is higher. Thus, relative to the simple cutoff policy of Proposition 1, the social planner would form all linkages such that (heuristically) $\lambda_1 \cdot \lambda_2$ is not excessively negative and would only form linkages with $\lambda_2 > 0$ as λ_1 grows large. Proposition 11 in the Appendix formalizes this intuition.

In this scenario, the consequences for consumer welfare relative to the outcome of regulation depend heavily on the distribution of bargaining power. In our stylized setting, where firm 1 has all the bargaining power, the consumer is as well off as under privacy for any (λ_1, λ_2). This outcome is weakly worse than that under mandatory consent for any (λ_1, λ_2), but the ranking relative to *laissez faire*, transparency, or no discrimination is sensitive to the specific values of λ_1, λ_2, and $\hat{\sigma}$.

6 Extensions

Our model of data markets makes several simplifying assumptions. We show in this section that the driving forces of equilibrium behavior are robust to two important features of real-world markets. In particular, we first allow the first-period firm to form linkages with multiple, heterogeneous second-period firms. We then allow for competition in the second period among firms with the same type but differentiated products. In both cases, we derive the unique linear equilibrium and discuss its properties, but the analysis of welfare-improving linkages also readily extends to these richer environments.

6.1 Multiple Data Uses

Consider a consumer who interacts with a single firm at $t = 1$ and with a continuum of heterogeneous firms at $t = 2$. We refer to λ_t as the type of firm t. While the type of the period-1 firm λ_1 is commonly known, the type of the each second-period firm λ_2 is drawn from a distribution F with support $\Lambda \subseteq [-1/2, \infty)$. Thus, once collected, the consumer's data can be used in a large number of ways. An alternative, equivalent interpretation is that the consumer faces uncertainty over the type of the period-2 firm.

Recall that the expected surplus of consumer θ when interacting with second-period firm λ_2 is given by (7), i.e.,

$$U_2^*(\theta, m, \lambda_2) = \frac{1}{2}(\theta + \lambda_2 m)^2.$$

Clearly, the firm's posterior belief m will vary depending on whether the firm has access to the period-1 outcome data.

We now characterize the equilibrium strategies and payoffs when the first-period outcome is observed by a measurable subset of period-2 firms $\Lambda^o \subseteq \Lambda$. Thus, all firms $\lambda_2 \in \Lambda^o$ observe (p_1, y_1, q_1) prior to setting their price and quality levels, while the remaining firms $\lambda_2 \in \Lambda \setminus \Lambda^o$ operate under the prior distribution only.

Upon receiving a first-period offer (p_1, y_1) and facing the prospect of firms $\lambda_2 \in \Lambda^o$

observing the first-period outcome, the consumer solves the following problem

$$\max_{q_1}\left[U_1(\theta,q_1,p_1,y_1,\lambda_1)+\int_{\Lambda^o}U_2^*(\theta,m(q_1),\lambda_2)\,dF(\lambda)+\int_{\Lambda\setminus\Lambda^o}U_2^*(\theta,\mu,\lambda_2)\,dF(\lambda)\right].$$

Proposition 9 characterizes the equilibrium strategies for an arbitrary "linked set" Λ^o.

Proposition 9 (Equilibrium with Multiple Uses)
For any linked set Λ^o, there exists a unique linear equilibrium of the game.

1. *In the first period, the consumer's demand function is given by*

$$q_1^*(\theta,p_1,y_1)=\alpha^*(\Lambda^o)\,\theta+b_1y_1-p_1,$$

where

$$\alpha^*(\Lambda^o) \triangleq \frac{1}{2}\left(1+\sqrt{4k(\Lambda^o)+1}\right),\ \textit{and} \tag{16}$$

$$k(\Lambda^o) \triangleq \int_{\Lambda^o}(1+\lambda)\,\lambda dF(\lambda). \tag{17}$$

2. *Firm 1 offers terms of trade $(p_1^*(\Lambda^o), y_1^*(\Lambda^o))$ that satisfy*

$$b_1y_1^*(\Lambda^o)-p_1^*(\Lambda^o)=\alpha^*(\Lambda^o)\,\lambda_1\mu.$$

3. *In the second period, all players follow the strategies in Proposition 1, with each firm λ forming its beliefs according to its information set.*

As in the case of a deterministic λ_2, the consumer's manipulation incentives introduce a distortion in her first-period behavior that affects only the weight of the consumer's type in the equilibrium quantity.[11] Furthermore, the first-period terms of trade effect is entirely unchanged: firms $\lambda_1 > 0$ raise prices and quality levels when the set of firm-2 linked firms Λ^o leads the consumer to manipulate upward, i.e., to set $\alpha^* > 1$.

However, the consumer's incentives to manipulate her behavior are more responsive to their true type when the future interaction is uncertain. To formalize this comparison, we rewrite the function $k(\Lambda^o)$ in (17) as

$$k(\Lambda^o)=F(\Lambda^o)\left[\mathbb{E}[\lambda\,|\Lambda^o]+\mathbb{E}[\lambda\,|\Lambda^o]^2+\mathrm{var}[\lambda\,|\Lambda^o]\right].$$

[11] The case of a single, deterministic λ_2 corresponds to the case where the distribution $F(\lambda_2)$ is degenerate. In that case, the right-hand side of (16) reduces to $\alpha(\lambda_2)=1+\lambda_2$, which is the expression in Proposition 2.

The consumer responds more aggressively to her type when the nature of the second-period interaction is stochastic, relative to the deterministic case in which $\text{var}[\lambda|\Lambda^o] = 0$. This occurs because the incentives to manipulate are related to the consumer's type through the product of two terms: first, the marginal value of a higher θ on the continuation value vis-à-vis firm λ is given by $1+\lambda$; second, the marginal value of manipulating the firm's belief is itself λ. Thus, the marginal benefit of manipulation is a convex function of λ.

6.2 Competing Firms

We now return to our baseline model, but we introduce competition in the second period. In particular, the consumer interacts with a monopolist firm of type λ_1 in the first period. She then faces two period-2 firms that sell differentiated products and compete in prices and qualities. The second-period firms share a common value of quality b_2. We let (p_{2j}, y_{2j}, q_{2j}) denote the second-period actions, with $j = 1, 2$. To maintain the assumption of linear demand, we let the utility function of the consumer in the second period be given by

$$U_2(p, y, q) \triangleq \frac{1}{2}\sum_{j=1}^{2}\left[(\theta + b_2 y_{2j} - p_{2j})q_{2j} - \frac{1}{2}q_{2j}^2\right] - sq_{21}q_{22}, \tag{18}$$

where $s \in [0, 1)$ captures with the degree of substitutability of the two products, i.e., the intensity of second-period competition.

We now characterize the unique linear equilibrium of the game in which the first-period firm has formed a linkage with both second-period competitors.

Proposition 10 (Equilibrium with Second-Period Competition)
For any $s \in [0, 1)$, there exists a unique linear equilibrium of the game.

1. *The consumer's t_1 demand function is given by*

$$q_1^*(\theta, p_1, y_1) = \alpha^*(s)\,\theta + b_1 y_1 - p_1,$$

where

$$\alpha^*(s) \triangleq \frac{1}{2} + \frac{1}{2}\sqrt{4\hat{\lambda}(s) + 1}, \tag{19}$$

$$\hat{\lambda}(s) \triangleq \frac{b_2^2 + s^2 - 1}{(2 - b_2^2 + s(1-s))^2}. \tag{20}$$

2. *Firm 1 offers terms of trade (p_1^*, y_1^*) that satisfy*

$$b_1 y_1^* - p_1^* = \alpha^*(s)\,\lambda_1\mu.$$

For moderately fierce competition in the second stage, the consumer's behavior is qualitatively identical to the case of a monopoly. Distortions again affect only the coefficient on θ, and the terms of trade effect is unchanged from the baseline case as a function of the coefficient α^*. Competition does, however, have a quantitative effect on the consumer's equilibrium behavior. As we can see from expressions (19) and (20), the equilibrium coefficient α^* is increasing in s for all b_2. Furthermore, α^* is larger than one for all $b_2 \geq \sqrt{1-s^2}$, which is strictly lower than the threshold $b_2 = 1$ in the case of monopoly. Intuitively, fiercer competition in the second period alleviates the ratchet effect. Conversely, competition creates a greater incentive for the consumer to be perceived as high type to receive higher-quality products at lower prices than under monopoly.

7 Conclusions

We have developed a simple model that provides a microfoundation for consumers' preferences over the collection and transmission of behavior data by heterogeneous firms. We have shown that the impact of data linkages on consumer surplus critically depends on the degree of similarity of the collecting and receiving firms. For instance, in markets where product quality is not a salient dimension, allowing the use of purchase histories for price discrimination need not harm sophisticated consumers. In contrast, a consumer can be harmed if purchase histories are used to determine whether she is worthy of high-quality products or customer service.

Our welfare results inform the evaluation of current privacy protection regulation both in the EU and in the US. In particular, we have stressed the importance of assigning property rights over personal information to consumers because these rights endow consumers with veto power over harmful linkages. Because firms still hold proposal power over linkage formation, however, recent policies implement mutual veto rights that may lead to a socially suboptimal level of data sharing. In other words, the imperfect instruments available to compensate consumers for the loss in privacy drive a wedge between the current regulatory environment and an efficient market for information.

As we have shown, our model extends to imperfectly competitive product markets, which opens the possibility for future work to analyze the welfare implications of mergers and acquisitions from a data sharing angle. Should merging firms operating in different markets be allowed to share the data they collect about their transactions with consumers? How would this affect consumer surplus directly (through the effects presented in this paper) and indirectly (through the impact of competition)? Finally, under which conditions do mergers foreclose competitors by limiting their access to consumers' transaction data?

A Appendix

Proof of Proposition 1. The realized utility, profits, and welfare are, respectively:

$$
\begin{aligned}
U(\theta, m, \lambda) &= \frac{(\theta + m\lambda)^2}{2} \\
\Pi(\theta, m, \lambda) &= \frac{m(2\theta - m)(1+\lambda)}{2} \\
W(\theta, m, \lambda) &= \frac{(\theta + m\lambda)^2}{2} + \frac{m(2\theta - m)(1+\lambda)}{2}
\end{aligned}
$$

Taking expectations over (θ, m), we obtain the consumer's ex ante welfare, the firm's profits and total welfare. If the firm learns that $m = \theta$, the expected utility, profits and welfare are given by

$$
\begin{aligned}
\mathbb{E}[U \mid \mathcal{I}^*] &= \frac{1}{2}\left(\mu^2 + \sigma^2\right)(1+\lambda)^2, \\
\mathbb{E}[\Pi \mid \mathcal{I}^*] &= \frac{1}{2}\left(\mu^2 + \sigma^2\right)(1+\lambda), \\
\mathbb{E}[W \mid \mathcal{I}^*] &= \frac{1}{2}\left(\mu^2 + \sigma^2\right)(1+\lambda)(2+\lambda).
\end{aligned}
$$

If instead the firm has only access to the prior distribution of θ, we have $m = \mu$, and the expected utility and profits are given by:

$$
\begin{aligned}
\mathbb{E}[U \mid \varnothing] &= \mathbb{E}_\theta\left[\frac{(\theta + \mu\lambda)^2}{2}\right] = \frac{\mu^2(1+\lambda)^2 + \sigma^2}{2} \\
\mathbb{E}[\Pi \mid \varnothing] &= \mathbb{E}_\theta\left[\frac{\mu(2\theta - \mu)(1+\lambda)}{2}\right] = \frac{1}{2}\mu^2(1+\lambda)
\end{aligned}
$$

(1.) The change in consumer surplus is

$$
\Delta U \triangleq \mathbb{E}[U \mid \mathcal{I}^*] - \mathbb{E}[U \mid \varnothing] = \frac{\sigma^2}{2}\lambda(\lambda + 2)
$$

which is positive iff $\lambda > 0$.

(2.) When the firm has exogenous information and $m = \theta$, the change in profits is

$$
\Delta \Pi \triangleq \mathbb{E}[\Pi \mid \mathcal{I}^*] - \mathbb{E}[\Pi \mid \varnothing] = \frac{\sigma^2(1+\lambda)}{2} > 0.
$$

(3.) Finally, the change in social welfare is

$$\Delta W \triangleq \Delta U + \Delta \Pi = \frac{\sigma^2}{2}\left(\lambda^2 + 3\lambda + 1\right).$$

Given the domain of $\lambda \in [-1/2, \infty)$, a data linkage improves social welfare for

$$\lambda \geq \lambda^* = -\left(3 - \sqrt{5}\right)/2,$$

which is strictly negative. ■

Proof of Proposition 2. We seek to construct an equilibrium where the consumer's first-period strategy takes the form

$$q_1 = \alpha\theta + \beta y_1 + \gamma p_1 + \delta. \tag{21}$$

With this linear demand function, the firm maximizes its expected profits,

$$\mathbb{E}_\theta\left[\Pi_1\right] = p_1\left(\alpha\mu + \beta y_1 + \gamma p_1 + \delta\right) - \frac{y_1^2}{2}.$$

The first-order conditions for the firm's problem with respect to (p_1, y_1) are given by

$$\begin{aligned} p_1\beta - y_1 &= 0, \\ 2\gamma p_1 + y_1\beta + \delta + \alpha\mu &= 0. \end{aligned}$$

Therefore, if firm 1 conjectures the demand as in (21), its optimal choices of price and quality are given by

$$p_1^* = -\frac{\delta + \alpha\mu}{\beta^2 + 2\gamma} \tag{22}$$

$$y_1^* = -\beta\frac{\delta + \alpha\mu}{\beta^2 + 2\gamma}. \tag{23}$$

Next, we solve the consumer's problem and derive the equilibrium values of the coefficients of her linear demand.

The consumer maximizes (9), i.e., the sum of her current flow utility U_1 and her expected second period utility, which is given by (7). Under first-period demand (21), firm 2 forms a degenerate posterior belief over the consumer's type,

$$m\left(q_1\right) = \frac{q_1 - \beta y_1 - \gamma p_1 - \delta}{\alpha}. \tag{24}$$

The consumer anticipates (24) and therefore, upon observing the choice of (p_1, y_1), she solves the following problem:

$$\max_{q_1} \left[U_1(\theta, q_1) + U_2^*(\theta, m(q_1)) \right].$$

The first-order condition with respect to q_1 is given by

$$\theta + b_1 y_1 - p_1 - q_1 + \lambda_2 m'(q_1)(\theta + \lambda_2 m(q_1)) = 0.$$

Under (24) above, this is a linear equation in q_1. Solving this condition for q_1 yields the following linear function of (θ, y_1, p_1):

$$q_1^* = \frac{\theta + b_1 y_1 - p_1 + \frac{\lambda_2}{\alpha}\left(\theta + \lambda_2 \frac{-\beta y_1 - \gamma p_1 - \delta}{\alpha}\right)}{1 - (\lambda_2)^2 / \alpha^2}.$$

Matching the coefficients to those in (21) we obtain a unique solution to the resulting system of linear equations, which pins down the equilibrium strategies:

$$\begin{aligned} \alpha^* &= 1 + \lambda_2 \\ \beta^* &= b_1 \\ \gamma &= -1 \\ \delta &= 0. \end{aligned}$$

Substituting into conditions (21)-(23) yields the equilibrium strategies in the statement. ■

Proof of Proposition 3. (1.) By Proposition 2, the consumer's first-period realized payoff with a data linkage can be written as

$$U(p_1^*, y_1^*, q_1^*) = \frac{1}{2}(\lambda_2 + 1)(\theta + \mu\lambda_1)\left[\theta(1 - \lambda_2) + \mu\lambda_1 + \mu\lambda_1\lambda_2\right].$$

Therefore, the expected first-period consumer surplus is

$$\mathbb{E}U_1 = \frac{\mu^2}{2}(\lambda_2 + 1)(1 + \lambda_1)(1 - \lambda_2 + \lambda_1 + \lambda_1\lambda_2) + \frac{\sigma^2}{2}\left(1 - \lambda_2^2\right)$$

Without a data linkage, the consumer's expected utility is given by

$$\mathbb{E}U_1^p = \frac{\mu^2}{2}(1 + \lambda_1)^2 + \frac{\sigma^2}{2}.$$

The value of the linkage is therefore equal to

$$\Delta U_1 \triangleq \mathbb{E}U_1 - \mathbb{E}U_1^p = \frac{\mu^2}{2}\left(\lambda_2\left(\lambda_1+1\right)\left(2\lambda_1-\lambda_2+\lambda_1\lambda_2\right)-\hat{\sigma}^2\lambda_2^2\right)$$

Solving the right hand side for λ_2 yields two roots:

$$\lambda_2 \in \left\{0, \lambda_1\frac{2\left(1+\lambda_1\right)}{\hat{\sigma}^2+1-\lambda_1^2}\right\}.$$

For all $\lambda_1 < \sqrt{1+\hat{\sigma}^2}$, the coefficient on the quadratic term in λ_2 is negative and the surplus-improving linkages are in between the two roots. For $\lambda_1 > \sqrt{1+\hat{\sigma}^2}$ the coefficient is positive and the second root negative, therefore all $\lambda_2 > 0$ improve consumer surplus.

(2.) For producer surplus, Proposition 2 implies that expected profits with and without a linkage are given by

$$\mathbb{E}\Pi_1 = \mathbb{E}_\theta\left[\Pi_1^*\right] = \frac{\mu^2}{2}\left(\lambda_2+1\right)^2\left(\lambda_1+1\right)$$

and

$$\mathbb{E}\Pi_1^p = \frac{\mu^2}{2}\left(1+\lambda_1\right),$$

respectively. The difference between these two then has the same sign as λ_2. ∎

Proof of Proposition 4. (1.) Summing the changes in consumer surplus in the two periods (Propositions 3 and proof of Proposition 1), the overall change due to the introduction of a linkage is proportional to

$$\Delta U = \left(\lambda_1+1\right)\lambda_2\left(2\lambda_1-\lambda_2+\lambda_1\lambda_2\right)+2\hat{\sigma}^2\lambda_2. \tag{25}$$

Solving for λ_2 we obtain the two roots

$$\lambda_2 \in \left\{0, \frac{2\left(\hat{\sigma}^2+\lambda_1\left(1+\lambda_1\right)\right)}{1-\lambda_1^2}\right\}.$$

If $\lambda_1 > 1$, the second root is negative and the quadratic term on λ_2 is positive in (25). Therefore values of λ_2 for which $\Delta U > 0$ are all $\lambda_2 > 0$. Conversely, if $\lambda_1 < 1$, all values for which $\Delta U > 0$ are in between the two roots.

(2.) Summing the changes in profits in the two periods, the overall effect of the introduction of a linkage is proportional to

$$\Delta\Pi = \left(1+\lambda_1\right)\lambda_2\left(\lambda_2+2\right)+\hat{\sigma}^2\left(1+\lambda_2\right). \tag{26}$$

Because $\lambda_2 > -1$, the right-hand side of expression (26) is increasing in λ_2. Solving for λ_2 and selecting the larger root yields

$$\lambda_2 \geq \sqrt{\left(\frac{\hat{\sigma}^2}{2(\lambda_1+1)}\right)^2+1}-\frac{\hat{\sigma}^2}{2(\lambda_1+1)}-1,$$

which is the condition for $\Delta\Pi \geq 0$ in the statement. ■

The proofs for Propositions 5-7 in Section 5 are given in the text.

Proof of Proposition 8. If the consumer gives consent, her expected utility is

$$\begin{aligned}\mathbb{E}_\theta U_1 &= \mathbb{E}_\theta\left\{(\theta+by_1-p_1)\left[(1+\lambda_2)\theta+by_1-p_1\right]-\frac{\left[(1+\lambda_2)\theta+by_1-p_1\right]^2}{2}\right\}\\ &= (\mu+by_1-p_1)\left[(1+\lambda_2)\mu+by_1-p_1\right]-\frac{\left[(1+\lambda_2)\mu+by_1-p_1\right]^2}{2}+\frac{\sigma^2\left(1-\lambda_2^2\right)}{2}\end{aligned}$$

in the first period and

$$\mathbb{E}_\theta U_2 = \frac{\left(\mu^2+\sigma^2\right)(1+\lambda_2)^2}{2}$$

in the second. If instead she denies consent, her expected utility is

$$\mathbb{E}_\theta U_1^p = \mathbb{E}_\theta\left[\frac{(\theta+by_1-p_1)^2}{2}\right]=\frac{(\mu+by_1-p_1)^2+\sigma^2}{2}$$

in the first period and

$$\mathbb{E}_\theta U_2^p = \frac{\mu^2(1+\lambda_2)^2+\sigma^2}{2}$$

in the second period. Giving consent is optimal iff

$$\mathbb{E}_\theta U_1+\mathbb{E}_\theta U_2-\mathbb{E}_\theta U_1^p-\mathbb{E}_\theta U_2^p \geq 0$$

The above condition can be simplified to

$$\frac{1}{2}\lambda_2\left(2\widehat{\sigma}^2-\lambda_2\right)\geq 0,$$

which establishes the result. ■

Proof of Proposition 9. We now characterize a linear equilibrium in which the consumer plays the first period strategy

$$q_1 = \alpha\theta + \beta y_1 + \gamma p_1 + \delta. \tag{27}$$

In the second period, firms set prices as in (4) and (5), with $m = \mu$ for $\lambda \notin \Lambda^o$ and $m = m(q)$ as in (24). The consumer accordingly uses her myopic demand function and obtains $U_2^*(\theta, \lambda_2, m)$ as in (7).

Under this period-2 conjecture, the consumer's period 1 can be written as

$$W(\theta) = \max_q \left[\begin{array}{c} (\theta + b_1 y_1 - p_1) q - \frac{q^2}{2} + \frac{1}{2}\int_{\Lambda^o} \left(\theta + \lambda \frac{q - (\beta y_1 + \gamma p_1 + \delta)}{\alpha}\right)^2 \mathrm{d}F(\lambda) \\ + \frac{1}{2}\int_{\Lambda \backslash \Lambda^o} (\theta + \lambda\mu)^2 \mathrm{d}F(\lambda). \end{array} \right] \tag{28}$$

If the strategy (27) is an equilibrium, then it satisfies the first-order condition for the consumer's problem (28)

$$\theta + b_1 y_1 - p_1 - q_1 + \int_{\Lambda^o} \frac{\lambda}{\alpha} \left(\theta + \lambda \frac{q_1 - (\beta y_1 + \gamma p_1 + \delta)}{\alpha}\right) \mathrm{d}F(\lambda) = 0$$

as well as (27). Substituting the latter into the f.o.c., we obtain

$$0 = \theta + b_1 y_1 - p_1 - (\alpha\theta + \beta y_1 + \gamma p_1 + \delta) + \frac{\theta}{\alpha} \int_{\Lambda^o} (1 + \lambda) \lambda \mathrm{d}F(\lambda).$$

Matching coefficients, we obtain the unique solution

$$\beta = b_1,\ \gamma = -1,\ \delta = 0,$$

and

$$1 - \alpha + \frac{k(\Lambda^o)}{\alpha} = 0,$$

where $k(\Lambda^o)$ is defined as in (17). We solve for α and select the unique positive root for α, so that the resulting prices and qualities in (29)-(30)

$$p_1^*(\Lambda^o) = \frac{\alpha^*(\Lambda^o)}{2 - b_1^2}\mu \tag{29}$$

$$y_1^*(\Lambda^o) = b_1 \frac{\alpha^*(\Lambda^o)}{2 - b_1^2}\mu \tag{30}$$

are non-negative. This yields equation (16), and completes the proof. ■

Proof of Proposition 10. We begin with the second-period behavior and equilibrium values. Taking the first order conditions in (18) with respect to (q_{21}, q_{22}) and solving for q_{21} and q_{22}, we obtain

$$q_{2i} = \frac{1}{1-s^2}\left((1-s)\theta + b_2\left(y_{2i} - sy_{2j}\right) - p_{2i} + sp_{2j}\right), \text{ for } i,j = 1,2 \text{ and } i \neq j.$$

Given a posterior belief m, each firm i at t_2 then maximizes

$$\Pi_{2i} = p_{2i}q_{2i} - \frac{1}{2}y_{2i}^2,$$

which yields the following first order conditions:

$$\begin{aligned} 0 &= \frac{1}{1-s^2}\left((1-s)m + b_2y_{2i} - sb_2y_{2j} - 2p_{2i} + sp_{2j}\right), \\ 0 &= \frac{1}{1-s^2} \cdot p_{2i}b_2 - y_{2i}. \end{aligned}$$

Solving for a symmetric equilibrium yields the following expressions:

$$\begin{aligned} y_2^* &= \frac{bm}{2 - b^2 + s\left(1-s\right)} \\ p_2^* &= \frac{(1-s^2)m}{2 - b^2 + s\left(1-s\right)}. \end{aligned}$$

Note that $b^2 \leq 2$ and $s^2 \leq s$ so the denominator is non-zero. The resulting terms of trade in period 2 are given by

$$b_2y_2^* - p_2^* = \frac{b^2 - (1-s^2)}{2 - b^2 + s\left(1-s\right)}m \triangleq \lambda_2\left(s\right)m. \tag{31}$$

We then compute the second-period utility of a consumer of type θ when interacting with a pair of firms with beliefs m, which is given by

$$U_2^*\left(\theta, m\right) = \frac{1}{2}\frac{\left(\theta + \lambda_2\left(s\right) \cdot m\right)^2}{1+s}.$$

Now suppose there was a linear equilibrium where

$$q_1 = \alpha\theta + \beta y_1 + \gamma p_1 + \delta.$$

Given the first period linear demand, the second period firms form a degenerate belief on the consumer's type

$$m\left(q_1\right)=\frac{q_1-\beta y_1-\gamma p_1-\delta}{\alpha}.$$

The consumer anticipates this and therefore solves the following problem

$$\max_{q_1}\left[U_1\left(\theta,q_1\right)+U_2^*\left(\theta,m\left(q_1\right)\right)\right],$$

the first order condition for which is given by

$$\theta+b_1y_1-p_1-q_1+m'\left(q_1\right)\frac{\lambda_2\left(s\right)}{1+s}\left(\theta+\lambda m\left(q_1\right)\right)=0.$$

Solving for q_1 yields

$$q_1^*=\frac{\theta+b_1y_1-p_1+\frac{\lambda_2(s)}{\alpha(1+s)}\left(\theta+\lambda_2\left(s\right)\frac{-\beta y_1-\gamma p_1-\delta}{\alpha}\right)}{1-\frac{\lambda_2(s)^2}{(1+s)\alpha^2}}$$

Matching the coefficients above we obtain a unique system of linear equations which pins down the strategies described in (19).

$$\begin{aligned}
\alpha^*&=\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{4\lambda_2\left(s\right)\left(\lambda_2\left(s\right)+1\right)}{s+1}}\\
\beta^*&=b_1\\
\gamma&=-1\\
\delta&=0
\end{aligned}$$

To complete the proof, we need to show that the term in the square root is always positive. Recall the definition of $\lambda_2\left(s\right)$ in (31), which implies

$$\frac{\lambda_2\left(s\right)\left(\lambda_2\left(s\right)+1\right)}{s+1}=\frac{b_2^2+s^2-1}{\left(2-b_2^2+s\left(1-s\right)\right)^2}$$

This expression is minimized at $b_2=s=0$, yielding a value of $-1/4$, so the square root is in fact always positive. ■

Proposition 11 (Social Welfare)

There exist two thresholds $\tilde{\lambda}_2(\lambda_1, \hat{\sigma})$ and $\tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma})$ satisfying $\tilde{\lambda}_2(\lambda_1, \hat{\sigma}) < 0$ for all $\lambda_1, \hat{\sigma} \geq 0$, and $\tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma}) > 0$ for all $\lambda_1 < 0 < \hat{\sigma}$, such that the following hold.

1. *For $\lambda_1 \geq 0$ all linkages with $\lambda_2 \geq \tilde{\lambda}_2(\lambda_1, \hat{\sigma})$ increase social welfare.*
2. *For $\lambda_1 < 0$, all linkages $\lambda_2 \in [\tilde{\lambda}_2(\lambda_1, \hat{\sigma}), \tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma})]$ increase social welfare.*

Proof of Proposition 11. The total change in social welfare ΔW due to the formation of a linkage can be obtained by adding lines (25) and (26):

$$\Delta W = \frac{\mu^2}{2}(\lambda_1 + 1)\lambda_2(2\lambda_1 + \lambda_1\lambda_2 + 2) + \frac{\sigma^2}{2}(3\lambda_2 + 1)$$

Dividing by μ^2, multiplying by 2, and rearranging, we obtain

$$\Delta W \propto \lambda_2(\lambda_2 + 2)\lambda_1^2 + \lambda_2(\lambda_2 + 4)\lambda_1 + (3\lambda_2 + 1)\hat{\sigma}^2 + 2\lambda_2. \tag{32}$$

This is a quadratic expression in λ_2 with a coefficient $\lambda_1(1 + \lambda_1)$ on the quadratic term. The two roots are given by

$$\tilde{\lambda}_2(\lambda_1, \hat{\sigma}) \triangleq \frac{-3\hat{\sigma}^2 - 2(1 + \lambda_1)^2 + \sqrt{-4\hat{\sigma}^2\lambda_1(1 + \lambda_1) + \left(3\hat{\sigma}^2 + 2(1 + \lambda_1)^2\right)^2}}{2\lambda_1(1 + \lambda_1)},$$

and

$$\tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma}) \triangleq \frac{-3\hat{\sigma}^2 - 2(1 + \lambda_1)^2 - \sqrt{-4\hat{\sigma}^2\lambda_1(1 + \lambda_1) + \left(3\hat{\sigma}^2 + 2(1 + \lambda_1)^2\right)^2}}{2\lambda_1(1 + \lambda_1)}.$$

Note that the term in the root is always positive. Furthermore, one can show the following properties.

Whenever $\lambda_1 \geq 0$, we have $0 > \tilde{\lambda}_2(\lambda_1, \hat{\sigma}) > -1/2 > \tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma})$ for all $\hat{\sigma} \geq 0$, and the expression (32) has a positive coefficient on the quadratic term. Therefore, all $\lambda_2 \geq \tilde{\lambda}_2(\lambda_1, \hat{\sigma})$ increase social welfare.

Whenever $\lambda_1 < 0$, we have $-1/2 < \tilde{\lambda}_2(\lambda_1, \hat{\sigma}) < 0 < \tilde{\tilde{\lambda}}_2(\lambda_1, \hat{\sigma})$ and (32) has a negative coefficient on the quadratic term. Therefore all $\lambda_2 \in [\tilde{\lambda}_2, \tilde{\tilde{\lambda}}_2]$ increase social welfare.

This ends the proof. ■

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019): "Too Much Data: Prices and Inefficiencies in Data Markets," Discussion paper, MIT.

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

Acquisti, A., and H. R. Varian (2005): "Conditioning prices on purchase history," *Marketing Science*, 24(3), 367–381.

Ali, S. N., G. Lewis, and S. Vasserman (2019): "Voluntary Disclosure and Personalized Pricing," Discussion Paper 26592, National Bureau of Economic Research.

Ball, I. (2020): "Scoring Strategic Agents," Discussion paper, Yale University.

Baye, M. R., and D. E. Sappington (2020): "Revealing transactions data to third parties: Implications of privacy regimes for welfare in online markets," *Journal of Economics & Management Strategy*, 29(2), 260–275.

Belleflamme, P., and W. Vergote (2016): "Monopoly price discrimination and privacy: The hidden cost of hiding," *Economics Letters*, 149, 141–144.

Bergemann, D., A. Bonatti, and T. Gan (2020): "The Economics of Social Data," Discussion Paper 2203R, Cowles Foundation for Research in Economics, Yale University.

Bonatti, A., and g. Cisternas (2020): "Consumer Scores and Price Discrimination," *Review of Economic Studies*, 87, 750–791.

Calzolari, G., and A. Pavan (2006): "On the optimality of privacy in sequential contracting," *Journal of Economic Theory*, 130(1), 168–204.

Choi, J., D. Jeon, and B. Kim (2019): "Privacy and Personal Data Collection with Information Externalities," *Journal of Public Economics*, 173, 113–124.

Conitzer, V., C. R. Taylor, and L. Wagman (2012): "Hide and seek: Costly consumer privacy in a market with repeat purchases," *Marketing Science*, 31(2), 277–292.

Cummings, R., K. Ligett, M. Pai, and A. Roth (2016): "The Strange Case of Privacy in Equilibrium Models," in *ACM-EC (Economics and Computation) 2016*.

Elvy, S.-A. (2017): "Paying for Privacy and the Personal Data Economy," *Columbia Law Review*, 117, 1369–1459.

European Commission (2020): *A European Strategy for Data.* European Commission, Brussels.

Fainmesser, I., A. Galeotti, and R. Momot (2020): "Digital Privacy," Discussion paper, Johns Hopkins University.

Frankel, A., and N. Kartik (2019): "Improving information from manipulable data," *arXiv preprint arXiv:1908.10330*.

Ichihashi, S. (2020): "The Economics of Data Externalities," Discussion paper, Bank of Canada.

Jann, O., and C. Schottmüller (2020): "An informational theory of privacy," *Economic Journal*, 130(625), 93–124.

Jullien, B., Y. Lefouili, and M. Riordan (2020): "Privacy Protection, Security, and Consumer Retention," Discussion paper, Toulouse School of Economics.

Laffont, J., and J. Tirole (1988): "The Dynamics of Incentive Contracts," *Econometrica*, 59, 1735–1754.

Liang, A., and E. Madsen (2020): "Data and Incentives," Discussion paper, Northwestern University.

Paes Leme, R., M. Pal, and S. Vassilvitskii (2016): "A Field Guide to Personalized Reserve Prices," in *Proceedings of the 25th international conference on world wide web*, pp. 1093–1102.

Taylor, C. R. (2004): "Consumer privacy and the market for customer information," *Rand Journal of Economics*, 35(4), 631–650.

Tirole, J. (2020): "Digital Dystopia," Discussion paper, Toulouse School of Economics.

Turow, J. (2008): *Niche envy: Marketing discrimination in the digital age.* MIT Press.