Data Markets with Privacy-Conscious Consumers

By ROSSELLA ARGENZIANO AND ALESSANDRO BONATTI*

A salient feature of online markets is that many platforms, firms, merchants, websites can collect consumer data. Once collected, these data can be used in a large number of ways. Some uses of data generate value for the consumer, for example by personalizing product features and offering tailored service quality. Some other uses are more adverse, for example personalized pricing and product steering.

As awareness of data collection increases and privacy becomes a more salient dimension of the policy debate, it becomes critical to understand how privacy-conscious consumers react to the possibility that their data be collected, traded, and ultimately used in a future transaction.

In this short paper, we explore how privacy-conscious consumers strategically react when they *know* their data will be used, but they face uncertainty as to exactly *how* it will be used. We focus on the equilibrium effects of a data market. In our model, consumers expect the terms of trade in future transactions to be informed by their current behavior. They then seek to manipulate the data-using firms' beliefs about their preference type by distorting their demand for the products of data-collecting firms.

We show that the direction of the consumer's behavior distortion depends on the distribution of data uses the consumer expects. We provide a microfoundation for collaborative vs. adverse uses in a static game, and we leverage this characterization in a dynamic game to show that the consumer's signaling incentives can lead to both *ratchet* effect and *niche envy* effect, depending on properties of the data uses distribution.

Under these rich equilibrium effects, it is not a priori clear that a market for consumer data is even profitable for all firms. For example, suppose the data using firm monetizes the information so gained through personalized prices. The resulting drop in consumer's propensity to buy–the ratchet effect–can erode the entire value of information, to the point that the data collecting firm would be better off committing to offering full privacy (Calzolari and Pavan, 2006).

Our approach explains the existence of active markets for information—proxied by positive net gains from trade of consumer data. We identify several forces that contribute to raising the firms' value of acquiring (and selling) the consumer's data. A necessary condition is that there exist limitations to the data using firms' pricing instruments. Other factors include a sufficiently favorable distribution of future uses; large uncertainty over future data use; and large uncertainty over the consumer's type.

Our paper joins a vast body of work on the economics of privacy and markets for information surveyed, e.g., by Acquisti, Taylor and Wagman (2016), and Bergemann and Bonatti (2019). Our model is most directly related to the behavior-based price discrimination literature (Fudenberg and Villas-Boas, 2006, 2012), with seminal contributions by Villas-Boas (1999), Taylor (2004), Acquisti and Varian (2005), Calzolari and Pavan (2006), Zhang (2011), and most recently Baye and Sappington (2020). Relative to these papers, our model allows for heterogeneous sources and heterogeneous uses of data. Our model also formalizes conditions on the distribution of firm and consumer types that make data trades profitable.

Our analysis is limited to a setting with very little regulatory control, where consumers are aware of data markets but do not influence data trades directly—they do so only indirectly though their strategic behavior. In a new regulatory regime, such as the one introduced by the EU GDPR and California's CPRA, consumers can specify which uses of their information they consent to. In our companion paper (Argenziano and Bonatti, 2022), we study how insti-

* Argenziano: Economics Department, University of Essex, Colchester CO4 3SQ, UK, rargenz@essex.ac.uk. Bonatti: MIT Sloan School of Management, Cambridge, MA 02142, bonatti@mit.edu. Argenziano acknowledges financial support through BA/Leverhulme grant SRG18R1–180705. Bonatti acknowledges financial support through NSF Grant SES–1948692. We thank our discussant Jidong Zhou and participants in the "Privacy in Markets" session at the 2023 ASSA meetings for helpful comments.

tutional details and property rights assignments affect the data markets that emerge and how they impact consumers' welfare.

I. Model

Consider a single consumer who lives for two periods and interacts with two firms sequentially: a *data-collecting* firm in the first period and a *data-using* firm in the second period. The active firm in each period $t = 1,2$ sets a quality level y_t and charges a unit price p_t for its product. The consumer, in turn, purchases a quantity q_t. The consumer's per-period utility is given by

$$U^*(\theta, p_t, y_t, q_t) = (\theta + b_t y_t - p_t) q_t - \frac{q_t^2}{2}, \quad (1)$$

and firm t's profits are given by

$$\Pi(p_t, y_t, q_t) = p_t q_t - \frac{c_t y_t^2}{2}. \quad (2)$$

The consumer's type θ captures her "baseline" willingness to pay per unit of the product, i.e., the intercept of her demand curve before accounting for the firm's investment in quality y_t.

The parameters b_t and c_t denote firm t's characteristics. In particular, b_t captures the relative salience of price and quality from *every* consumers' perspective. In particular, the case $b_t = 0$ corresponds to a pure price-setting firm.

Each firm t has a constant marginal cost of producing quantity q_t that we normalize to zero and a quadratic cost of producing quality y_t that is scaled by c_t. We assume that the sensitivity of the consumer's utility to quality satisfies $b_t \in [0, \sqrt{2c_t})$ in each period $t = 1,2$.

The consumer's type θ is fixed over time. It is distributed on a compact set $\Theta \subset \mathbb{R}_+$ with mean μ and variance σ^2. The consumer privately observes the realized type θ at the beginning of period 1. The characteristics (b_1, c_1) of firm 1 are commonly known at the onset of the game. In contrast, the characteristics (b_2, c_2) of firm 2 are unknown to the first-period firm and to the consumer. They are drawn from a known distribution and observed by all players at the beginning of the second period. We interpret this draw as the realization of the consumer's period-2 *need*, which is unknown to all players in period 1.

The two firms differ in their information structure: firm 1 sets (p_1, y_1) on the basis of the prior distribution, while firm 2 observes the outcome of the first-period transaction (p_1, y_1, q_1) before interacting with the consumer.

The timing of our game is the following:

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

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

3) Firm 2 observes (p_1, y_1, q_1) and offers price p_2 and quality level y_2 to the consumer.

4) The consumer observes firm 2's characteristics (b_2, c_2) and selects a quantity q_2.

We focus on *linear equilibria*, as defined in Ball (2020). These are (fully separating) Bayesian Nash equilibria in which the consumer's strategy is linear in her type and the second-period firm's strategy is linear in the first-period outcome variables.

II. The Static Game

Consider a benchmark static model with a single firm with characteristics (b, c). The consumer observes the firm's offer (p, y) and maximizes the current-period utility (1). Thus, she chooses the following quantity:

$$q^*(\theta, p, y) = \theta + by - p. \quad (3)$$

For any choice of quality y, integrating (3) over the consumer's types yields the demand curve

$$\mathbb{E}[q^*(\theta, p, y)] = \mu + by - p. \quad (4)$$

The firm then chooses the monopoly price that maximizes its expected profits (2)

$$p(y) = \frac{1}{2}(\mu + by).$$

Viewed through this lens, the firm's choice of quality y is a costly investment in quality that shifts out the demand curve. The firm's problem then consists of identifying the optimal investment y^* given the distribution of consumer types and considering that monopoly pricing enables the firm to appropriate only a fraction of the surplus it generates.

PROPOSITION 1 (Static Equilibrium): *The static equilibrium quality and price are given by*

$$y^*(\mu,b,c) = \frac{b}{2c-b^2}\mu, \tag{5}$$

$$p^*(\mu,b,c) = \frac{1}{2c-b^2}\mu. \tag{6}$$

Because the firm's optimal actions are linear in the expectation of the consumer's type μ, the optimal price and quality increase in μ.

Intuitively, the firm invests more and charges a higher price when it expects that the consumer will buy more units. The net impact of the firm's beliefs on its offer to the consumer is summarized by the *terms of trade*, which we define as the price-adjusted quality level:

$$by^*(\mu,b,c) - p^*(\mu,b,c) = \lambda(b,c)\mu, \tag{7}$$

$$\text{with} \quad \lambda(b,c) \triangleq \frac{b^2-c}{2c-b^2}. \tag{8}$$

Because the parameters satisfy $b \in [0,\sqrt{2c})$, the function λ takes values in $[-1/2,\infty)$.

When the effect of the firm's quality on consumer demand b is high relative to the marginal cost of investment c, i.e. when $\lambda(b,c) > 0$, the firm offers better terms of trade when its their prior beliefs on θ improve: their optimal investment in quality y increases faster than the monopoly price p, which benefits consumers.

Substituting (7) into the demand function (3) and using the definition of λ in (8) above, we obtain the realized quantity

$$q^*(\theta,p^*(\mu,\lambda),y^*(\mu,\lambda)) = \theta+\lambda\mu,$$

and the realized consumer utility for type θ,

$$\begin{aligned} U(\theta,\mu,\lambda) =& \frac{1}{2}q^*(\theta,p^*(\mu,\lambda),y^*(\mu,\lambda))^2 \\ =& \frac{1}{2}(\theta+\lambda\mu)^2. \end{aligned} \tag{9}$$

Therefore, the sign of the firm's type $\lambda(b,c)$ determines how both the equilibrium terms of trade and the equilibrium consumer surplus respond to changes in the firm's beliefs. Hence, we shall refer to λ_t as the period-t *firm's type*.

III. The Dynamic Game

We turn to our dynamic model where the consumer faces uncertainty over the type of the firm that will use her data. The type λ_1 of the first-period firm is commonly known, while the type of the second-period firm λ_2 is drawn from a distribution F with support $\Lambda \subseteq [-1/2,\infty)$. Recall from (9) that the expected surplus of consumer θ when interacting with a second-period firm of type λ_2 that holds beliefs $m = \mathbb{E}[\theta]$ is given by

$$U_2^*(\theta,m,\lambda_2) = (\theta+\lambda_2 m)^2/2.$$

The second-period firm's posterior mean m depends on the observed first-period transaction and on the consumer's conjectured strategy.

Suppose the consumer receives a first-period offer (p_1,y_1), fix the second-period firm's conjecture, and let $m(q_1)$ denote the firm's beliefs as a function of the purchased quantity. The consumer solves the following problem

$$\begin{aligned} \max_{q_1} \big[& U_1(\theta,p_1,y_1,q_1;\lambda_1) \\ & + \int_\Lambda U_2^*(\theta,m(q_1),\lambda_2)\,dF(\lambda_2). \end{aligned}$$

Differentiating the consumer's objective with respect to the second period firms' beliefs and evaluating at $m=\theta$, we obtain the consumer's incentive to distort the first-period quantity:

$$\frac{\partial \int_\Lambda U_2^*(\theta,\theta,\lambda_2)\,dF(\lambda_2)}{\partial m} = \theta\kappa. \tag{10}$$

where

$$\kappa \triangleq \mathbb{E}_F[\lambda_2(1+\lambda_2)]. \tag{11}$$

The expression in (10) highlights three critical properties of our model. First, the consumer's incentives to manipulate the second-period firm's beliefs are proportional to her type, because high-θ consumers buy more units and benefit more from an improvement in the *terms of trade*. Second, the *direction* of the consumer's manipulation depends on the sign of κ. Loosely, if the consumer assigns a large probability to interacting with firms with $\lambda_2 < 0$, she will be wary of the *ratchet effect* (Laffont and Tirole, 1988) and distort her purchases downward; conversely, she will exhibit *niche envy* (Turow, 2008) and distort her purchases upward. Third, because the marginal benefit of manipulating a given firm's beliefs is quadratic in λ_2, the statistic κ is a convex function of F. Therefore, the consumer has a stronger incentive to manipulate upward when the nature of the second-period interaction is more uncertain.

PROPOSITION 2 (Dynamic Equilibrium): *There exists a unique linear equilibrium.*

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

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

where

$$\alpha^* \triangleq (1+\sqrt{4\kappa+1})/2. \tag{12}$$

2) *In the first period, firm 1 offers terms of trade (p_1^*, y_1^*) that satisfy*

$$b_1 y_1^* - p_1^* = \alpha^* \lambda_1 \mu.$$

3) *In the second period, players follow the strategies in Proposition 1, with the firm's beliefs given by $m(q_1^*(\theta)) = \theta$.*

Proposition 2 shows that the consumer's manipulation incentives influence both the sensitivity of the first-period quantity to θ and the first-period terms of trade. The former is larger (smaller) than in the static game depending on the sign of κ. The magnitude of the latter is magnified by α^* (e.g., larger than in the static equilibrium when $\kappa > 0$) but its sign is still determined by the first-period firm's type λ_1.

IV. The Market for Consumer Data

We now turn to the implications of our dynamic equilibrium for the profitability of trading consumer-level transaction data. In particular, we examine whether a data transfer agreement between the two firms is profitable ex ante (i.e., before the consumer's type θ and the second period firm type λ_2 are realized).

Several equivalent interpretations for this arrangement are possible. First, the two firms may trade information before knowing whether firm λ_2 will meet the consumer in the second period. Second, the consumer may interact with a single firm at $t = 1$ and with a continuum of "small" heterogeneous firms at $t = 2$, and the first-period firm negotiates with all second-period firms jointly. Third, the second-period firm may be a multiproduct firm that faces uncertainty over which good the consumer needs.

In all these settings, a necessary condition for the trade of consumer data to be profitable is that it raises aggregate producer surplus. We therefore consider whether the firm's intertemporal profits are larger in the dynamic equilibrium relative to the appropriate static benchmark.

An immediate implication of Proposition 1 is that, in the absence of a data transfer, the expected profit of any (first- or second-period) firm with type λ is given by

$$\mathbb{E}[\Pi(\lambda) \mid \varnothing] = \frac{\mu^2}{2}(1+\lambda). \tag{13}$$

When the consumer's data is traded, however, the second-period firm operates under complete information I^*. Its profits then increase to

$$\mathbb{E}[\Pi(\lambda) \mid \mathrm{I}^*] = \frac{\mu^2+\sigma^2}{2}(1+\lambda). \tag{14}$$

Finally, the first-period firm's profits in the dynamic equilibrium reflect the consumer's manipulation incentives and the adjustment in the terms of trade. As a function of the period-1 firm's type, these profits are given by

$$\mathbb{E}[\Pi^*(\lambda_1)] = \frac{\mu^2}{2}(\alpha^*)^2(1+\lambda_1). \tag{15}$$

Combining these terms, we obtain the total gains from trading information for the firms:

$$\Delta\Pi = \frac{\mu^2}{2}\left(\alpha^{*2}-1\right)(1+\lambda_1) + \frac{\sigma^2}{2}(1+\mathbb{E}_F[\lambda_2]). \tag{16}$$

Recall that $\alpha^* > 1$ in (12) if and only if $\kappa > 0$ in (11), which means $\mathbb{E}_F[\lambda_2] + \mathbb{E}_F[\lambda_2^2] > 0$. Therefore, we can identify three factors that are conducive to an active market for transaction data.

PROPOSITION 3 (Market for Data): *If firms bargain efficiently, trading consumer data is profitable when either:*

1) *the expected type of the data-using firm satisfies $\mathbb{E}_F[\lambda_2] > 0$;*

2) *the uncertainty over the use of this information $\mathrm{var}_F[\lambda_2]$ is large enough; or*

3) *the uncertainty over the consumer's type σ^2 is large enough.*

Note that, unlike the expected type of data-using firm, a higher type for the data collecting firm does not necessarily facilitate the market for data. In particular, a higher λ_1 increases the gains from trade only when $\alpha^* > 1$ and the consumer tries to manipulate beliefs upward by buying more. Conversely, when $\alpha^* < 1$, a higher-λ_1 firm stands to lose even more from the consumer distorting her demands downward and would prefer not to sell the information.

V. Conclusion

We have developed a tractable model of consumer behavior in the presence of data markets. Privacy-conscious consumers distort their purchases from a data-collecting firm to manipulate the data-using firm's beliefs over their willingness to pay. The direction of the consumer's desired manipulation can be upward (for data-using firms that personalize quality) or downward (for data-using firms that personalize prices). As such, our framework captures both the *ratchet* and the *niche envy* effect.

The availability of transaction-level information enables the data-using firm to better tailor its strategy to the consumer's type. This firm always has a positive value of information. In contrast, the strategic behavior of privacy-conscious consumers has rich implications for the equilibrium profits of the data-collecting firm. Combining these two forces, we have identified conditions under which data linkages increase total producer surplus. Therefore, if firms can trade data efficiently, our setting with limited second-period pricing instruments provides a rationale for the existence of data markets even in the presence of privacy-conscious consumers.

In parallel work (Argenziano and Bonatti, 2022), we extend this model to study how different regulatory regimes affect the emergence of some, but not all, data markets as well as their implications for consumers' welfare.

REFERENCES

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

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

Argenziano, Rossella, and Alessandro Bonatti. 2022. "Data Linkages and Privacy Regulation." Essex and MIT.

Ball, I. 2020. "Scoring Strategic Agents." Yale University.

Baye, Michael R, and David 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.

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

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

Fudenberg, Drew, and J Miguel Villas-Boas. 2006. "Behavior-Based Price Discrimination and Customer Recognition." *Handbook on Economics and Information Systems.*

Fudenberg, Drew, and J Miguel Villas-Boas. 2012. "Price Discrimination in the Digital Economy." In *The Oxford Handbook of the Digital Economy*. , ed. M. Peitz and J. Waldfogel. Oxford University Press.

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

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

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

Villas-Boas, J Miguel. 1999. "Dynamic Competition with Customer Recognition." *RAND Journal of Economics*, 604–631.

Zhang, Juanjuan. 2011. "The Perils of Behavior-Based Personalization." *Marketing Science*, 30(1): 170–186.

Mathematical Appendix

PROOF OF PROPOSITION 1:

In a static game, the consumer's demand function is given by (3). Substituting the expected demand function (4) into the firm's profit (2) and maximizing with respect to p and y yields the result.

PROOF OF PROPOSITION 2:

We now characterize a linear equilibrium in which the consumer plays the first period strategy

$$q_1 = \alpha\theta + \beta y_1 + \gamma p_1 + \delta. \tag{A1}$$

In the second period, the firm with realized type λ_2 firms set prices as in (5) and (6), where

$$m(q_1) = \frac{q_1 - (\beta y_1 + \gamma p_1 + \delta)}{\alpha} \tag{A2}$$

replaces μ. The consumer uses her static demand function and obtains $U_2^*(\theta, m(q_1), \lambda_2)$ as in (9).

Given that the period-2 firm's updates its beliefs according to (A2), the consumer solves

$$\max_q \left[(\theta + b_1 y_1 - p_1)\, q - \frac{q^2}{2} + \frac{1}{2} \int_\Lambda (\theta + \lambda_2 m(q_1))^2 \, \mathrm{d}F(\lambda_2) \right].$$

The first-order condition for the consumer's period-1 problem is then given by

$$\theta + b_1 y_1 - p_1 - q_1 + \int_\Lambda \frac{\lambda_2}{\alpha} \left(\theta + \lambda_2 \frac{q_1 - (\beta y_1 + \gamma p_1 + \delta)}{\alpha} \right) \mathrm{d}F(\lambda_2) = 0. \tag{A3}$$

Substituting the period-2 firm's conjecture (A1) into (A3) and matching coefficients, we obtain

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

and

$$1 - \alpha + \frac{\kappa}{\alpha} = 0,$$

where κ is defined as in (11). Selecting the unique positive root yields α^* as in (12). Finally, solving the period-1 firm's problem, the equilibrium terms of trade follow from the optimal price and quality

$$p_1^* = \frac{1}{2 - b_1^2} \alpha^* \mu \quad \text{and} \quad y_1^* = \frac{b_1}{2 - b_1^2} \alpha^* \mu.$$

PROOF OF PROPOSITION 3:

Notice first that from equations (13)-(15), we can write the difference in profits $\Delta\Pi(\lambda_1)$ as

$$\Delta\Pi(\lambda_1) = \frac{\mu^2}{2} (\alpha^*)^2 (1 + \lambda_1) + \frac{\mu^2 + \sigma^2}{2} (1 + \mathbb{E}[\lambda_2]) - \frac{\mu^2}{2} (2 + \lambda_1 + \mathbb{E}[\lambda_2]).$$

Simplifying then yields (16). Part (1.) of the statement follows from the fact that $\operatorname{supp} F \subset [-1/2, \infty)$ and hence $1 + \mathbb{E}_F[\lambda_2] > 0$ for all F. Therefore, if $\mathbb{E}_F[\lambda_2] > 0$ then $\alpha^* > 1$ and both terms in (16) are positive.

Part (2.) uses the facts that $1 + \lambda_1 > 0$ and that α^* in (12) increases without bound as $\kappa \to \infty$. Moreover, we can rewrite (11) as

$$\kappa = \mathbb{E}_F[\lambda_2] + \mathbb{E}_F[\lambda_2]^2 + \operatorname{var}_F[\lambda_2].$$

Part (3.) follows from the observation that the right-hand side of (16) is linear in σ.