Personalization and Privacy Choice[*]

Andrew Rhodes[†]   Jidong Zhou[‡]

April 2025

Abstract

This paper studies consumers' privacy choices when firms can use their data to make personalized offers. We first introduce a general framework of personalization and privacy choice, and then apply it to personalized recommendations, personalized prices, and personalized product design. We argue that due to firms' reaction in the product market, consumers who share their data often impose a negative externality on other consumers. Due to this privacy-choice externality, too many consumers share their data relative to the consumer optimum; moreover, more competition, or improvements in data security, can lower consumer surplus by encouraging more data sharing.

Keywords: personalization, consumer data, privacy, personalized pricing, personalized recommendations, personalized product design

JEL classification: D43, D82, L13

1 Introduction

Two important trends in digital markets are increased personalization and heightened concerns around consumer privacy. Firms increasingly have access to rich consumer-level data, as well as sophisticated AI prediction tools. This enables them to offer consumers a personalized shopping experience—in the form of tailored ads and recommendations, personalized prices, and even personalized products.[1] For instance, around 1.8 million US websites offer personalized recommendations (Donnelly, Kanodia, and Morozov, 2024), and recommendations drive 30% of sales on Amazon, 70% of views on YouTube, and 80% of engagement on Pinterest (CDEI, 2020). However, largescale collection and processing of data also raises privacy concerns. For example, it is estimated that ad tech firms observe over 90% of a typical consumer's browsing history (CDEI, 2020). Similarly, a consumer's voice assistant listens in on her conversations, her telephone tracks her real-time location, and social media sites infer her emotional state from her photos (Stucke and Ezrachi, 2017). Concerns about consumer privacy have led to initiatives that give consumers more control over what data is collected and how it is used, such as the EU's GDPR and California's CCPA, as well as Apple's app tracking transparency policy.

Personalization and privacy choice are clearly intertwined: it is easier for firms to personalize when consumers share their data, but a consumer's decision of whether to share her data also depends on how personalization affects her utility. In this paper we provide a simple framework to study the interaction between data-driven personalization and consumer privacy choice. Each consumer first decides independently whether to disclose her data, and a market game between firms and consumers then follows. Consumers have rational expectations about how their data-sharing decisions will affect their consumption utility from the product market; they also face heterogeneous intrinsic privacy costs (e.g., due to data security concerns) or benefits (e.g., due to better service or compensation) from data sharing.[2]

---

[1]According to Deloitte (2018) these are currently the main sources of AI-driven personalization. For examples of personalization, see, e.g., https://bit.ly/48KT28R.

[2]An emerging empirical literature seeks to quantify how much consumers value privacy. For example, Prince and Wallsten (2022) offer survey evidence that people value privacy differently depending on the country and the type of data in question. (For instance, Germans value privacy more than people in the other five countries in their survey. Banking information such as balance is worth about \$8.5/month and browsing history is worth about \$3.75.) Lin (2022) documents lab evidence that (intrinsic) privacy cost is relatively small (per demographic variable) but highly heterogeneous across consumers, while Tang (2019) shows in a large-scale field experiment that privacy cost in an online lending market is much larger. See

The core economic force in our model is a privacy-choice externality across consumers: when some consumers share their data, this affects not only the offers that firms make to them, but may also affect the offers made to other consumers. If the payoffs of sharing and anonymous consumers both decrease as more other consumers share their data, we show that the privacy choice equilibrium features too much data sharing relative to the consumer optimum. (This is true even though aggregate consumer surplus can vary non-monotonically with the number of sharing consumers.) In contrast, the equilibrium features too little data sharing if sharing and anonymous consumers' payoffs increase as more other consumers share. We then study three applications of the framework, and argue that the former case often arises. In that case, we further demonstrate that more competition (due to an increase in the number of competitors in the product market) or improved data security (which decreases consumer privacy costs) can have a perverse effect on aggregate consumer welfare by increasing overall data sharing in the market.

We now briefly explain the privacy-choice externality in each of the three applications.

Personalized recommendations. As already mentioned, online firms routinely offer personalized recommendations. We consider a set-up in which consumers can search firms on a platform to learn about prices and match values, but if they share their data, the platform will recommend (without bias) their best-matched product. One might expect consumers to benefit from such recommendations because they no longer need to search. However, anticipating that consumers who receive recommendations become less likely to search any non-recommended product, firms have an incentive to raise their prices. This obviously harms consumers who hide their data and so do not receive recommendations. Even for consumers who receive recommendations, this adverse price effect can also outweigh the informational benefit.[3] Moreover, we show that as the number of competing firms increases, more consumers choose to share their data, which leads to higher prices and so potentially lower consumer welfare.

Personalized pricing. Another common use of consumer data is personalized pricing, i.e., charging different consumers different prices based on estimates of their willingness-to-pay.[4] To avoid a consumer backlash, firms often implement personalized prices by offering

---

also Goldfarb and Tucker (2012), Athey, Catalini, and Tucker (2017), and Kummer and Schulte (2019).

[3]Our analysis in this part is also related to targeted advertising. Specifically, ad exchanges provide detailed information about consumers (who shared their data) to advertisers, who then bid for the right to display an ad. Assuming the advertiser with the best match for a consumer wins the auction (and so is seen by the consumer first), as more consumers share their data this creates a similar negative externality.

[4]See https://bit.ly/3A4Rk10 and https://bit.ly/38Ygzq6 for a history of personalized pricing,

targeted discounts off a public price via emails or smartphone app.[5] We consider a set-up in which firms charge a public list price to consumers who hide their data, and offer personalized discounts to consumers who share their data. A higher list price reduces demand from the former consumers, but allows for more flexible pricing to the latter consumers. Hence, as more consumers share their data, firms raise their list prices. This harms consumers who hide their data, as well as sharing consumers who are not offered a discount. We again show that more competition need not always benefit consumers, if it raises their incentives to share data and hence leads to a higher list price.

Personalized product design. In many industries such as apparel, furniture, and health-care, there is a trend towards firms offering products that are individually designed for consumers. By doing this, firms create additional value which they can then extract via a high price; at the same time, firms are constrained by the fact that consumers can often also access publicly-available and less-personalized products which are offered to consumers who have not shared their data. We construct a model with this feature, and show that to extract more surplus from consumers who share their data, firms degrade their public offering by distorting both product designs and prices. This again implies that consumers who choose to share their data impose a negative externality on others.

In this paper we highlight the importance of accounting for privacy-choice externalities across consumers when assessing the impact of privacy policies. For example, a policy like GDPR which enables consumers to costlessly hide their data often improves consumer welfare compared to the case where firms have free access to consumer data. However, due to the privacy-choice externality, there are still too many consumers who share their data relative to the consumer optimum. There are of course other possible reasons for suboptimal consumer privacy choices. For example, this can happen if consumers are not perfectly rational in their privacy decisions (e.g., if consumers put too much weight on the immediate satisfaction from better service after sharing data but underestimate the future cost due to, say, data security problems). Another possible reason is the so-called "social data externality" as advocated in a sequence of recent articles, including Choi, Jeon, and Kim (2019), Ichihashi (2021), Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022), and Bergemann, Bonatti, and Gan (2022). The main idea in these papers is that when a group of consumers have correlated preferences, one consumer's data sharing diminishes the value of other consumers' data. When each consumer makes their data-

---

and OECD (2018), Which? (2018) and http://bit.ly/3E4nDBT for detailed examples.

[5]E.g., see https://bit.ly/37OftAc for how Kroger uses its mobile app to offer personalized coupons.

sharing decision independently, they do not take this externality into account; this results in too much data sharing, and enables data intermediaries to acquire consumer data too cheaply. This externality, however, is purely at the data market level and is independent of the product market channel that we emphasize in our paper.[6]

Some other literature. The economics literature on consumer privacy is exploding (see, e.g., recent surveys by Acquisti, Taylor, and Wagman, 2016, and Bergemann and Bonatti, 2019). Our paper contributes to the study of consumer privacy choice when consumer data is used by firms to make personalized offers. We do not follow the "first-party data" approach in the large literature on purchase-behavior-based discrimination (see, e.g., the survey by Fudenberg and Villas-Boas, 2007) where a firm can only obtain the data of consumers who previously bought from it. Instead we consider broader sources of consumer data: if a consumer shares information about herself with non-merchant sites (such as social media and service apps), firms may be able to purchase this data and hence learn about the preferences of such a consumer.

For each of our applications, there are some closely related works. (Other related papers will be discussed later in the paper after each application.) Anderson and Renault (2000) consider a duopoly version of Wolinsky (1986) in which some consumers perfectly know their product valuations before they search. Since these informed consumers do not actively search, their presence induces firms to raise their prices, similar to sharing consumers (who receive a perfect recommendation) in our application of personalized recommendations. Anderson, Baik, and Larson (2023) study consumer privacy choice in the context of personalized pricing in the second part of their paper, and observe a similar externality of sharing consumers on others, but their setup of personalized pricing is different from ours. Neither Anderson and Renault (2000) nor Anderson, Baik, and Larson (2023) discuss the potential perverse effect of more competition. Our application to personalized product design is most closely related to Bergemann and Bonatti (2024) and Section 7 of Vaidya (2023). These papers address very different research questions, e.g., the first one does not study consumer privacy choice, while the second one focuses on a regulator's choice of what information to allow consumers to disclose. Nevertheless,

---

[6]In addition, at the data market level, Miklós-Thal, Goldfarb, Haviv, and Tucker (2024) study the implications of correlation between different dimensions of user data: when a firm accumulates more consumer data, it is better able to infer sensitive data from non-sensitive data. In this case, data sharing by early users imposes a negative externality on later users even if they share only non-sensitive data. As a result, later users have incentives to withhold all data.

both models eventually boil down to a variant of Mussa and Rosen (1978) where some consumers' preferences are perfectly observable to the firm so that it can offer a personalized product (which is also used in our paper). We discuss these papers in more detail in the section on applications, but they either do not consider the problem of consumer privacy choice explicitly or consider it in the context of a particular type of personalization. The general framework offered in this paper not only helps connect these works but also delivers new insights about the interaction between personalization and consumer privacy choice in general.

Some other recent papers also find privacy-choice externalities across consumers, but in different contexts.[7] For example, Fainmesser, Galeotti, and Momot (2023) study data protection. In their model, a firm chooses how much data to collect from consumers and how much to protect it from hackers, while consumers decide how much data to give the firm. When consumers share more data, they exert both positive externalities (due to better service) and negative externalities (by making the firm a more attractive target for hackers) on other consumers. Unlike in our paper, these externalities are not driven by a market channel. Galperti, Liu, and Perego (2024) study intermediated data markets. (See also Galperti, Levkun, and Perego, 2023.) In their model, a platform first buys data from consumers and then sells it to a firm that can price discriminate. If the platform cares enough about consumers, it pools some consumer types together to avoid perfect discrimination by the seller. A consumer's privacy choice can then influence the seller's belief about the composition of consumers in each market segment, and thus also influence the discriminatory prices it offers. Like in our paper, this generates a cross-consumer externality via a product market channel.[8] However, although they also have externalities across consumers, these papers address different research questions from ours.

Finally, in this paper we assume that consumers make their privacy choices after learning their privacy costs but before learning their preferences over different products. Consequently, a consumer's privacy choice does not convey any information about the latter. This is different from much of the existing literature on privacy choice, such as Belleflamme and Vergote (2016), Montes, Sand-Zantman, and Valletti (2019), Chen,

---

[7]See also pp. 445-446 in Acquisti, Taylor, and Wagman (2016) for a discussion of other types of externality that can arise due to consumer privacy choices.

[8]Chen (2025) studies consumer privacy choice when data generates both data-monetization revenue for the firm and data-driven quality improvement for consumers. His model has a positive cross-consumer externality, because more data sharing induces the firm to invest more in product quality and also induces a higher data-driven quality enhancement, both of which benefit consumers who do not share data.

Choe, and Matsushima (2020), Ichihashi (2020), Hidir and Vellodi (2021), Ali, Lewis, and Vasserman (2023), and Chen (2025) which assumes that consumers know their preferences when deciding whether to share their data. Our assumption captures the idea that when consumers decide whether to share their data—such as by enabling cookies on a news website or allowing tracking on a gaming app—they are aware that the data may later be used to make them personalized offers. However, prior to entering a specific product market, they remain uncertain about their preferences in that market. We elaborate on this assumption in more detail after presenting the model. Two papers that make the same timing assumption as us are Anderson, Baik, and Larson (2023) and Argenziano and Bonatti (2023).

2 Framework

Consider a (representative) market with $n \geq 1$ firms and a unit mass of consumers. A consumer is characterized by her consumption type $\theta \in \Theta$ and her privacy type $\tau \in [\underline{\tau}, \overline{\tau}]$. The consumption type θ captures a consumer's preferences over the firms' products. The privacy type τ is a consumer's net cost from sharing her data; it can be positive or negative, and reflects all costs and benefits from sharing data which are not related to product consumption (e.g., concerns about data security, and benefits from better website/app functionality). Each consumer's θ and τ are drawn independently according to differentiable CDFs $F(\theta)$ and $T(\tau)$ respectively.

We consider the following privacy choice game. At the first stage, each consumer learns her privacy type τ, and decides whether to share her data based on a rational expectation of how her data sharing decision will affect the offers that she will receive. (For simplicity, we do not allow consumers to choose how much or what data to share.) If a consumer shares her data, she pays her privacy cost τ. At the second stage, consumers who shared their data receive a personalized offer (e.g., a recommendation, price, or product) which depends on their θ. Consumers who did not share their data receive a uniform offer which does not depend on their θ. Finally, at the third stage, consumers learn their consumption type θ (which may require costly information acquisition in some applications) and decide which offer (if any) to accept.

We allow for externalities in sharing decisions. Specifically, let σ be the fraction of consumers who share their data. At the third stage, a consumer of type θ gets a consumption surplus $v_s(\theta, \sigma)$ if she shared her data and $v_a(\theta, \sigma)$ if she did not share her

data (and is thus anonymous).[9] Since consumers make their privacy choice before learning their θ, what matters at the first stage is their ex-ante expected consumption surplus. To this end, we let $V_s(\sigma) = \mathbb{E}_\theta[v_s(\theta, \sigma)]$ and $V_a(\sigma) = \mathbb{E}_\theta[v_a(\theta, \sigma)]$ be the expected consumption surpluses for a sharing and an anonymous consumer respectively. We assume that $V_s(\sigma)$ and $V_a(\sigma)$ are finite, and also continuous and differentiable in σ. (We are implicitly assuming here that for a given σ the market equilibrium in the second stage is unique. This is true in all of our applications.) Throughout this paper, we say that sharing consumers exert *negative* externalities if $V_s'(\sigma) < 0$ and $V_a'(\sigma) < 0$ for any $\sigma \in [0, 1]$, and *positive* externalities if $V_s'(\sigma) > 0$ and $V_a'(\sigma) > 0$ for any $\sigma \in [0, 1]$.[10]

Using the above framework, we will examine whether consumers share too much or too little relative to the consumer optimum. We will also study whether more competition (due to an increase in n) or better data security (due to a decrease in the distribution of τ) necessarily benefits consumers when privacy choices are endogenous.[11]

Remarks: Before solving the model, we briefly discuss some of our assumptions. (Several other modeling issues will be discussed later in Section 4.)

(i) Third-party data approach and timing. As mentioned earlier, our model adopts the "third-party data" approach. Consider a consumer who browses various types of websites or uses a wide range of social media platforms and apps. We assume that she needs to make a uniform privacy choice—either allowing her data to be tracked or declining permission altogether. If she consents to tracking, her data becomes accessible to all relevant firms, for instance, through the data market. If she declines, her data remains unavailable to firms.

When making this privacy choice, the consumer understands that her data may influence the offers she will receive in future product markets. However, at this stage—before entering any particular market and beginning the shopping process—she remains uncertain about her specific preferences θ over the products in each market. As a result, her privacy decision is based on expected utility, calculated over all possible realizations of θ in each market and also aggregated across different markets if data will be used in

---

[9]The externality is assumed to depend only on the total number of sharing consumers but not their composition. This is the case in all of our applications below.

[10]Henceforth, whenever we write $V_i'(\sigma) < 0$ or $V_i'(\sigma) > 0$, $i \in \{s, a\}$, we mean it holds for all $\sigma \in [0, 1]$.

[11]Developing this reduced-form framework first helps unify (and also simplify) our analysis in the later applications. We note, however, that our framework applies beyond privacy choice: it also applies to contexts where consumers make heterogeneous choices (e.g., whether to install an ad blocker by paying a cost) and impose externalities on each other via a product market channel (e.g., price competition).

multiple contexts. For simplicity, our model implicitly assumes that these markets are ex ante symmetric, allowing us to focus on a representative market. Under such a timing, a consumer's privacy choice does not convey any information about her true θ. We discuss how this timing assumption can be relaxed in Section 4.2.

(ii) Externalities. Externalities across consumers play a crucial role in our analysis, and we provide several different microfoundations for them in Section 3. As an example, suppose data is used to make personalized recommendations to consumers who face search frictions, but firms must offer the same price to both sharing and anonymous consumers. Suppose sharing consumers use recommendations to modify their search behavior. It is natural that as more consumers share, firms face a different demand elasticity, and so adjust their prices. Sharing consumers then exert an externality on other consumers.

2.1 Equilibrium privacy choice

We start by solving for consumers' equilibrium privacy choices. To this end, denote by $\Delta(\sigma)$ the (expected) consumption benefit or loss from sharing data, where

$$\Delta(\sigma) \equiv V_s(\sigma) - V_a(\sigma). \tag{1}$$

Consumers follow a cut-off rule: if a fraction σ of consumers are expected to share, those with $\tau < \Delta(\sigma)$ optimally share their data whereas all the others optimally hide their data. As a result, σ^* is an equilibrium of the privacy game if and only if

$$\sigma^* = T(\Delta(\sigma^*)) . \tag{2}$$

It follows immediately that:

Proposition 1. *The privacy choice game has at least one equilibrium. Moreover:*
(i) No consumers sharing (i.e., $\sigma^* = 0$) is an equilibrium if and only if $\Delta(0) \leq \underline{\tau}$. All consumers sharing (i.e., $\sigma^* = 1$) is an equilibrium if and only if $\Delta(1) \geq \overline{\tau}$.
(ii) All equilibria are interior (i.e., $0 < \sigma^* < 1$) if and only if $\Delta(0) > \underline{\tau}$ and $\Delta(1) < \overline{\tau}$.
(iii) A sufficient condition for uniqueness of equilibrium is that $\Delta'(\sigma) \leq 0$.

Equilibrium existence follows from Brouwer's fixed point theorem, because the right-hand side of equation (2) is a continuous mapping from $[0, 1]$ onto itself. Parts (i) and (ii) of the proposition provide simple conditions for "corner" (i.e., $\sigma^* = 0$ or $\sigma^* = 1$) or "interior" equilibria. For instance, if $\underline{\tau}$ is sufficiently small and $\overline{\tau}$ is sufficiently large, then any privacy choice equilibrium must be interior with some consumers sharing their data and

others hiding it. Part (iii) of the proposition shows that equilibrium is unique whenever $\Delta(\sigma)$ is weakly decreasing, such that there is "substitutability" in sharing decisions. This is illustrated in Figure 1a. On the other hand, multiple equilibria can arise when sharing decisions are complementary and $\Delta(\sigma)$ is increasing. This is illustrated in Figure 1b where there are two corner equilibria and one interior equilibrium.
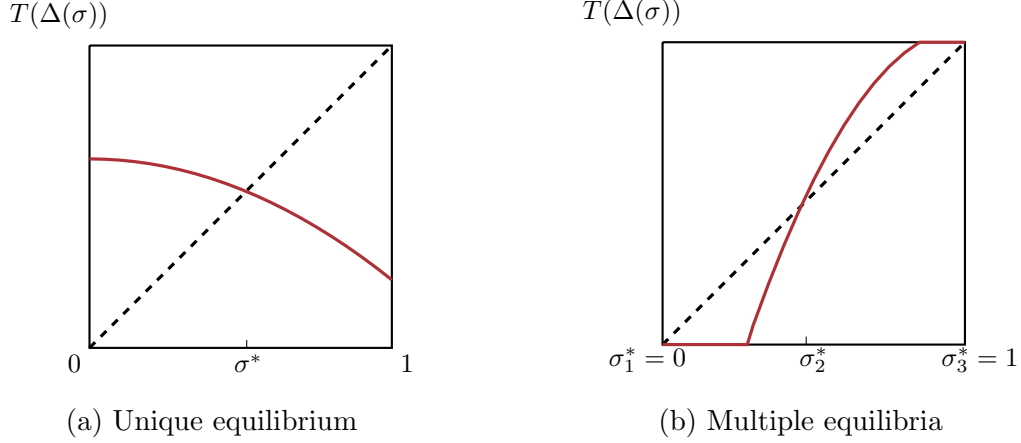

(a) Unique equilibrium	(b) Multiple equilibria

Figure 1: Privacy choice equilibrium

The following simple observation is useful in our subsequent analysis:

Lemma 1. *Suppose σ^* is an equilibrium of the privacy choice game. If sharing consumers exert negative externalities (i.e., $V_s'(\sigma) < 0$ and $V_a'(\sigma) < 0$), each consumer prefers this equilibrium over any market situation with more sharing consumers (i.e., with some $\tilde{\sigma} > \sigma^*$), regardless of whether that situation is an equilibrium or not. In contrast, if sharing consumers exert positive externalities (i.e., $V_s'(\sigma) > 0$ and $V_a'(\sigma) > 0$), each consumer prefers this equilibrium over any market situation with fewer sharing consumers (i.e., with some $\tilde{\sigma} < \sigma^*$).*

Proof. A consumer with privacy cost τ gets expected surplus $\max\{V_a(\sigma^*), V_s(\sigma^*) - \tau\}$ in the σ^* equilibrium. Meanwhile in a situation with $\tilde{\sigma} \neq \sigma^*$, the same consumer's expected surplus is *at most* $\max\{V_a(\tilde{\sigma}), V_s(\tilde{\sigma}) - \tau\}$ (because if $\tilde{\sigma}$ is not an equilibrium, some consumers make a suboptimal privacy choice). The former strictly exceeds the latter if (i) $\tilde{\sigma} > \sigma^*$, $V_s'(\sigma) < 0$ and $V_a'(\sigma) < 0$, or if (ii) $\tilde{\sigma} < \sigma^*$, $V_s'(\sigma) > 0$ and $V_a'(\sigma) > 0$. □

The result in this lemma follows immediately if a consumer makes the same privacy choice at $\tilde{\sigma}$ and σ^*. Using a revealed preference argument, the proof demonstrates that it also holds even if the consumer makes a different privacy choice in the two cases.

Lemma 1 can be used to provide a ranking when there are multiple equilibria:

Corollary 1. *Suppose that σ_1^* and σ_2^* are both equilibria of the privacy game, with $\sigma_1^* < \sigma_2^*$. Every consumer is better off in the σ_1^* equilibrium if sharing consumers exert negative externalities. The opposite is true if sharing consumers exert positive externalities.*

Lemma 1 can also be used to evaluate the impact of privacy policies such as GDPR in the EU and CCPA in California. Specifically, imagine that initially consumers have no control over their data, which is therefore shared with the firms; in terms of the above discussion, this is equivalent to having $\sigma = 1$. Then suppose there is a privacy policy, which allows consumers to costlessly hide their data. The following result is immediate:

Corollary 2. *Suppose sharing consumers exert negative externalities. A privacy policy that enables consumers to costlessly hide their data strictly benefits every consumer if it induces an equilibrium with $\sigma^* < 1$.*

On the other hand, when sharing consumers exert positive externalities, a privacy policy can backfire and harm every consumer. For instance, this happens when $\Delta(\sigma) \leq \underline{\tau}$ for every σ (such that the privacy policy induces all consumers to become anonymous) and in addition $V_a(0) < V_s(1) - \overline{\tau}$ (such that even the most privacy-conscious consumer prefers the situation where everyone shares). Later we provide an example where these conditions are satisfied.

2.2 Comparison with consumer optimum

We now consider a social planner that wishes to maximize aggregate consumer surplus, and can decide which consumers share their data and which consumers are anonymous. Clearly, conditional on having σ sharing consumers, the social planner chooses those with the lowest privacy type, which gives rise to aggregate consumer surplus of

$$V(\sigma) = \sigma V_s(\sigma) + (1 - \sigma)V_a(\sigma) - \int_{\underline{\tau}}^{T^{-1}(\sigma)} \tau dT(\tau) . \tag{3}$$

The derivative of this expression with respect to σ is equal to

$$V'(\sigma) = \sigma V_s'(\sigma) + (1 - \sigma)V_a'(\sigma) + \Delta(\sigma) - T^{-1}(\sigma) . \tag{4}$$

Notice that at an interior equilibrium (i.e., when $0 < \sigma^* < 1$), the marginal sharing consumer's privacy type $T^{-1}(\sigma^*)$ is exactly equal to the equilibrium consumption benefit $\Delta(\sigma^*)$. Hence the last two terms in equation (4) cancel, and we can write that:

$$V'(\sigma^*)|_{0<\sigma^*<1} = \sigma^* V_s'(\sigma^*) + (1 - \sigma^*)V_a'(\sigma^*) . \tag{5}$$

This tells us whether a local change in σ raises or lowers aggregate consumer surplus, while Lemma 1 can be used to look at the effect on a non-local change in σ. Therefore combining the two, we find that:

Proposition 2. *Suppose the privacy choice game has an interior equilibrium (i.e., $0 < \sigma^* < 1$). If sharing consumers exert negative externalities, there are strictly too many sharing consumers relative to the consumer optimum; the opposite is true if sharing consumers exert positive externalities.*

Proof. Consider the case with $V'_s(\sigma) < 0$ and $V'_a(\sigma) < 0$. It is clear from equation (5) that $V'(\sigma^*) < 0$. In addition, Lemma 1 implies that $V(\sigma^*) > V(\tilde{\sigma})$ for any $\tilde{\sigma} > \sigma^*$, regardless of whether $\tilde{\sigma}$ is an equilibrium or not. Hence $\arg\max_\sigma V(\sigma) < \sigma^*$. The opposite case with $V'_s(\sigma) > 0$ and $V'_a(\sigma) > 0$ can be proved in the same way. $\square$

This result may seem trivial because, for example, when sharing consumers exert negative externalities, a higher σ reduces both $V_s(\sigma)$ and $V_a(\sigma)$ and so must harm consumers in aggregate. Notice, however, that $V_s(\sigma)$ can also exceed $V_a(\sigma)$, i.e., a sharing consumer can obtain more consumption surplus than an anonymous consumer; this is a countervailing force which favors having more sharing consumers.[12] We also note that due to this countervailing force, $V(\sigma)$ may not be globally decreasing in σ when there are negative externalities: the marginal consumer's gain from sharing (i.e., the last two terms in (4)) can be positive and outweigh the negative impact on other consumers' surplus (i.e., the first two terms in (4)). We will demonstrate these points in later applications.

2.3 The effect of more competition or improved data security

Fixing consumers' privacy choices, one would usually expect more competition (i.e., higher n) or improved data security (i.e., a FOSD decrease in τ) to raise consumer welfare. We now show that this may not hold with endogenous privacy choices. Hence a competition or consumer protection policy can have a perverse effect on consumers.

To illustrate this as simply as possible, we focus on the case where the privacy game has a unique equilibrium. As a preliminary step, we find that:

---

[12]We note that Proposition 2 may not hold if the privacy choice game has a corner equilibrium. In this case, the privacy choice equilibrium features only weak over-/under-sharing respectively. To illustrate, suppose for each $\sigma \in [0,1]$ that $V'_s(\sigma) < 0$ and $V'_a(\sigma) < 0$ but $\Delta(\sigma) > \overline{\tau}$. The privacy choice game has a unique equilibrium with $\sigma^* = 1$. Moreover, since the last two terms in equation (4) are now strictly positive, it is possible that $V'(\sigma) > 0$ for all $\sigma \in [0,1]$, such that $\sigma = 1$ is also the consumer optimum.

Lemma 2. *Suppose σ^* is unique, and either (i) n increases and this raises $\Delta(\sigma)$ for each σ or (ii) $T(\tau)$ decreases in the sense of FOSD. Then σ^* weakly increases.*

Proof. Consider a shift from $\{\Delta(\sigma), T(\tau)\}$ to $\{\tilde{\Delta}(\sigma), \tilde{T}(\tau)\}$. Let σ^* and $\tilde{\sigma}^*$ be the (unique) equilibrium associated with the former and latter, respectively. Assume $\tilde{\Delta}(\sigma) \geq \Delta(\sigma)$ for each σ, and $\tilde{T}(\tau) \geq T(\tau)$ for each τ. Clearly if $\sigma^* = 0$ then it follows immediately that $\tilde{\sigma}^* \geq \sigma^*$. If instead $\sigma^* > 0$ then we must have $T(\Delta(\sigma)) > \sigma$ for all $\sigma < \sigma^*$, which implies that $\tilde{T}(\tilde{\Delta}(\sigma)) > \sigma$ for all $\sigma < \sigma^*$, which in turn implies that $\tilde{\sigma}^* \geq \sigma^*$. ☐

If more competition raises the benefit of sharing data, or if better data security leads to a FOSD reduction in privacy costs, then in equilibrium more consumers share their data.[13]

We first argue that an increase in competition can have a perverse effect on consumer welfare. Specifically, abusing notation, the effect on aggregate consumer welfare of adding an extra firm is $V(\sigma^*(n+1); n+1) - V(\sigma^*(n); n)$, which can be rewritten as

$$\underbrace{V(\sigma^*(n); n+1) - V(\sigma^*(n); n)}_{\text{Direct effect of more competition}} + \underbrace{V(\sigma^*(n+1); n+1) - V(\sigma^*(n); n+1)}_{\text{Indirect effect due to endogenous privacy choice}} . \qquad (6)$$

The first part captures the effect of more competition for given privacy choices, and is typically positive. The second part captures the effect of more competition through a change in privacy choice: it is negative if, for example, an increase in n raises $\Delta(\sigma)$ and thus σ^* (by the above lemma), and V is decreasing in σ. We will show in later applications that the second part can indeed be negative and can also dominate the first part.[14]

We also argue that a reduction in privacy costs can similarly have a perverse effect on consumer welfare. To illustrate this in a simple way, consider the case with $V_s(\sigma) > V_a(\sigma)$ for any σ, i.e., there is always a consumption benefit from sharing data. Suppose that initially the privacy cost is so high that all consumers choose to be anonymous, leading to consumer welfare of $V_a(0)$. Suppose data security improves so much that the privacy cost drops to zero. Given that $\Delta(\sigma) > 0$, all consumers share their data, leading to consumer welfare of $V_s(1)$. Consumers are then worse off whenever $V_s(1) < V_a(0)$. We provide applications later on where this condition is satisfied.

---

[13]When the privacy game has multiple equilibria, the same observation applies to all stable interior equilibria (in which $T(\Delta(\sigma))$ crosses σ from above) as well as any corner equilibria.

[14]Note that, for simplicity, we assume here that any increase in competition does not affect privacy costs. If privacy costs increase in n (e.g., because there is more chance that a consumer's data is mishandled), those consumers who still choose to share their data pay an extra privacy cost; on the other hand, this induces fewer consumers to share, and so may mitigate any perverse effect of competition.

3   Applications

We now apply our framework to personalized recommendations, prices, and products.

3.1   Personalized recommendations

As discussed in the Introduction, many online platforms use data to generate personalized recommendations for consumers. In this section we study the interaction between *unbiased* recommendations and privacy choice.[15] Such recommendations only matter if consumers have imperfect information and find it costly to discover their preferred product on their own. In order to capture this, we build on the canonical search framework with product differentiation developed by Wolinsky (1986) and Anderson and Renault (1999).

Primitives.   Consider a discrete-choice framework with n firms. Let v_i, $i \in \{1, ..., n\}$, denote a consumer's valuation for firm i's product. We assume that the v_i's are IID across firms and consumers, and drawn from a common CDF $F(v)$ with support $[\underline{v}, \overline{v}]$. Lef $f(v)$ be its density function, and assume it is log-concave. The consumption type is then a consumer's vector of valuations for the n products, i.e., $\theta = (v_1, \dots, v_n)$.

Each consumer is initially uninformed about her valuations for the n products as well as their prices, but can learn this information via a standard sequential search process. Specifically, if a consumer visits a firm, she learns her valuation for its product and its price, and then decides whether to purchase its product immediately or continue searching. We assume that the first visit is costless, but that visiting any additional firm incurs a search cost $s > 0$; consumers may costlessly recall any firm they searched previously.

If a consumer shares her data (e.g., with a platform that hosts the n firms), she receives a personalized recommendation informing her about which product gives her the highest valuation. (Later on we discuss recommendations based on net surplus instead of match value.) If a consumer does not share her data, she does not receive any recommendation. Suppose that firms do not have access to data and so cannot price discriminate, e.g., they do not know whether their products have been recommended to a consumer or not.

---

[15]We therefore sidestep the concern that sellers may pay a platform to obtain a biased recommendation; there is already an extensive literature studying this (see, e.g., Armstrong and Zhou, 2011, Inderst and Ottaviani, 2012, de Cornière and Taylor, 2019, and Teh and Wright, 2022). Needless to say, an interesting question for future research would be the interplay between privacy choice and data regulation, and sellers' incentives to pay for recommendations.

The timing is as follows. Each consumer first learns her privacy type τ, and then independently chooses whether or not to share her data. At the same time, firms form a rational expectation about σ, and then set their prices simultaneously to maximize their own profit. Consumers search optimally (with or without receiving recommendations according to their privacy choice), holding a rational expectation about firms' pricing strategies. Since there are no correlated shocks across firms, we make the usual assumption of passive beliefs, i.e., upon seeing an off-equilibrium price at some firm, consumers believe that other unsampled firms still charge their equilibrium prices. We look for a symmetric perfect Bayesian equilibrium where all firms set the same price.

Pricing equilibrium with a fixed σ. We first study price competition with an exogenous fraction σ of sharing consumers. We look for a symmetric equilibrium where each firm charges the same price p.

Let us first derive the demand for firm i if it unilaterally deviates and charges a price p_i (which is not observable to consumers before they search). We begin with demand from sharing consumers. Firm i is recommended to a sharing consumer with probability $\frac{1}{n}$. Since all firms are expected to charge the same price, any sharing consumer who is recommended firm i will visit that firm first. Suppose p_i is a small local deviation (e.g., $p_i < p + s$) so that a sharing consumer who is recommended firm i has no incentive to search beyond firm i after seeing its deviation price. (We discuss non-local deviations later.) In this case firm i competes only with the outside option, and so its expected demand from a sharing consumer is

$$q_s(p_i) \equiv \frac{1}{n}[1 - F(p_i)^n] \, , \tag{7}$$

because conditional on being recommended—and therefore being the best of n products—product i's valuation has a CDF $F(v)^n$. Note that the recommended firm is therefore like a multiproduct monopolist that charges the same price on each of its n products.

Now consider demand from anonymous consumers. In a symmetric equilibrium where firms charge the same price p, anonymous consumers search randomly among firms as in the usual Wolinsky-Anderson-Renault model. Their optimal stopping rule is characterized by a reservation surplus $r - p$ where r is the standard reservation match utility that solves

$$\mathbb{E}[\max\{0, v - r\}] = \int_r^{\overline{v}} [1 - F(v)]dv = s.$$

(The left-hand side is the expected benefit from sampling one more product when prices are symmetric.) After visiting firm i for the first time, a consumer buys its product

immediately if and only if firm i offers her a surplus greater than $r - p$. We assume for now that $r - p > 0$ such that in equilibrium some anonymous consumers search beyond the first firm they sample, and we provide a primitive condition for it to hold below.

When firm i deviates to price p_i but other firms stick to the equilibrium price p, firm i's demand from anonymous consumers has two sources. One is from those who stop searching and buy immediately after visiting firm i for the first time. Since anonymous consumers search randomly (as they expect all firms to charge the symmetric equilibrium price p), firm i is in the kth position of a consumer's search process with probability $\frac{1}{n}$. In that case, the consumer will come to visit it for the first time when the previous $k - 1$ products each have a match utility less than r, which happens with probability $F(r)^{k-1}$. The consumer will then buy immediately if $v_i - p_i > r - p$, which happens with probability $1 - F(r - p + p_i)$. Therefore, the first portion of the demand is

$$\frac{1}{n} \sum_{k=1}^{n} F(r)^k [1 - F(r - p + p_i)] = \frac{1}{n} \frac{1 - F(r)^n}{1 - F(r)} [1 - F(r - p + p_i)] .$$

The other source is from those consumers who first choose to continue searching but ultimately return to buy after discovering worse options elsewhere. Regardless of firm i's position in a consumer's search process, the consumer will choose to search on if $v_i - p_i < r - p$ but will return to buy if $v_i - p_i > \max_{j \neq i} \{0, v_j - p\}$. (The decomposition is valid even when firm i is in the last position.) This happens with probability

$$\int_{p_i}^{r-p+p_i} F(v_i - p_i + p)^{n-1} dF(v_i) = \int_{p}^{r} F(v)^{n-1} f(v - p + p_i) dv ,$$

where the equality uses a change of variables.

In sum, firm i's deviation demand from an anonymous consumer is

$$q_a(p_i) = \frac{1}{n} \frac{1 - F(r)^n}{1 - F(r)} [1 - F(r - p + p_i)] + \int_{p}^{r} F(v)^{n-1} f(v - p + p_i) dv . \qquad (8)$$

Notice that $q_a(p) = q_s(p) = \frac{1}{n}[1 - F(p)^n]$, i.e., a firm's equilibrium demand from an anonymous consumer is the same as that from a sharing consumer. This is because an anonymous consumer buys something provided at least one of her valuations exceeds p, and on average each firm gets a $\frac{1}{n}$ share of her demand. However, as we discuss in more detail shortly, anonymous and sharing consumers have different demand elasticities.

Firm i's problem is then

$$\max_{p_i} (p_i - c)[\sigma q_s(p_i) + (1 - \sigma) q_a(p_i)] . \qquad (9)$$

16

We focus on the case where this profit function is well-behaved, such that the equilibrium price is determined by the first-order condition. (We will further discuss this issue after the next lemma.) The equilibrium price p then solves

$$\frac{1 - F(p)^n}{n} \frac{1}{p - c} = \sigma |q_s'(p)| + (1 - \sigma)|q_a'(p)|$$

$$= \sigma F(p)^{n-1} f(p) + (1 - \sigma) \left[\frac{f(r)}{n} \frac{1 - F(r)^n}{1 - F(r)} - \int_p^r F(v)^{n-1} f'(v) dv \right].$$

$$(10)$$

Lemma 3. *The first-order condition (10) has a unique solution $p \in (c, r)$ and it increases in σ. That is, if the equilibrium price is determined by (10), it increases in the fraction of sharing consumers.*

Anonymous consumers are more price-sensitive than sharing consumers. The reason is that anonymous consumers search randomly, and so when a seller raises its price, some of these consumers search on, and not all of them return later and purchase. In contrast, sharing consumers are recommended the product with the highest valuation $\max_i\{v_i\}$, and in equilibrium they either buy it or take the outside option—and so firms are like "multiproduct monopolists" over these consumers. Therefore, as σ increases, firms face a less elastic demand and charge a higher price. Indeed, as σ approaches 1, firms charge the multiproduct monopoly price $\arg\max_p(p - c)q_s(p)$.

Remarks. We clarify a few technical issues before proceeding. First, it is hard to derive simple primitive conditions for the deviation profit function in (9) to be well-behaved, so that the first-order condition (10) is sufficient for defining the equilibrium price. This is the case even in the standard Wolinsky model (i.e., the case with $\sigma = 0$); having the extra demand component from the sharing consumers makes the problem more complicated. In the numerical examples used below, we have verified that the deviation profit is single-peaked.

Second, when firm i sets a high deviation price, even sharing consumers who were recommended it may choose to search other products in the market. The firm's demand from sharing consumers is then lower than $q_s(p_i)$ in (7). This, however, does not affect the above equilibrium analysis: if a firm has no incentive to deviate under $q_s(p_i)$ in (7), it also has no incentive to deviate under a demand function that is smaller at high p_i values.

Finally, the above analysis is predicated on $p < r$, such that in equilibrium some anonymous consumers search beyond the first firm that they sample.[16] Since p is capped

---

[16]When instead $p \geq r$, in a symmetric pure-strategy equilibrium anonymous consumers do not search

by the multiproduct monopoly price $\arg\max_p(p-c)q_s(p)$, a sufficient condition is that the latter is less than r, or equivalently

$$r - c > \frac{1 - F(r)^n}{nF(r)^{n-1}f(r)} \ . \tag{11}$$

In the appendix, we show that (11) holds when r is above (or s is below) a certain threshold. This condition is also verified in the numerical examples we use below.

Now consider the surplus enjoyed by the two consumer types. A sharing consumer buys if and only if her favorite product has a value above p, so her expected surplus is

$$V_s(\sigma) = \int_p^{\bar{v}} (v - p)dF(v)^n = \int_p^{\bar{v}} [1 - F(v)^n]dv \ . \tag{12}$$

An anonymous consumer's expected surplus can be shown to equal[17]

$$V_a(\sigma) = \int_p^r [1 - F(v)^n]dv + s \ . \tag{13}$$

It then follows immediately from Lemma 3 that:

Corollary 3. *Sharing consumers exert negative externalities: $V_a'(\sigma) < 0$ and $V_s'(\sigma) < 0$.*

When there are more sharing consumers, this relaxes price competition, to the detriment of both anonymous and sharing consumers.

Privacy choice equilibrium. Using equations (12) and (13), as well as the definition of r, we can write the consumption benefit from sharing data as

$$\Delta(\sigma) = V_s(\sigma) - V_a(\sigma) = \int_r^{\bar{v}} [F(v) - F(v)^n]dv \ . \tag{14}$$

Note that $\Delta(\sigma)$ is strictly positive for all $n \geq 2$. The reason is that sharing consumers pay the same price as anonymous consumers, but are recommended their best product without having to search. Note also that $\Delta(\sigma)$ is independent of σ. Intuitively, both sharing and anonymous consumers make a purchase if and only if they value (at least) one product more than p, so changes in p affect their surpluses in the same way. It then follows from Proposition 1 that the privacy choice game has a unique equilibrium σ^*.

---

beyond the first visited firm. In that case, each firm's demand is a weighted sum of standard single-product and standard multiproduct monopoly demands. Since the multiproduct monopoly price is greater than the single-product monopoly price, the equilibrium price also increases in σ just as in Lemma 3.

[17]If the first search also costs s, consumer surplus is the first term in (13). This can be derived by using, e.g., Lemma 3 in Rhodes and Zhou (2019). The second term is needed because the first search is assumed to be free in our model.

Too much data sharing. Proposition 2 and Corollary 3 together imply that there is too much data sharing in any (interior) equilibrium relative to the consumer optimum. Figure 2 illustrates this for the case where $n = 2$, $r = 2/3$, valuations are uniform on $[0, 1]$, $c = 1/4$, and τ follows a Beta $(1/2, 10)$ distribution. The privacy choice game has a unique equilibrium with $\sigma^* = 0.646$, whereas aggregate consumer surplus $V(\sigma)$ is hump-shaped and maximized at $\hat{\sigma} = 0.409$. At this optimum, total consumer surplus is around 3% higher compared to in the privacy choice game.
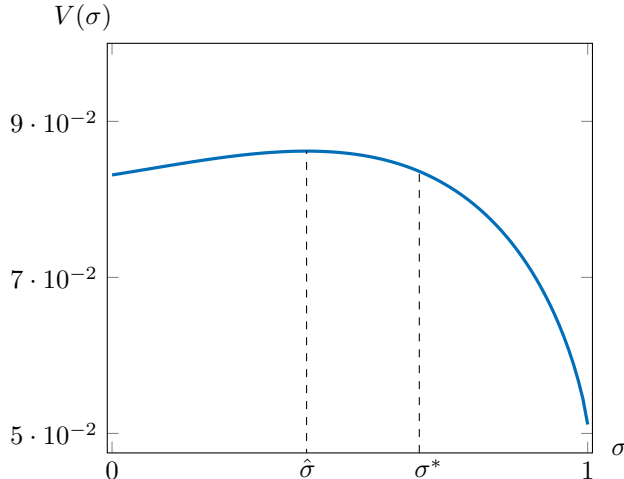

Figure 2: Aggregate consumer surplus as a function of σ

Perverse effect of more competition. It is clear from equation (14) that $\Delta(\sigma)$ increases in n. (Intuitively, when n is higher, the best product has a higher match, and the benefit of being recommended it rather than having to search among a larger number of firms is also higher.) It then follows from Lemmas 2 and 3 that an increase in n raises σ^* and relaxes competition. We now show via an example that having more firms can relax competition so much that it ends up harming consumers.

Example. Suppose that $c = 0$, valuations are uniformly distributed on $[0, 1]$, τ is uniformly distributed on $[0.025, 0.055]$, and the search cost is such that $r = 0.8$. Figure 3 depicts the impact of changes in n on equilibrium outcomes. The red dotted curves are for the case where all consumers *exogenously* hide their data (i.e., the standard Wolinsky model): as n increases, price falls and consumers are better off. The blue solid curves are for the case with *endogenous* privacy choice: (i) the fraction of consumers that share is zero for $n = 1$ and $n = 2$, intermediate for $n = 3$ and $n = 4$, and one for all $n \geq 5$, (ii) the equilibrium price is U-shaped in n and minimized at $n = 2$, and (iii) aggregate consumer surplus is

19

hump-shaped in n and maximized at $n = 3$. Intuitively, when $n = 1$ recommendations have no value, and for $n = 2$ they have only limited value, so no consumers share; the red and blue curves therefore coincide. However, as n increases further, recommendations become more valuable and some consumers start sharing, which makes demand less elastic and induces firms to charge a higher price, which for $n > 3$ dominates the improvement in match utilities and so harms consumers.

(a) Sharing consumers   (b) Equilibrium Price   (c) Aggregate Consumer Surplus
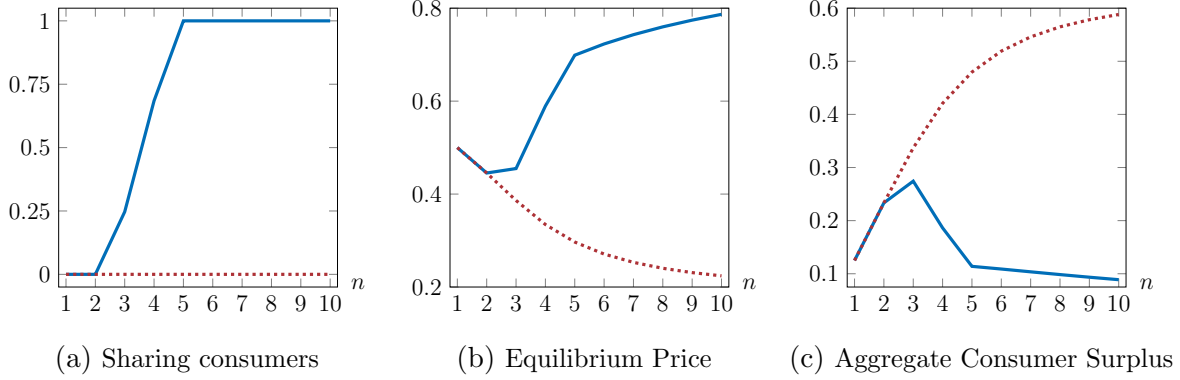
Figure 3: The impact of competition with personalized recommendations
(The solid blue curve depicts equilibrium outcomes, and the dotted red curve depicts outcomes when all consumers exogenously hide their data)

Perverse effect of improved data security. We now fix $n \geq 2$ firms and consider a change in privacy costs. In Section 2 we gave an example where improvements in data security that reduce privacy costs to zero induce all consumers to switch from being anonymous to sharing their data. We argued that this would harm consumers if $V_s(1) < V_a(0)$: in the current model of personalized recommendations, this condition becomes

$$\int_{p_n^M}^{\overline{v}} [1 - F(v)^n]dv < \int_{p_W}^{r} [1 - F(v)^n]dv + s \,,$$

where p_n^M is the multiproduct monopoly price with n products and p_W is the standard Wolinsky price. Since $p_W < p_n^M$, it is evident that the above condition holds if the search cost is small and hence r is close to $\overline{v}$. Intuitively, improved data security encourages consumers to share and use recommendations: this benefits consumers via lower search costs, but harms them through a higher price, and when the search cost is small anyway the latter effect dominates.[18]

---

[18]The above condition also holds for $n = 2$ in the numerical example depicted in Figure 3. Before the improvement in data security consumers pay 0.445 and get $V_a(0) = 0.233$, whereas after the improvement they pay 0.577 and get $V_s(1) = 0.153$.

Discussion: surplus-based recommendations. So far we have assumed that sharing consumers are recommended the product with the highest *match value*. We now consider an alternative scenario where they are recommended the product with the highest *surplus* (i.e., match value minus price), and show that this can lead to different conclusions.[19]

Sharing consumers now buy the recommended product as long as it has a positive surplus. Competition for these consumers is therefore the same as in the frictionless Perloff-Salop model with a zero outside option. Hence if firm i unilaterally deviates to price p_i, its expected demand from a sharing consumer is

$$q_s(p_i) = \Pr[v_i - p_i > \max_{j \neq i}\{0, v_j - p\}] = \int_{p_i}^{\overline{v}} F(v - p_i + p)^{n-1} dF(v) .$$

However firm i's demand from an anonymous consumer is exactly the same as in equation (8) from earlier. Therefore, assuming the first-order condition is sufficient to determine the equilibrium,[20] the equilibrium price p now solves

$$
\begin{aligned}
\frac{1 - F(p)^n}{n} \frac{1}{p - c} &= \sigma[F(p)^{n-1}f(p) + \int_p^{\overline{v}} f(v)dF(v)^{n-1}] \\
&\quad + (1 - \sigma)\left[\frac{f(r)}{n}\frac{1 - F(r)^n}{1 - F(r)} - \int_p^r F(v)^{n-1}f'(v)dv\right] . \quad (15)
\end{aligned}
$$

As detailed in the appendix, given our log-concavity condition, the equilibrium price is unique and now *decreases* in σ.[21] Intuitively, sharing consumers here are just like anonymous consumers but with a zero search cost; since search costs tend to make consumers less price-sensitive, a higher σ makes demand more elastic and so reduces the equilibrium price. Hence sharing consumers now impart a *positive* externality on others.

The expressions for $V_s(\sigma)$ and $V_a(\sigma)$ are obviously the same as before. However now they *increase* in σ. According to Proposition 2, this implies that in any interior privacy choice equilibrium there is now *too little* data sharing relative to the consumer optimum. Similarly, now there is *no* perverse effect from more competition or improved data security. For instance, since $\Delta(\sigma)$ still increases in n, more competition again induces

---

[19] Of course we note that in practice it would be easier for a platform to learn a consumer's relative preferences across products (which is enough for match-based recommendations) rather than her exact valuations for each product (which is needed for surplus-based recommendations off equilibrium path).

[20] For the same reason as in the case of match-based recommendations, it is hard to find simple primitive conditions for this, but we verify that it holds in the numerical example below.

[21] As with match-based recommendations, our analysis here is predicated on $r > p$. A sufficient condition for this is that the Wolinsky price is below r, or equivalently, $r - c > \frac{1-F(r)}{f(r)}$. Note that this condition is less stringent than (11), and holds for r sufficiently large, or equivalently s sufficiently small.

more consumers to share their data, but now this reduces the equilibrium price and so unambiguously benefits consumers.

Finally, it is now possible that a privacy policy which allows consumers to hide their data can backfire. We illustrate this through the following simple example.

Example: perverse effect of privacy policies. Suppose $n = 2$, $c = 0$, and that valuations are uniformly distributed on $[0, 1]$. Suppose also that $\underline{\tau} = 0.06$ and $\overline{\tau} = 0.07$, and that the search cost s is such that $r = 0.6$. Absent a privacy policy all consumers' data is shared (i.e., $\sigma = 1$); firms charge the standard Perloff-Salop price $p = \sqrt{2} - 1 \approx 0.414$, and each consumer's surplus is at least $V_s(1) - \overline{\tau} \approx 0.206$. A privacy policy that enables consumers to costlessly hide their data harms *every* consumer. Specifically, no consumer shares (i.e., $\sigma^* = 0$) because $\Delta(\sigma) \approx 0.059 < \underline{\tau}$; this induces firms to raise their price to $p \approx 0.481$, which lowers consumer surplus to $V_a(0) \approx 0.164$.

Other related literature. Our application to personalized recommendations is most closely related to Anderson and Renault (2000). They consider a duopoly model in which some consumers need to search the firms to learn their product matches, while other consumers are fully informed of their matches and so in equilibrium only search their favorite firm. The informed consumers in their model are the same as the sharing consumers in our model. However, Anderson and Renault (2000) assume that the market is fully covered, and so demand from informed consumers is completely inelastic, and prices are capped by consumers' budget constraints. Like us, they show that having more informed consumers raises the market price and so imposes a negative externality on other consumers.[22] However, we consider a more general oligopoly model which enables us to examine how increased competition affects consumer privacy choice and market performance. We also consider surplus-based recommendations, and show that the externality works in the opposite direction.

Personalized recommendations lead to a trade-off between match quality and price. Such a trade-off is also present in other works on privacy choice such as de Cornière and de Nijs (2016), Ichihashi (2020), and Hidir and Vellodi (2021). In de Cornière and de Nijs (2016), if an ad exchange platform provides consumer preference information to advertisers, for each consumer the advertiser which best matches her preferences wins the ad auction. They assume that consumers have a less elastic demand for a better-matched

---

[22]They also argue that when consumers can pay a presearch cost (which is like our privacy cost) to become informed, there are too many informed consumers from a collective viewpoint (see their page 734), and that reducing the information acquisition cost can harm consumers (see their footnote 17).

product, so this leads to a higher market price. They focus on the platform's privacy choice instead of consumers' privacy choice. In Ichihashi (2020) and Hidir and Vellodi (2021), there is a multiproduct monopolist which sells several varieties of a product. If consumers reveal their preference information, they are provided with the best matched variety, but at the same time the firm can extract more surplus via personalized pricing.

There is also growing empirical research on personalized recommendations/rankings on e-commerce platforms. See, for example, Donnelly, Kanodia, and Morozov (2024) for a study on Wayfair.com and Zhou, Lin, Xiao, and Fang (2023) for a study on TaoBao.com. They mainly focus on the impact of personalized recommendations on consumer search and purchase behavior,[23] but do not consider the potential impact on product pricing or endogenous consumer privacy choice. Minaev (2021) constructs a structural model to study the impact of personalized rankings on both the demand and the supply side. Using Expedia hotel data, he shows in a counterfactual that personalized rankings save consumers on search costs, help them find better-matched products, but raise market prices. On average consumers suffer from personalized rankings.[24]

3.2 Personalized pricing

As explained in the Introduction, firms are increasingly using data to offer consumers personalized prices. This is often implemented via targeted discounts off a regular price. In this section we study the interaction between personalized prices and privacy choice.

Primitives. We use the same discrete-choice framework as in the previous application, i.e., a consumer's valuation v_i for product i is drawn IID using a log-concave density $f(v)$ with support $[\underline{v}, \overline{v}]$. However, here we assume that consumers automatically learn their consumption type $\theta = (v_1, \ldots, v_n)$ after making their privacy choice.

If a consumer does not share her data, she is anonymous and each firm offers her a public "list" price. If instead a consumer shares her data, all firms learn her consumption

---

[23]Donnelly, Kanodia, and Morozov (2024) find that personalized rankings on Wayfair.com induce more consumers to search and improve purchase diversity (i.e., shifting demand from bestsellers to niche items); while Zhou, Lin, Xiao, and Fang (2023) find that when TaoBao.com returns more targeted search results to consumers, they search less and buy the featured products more often, but meanwhile they also spend less time in exploring other product categories, reducing unplanned purchases.

[24]Calvano, Calzolari, Denicoló, and Pastorello (2023) use a computational model to study personalized recommendations in a search market. When a collaborative-filtering algorithm generates recommendations, the average market price increases, but consumers benefit overall due to better product matches.

type θ and can offer her personalized prices; however, importantly, we assume that sharing consumers always have the option to buy at firms' list prices. For example, the list price could be freely available on a firm's website, which sharing consumers can consult before choosing whether or not to accept their personalized price.[25] One can therefore interpret personalized prices as targeted discounts off the list price.

The timing is as follows. At the first stage, each consumer learns her privacy type τ. Then, before learning her consumption type θ, each consumer independently decides whether or not to share her data. At the second stage, each consumer's θ is realized. Firms observe a consumer's θ if and only if she shared her data. Firms then simultaneously choose a list price for anonymous consumers, and a (weakly lower) personalized price for each consumer who shared her data. At the third stage, consumers decide which product (if any) to buy. We normalize a consumer's outside option from purchasing nothing to zero. Assume that firms have the same constant marginal cost c.

Pricing equilibrium with a fixed σ. We first study price competition with an exogenous fraction σ of sharing consumers. We look for a symmetric equilibrium where each firm uses the same list price $p(\sigma)$; when there is no confusion we simply denote this list price by p. (We note that Rhodes and Zhou, 2024 study two (exogenous) limit cases of our game: the case $\sigma = 0$, as well as the case $\sigma = 1$ when the list price is fixed at $\overline{v}$.)

Suppose firm i unilaterally deviates to a list price $p_i > c$. An anonymous consumer buys from firm i if and only if $v_i - p_i > \max_{j \neq i}\{0, v_j - p\}$. Following Rhodes and Zhou (2024), let $x_p \equiv v_i - p - \max_{j \neq i}\{0, v_j - p\}$ denote the relative preference for product i when all products are sold at price p, and let $H_p(\cdot)$ be its CDF. Then firm i's expected demand from an anonymous consumer is $1 - H_p(p_i - p)$, which leads to an expected profit

$$\pi_a(p_i, p) \equiv (p_i - c)[1 - H_p(p_i - p)] . \tag{16}$$

Competition for a sharing consumer is just a Bertrand game, but with (generically) asymmetric valuations $(v_1, \ldots, v_n)$, and with the constraint that each firm's personalized price is capped by its list price. Therefore, using standard arguments, firm i wins a consumer if and only if the consumer's valuation for its product is the highest and it exceeds c (i.e., if $v_i - \max_{j \neq i}\{c, v_j\} > 0$). Note that *absent* the list price constraint, firm i would charge the consumer a price which makes her indifferent to her next best alternative in the market. Specifically, if $\max_{j \neq i}\{v_j\} \leq c$, the next best alternative would

---

[25]Note that if sharing consumers could not access the public list price, there would be no interaction between sharing and anonymous consumers and hence no externalities.

be the outside option, so firm i would charge v_i; if instead $\max_{j \neq i}\{v_j\} > c$, the next best alternative would be to buy the second-best product at cost, so firm i would charge $c + v_i - \max_{j \neq i}\{v_j\}$.[26] Summing up, without the list price constraint, firm i would charge $c + v_i - \max_{j \neq i}\{c, v_j\}$. However, since firm i *is* constrained by its list price, it drives the consumer as close as possible to her next best alternative, and therefore charges her the minimum of $c + v_i - \max_{j \neq i}\{c, v_j\}$ and p_i. As a result, firm i's profit margin when it wins a consumer is

$$p(v_i, \mathbf{v}_{-i}) - c = \min\{v_i - \max_{j \neq i}\{c, v_j\}, p_i - c\} \ .$$

Letting $H_c(\cdot)$ be the CDF of $x_c = v_i - \max_{j \neq i}\{c, v_j\}$, firm i's expected profit from a sharing consumer can then be written as

$$\pi_s(p_i) \equiv \int_0^{p_i - c} x \, dH_c(x) + (p_i - c)[1 - H_c(p_i - c)] = \int_0^{p_i - c} [1 - H_c(x)] dx \ , \qquad (17)$$

where the second equality follows from integration by parts. Notice that $\pi_s(p_i)$ is increasing in firm i's list price. This is because an increase in p_i enables firm i to charge a higher personalized price to each consumer for whom the list price is binding. Hence firm i faces a trade-off: as it increases its list price, it earns more profit from sharing consumers, but loses demand from anonymous consumers.

Using the above, firm i's deviation profit from charging a list price p_i is

$$\sigma \pi_s(p_i) + (1 - \sigma) \pi_a(p_i, p) \ . \qquad (18)$$

When $\sigma = 1$, firms set their list price (weakly above) $\overline{v}$ because $\pi_s(p_i)$ is strictly increasing in $p_i < \overline{v}$. Hence in this case the list price never binds, so firms set the same personalized prices as in Rhodes and Zhou (2024). In the following we focus on the case $\sigma < 1$.

Given our assumption that f is log-concave, the deviation profit (18) can be shown to be quasi-concave in p_i and $p \geq c$.[27] We then obtain the following result:

Lemma 4. *For any $\sigma < 1$ the symmetric equilibrium list price $p(\sigma)$ uniquely solves*

$$p - c = \frac{1 - H_p(0)}{h_p(0)} + \frac{\sigma}{1 - \sigma} \frac{1 - H_c(p - c)}{h_p(0)} \ , \qquad (19)$$

and it increases in σ.

---

[26] We follow the usual tie-break rule that when indifferent between multiple options, the consumer chooses the one that maximizes total welfare.

[27] Note that $\pi_a(p_i, p)$ in (16) is log-concave in p_i given the log-concavity of f, while $\pi_s(p_i)$ in (17) is concave in p_i. However it is not immediate that a linear combination of them is quasi-concave.

Intuitively, as more consumers share their data, firms optimally raise their list price in order to expand the range of personalized prices they can offer. In turn, this harms anonymous consumers, as well as sharing consumers for whom the list price binds. To see this last point, note that the expected surplus of an anonymous consumer is

$$V_a(\sigma) \equiv \mathbb{E}[\max\{v_{n:n} - p, 0\}] , \tag{20}$$

where $v_{n:n}$ is the highest order statistic of $\{v_1, ..., v_n\}$, while for a sharing consumer it is

$$V_s(\sigma) \equiv \mathbb{E}[\max\{v_{n:n} - p, v_{n-1:n} - c, 0\}] , \tag{21}$$

where $v_{n-1:n}$ is the second highest order statistic of $\{v_1, ..., v_n\}$. The latter is explained as follows. A sharing consumer buys if and only if $v_{n:n} \geq c$. If, in addition, $v_{n-1:n} < c$, the firm with the best product is a monopolist, and so it extracts as much surplus as possible from the consumer; in particular, it charges $\min\{v_{n:n}, p\}$, leaving the consumer with surplus $\max\{v_{n:n} - p, 0\}$. If instead $v_{n-1:n} \geq c$, then as explained earlier, the firm with the best product extracts as much of the additional surplus as possible that it creates compared to if the consumer bought the next best product at cost; in particular, it charges $\min\{c + v_{n:n} - v_{n-1:n}, p\}$, leaving the consumer with surplus $\max\{v_{n:n} - p, v_{n-1:n} - c\}$.

Lemma 4 immediately implies the following observation:

Corollary 4. *Sharing consumers exert negative externalities: $V_a'(\sigma) < 0$ and $V_s'(\sigma) < 0$.*

Privacy choice equilibrium. We now solve for equilibrium privacy choices at the first stage of the game, starting with the monopoly case and then the competitive case.

Under monopoly, $\Delta(\sigma) = V_s(\sigma) - V_a(\sigma) = 0$, i.e., the consumption benefit from sharing data is zero. Intuitively, a consumer whose valuation exceeds p buys at a price of p irrespective of whether or not she shares her data—because personalized prices are capped at p. Meanwhile a consumer whose valuation is less than p always gets zero surplus: when she is anonymous she does not buy, and when she shares she buys but the monopolist fully extracts her willingness-to-pay. Equation (2) then implies a unique equilibrium of the privacy choice game, with $\sigma^* = T(0)$ sharing consumers.

Under competition, $\Delta(\sigma) > 0$ and $\Delta'(\sigma) > 0$, i.e., the consumption benefit from sharing data is positive and increasing in σ.[28] Intuitively, competition for sharing consumers that have relatively close valuations for two products leads firms to offer them discounts.

---

[28]To see why $\Delta'(\sigma) > 0$, notice that when p is larger (which is the case when σ is higher), it is more likely that $v_{n-1:n} - c$ exceeds $v_{n:n} - p$, and so $\Delta(\sigma)$ is also larger.

Moreover, from Lemma 4, as σ increases, firms increase their list prices; this increases the size of the discounts that they offer, which makes sharing even more attractive. As discussed earlier, this "complementarity" in privacy choices can lead to multiple equilibria.[29] Moreover, since sharing consumers generate negative externalities, according to Corollary 1 each consumer prefers the equilibrium with the lowest σ^*.

Too much data sharing. Since sharing consumers exert negative externalities, Proposition 2 says that there is too much data sharing in any interior equilibrium relative to the consumer optimum. Figure 4 illustrates this for the case where $n = 2$, valuations are uniform on $[0, 1]$, $c = 0$, and τ follows a Beta $(1/2, 5)$ distribution. The privacy choice game has a unique equilibrium with $\sigma^* = 0.912$, whereas aggregate consumer surplus $V(\sigma)$ is hump-shaped and maximized at $\hat{\sigma} = 0.465$. At this optimum, total consumer surplus is around 9% higher compared to in the privacy choice game.
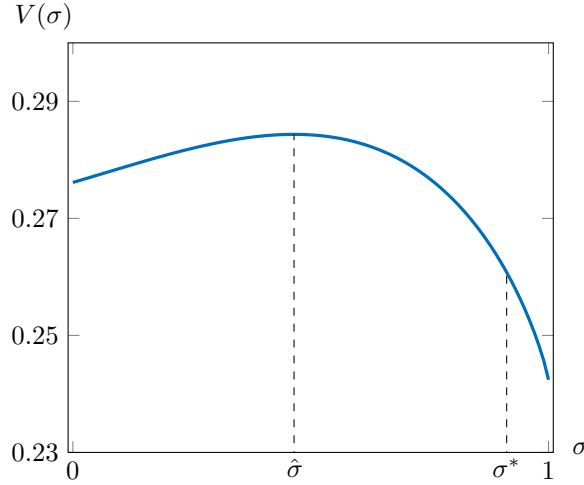

Figure 4: Aggregate consumer surplus as a function of σ

Perverse effect of more competition. As explained earlier, when privacy choices are endogenous, more competition can harm consumers. To illustrate this in a simple way, suppose that $\underline{\tau} \geq 0$. Then, as explained above, under monopoly $\sigma^* = 0$, i.e., no consumer shares. Hence equilibrium consumer surplus under monopoly is

$$V_m = \int_{p_m}^{\overline{v}} (v - p_m) dF(v) = \int_{p_m}^{\overline{v}} [1 - F(v)] dv \ , \tag{22}$$

---

[29]In particular, $\sigma^* = 1$ is always an equilibrium if $\mathbb{E}[\max\{v_{n-1:n} - c, 0\}] \geq \overline{\tau}$. If all consumers share their data, firms optimally set a list price $p = \overline{v}$; comparing (20) and (21), each consumer then indeed prefers to share under the condition. Note also that the condition is easier to satisfy when n is larger.

where $p_m = \arg\max_p (p - c)[1 - F(p)]$ is the standard monopoly price. We now show that this can exceed consumer surplus under competition. Specifically, suppose that $\overline{\tau} \leq \Delta(0)$ for some $n \geq 2$. We then infer from equation (2) that with competition $\sigma^* = 1$, i.e., all consumers share. As explained above firms then set a list price $p(1) = \overline{v}$. Hence equilibrium consumer surplus with $n \geq 2$ firms is $V_s(1) - \mathbb{E}[\tau]$, where

$$V_s(1) = \mathbb{E}[\max\{v_{n-1:n} - c, 0\}] = \int_c^{\overline{v}} [1 - F_{(n-1)}(v)]dv \tag{23}$$

with $F_{(n-1)}(v) = F(v)^n + n(1 - F(v))F(v)^{n-1}$ being the CDF of $v_{n-1:n}$. Consumers are definitely worse off under competition if $V_s(1) < V_m$, i.e., if competitive personalized pricing gives them less surplus than uniform pricing in monopoly. This can arise as demonstrated in the following example:

Example. Suppose the v_i's are uniformly distributed on $[0, 1]$ and $c < 1$. The monopoly price is $p_m = \frac{1+c}{2}$ and so $V_m = \frac{(1-c)^2}{8}$. In addition

$$V_s(1) = c^n - c + \frac{n-1}{n+1}(1 - c^{n+1}) .$$

For instance, when $n = 2$ we find that $V_s(1) < V_m$ for all $c > \frac{5}{8}$, while when $n = 3$ we find that $V_s(1) < V_m$ for all $c > \frac{\sqrt{3}}{2}$. For a higher c, $V_s(1) < V_m$ for a wider range of n.[30]

The perverse effect of competition can also arise when privacy choice is interior. To illustrate this, consider Figure 5, which depicts market outcomes for the case where valuations are uniform on $[0, 1]$, $c = 3/4$, and τ is uniform on $[0, 0.04]$. The red dotted curve shows outcomes when all consumers exogenously hide their data: as n increases, firms reduce their (list) price, and consumers benefit due to the lower price and higher variety. The solid blue curve shows outcomes when privacy choices are endogenous. As explained earlier, when $n = 1$ no consumer shares her data. However, for $n \geq 2$, sharing consumers receive personalized discounts, which encourages those with low τ to share. Indeed, the equilibrium σ^* grows monotonically in n, and for $n \geq 6$ all consumers share.[31] Due to this endogenous sharing, the equilibrium list price increases monotonically in n. The increase

---

[30] In fact, for a general valuation distribution, $V_s(1) < V_m$ holds if c is sufficiently high (and the privacy cost condition $\overline{\tau} \leq \Delta(0)$ is satisfied). Under monopoly, all consumers hide (assuming still that $\underline{\tau} \geq 0$) and so receive the same uniform price, which allows high-valuation consumers to get some surplus. Under competition, however, all consumers share. As explained in Rhodes and Zhou (2024), if c is high, few consumers have a valuation above c for more than one product, so each firm acts almost like a monopolist. This means that consumers are almost fully extracted, and so are worse off than under monopoly.

[31] The privacy choice game has a unique equilibrium σ^* for all n values considered in Figure 5. However, for (much) larger values of n there are multiple equilibria. For large n there is always an equilibrium with

in σ^* and hence also in the list price is particularly sharp at $n = 6$, which explains why an increase from $n = 5$ to $n = 6$ actually *reduces* consumer surplus. (Moreover, notice that due to the negative externalities exerted by sharing consumers and the privacy cost they pay, for each $n \geq 2$ consumers would be weakly better off if they all hid their data.)

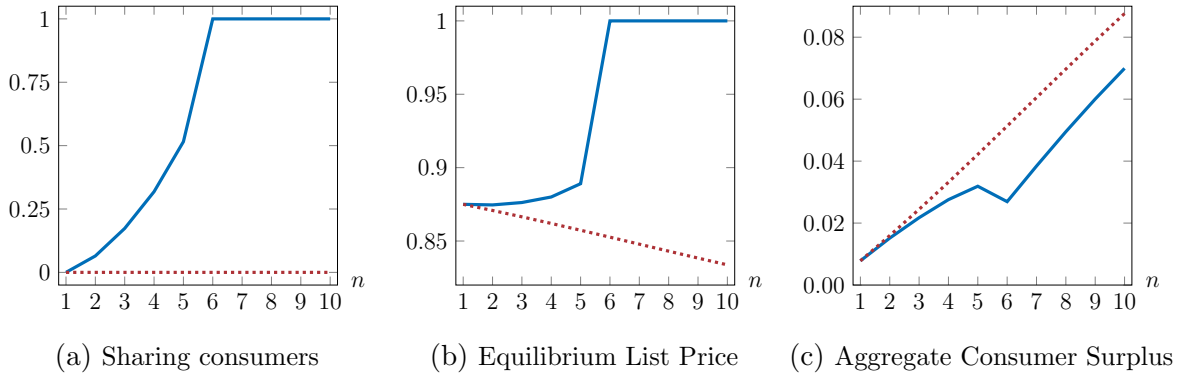(a) Sharing consumers   (b) Equilibrium List Price   (c) Aggregate Consumer Surplus

Figure 5: The impact of competition with personalized prices

(The solid blue curve depicts equilibrium outcomes, and the dotted red curve depicts outcomes when all consumers exogenously hide their data)

Perverse effect of improved data security. Improvements in data security can also harm consumers when privacy choices are endogenous. We gave a simple example where better data security switched all consumers from anonymous to sharing, and noted that consumers would be harmed if $V_s(1) < V_a(0)$. This condition exactly corresponds to the condition in Rhodes and Zhou (2024) for competitive personalized pricing to harm consumers in aggregate relative to uniform pricing. As shown there, this cannot happen if the market is fully covered under uniform pricing, but often happens when market coverage is low, and always happens when the production cost c is sufficiently high.

Other related literature. There is a substantial literature on personalized pricing and consumer privacy choice. Most papers consider purchase-history based price discrimination as surveyed in Fudenberg and Villas-Boas (2007).[32] Our application to personalized pricing is most closely related to Section 6 in Anderson, Baik, and Larson (2023). They

---

$\sigma^* = 1$; given our assumptions on $c, \overline{v}, \overline{\tau}$ in this example, the explanation follows from footnote 29. For large n there are also interior equilibria: intuitively, if few consumers share, competition between firms leads to p close to c, which induces only those consumers with the very lowest τ close to zero to share.

[32]Interesting externalities across consumers can also arise in that context. For example, consider the two-period monopoly model in Conitzer, Taylor, and Wagman (2012), where consumers who buy in the first period can hide (at some cost) their purchase history, and in the second period the firm can price

also consider an oligopoly model where personalized pricing is implemented as individualized discounts off a list price, and where consumers make their privacy choice before learning their product valuations. There are two key differences. Firstly, in their model it is costly to send discounts (via targeted ads) to consumers. This induces a more complicated analysis of equilibrium discounts and advertising since they need to deal with mixed-strategy equilibrium. Secondly, they assume full market coverage. This simplifies the analysis of equilibrium list prices. Like us, they show that having more sharing consumers increases firms' list prices and so imposes a negative externality on other consumers. They also point out that there can be too much data sharing, and that making data sharing more costly can benefit consumers in aggregate. However, they do not discuss the potential perverse effect of more competition due to the way it can endogenously cause more consumers to share.

Belleflamme and Vergote (2016) also make a similar point, but in a different setup, that having more sharing consumers is associated with a higher price offered to anonymous consumers. They consider a setting where consumers can hide their data (at a cost), and a monopoly firm can commit to the price it will charge anonymous consumers. The firm then has an incentive to raise this anonymous-consumer price, so as to reduce consumers' incentives to hide their data. Therefore, in their model price is strategically distorted upwards to encourage sharing. This is very different from our setup where a higher anonymous-consumer price is a consequence of having more sharing consumers.

Ali, Lewis, and Vasserman (2023) also study consumer privacy choice and personalized pricing. However, they consider a different game where consumers know their preference types before their privacy choices, they do not face any intrinsic privacy costs, and they can disclose different information to different firms. They do not study privacy-choice externalities explicitly, though there is an externality across consumers via a market segmentation effect as also studied in Galperti, Liu, and Perego (2024).

---

discriminate consumers who did not hide. When the cost of anonymity increases, more consumers share their data (other things equal); the firm optimally responds by cutting its first-period price, so as to sell to more consumers who it can price discriminate later on. Unlike in our model, this generates a positive externality on anonymous consumers. See also Montes, Sand-Zantman, and Valletti (2019) for a similar positive externality of sharing consumers on anonymous consumers.

3.3 Personalized product design

As explained in the Introduction, improvements in technology are also making it easier for firms to offer consumers personalized products. In this section we study the interaction between personalized product design and privacy choice. To do this, we build on the canonical monopoly screening model of Mussa and Rosen (1978).

Primitives   There is a measure one of consumers with different quality preferences. If a consumer of type $\theta \in [0, \bar{\theta}]$ buys a product of quality q at price p, she obtains a surplus $\theta q - p$. Let $F(\theta)$ be the CDF of consumer types in the population, and assume that its density function $f(\theta)$ is strictly positive everywhere and that $1 - F(\theta)$ is log-concave. There is a monopoly firm in the market. Its cost of producing a product of quality q is $c(q)$. Suppose that $c(q)$ is strictly convex, $c'(0) = 0$ and $c'(q) > \bar{\theta}$ for sufficiently large q.

   If a consumer does not share her data, her type θ is her private information, and the firm offers her a "public" menu of products. If a consumer does share her data, then the firm perfectly learns her type θ, and offers her a personalized price-quality pair (which other consumers cannot access). Like in our second application, we assume that a sharing consumer can still buy from the public menu of products if she wishes. When she is indifferent between the personalized offer and the best option in the public menu, we assume she buys the former.

   We again assume that consumers make their privacy choice before learning their type θ (e.g., because their quality preferences are market-specific). However, as we discuss later, in this application nothing would change if consumers knew their θ when making privacy choices.

Optimal product design with a fixed σ.   We first study optimal product design with an exogenous fraction σ of sharing consumers. Anonymous consumers can only buy from the public menu, which we denote by $\{(q_a(\theta), p_a(\theta))\}_\theta$. Sharing consumers can buy from the public menu, but also receive a personalized offer which we denote by $(q_s(\theta), p_s(\theta))$. (For notational simplicity, we suppress the dependence of these variables on σ.) It is evident that the firm is (weakly) better off selling to a sharing consumer via a personalized offer, since it can at least mimic the public offer.

   A type-θ anonymous consumer buys the product $(q_a(\theta), p_a(\theta))$ designed for her if and only if both the standard IC conditions $\theta q_a(\theta) - p_a(\theta) \geq \theta q_a(\theta') - p_a(\theta')$ and the IR

conditions $\theta q_a(\theta) - p_a(\theta) \geq 0$ hold for any $\theta, \theta' \in [0, \bar{\theta}]$. Define

$$v_a(\theta) \equiv \max_{\theta'} \theta q_a(\theta') - p_a(\theta')$$

as a type-θ anonymous consumer's equilibrium surplus. Then $v_a(\theta)$ must be increasing and convex, and $v_a'(\theta) = q_a(\theta)$ almost everywhere. Hence we can write $v_a(\theta)$ as

$$v_a(\theta) = \int_{\hat{\theta}}^{\theta} q_a(t)dt \ , \tag{24}$$

where $\hat{\theta}$ is the critical type such that types below it are excluded from the market. It is well-known that the IC conditions are equivalent to (24) and $q_a(\theta)$ being increasing.

Since sharing consumers have access to both the public menu and their personalized offer $(q_s(\theta), p_s(\theta))$, they have an additional IC constraint given by

$$v_s(\theta) \equiv \theta q_s(\theta) - p_s(\theta) \geq \theta q_a(\theta) - p_a(\theta) = v_a(\theta) \tag{25}$$

for any θ, where $v_a(\theta) = 0$ for $\theta \leq \hat{\theta}$.

The seller's problem is to maximize

$$\sigma \int_0^{\bar{\theta}} [p_s(\theta) - c(q_s(\theta))]dF(\theta) + (1 - \sigma) \int_{\hat{\theta}}^{\bar{\theta}} [p_a(\theta) - c(q_a(\theta))]dF(\theta)$$

subject to (24), the monotonicity constraint that $q_a(\theta)$ is increasing, $v_a(\hat{\theta}) = 0$, and (25).

We first solve the relaxed problem by ignoring the monotonicity constraint. It is evident that the constraint (25) must bind for each type (otherwise the firm could always improve its profit by increasing the personalized price $p_s(\theta)$). Hence, we have

$$p_s(\theta) = \theta q_s(\theta) - v_a(\theta) \ . \tag{26}$$

From the definition of $v_a(\theta)$ we also have

$$p_a(\theta) = \theta q_a(\theta) - v_a(\theta) \ . \tag{27}$$

Substituting them into the objective function and then using (24) and integration by parts, the seller's objective function simplifies to

$$\sigma \int_0^{\bar{\theta}} [\theta q_s(\theta) - c(q_s(\theta))]dF(\theta) + (1-\sigma) \int_{\hat{\theta}}^{\bar{\theta}} \left[\left(\theta - \frac{1}{1-\sigma} \frac{1 - F(\theta)}{f(\theta)} \right) q_a(\theta) - c(q_a(\theta)) \right] dF(\theta) \ .$$

This optimization problem can be solved pointwise. The optimal $q_s(\theta)$ solves $\theta = c'(q_s(\theta))$, which means that each sharing consumer's personalized product has the efficient quality level. To solve for the public menu, let $\hat{\theta}$ be the unique solution to

$$\hat{\theta} = \frac{1}{1 - \sigma} \frac{1 - F(\hat{\theta})}{f(\hat{\theta})} \ . \tag{28}$$

Then the optimal $q_a(\theta)$ is zero for $\theta < \hat{\theta}$, and otherwise is the unique $q_a(\theta)$ that solves

$$\theta - \frac{1}{1-\sigma}\frac{1-F(\theta)}{f(\theta)} = c'(q_a(\theta)) . \tag{29}$$

Given the log-concavity of $1 - F$, the $q_a(\theta)$ that we have solved for is indeed increasing in θ.[33] Hence the solution to the relaxed problem is also the real optimal solution.

Lemma 5. *Suppose that an exogenous fraction σ of consumers share their data.*
(i) The personalized products designed for sharing consumers are efficient, and are sold at prices $p_s(\theta)$ defined in (26) which increase in σ.
(ii) The products designed for anonymous consumers with $\theta \geq \hat{\theta}$ have qualities $q_a(\theta)$ solving (29) and are sold at prices $p_a(\theta)$ defined in (27). (Anonymous consumers with $\theta < \hat{\theta}$ leave the market without purchasing a product.)
(iii) $q_a(\theta)$ is distorted downward (except at the top) and decreases in σ; if $c'''(q) \geq 0$, for a given σ, there exists a $\tilde{\theta} \in (\hat{\theta}, \bar{\theta})$ such that $\frac{dp_a(\theta)}{d\sigma}$ is positive for $\theta > \tilde{\theta}$ and negative for $\theta < \tilde{\theta}$.

The firm offers a sharing consumer an efficient personalized product, so as to maximize the surplus generated by her. The firm then prices that product in such a way that the consumer is indifferent between buying it or her favorite product from the public menu.

The design of the public menu (which in equilibrium only caters to anonymous consumers) is more interesting. First, as usual, product quality $q_a(\theta)$ is distorted downwards for all but the highest type $\bar{\theta}$. Second, it is clear from equation (29) that $q_a(\theta)$ decreases in σ, i.e., the distortion is more severe when more consumers share. Third, it is also evident from equation (28) that $\hat{\theta}$ increases in σ, i.e., more anonymous consumers are excluded from the market when there are more sharing consumers. Intuitively, to extract more surplus from sharing consumers, the firm deteriorates the public menu by offering worse products and excluding more consumers. Moreover, recalling the expression for $v_a(\theta)$ in equation (24), the previous two observations imply that anonymous consumers are made worse off as more consumers share their data, i.e., $v_a(\theta)$ decreases in σ for each θ. Since the optimal solution has $v_s(\theta) = v_a(\theta)$, sharing consumers are also made worse off as σ increases. Indeed, in the limit as $\sigma \to 1$, the firm excludes all anonymous consumers from purchasing so that it can fully extract surplus from sharing consumers, in which case every consumer gets a zero surplus. In summary:

---

[33]Note that the usual regularity condition that the virtual type $\theta - \frac{1-F(\theta)}{f(\theta)}$ is increasing in θ is not sufficient to ensure an increasing $q_a(\theta)$ for any σ.

Corollary 5. *Sharing consumers exert negative externalities: $V_a'(\sigma) < 0$ and $V_s'(\sigma) < 0$.*

Continuing with properties of the firm's optimal product design, note that since sharing consumers' fallback options $v_a(\theta)$ decrease in σ, the firm charges more for personalized products when more consumers share their data. The impact of σ on the prices for anonymous consumers is subtler. Intuitively, for high-type consumers, distorting quality design is rather costly, so the distortion is small; in order to deteriorate the offers to them, the firm raises the prices. For low-type consumers (with $\theta \geq \hat{\theta}$), however, the quality distortion is severe, so the firm needs to charge them less to induce them to still purchase.

We illustrate some of these properties of the optimal public menu in Figure 6. Suppose the type θ is uniformly distributed on $[0, 1]$, and that the cost of quality is $c(q) = q^3$. The figure shows that as σ increases, the qualities offered to anonymous consumers fall and more lower types are excluded from the market (left panel), while for those types that are still served, the offered price falls for relatively low types but increases for relatively high types (right panel).
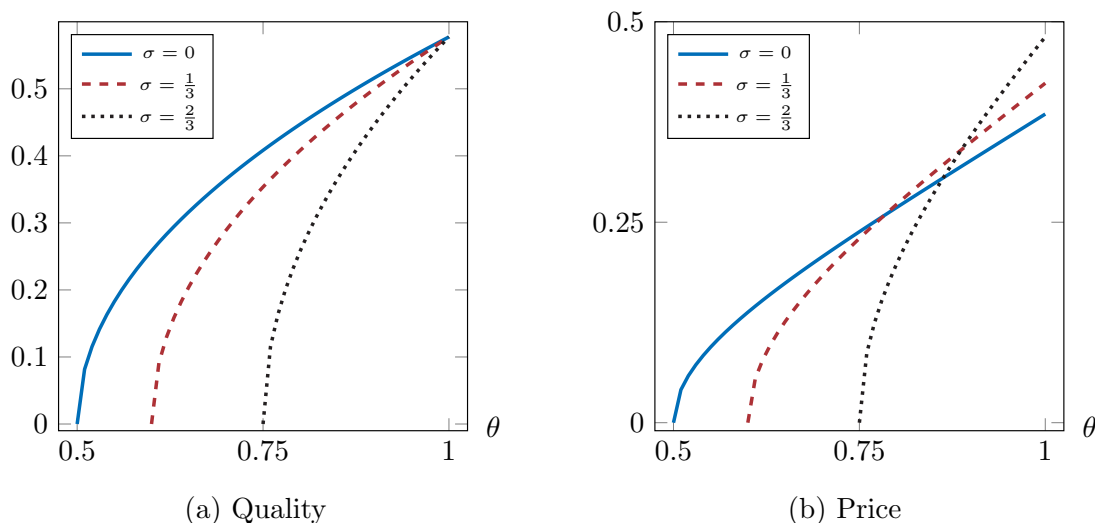

(a) Quality   (b) Price

Figure 6: The menu offered to anonymous consumers of type θ, for different values of σ.

Privacy choice equilibrium. We now solve for equilibrium privacy choices at the first stage of the game. Recall from above that any type obtains the same surplus regardless of whether she shares her data or not (i.e., $v_s(\theta) = v_a(\theta)$). As a result, the ex-ante expected surplus is also the same (i.e., $V_s(\sigma) = V_a(\sigma)$). Therefore a consumer shares her data if and only if $\tau < 0$, so there is a unique equilibrium with $\sigma^* = T(0)$. Notice also that because

$v_s(\theta) = v_a(\theta)$ for any θ, in this application, whether a consumer makes her privacy choice before or after learning θ does not affect the privacy-choice equilibrium.

Too much data sharing. Given sharing consumers exert negative externalities, by Proposition 2 there is too much data sharing in equilibrium (if $\underline{\tau} < 0$) relative to the consumer optimum. Unlike the previous two applications, here $\sigma V_s(\sigma) + (1 - \sigma)V_a(\sigma)$ always decreases in σ because $V_s(\sigma) = V_a(\sigma)$ and both decrease in σ. However, this does not necessarily imply that $V(\sigma)$ decreases in σ if $\underline{\tau} < 0$. In that case, when σ is small, sharing consumers all pay a *negative* privacy cost, and so a higher σ can lead to higher consumer surplus.

Perverse effect of improved data security. Since $\sigma^* = T(0)$, a decrease in the distribution of privacy types (weakly) increases σ^*; given that $v_s(\theta) = v_a(\theta)$ decreases in σ, this leads to a reduction in each individual's consumption surplus. After accounting for the reduction in privacy cost for some consumers who share, it is still possible that overall consumer surplus decreases. To illustrate this in a simple way, suppose that initially $\underline{\tau} > 0$, such that nobody shares their data, in which case total consumer surplus is strictly positive. Suppose that after the improvement in data security $\overline{\tau} < 0$, such that everybody shares; consumers now receive no consumption surplus, so if $\mathbb{E}[\tau]$ is sufficiently close to zero, total consumer surplus is lower than before the improvement in data security.

Discussion: competition. Throughout this section we have focused on monopoly. An oligopoly version of the Mussa-Rosen model is known to be complicated to deal with, and in general very limited analytical progress can be made (see, e.g., Rochet and Stole, 2002). One simple case is when all IC constraints become slack due to competition, which happens, for example, in the standard Hotelling model when the market is fully covered (see, e.g., Armstrong and Vickers, 2001). In that case, however, even the public menu would be efficient, and so for our purposes it would not be interesting since varying σ would have no effect on market equilibrium. Without full market coverage, the public menu is usually not efficient but it is complicated to characterize (see, e.g., the duopoly model explored in Yang and Ye, 2008). This is a challenging problem that we hope to revisit in future work.

Other related literature. Our application to personalized product design is most closely related to Bergemann and Bonatti (2024) and Vaidya (2023). The first paper

does not study consumer privacy choice, so is different from ours in terms of the research question. In their model some consumers use a platform to find a product, while others look directly in an offline market. For the former consumers, the platform perfectly observes their preferences and always steers them to (and also shares their preference information with) the best-matched firm. These consumers, however, still have the option to buy in the offline market. Due to the informational assumption in their paper, a Diamond (1971) paradox result applies such that no consumers search actively. As a result each firm acts as a monopolist (with an updated consumer type distribution). Since each firm offers a menu of products with different qualities, their model essentially boils down to Mussa and Rosen (1978) where some consumers' preferences are perfectly observable to the firm. Like us, they show that as more consumers use the platform, firms further distort their product design for offline consumers. The second paper studies a regulator's choice of what information consumers can disclose to a monopoly firm. The extension in its Section 7 also uses the Mussa-Rosen framework, but the setting is otherwise quite different from ours: there is no privacy cost, consumers are informed ex ante of their taste for quality, but only some are (exogenously) able to disclose it, and the firm can commit ex ante to a menu of price-quality pairs. The optimal menu induces disclosure by all those consumers who are able to do it, and so interestingly turns out to be the same as in our model (for a fixed σ). Both these papers, therefore, have the same negative externality that we highlight in our paper (although neither of them explores the impact of sharing/disclosure on prices). However, unlike us, they do not endogenize the fraction of sharing consumers through a privacy choice game and so, e.g., do not consider the possible perverse effect of improvements in data security.

Doval and Skreta (2023) also study consumer data and product design but in a very different setting. They consider a two-period model: in the first period, an upstream firm offers a menu of products to screen consumers as in Mussa and Rosen (1978); in the second period, a downstream firm sells a single product, and part of its profit is transferred to the upstream firm (e.g., because it buys data on consumer purchase history from the upstream firm). Importantly, the downstream firm can use first-period purchase information to infer consumers' willingness-to-pay for its product and do price discrimination. Anticipating this, consumers are less willing to buy in the first period when the product line is richer (because buying in the first period reveals more information). As a result, the upstream firm prunes its product line compared to the standard Mussa-Rosen case. Although this product-line distortion arises due to consumer data, the mechanism is very different from

ours, and unlike in our paper there is also no personalized product design.

Argenziano and Bonatti (2023) also consider a setting in which consumers trade sequentially with two firms, and differ in how much they value additional units of the two firms' goods. When the second firm has access to data about the consumer's consumption choice at the first firm, it is able to infer the consumer's type and offer her a personalized product. They examine whether the first firm should be allowed to degrade the offer it makes to consumers who forbid it from sharing their data. Like us, they assume that a consumer decides whether or not to share her data before she learns her consumption type. However, unlike us, they assume that consumers are unable to access a public offer, and so sharing consumers do not exert any externality on other consumers.

Krähmer and Strausz (2023) study a variant of Mussa and Rosen (1978) in which some consumers are data-sensitive and averse to having their types learned by the firm, while other consumers are standard and unconcerned about privacy. Unlike in our model, consumers in their framework learn both their privacy and consumption types at the outset, requiring the firm to screen them on both dimensions. To incentivize data-sensitive consumers to purchase, the firm must offer them the same product regardless of their consumption type heterogeneity. However, since standard consumers also have the option to buy the product designed for data-sensitive consumers, the firm strategically degrades the quality of the product intended for data-sensitive consumers. As a result, the presence of standard consumers imposes negative externalities on data-sensitive consumers. This mirrors the negative externality that sharing consumers impose on anonymous consumers in our model, as standard consumers in Krähmer and Strausz (2023) ultimately reveal their consumption types in equilibrium while data-sensitive consumers do not.

4 Discussion

We now briefly discuss some of our assumptions, and how they could be relaxed.

4.1 Imperfect learning of consumer types

We have implicitly assumed that once a consumer shares her data, firms perfectly learn her consumption type θ. In reality, however, learning may be imperfect. Moreover, the accuracy of learning may depend on how many other consumers share their data—e.g., because a larger dataset may improve the effectiveness of prediction algorithms.

Considering imperfect learning will not affect our analysis of the general framework

in Section 2. This is because our definitions of $V_s(\sigma)$ and $V_a(\sigma)$ in that section do not rely on perfect learning. (Note that with imperfect learning $v_s(\theta, \sigma)$ becomes an *expected* consumption surplus, because a consumer who shares her data generates a distribution of signals that firms use to make their offers.)

However, for the applications in Section 3, allowing for imperfect learning would make the analysis less tractable. For instance, in the application to personalized recommendations, a sharing consumer's optimal stopping rule becomes more challenging to characterize, and may not follow a simple cutoff structure.[34] Similarly, in the application to personalized pricing, imperfect learning complicates the study of price competition, which may no longer admit a pure-strategy equilibrium.[35] Moreover, in the application to personalized products, the firm would need to offer not only a public menu, but also a private menu to each sharing consumer so as to screen information about their type which is not revealed by the noisy signal. Nevertheless, in each application the main economic force that sharing consumers impose negative externalities on other consumers should remain unchanged. Consumers who are recommended a better-matched product—even if it is not a perfect match—should have less incentive to continue searching. Similarly, setting a higher list price still increases the flexibility of price discrimination among sharing consumers, even if the discrimination is no longer perfect. Moreover, degrading the public menu should still enable a firm to extract more surplus from sharing consumers.

4.2   Type-dependent privacy choice

So far we have assumed that consumers do not know their consumption type when making their privacy choice. As argued earlier, in many contexts this is a good approximation. Nevertheless, we now consider the possibility that consumers know their θ when choosing whether to share their data.[36] While a fully-fledged analysis can be quite complicated—

---

[34]To illustrate, consider $n = 3$. With imperfect learning a consumer who receives a relatively low valuation v_1 from the recommended product may search a second time. If the second search yields a product with $v_2 < v_1$ the consumer becomes more confident about the initial recommendation and may stop searching. However, if the second search yields $v_2 > v_1$, the consumer becomes less confident about the initial recommendation and, unless v_2 is very high, may search the third product.

[35]For example, consider the price competition game for a sharing consumer, and note that standard regularity conditions that ensure existence of a pure-strategy equilibrium need not hold for an arbitrary mapping from valuations to signals.

[36]Some policies may also make this case more relevant. For example, aside from giving consumers more control over their data, the GDPR requires firms to disclose how they use data. This increased transparency may help consumers better judge their consumption type at the privacy choice stage.

not least because a consumer's privacy choice may now (partially) signal her consumption type—we argue that our basic insights from earlier should remain largely unchanged.

Start with the application to personalized recommendations. By definition, in our search model consumers do not observe their product valuations *ex ante*, so it is impossible to have type-dependent privacy choices. Consider, however, a slightly different set-up in which $v_i = \lambda \tilde{v}_i$: suppose consumers know *ex ante* their λ (which indicates, e.g., their income level) but not their $\tilde{v}_i$'s (which can be learned through search after making their privacy choice). In this case, conditional on the privacy cost, consumers with higher λ are "choosier" and so have more incentive to share data and get recommendations; consumers who remain anonymous have smaller λ and so are less willing to search. Intuitively, this is an additional force for firms to raise their price when some additional consumers share their data (which in turn imposes a negative externality on other consumers).

Now consider the application to personalized pricing. Conditional on the privacy cost, consumers with stronger preferences (i.e., those with a larger gap between their highest and second-highest valuations) have less incentive to share their data. This implies that anonymous consumers are more likely to have strong preferences. Moreover, as more consumers share their data, the remaining anonymous consumers have even stronger preferences, which gives firms more incentive to raise their public list prices. (However the privacy choice game will not unravel, provided some consumers have a sufficiently high privacy cost that they never share irrespective of their consumption type.) This is an additional force for sharing consumers to exert a negative externality on other consumers.

Finally, as noted earlier, in the application to personalized product design, even if a consumer knew her θ when making her privacy choice, it would not affect her decision of whether or not to share. This is because personalized products are priced so as to fully extract a sharing consumer's additional surplus relative to the public menu.

4.3   Correlation between consumption and privacy types

For simplicity, we have also assumed that a consumer's privacy type τ is independent of her consumption type θ. However in practice they may be correlated, e.g., if higher-income consumers are more concerned about data security, and also have systematically different product valuations compared to lower-income consumers. When τ and θ are correlated, a consumer will update her belief about the latter when she learns the former. This causes two differences compared to our baseline model. First, there is now a signaling effect in the privacy choice stage, as was also the case in Section 4.2. This affects firms' beliefs

about the distribution of consumption types among anonymous consumers. Second, the expected consumption surplus associated with each privacy choice now depends on the privacy cost type τ. More precisely, we now have $V_s(\sigma|\tau) = \mathbb{E}_\theta[v_s(\theta, \sigma)|\tau]$ and $V_a(\sigma|\tau) = \mathbb{E}_\theta[v_a(\theta, \sigma)|\tau]$, and hence a consumer with privacy cost τ shares her data if and only if $\tau < V_s(\sigma|\tau) - V_a(\sigma|\tau)$. This makes the analysis more challenging, because it implies that consumer privacy choices may no longer follow a simple cutoff structure.[37]

Moreover, even if consumer privacy choices still follow a cutoff, introducing correlation may make privacy choice externalities harder to sign. For example, consider the application to personalized recommendations. If a higher τ is associated with less dispersed product valuations, we still have a cutoff result: consumers should choose to share if and only if their privacy cost is below a threshold. Now suppose that, for reasons orthogonal to the change in the correlation structure, more consumers share. Firms then infer that anonymous consumers have more dispersed product valuations and so are more likely to search. This is a force towards firms setting lower prices, which works against—though not necessarily reverses—the negative externality from our baseline analysis. We believe this would be an interesting avenue to explore in future research.

4.4 Behavioural consumers and educative policies

So far consumers have been assumed to be perfectly rational in their privacy choices. We now allow for some "behavioral" consumers who always hide or share their data.

Suppose a fraction $m_s \in [0, 1)$ of consumers (greatly) underestimate their privacy cost and so always share, while a fraction $m_a \in [0, 1 - m_s)$ of consumers (greatly) overestimate their privacy cost and so always hide their data. The remaining fraction $1 - m_s - m_a$ of ("rational") consumers use their true privacy cost to decide whether or not to share; as in the main model, they share if and only if $\tau \leq \Delta(\tilde{\sigma})$, where $\tilde{\sigma}$ is the total fraction of consumers in the market who share data. In equilibrium, we then have that

$$\tilde{\sigma}^* = m_s + (1 - m_s - m_a) \, T(\Delta(\tilde{\sigma}^*)). \tag{30}$$

Closely following earlier analysis, there is always at least one equilibrium $\tilde{\sigma}^*$, and all the main results from Section 2 carry over (including, e.g., the impact of privacy policies).

One natural policy intervention here is "education", which helps consumers learn the true consequences of their sharing decisions. For brevity, consider a policy that converts

---

[37]For example, consider the application to personalized recommendations, and suppose that a higher τ is associated with more dispersed product valuations. It is possible that consumers with high or low τ choose to share their data, while those with intermediate τ choose to hide their data.

some of the m_s behavioral consumers into rational consumers. Assuming for simplicity that the privacy game has a unique equilibrium, it is straightforward to show that $\tilde{\sigma}^*$ decreases. Hence, when sharing externalities are negative (i.e., $V_s'(\sigma) < 0$ and $V_a'(\sigma) < 0$) the policy benefits all consumers, but when the externalities are positive the opposite is true. Both possibilities are illustrated by the following example.

Example: the effect of educative policies. Consider the application to personalized recommendations, in which $\Delta(\tilde{\sigma}) \equiv \Delta > 0$ is constant in $\tilde{\sigma}$. Suppose the true privacy cost is distributed on $[\underline{\tau}, \overline{\tau}]$, with $\underline{\tau} = \Delta$ (so that no rational consumer will share their data). Suppose $m_s \in (0, 1)$ behavioral consumers falsely believe that $\tau = 0$, while the remaining $1 - m_s$ rational consumers use the correct privacy cost. Without an educative policy, behavioral consumers share and so $\tilde{\sigma}^* = m_s$; rational consumers get surplus $V_a(m_s)$, while behavioral consumers get surplus $V_s(m_s) - \mathbb{E}[\tau] < V_a(m_s)$ given $\underline{\tau} = \Delta = V_s(m_s) - V_a(m_s)$. With an educative policy, behavioral consumers learn their true privacy cost and do not share, hence $\tilde{\sigma}^* = 0$; all consumers now get surplus $V_a(0)$. When consumers are recommended their best product, $V_s'(\sigma) < 0$ and $V_a'(\sigma) < 0$, so all consumers benefit from the policy. However when recommendations are surplus-based, $V_s'(\sigma) > 0$ and $V_a'(\sigma) > 0$, so for $\overline{\tau}$ sufficiently close to $\underline{\tau}$ all consumers lose out from the policy.

Remark. Educative policies can be relevant even with rational consumers. For instance, suppose consumers are rational but receive only a noisy signal of their privacy cost. Consider an educative policy that enables them to fully learn their τ, resulting in a more dispersed distribution in the sense of a mean-preserving spread. The impact of this policy depends on the specifics of the privacy cost distribution. To illustrate, consider the application to personalized recommendations, and again recall that $\Delta(\sigma) = \Delta > 0$ is constant in σ. Suppose for simplicity that, absent education, all consumers lack information on their privacy cost and perceive it to equal $\mathbb{E}[\tau]$. First, notice that if $\mathbb{E}[\tau] < \Delta$ and the true distribution of τ is wide enough, the policy causes a shift from a corner equilibrium where everybody shares, to an interior equilibrium where only some consumers share. Consumers make better informed decisions, which is a force towards higher consumer surplus; the reduction in sharing reinforces this if externalities are negative, but works against it if externalities are positive. Second, notice that if $\mathbb{E}[\tau] > \Delta$ and the true distribution of τ is wide enough, the policy causes a shift from a corner equilibrium where nobody shares, to an interior equilibrium where some consumers share. Consumers are unambiguously better off in this case if externalities are positive.

4.5 Data acquisition by firms

In practice, the amount of data that firms possess depends not only on how much data consumers are willing to share but also on how much data firms choose to acquire. Our general framework does not specify the market stage game and thus is not well suited to directly study firms' data acquisition. However, this issue could be separately dealt with in each application. For instance, in the context of personalized recommendations, we can consider a strategic platform that chooses how much data to acquire. If its profit is aligned with sellers' profits, the platform has a strong incentive to acquire data from as many consumers as possible—subject to costs and data availability—since recommendations to more consumers soften price competition among sellers. A similar logic applies in the product design application, where the monopoly firm benefits from learning more consumers' quality preferences. The case of personalized pricing is more complex due to strategic interactions among competing firms. Since personalized pricing among more consumers can intensify overall price competition, firms may have a collective incentive to limit data acquisition; however, each firm's individual incentive to gain a competitive advantage through offering personalized prices to more consumers may lead to a prisoner's dilemma (see, e.g., Thisse and Vives, 1988 for an example of this when data is available on all consumers).

5 Conclusion

This paper offers a simple framework to study the interaction between data-driven personalization and consumers' incentives to share their data. We highlighted a novel externality, whereby sharing by some consumers affects the payoff of others via its impact on firms' behavior. Depending on whether this externality is positive or negative, we showed that consumers may share too much or too little, and privacy policies such as GDPR could benefit or harm consumers. We then applied the framework to understand the role of personalized recommendations, prices, and product design respectively, and argued that often sharing consumers impose a negative externality on others. Moreover, we showed that due to this negative externality, more competition, or improvements in data security, can harm consumers by incentivizing more of them to share their data.

Appendix: Omitted Proofs and Details

Proofs for Section 3.1

Proof of Lemma 3. It is straightforward to check that the left-hand side of (10) is greater than the right-hand side as $p \to c$, and the opposite is true as $p \to r$ under condition (11).

The next step is to show that the right-hand side of (10) divided by $\frac{1-F(p)^n}{n}$ increases in p, which implies uniqueness of the solution. When f is log-concave, so is $1 - F(p)^n$, which implies that $\frac{F(p)^{n-1}f(p)}{1-F(p)^n}$ increases in p. Now consider

$$\frac{f(r)\frac{1-F(r)^n}{1-F(r)} - n\int_p^r F(v)^{n-1}f'(v)dv}{1 - F(p)^n} .$$

It is increasing in p if and only if

$$[1 - F(p)^n]\frac{f'(p)}{f(p)} + f(r)\frac{1 - F(r)^n}{1 - F(r)} - n\int_p^r F(v)^{n-1}f'(v)dv \geq 0 . \tag{31}$$

Notice that when f is log-concave, $\frac{f'}{f}$ is decreasing, and so

$$n\int_p^r F(v)^{n-1}f'(v)dv = \int_p^r \frac{f'(v)}{f(v)}dF(v)^n \leq \frac{f'(p)}{f(p)}[F(r)^n - F(p)^n] .$$

Therefore, the left-hand side of (31) is greater than

$$[1 - F(r)^n]\frac{f'(p)}{f(p)} + f(r)\frac{1 - F(r)^n}{1 - F(r)} \geq [1 - F(r)^n]\frac{f'(r)}{f(r)} + f(r)\frac{1 - F(r)^n}{1 - F(r)} \geq 0 ,$$

where the first inequality used $\frac{f'(p)}{f(p)} \geq \frac{f'(r)}{f(r)}$, and the second is because the log-concavity of $1 - F$ implies $[1 - F(r)]f'(r) + f(r)^2 \geq 0$.

To show p increases in σ, it then suffices to show that

$$F(p)^{n-1}f(p) < \frac{f(r)}{n}\frac{1 - F(r)^n}{1 - F(r)} - \int_p^r F(v)^{n-1}f'(v)dv ,$$

so that the right-hand side of (10) decreases in σ. By integration by parts, the above inequality can be written as

$$F(p)^{n-1}f(p) < \frac{f(r)}{n}\frac{1 - F(r)^n}{1 - F(r)} - F(r)^{n-1}f(r) + F(p)^{n-1}f(p) + \int_p^r f(v)dF(v)^{n-1} .$$

This must be true as $\frac{1-F(r)^n}{1-F(r)} > nF(r)^{n-1}$. $\qquad\square$

We now prove the following claim about condition (11).

Claim 1. *Condition* (11) *holds when r is above a certain threshold.*

Proof. Given our assumption that f is log-concave, $1 - F(r)^n$ is also log-concave in r and so the right-hand side of (11) is decreasing in r. Then it suffices to show that (11) holds when r is sufficiently close to $\bar{v}$. To see this, note that

$$\lim_{r \to \bar{v}} \frac{1 - F(r)^n}{nF(r)^{n-1}f(r)} = \lim_{r \to \bar{v}} \frac{1 - F(r)^n}{1 - F(r)} \frac{1 - F(r)}{nf(r)} = \lim_{r \to \bar{v}} \frac{1 - F(r)}{f(r)} .$$

Given $\frac{1-F(r)}{f(r)}$ is decreasing in r, its limit is clearly less than $\lim_{r \to \bar{v}}(r - c)$ if $\bar{v} = \infty$, or if $\bar{v} < \infty$ and $f(\bar{v}) > 0$. If instead $\bar{v} < \infty$ and $f(\bar{v}) = 0$, then $f(v)$ must be strictly decreasing in a neighborhood of $\bar{v}$, and hence for r close to $\bar{v}$ we have

$$\frac{1 - F(r)}{f(r)} = \frac{\int_r^{\bar{v}} f(v)dv}{f(r)} < \frac{f(r)(\bar{v} - r)}{f(r)} = \bar{v} - r < r - c .$$

$\square$

We now prove the following claim from the discussion on surplus-based recommendations.

Claim 2. *The first-order condition* (15) *has a unique solution p and it decreases in σ.*

Proof. First, we prove that the right-hand side of (15) increases in σ. This requires that

$$F(p)^{n-1}f(p) + \int_p^{\bar{v}} f(v)dF(v)^{n-1} > \frac{f(r)}{n} \frac{1 - F(r)^n}{1 - F(r)} - \int_p^r F(v)^{n-1}f'(v)dv .$$

Integrating the left-hand side by parts, this is equivalent to

$$f(\bar{v}) - \int_p^{\bar{v}} F(v)^{n-1}f'(v)dv > \frac{f(r)}{n} \frac{1 - F(r)^n}{1 - F(r)} - \int_p^r F(v)^{n-1}f'(v)dv .$$

Note that as $r \to \bar{v}$ the right-hand side equals the left-hand side. It is therefore sufficient to prove that the right-hand side is strictly increasing in r, or equivalently that

$$\left[f'(r) + \frac{f(r)^2}{1 - F(r)} \right] \left[\frac{1}{n} \frac{1 - F(r)^n}{1 - F(r)} - F(r)^{n-1} \right] \geq 0.$$

This holds because f log-concave implies $f/[1 - F]$ is increasing, hence the first bracketed term is positive, while the second bracketed term is positive because $\frac{1-F(r)^n}{1-F(r)} > nF(r)^{n-1}$.

Next, it is clear that as $p \to c$ the left-hand side of (15) exceeds the right-hand side. In addition, given that a lower bound on the right-hand side of (15) is obtained by setting $\sigma = 0$, the condition in footnote 21 ensures that as $p \to r$ the right-hand side of (15) exceeds the left-hand side.

The next step is to show that the right-hand side divided by $1 - F(p)^n$ increases in p and so the solution is unique. We showed that the second term is increasing in the proof of Lemma 3. Consider the first term, and rewrite it (ignoring the σ part) using integration by parts as

$$\frac{f(\overline{v}) - \int_p^{\overline{v}} F(v)^{n-1} f'(v) dv}{1 - F(p)^n} .$$

Its derivative with respect to p is

$$\frac{F(p)^{n-1} f'(p)}{1 - F(p)^n} + \frac{f(\overline{v}) - \int_p^{\overline{v}} F(v)^{n-1} f'(v) dv}{[1 - F(p)^n]^2} n F(p)^{n-1} f(p) \geq 0 ,$$

where the inequality uses $f(\overline{v}) \geq 0$ and $f'(v) f(p) \leq f'(p) f(v)$ for $v > p$ (since $f(v)$ is log-concave).

Finally, we have already shown that the right-hand side of (15) increases in σ. It is then immediate that the equilibrium price decreases in σ. $\qquad\square$

Proof for Section 3.2

We begin by stating and proving Claims 3 and 4, which are required to prove Lemma 4.

Claim 3. *The profit function (18) is quasi-concave in p_i for any $p \geq c$.*

Proof. When $\sigma = 1$, profit equals $\pi_s(p_i)$ which is concave and hence also quasi-concave. When $\sigma = 0$, we have the uniform pricing regime studied in Rhodes and Zhou (2024); the results there imply that profit here is quasi-concave.

The remainder of the proof deals with the case $\sigma \in (0, 1)$ and $n \geq 2$ (the monopoly case is straightforward and so is omitted). To prove the result, we can focus on $p_i \geq c$ for which $1 - H_p(p_i - p) > 0$. Notice that the derivative of (18) with respect to p_i is proportional to (i.e., has the same sign as)

$$1 - (p_i - c) \frac{h_p(p_i - p)}{1 - H_p(p_i - p)} + \frac{\sigma}{1 - \sigma} \frac{1 - H_c(p_i - c)}{1 - H_p(p_i - p)} . \tag{32}$$

Therefore, if (32) is decreasing in p_i, (18) must be quasi-concave in p_i. From Rhodes and Zhou (2024), we already know that $1 - H_p(p_i - p)$ is log-concave in p_i given that f is log-concave, and so the first two terms in (32) are decreasing in p_i. Therefore, it suffices to show that the final term is decreasing in p_i. A sufficient condition for this is that $1 - H_y(x - y)$ be *totally positive of order 2* (TP_2) in (x, y). TP_2 implies that for x' and $x'' < x'$ and also y' and $y'' < y'$, we have

$$[1 - H_{y'}(x' - y')][1 - H_{y''}(x'' - y'')] \geq [1 - H_{y'}(x'' - y')][1 - H_{y''}(x' - y'')] ,$$

and so provided that $1 - H_{y'}(x' - y') > 0$,

$$\frac{1 - H_{y''}(x'' - y'')}{1 - H_{y'}(x'' - y')} \geq \frac{1 - H_{y''}(x' - y'')}{1 - H_{y'}(x' - y')} .$$

Then the desired result follows by setting $y'' = c$ and $y' = p$.

To prove the TP_2 property, we invoke the following theorem from Karlin (1968):

Theorem 1 (Theorem 5.2 on p. 124 of Karlin (1968)). *Let $f(\lambda, x)$ and $g(\lambda, x)$ be defined for $\Lambda \times X$, where Λ is linearly ordered and X is $(-\infty, \infty)$ (or the set of all integers). Suppose f and g are TP_2 in the variables $\lambda \in \Lambda$ and $x \in X$, and are PF_2 with respect to the x variable; i.e., $f(\lambda, x - \zeta)$ is TP_2 $(-\infty < x, \zeta < \infty)$ for fixed λ. Assume that*

$$h(\lambda, x) = \int_{-\infty}^{\infty} f(\lambda, x - \zeta)g(\lambda, \zeta)d\zeta \qquad -\infty < x < \infty; \lambda \in \Lambda$$

is well defined. Then h is TP_2 in the variables λ and x.

Notice that under our IID assumption,

$$1 - H_y(x - y) = \int_x^{\infty} G(v - x + y)dF(v) = \int_{-\infty}^{\infty} G(y - (x - v))\mathbf{1}_{x - v < 0}f(v)dv , \quad (33)$$

where $G(v) = F(v)^{n-1}$ is the CDF of $\max_{j \neq i} v_j$. First, $f(v)$ is trivially TP_2 in (y, v), and is PF_2 in v given that f is log-concave. Second, we prove that $G(y - (x - v))\mathbf{1}_{x - v < 0}$ is TP_2 in (y, x) for fixed v, and also TP_2 in (x, v) for fixed y. Since $F(v)$ is log-concave, $G(v)$ is log-concave and so PF_2 in v. Using the fact that if $k(x)$ is PF_2 then $k(x - y)$ is TP_2 in (x, y), it follows that $G(y - (x - v))$ is TP_2 in (y, x) and also in (x, v). Indicator functions are TP_2. Products of TP_2 functions are also TP_2. Hence the result follows. $\square$

Claim 4. *Both $\frac{1 - H_p(0)}{h_p(0)}$ and $\frac{1 - H_c(p - c)}{h_p(0)}$ are non-increasing in p.*

Proof. The first result is shown in Rhodes and Zhou (2024). For the second, notice that

$$\frac{1 - H_c(p - c)}{h_p(0)} = \frac{\int_p^{\bar{v}} G(v - p + c)dF(v)}{G(p)f(p) + \int_p^{\bar{v}} g(v)dF(v)} , \quad (34)$$

where $G(v) = F(v)^{n-1}$ and $g(v) = (n-1)F(v)^{n-2}f(v)$. One can show that (34) is decreasing in p if and only if

$$[G(p)f(p) + \int_p^{\bar{v}} g(v)dF(v)][G(c)f(p) + \int_p^{\bar{v}} g(v - p + c)dF(v)]$$

$$+ G(p)f'(p)\int_p^{\bar{v}} G(v - p + c)dF(v) \geq 0 .$$

46

It is immediate that this condition holds if $p \leq \underline{v}$. In the remainder of the proof suppose $p > \underline{v}$, in which case dividing through by $G(p)f(p)$ yields

$$\frac{G(p)f(p) + \int_p^{\bar{v}} g(v)dF(v)}{G(p)f(p)}[G(c)f(p) + \int_p^{\bar{v}} g(v - p + c)dF(v)]$$
$$+ \frac{f'(p)}{f(p)} \int_p^{\bar{v}} G(v - p + c)dF(v) \geq 0 .$$

When f is log-concave, f'/f is decreasing. Hence a sufficient condition for the above inequality to hold is

$$\frac{G(p)f(p) + \int_p^{\bar{v}} g(v)dF(v)}{G(p)f(p)}[G(c)f(p) + \int_p^{\bar{v}} g(v - p + c)dF(v)]$$
$$+ \int_p^{\bar{v}} G(v - p + c)f'(v)dv \geq 0 .$$

Applying integration by parts to the last term, and noticing that the first fraction term is greater than 1, it is straightforward to show that this inequality is satisfied. □

Proof of Lemma 4. This follows immediately from Claims 3 and 4, because the right-hand side of equation (19) is increasing in σ. □

Proofs for Section 3.3

Proof of Lemma 5. All the results are straightforward to see except for the impact of σ on $p_a(\theta)$. From (24) and (27) and using the fact $q_a(\hat{\theta}) = 0$, one can show that

$$\frac{dp_a(\theta)}{d\sigma} = \theta \frac{dq_a(\theta)}{d\sigma} - \int_{\hat{\theta}}^{\theta} \frac{dq_a(t)}{d\sigma} dt , \qquad (35)$$

where, using equation (29), we have

$$\frac{dq_a(\theta)}{d\sigma} = -\frac{1}{(1-\sigma)^2} \frac{1 - F(\theta)}{f(\theta)} \frac{1}{c''(q_a(\theta))}. \qquad (36)$$

At $\theta = \hat{\theta}$, (35) must be negative; on the other hand, $q_a(\bar{\theta})$ is efficient and so independent of σ, and therefore at $\theta = \bar{\theta}$ (35) must be positive. One can further show that

$$\frac{1}{\theta} \frac{dp'_a(\theta)}{d\sigma} = \frac{dq'_a(\theta)}{d\sigma} \propto -\left(\frac{1 - F(\theta)}{f(\theta)}\right)' \frac{1}{c''(q_a(\theta))} + \frac{1 - F(\theta)}{f(\theta)} \frac{c'''(q_a(\theta))q'_a(\theta)}{c''(q_a(\theta))^2} . \qquad (37)$$

Given the log-concavity of $1 - F$ and convexity of $c(q)$, this must be positive if $c'''(q) \geq 0$. In that case, we have the stated cutoff result regarding how σ affects prices offered to anonymous consumers. □

References

ACEMOGLU, D., A. MAKHDOUMI, A. MALEKIAN, AND A. OZDAGLAR (2022): "Too Much Data: Prices and Inefficiencies in Data Markets," *American Economic Journal: Microeconomics*, 14(4), 218–56.

ACQUISTI, A., C. TAYLOR, AND L. WAGMAN (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54(2), 442–92.

ALI, N., G. LEWIS, AND S. VASSERMAN (2023): "Voluntary Disclosure and Personalized Pricing," *Review of Economic Studies*, 90(2).

ANDERSON, S., A. BAIK, AND N. LARSON (2023): "Price Discrimination in the Information Age: Prices, Poaching, and Privacy with Personalized Targeted Discounts," *Review of Economic Studies*, 90(5).

ANDERSON, S., AND R. RENAULT (1999): "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *RAND Journal of Economics*, 30(4), 719–735.

————— (2000): "Consumer Information and Firm Pricing: Negative Externalities from Improved Information," *International Economic Review*, 41(3), 721–742.

ARGENZIANO, R., AND A. BONATTI (2023): "Data Linkages and Privacy Regulation," mimeo.

ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *RAND Journal of Economics*, 32(4), 579–605.

ARMSTRONG, M., AND J. ZHOU (2011): "Paying for Prominence," *Economic Journal*, 121(556), F368–F395.

ATHEY, S., C. CATALINI, AND C. TUCKER (2017): "The Digital Privacy Paradox: Small Money, Small Costs, Small Talk," NBER working paper.

BELLEFLAMME, P., AND W. VERGOTE (2016): "Monopoly Price Discrimination and Privacy: The Hidden Cost of Hiding," *Economics Letters*, 149, 141–144.

BERGEMANN, D., AND A. BONATTI (2019): "Markets for Information: An Introduction," *Annual Review of Economics*, 11(1), 85–107.

————— (2024): "Data, Competition, and Digital Platforms," *American Economic Review*, 114(8), 2553–95.

BERGEMANN, D., A. BONATTI, AND T. GAN (2022): "The Economics of Social Data," *RAND Journal of Economics*, 53(2), 263–296.

CALVANO, E., G. CALZOLARI, V. DENICOLÓ, AND S. PASTORELLO (2023): "Artificial Intelligence, Algorithmic Recommendations and Competition," mimeo.

CDEI (2020): "Online targeting: Final report and recommendations," Discussion paper, Centre for Data Ethics and Innovation.

CHEN, Z. (2025): "Paying Consumers for Their Data: An Economic Analysis of Data Acquisition and Digital Privacy," mimeo.

CHEN, Z., C. CHOE, AND N. MATSUSHIMA (2020): "Competitive Personalized Pricing," *Management Science*, 66(9), 4003–4023.

CHOI, J. P., D.-S. JEON, AND B.-C. 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.

DE CORNIÈRE, A., AND R. DE NIJS (2016): "Online advertising and privacy," *RAND Journal of Economics*, 47(1), 48–72.

DE CORNIÈRE, A., AND G. TAYLOR (2019): "A model of biased intermediation," *RAND Journal of Economics*, 50(4), 854–882.

DELOITTE (2018): "Consumer Experience in the Retail Renaissance: How Leading Brands Build a Bedrock with Data," Deloitte Digital.

DIAMOND, P. (1971): "A Model of Price Adjustment," *Journal of Economic Theory*, 3(2), 156–168.

DONNELLY, R., A. KANODIA, AND I. MOROZOV (2024): "Welfare Effects of Personalized Rankings," *Marketing Science*, 43(1), 92–113.

DOVAL, L., AND V. SKRETA (2023): "Purchase History and Product Personalization," *RAND Journal of Economics*, forthcoming.

FAINMESSER, I., A. GALEOTTI, AND R. MOMOT (2023): "Digital Privacy," *Management Science*, 69(6), 3157–3173.

FUDENBERG, D., AND J. M. VILLAS-BOAS (2007): "Behavior-Based Price Discrimina-

tion and Customer Recognition," in *Handbook on Economics and Information Systems (Vol. 1)*, ed. by T. Hendershott. Elsevier.

GALPERTI, S., A. LEVKUN, AND J. PEREGO (2023): "The Value of Data Records," *Review of Economic Studies*, 90(5), 2674.

GALPERTI, S., T. LIU, AND J. PEREGO (2024): "Competitive Markets for Personal Data," mimeo.

GOLDFARB, A., AND C. TUCKER (2012): "Shifts in Privacy Concerns," *American Economic Review*, 102(3), 349–53.

HIDIR, S., AND N. VELLODI (2021): "Privacy, Personalization, and Price Discrimination," *Journal of the European Economic Association*, 19(2), 1342–1363.

ICHIHASHI, S. (2020): "Online Privacy and Information Disclosure by Consumers," *American Economic Review*, 110(2), 569–595.

——— (2021): "The Economics of Data Externalities," *Journal of Economic Theory*, 196, 105316.

INDERST, R., AND M. OTTAVIANI (2012): "Competition through Commissions and Kickbacks," *American Economic Review*, 102(2), 780–809.

KARLIN, S. (1968): *Total Positivity, Vol. I.* Stanford University Press.

KRÄHMER, D., AND R. STRAUSZ (2023): "Optimal Nonlinear Pricing with Data-Sensitive Consumers," *American Economic Journal: Microeconomics*, 15(2), 80–108.

KUMMER, M., AND P. SCHULTE (2019): "When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications," *Management Science*, 65(8), 3470–3494.

LIN, T. (2022): "Valuing Intrinsic and Instrumental Preferences for Privacy," *Marketing Science*, 41(4), 663–681.

MIKLÓS-THAL, J., A. GOLDFARB, A. HAVIV, AND C. TUCKER (2024): "Frontiers: Digital Hermits," *Marketing Science*, 43(4), 697–708.

MINAEV, A. (2021): "Consumer Data and Consumer Welfare: Evidence from the Hotel Booking Market," mimeo.

MONTES, R., W. SAND-ZANTMAN, AND T. VALLETTI (2019): "The Value of Personal Information in Online Markets with Endogenous Privacy," *Management Science*, 65(3),

1342–1362.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

OECD (2018): "Personalised Pricing in the Digital Era," Discussion paper, Organisation for Economic Co-operation and Development.

Prince, J. T., and S. Wallsten (2022): "How much is privacy worth around the world and across platforms?," *Journal of Economics & Management Strategy*, 31(4), 841–861.

Rhodes, A., and J. Zhou (2019): "Consumer Search and Retail Market Structure," *Management Science*, 65(6), 2607–2623.

Rhodes, A., and J. Zhou (2024): "Personalized Pricing and Competition," *American Economic Review*, 114(7), 2141–70.

Rochet, J.-C., and L. Stole (2002): "Nonlinear Pricing with Random Participation," *Review of Economic Studies*, 69(1), 277–311.

Stucke, M. E., and A. Ezrachi (2017): "How Digital Assistants Can Harm Our Economy, Privacy, and Democracy," *Berkeley Technology Law Journal*, 32(3), 1239–1300.

Tang, H. (2019): "The Value of Privacy: Evidence from Online Borrowers," mimeo.

Teh, T., and J. Wright (2022): "Intermediation and Steering: Competition in Prices and Commissions," *American Economic Journal: Microeconomics*, 14(2), 281–321.

Thisse, J.-F., and X. Vives (1988): "On The Strategic Choice of Spatial Price Policy," *American Economic Review*, 78(1), 122–137.

Vaidya, U. (2023): "Regulating Disclosure: The Value of Discretion," mimeo.

Which? (2018): "Control, Alt or Delete? The Future of Consumer Data," Policy Report.

Wolinsky, A. (1986): "True Monopolistic Competition as a Result of Imperfect Information," *Quarterly Journal of Economics*, 101(3), 493–511.

Yang, H., and L. Ye (2008): "Nonlinear Pricing, Market Coverage, and Competition," *Theoretical Economics*, 3(1), 123–153.

Zhou, W., M. Lin, M. Xiao, and L. Fang (2023): "Higher Precision is Not Always Better: Search Algorithm and Consumer Engagement," mimeo.