The Platform Dimension of Digital Privacy

Alessandro Bonatti*

NBER Tutorial on the Economics of Privacy

Draft – February 3, 2023

Abstract

Recent advances in economic theory examine the practices of large digital platforms in collecting data about individual users and monetizing it by selling targeted advertising. A *platform* dimension of digital privacy arises, where the behavior of *all* users and advertisers influences the amount of information available about each individual user. The acquisition of information by platforms is facilitated by the data externalities arising from the correlation in different users' preferences. Balancing consumer privacy and product-market competition is challenging, as platforms use their data to both improve match quality and increase advertisers' market power. These findings highlight the complex relationship between privacy, competition, and regulation.

*Sloan School of Management, MIT, Cambridge, MA 02142, bonatti@mit.edu. I would like to thank Dirk Bergemann for many years of joint work on this topic and the participants in the 2022 NBER Privacy Tutorial for helpful comments.

1 Introduction

The past two decades have witnessed the collection and diffusion of individual-level data at an unprecedented scale. At a broad level, large digital platforms such as Amazon, Facebook, Google, Alibaba, JD, and Tencent collect users' data through their engagement with the platform (often through nominally free services). They then monetize the data by matching users to advertisers, merchants, and content producers, i.e., by selling access to qualified consumers' eyeballs. This has raised concerns by academics and policymakers alike over the implications for individual privacy. These concerns have eventually led to regulatory interventions such as the GDPR and the CPRA, to name but two examples.

At a closer look, the problem of protecting individual privacy in today's digital markets takes on a new dimension, where both the equilibrium level of privacy and its welfare consequences depend on the mechanisms through which two-sided platforms mediate the exchange of consumer data. In particular, the network effects from many users (on one side) and advertisers (on the other side) and the potential for competition among platforms themselves jointly determine the scale and granularity of consumer data intermediation.

In this paper, I review recent advances in economic theory that uncover the platform dimension of digital privacy, investigate potential sources of market failure, and suggest open areas for future research. The economic theory of privacy is decades old, beginning with the classic work of Stigler (1980) and Posner (1981), and more recently surveyed in the comprehensive work of Acquisti, Taylor, and Wagman (2016). However, the platform dimension of privacy and the dual role of digital platforms as gatekeepers of information and competition (Bergemann and Bonatti, 2022) introduce new challenges and require new modeling tools.[1]

As a consequence of the platform dimension, privacy has become a social issue, a competition issue, and a regulation issue. Throughout the paper, I focus on three separate questions: (1) How do different consumers' choices of privacy interact with one another? (2) Is there a tradeoff between privacy and competition? In other words, does keeping consumer data private also result in limited competition for the consumer? (3) How do regulatory interventions help, and how can they backfire?[2]

I argue that data acquisition by the platform is critically facilitated by data externaliti-

[1]Enormous amounts of attention have been devoted to privacy in several fields, including law, political science, and computer science. A common theme is that improvements in information and communications technology facilitate individual-level data collection and naturally introduce concerns. These concerns are not limited to big tech datasets and market power but extend to the role, for example, of government tracking and surveillance. The analysis in this chapter is highly specialized and complementary to those perspectives.

[2]The regulation dimension of privacy is examined by Johnson (2022), and I refer the reader to that paper for an in-depth analysis of the GDPR.

esthe effect that other consumers' data have on an individual user's decision to share their own data. When consumers' traits are positively correlated, I show conditions under which very little stands in the way of a large platform collecting vast amounts of individual data. This is true even if consumers had full control over their privacy because the marginal cost of acquiring each individual user's data is small relative to the aggregate value of a dataset. I also discuss whether, due to the very nature of information goods, competition by platforms for acquiring user data is likely to yield large welfare gains.

Turning to data monetization, I show how a digital platform with market power is able to transfer that power downstream to advertisers by selling monopoly (exclusive) access to consumers. In particular, the platform leverages its data from past and concurrent transactions to create surplus through better matching of consumers and sellers. At the same time, the possibility of awarding "de facto monopoly positions" (Cremèr, de Montjoye, and Schweitzer, 2019) to advertisers limits the diffusion of consumer data (which may be viewed as privacy protecting) but opens the door to exploitation through surplus extraction through personalized offers (i.e., price discrimination and product steering).[3] The resolution of this trade-off by a monopolist seller is then critical to understanding the welfare implications of market power by digital platforms and hence the relationship between privacy and competition for the consumer.

The two aspects of data intermediation interact. In particular, the growth of a platform's database (through the participation of more users) influences its ability to match products to tastes but also reduces each consumer's outside optiona new form of data externality. In this sense, the contribution of the data-acquisition model is that consumers' privacy choices interact, even if regulation such as the GDPR and CPRA intends to assign formal control rights over data to individual users. The contribution of the second model is that the optimal mechanisms for monetizing data put privacy at odds with competition for the consumer.

The overall picture that emerges is one where data externalities lead to economies of scale on the data-acquisition side, and market power on the monetization side leads to the sale of exclusive access to each consumer's attention. Under these conditions, the signs of data and participation externalities ultimately depend on the type of firms that gain access to the consumer—does this firm use information primarily to create or to extract value?

To address these questions, Section 2 introduces a model of a two-sided platform as a monopolist data intermediary and examines the economics of privacy through this lens; Section 3 focuses on the data-acquisition side; and Section 4 examines the data-monetization side. Section 5 concludes and suggests open areas for future research.

[3] The availability of granular individual-level data can, of course, introduce other concerns, including government surveillance, data leakages, fraud, misinformation campaigns, and addictive social media.

2 Basic Framework

The basic role of any digital platform is to intermediate large numbers of users and producers. Here, we focus on some key dimensions of the platform data intermediation. First, any information it acquires must be obtained from multiple users. Second, any data it has acquired can be monetized through multiple producers or firms of merchants. Third, consumers and producers may have outside access opportunities or the ability to meet off the platform. As we will see, a critical determinant of the platform's bargaining power is whether it is instrumental for a match between consumers and producers, or merely enables this match to occur under better complete information. Figure 1 illustrates the basic idea.

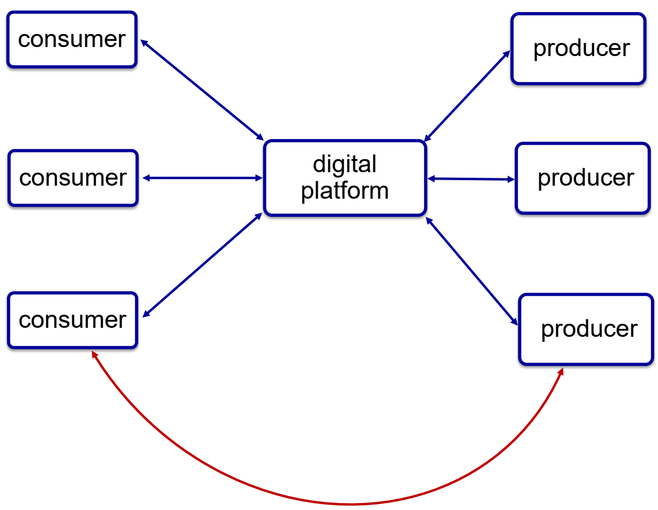

FIGURE 1: Platform Interaction with a Direct Channel

2.1 Value of Privacy

We begin this section with a simple framework to think about consumer privacy as private information about preferences. We develop a first model where a representative consumer interacts with a single producer (or "firm"). We later augment the framework by introducing multiple users, multiple producers, and potentially other agents (e.g., governments or platforms) interested in learning about the consumer.

The consumer has a preference type $\theta \in \Theta \subset \mathbb{R}$ that parametrizes their utility function. When the firm chooses action a, the consumer obtains utility

$$u(\theta, a).$$

For simplicity, we assume here that the firm chooses the action a (e.g., advertising, message, product, or price) to match consumer type:

$$a^* = \mathbb{E}[\theta].$$

The consumer knows their true type θ. The firm knows the prior distribution $F_0(\theta)$ of the consumer's type. In addition, the firm receives an informative signal $s \in S$. The signal structure induces a *segmentation* as in Yang (2022). Here we follow the exposition in Bonatti and Villas-Boas (2022). A segmentation

$$\mathcal{S} = \{(\pi_s, F_s)\}_{s \in S}$$

is a mixture distribution with weights π_s over individual distributions F_s that integrates to the prior, i.e.,

$$\int_s F_s(\theta)\pi_s \mathrm{d}s = F_0(\theta), \quad \forall \theta \in \Theta.$$

In particular, any segmentation $\mathcal{S}$ is a mean-preserving spread of the prior distribution F_0.

Any segmentation $\mathcal{S}$ admits two equivalent interpretations. In particular, π_s denotes probability of the signal realization s, and F_s denotes the distribution of the firm's posterior beliefs upon observing s. Equivalently, the signal structure induces a partition of the consumer types (i.e., a market segmentation) where the measure of each segment is given by π_s and the composition of each segment is given by $F_s(\theta)$.

Because all consumers in segment s (i.e., conditional on the firm observing signal s) receive the same action, we can write their aggregate surplus as

$$V(F_s) = \int_\theta u(\theta, a^*(F_s))\mathrm{d}F_s(\theta).$$

Simply averaging over segments (i.e., signal realizations) yields the expected (ex ante) consumer surplus under segmentation $\mathcal{S}$ as

$$U(\mathcal{S}) \triangleq \mathbb{E}_s[V(F_s)] = \int_s V(F_s)\pi_s \mathrm{d}s. \tag{1}$$

For example, the expected consumer surplus under prior information (i.e., full privacy) is given by

$$U(\varnothing) \triangleq V(F_0).$$

This formulation for consumer surplus allows for an immediate characterization of utility functions for which consumers unambiguously (i.e., for all segmentations) like or hate privacy.

Proposition 1 (Value of Privacy)**.**
If $V(\cdot)$ is concave (convex), consumers like (dislike) privacy.

The proof (which is a simple application of Jensen's inequality) and more details are in Bonatti and Villas-Boas (2022). Under the conditions of Proposition 1, the consumer's ideal segmentation is $\mathcal{S} \in \{\varnothing, \mathcal{S}^*\}$, where $\mathcal{S}^*$ is the full information segmentation consisting of a collection of degenerate random variables ($s = \theta$). We now illustrate the usefulness of this compact representation for the value of privacy through a parametrized example.

2.2 Application

Let the consumer's utility function be given by

$$u(\theta, a) = (\theta + \lambda a)^2$$

where $\lambda \in [-1, 1]$ is a parameter intuitively capturing the value creation vs. extraction role of the firm's action a, e.g., product quality vs. price as in Argenziano and Bonatti (2021). The case of $\lambda = -1/2$ is outcome-equivalent to the case of linear price discrimination, where consumer type θ facing a unit price of p obtains the following indirect utility. In that case, we obtain

$$u(\theta, p) = \max_q \{\theta q - pq - q^2/2\} = (\theta - p)^2/2 \text{ and}$$
$$p^* = \arg\max_p \{p(\theta - p)\} = \mathbb{E}[\theta]/2,$$

which are captured in Figure 2.

To illustrate how the basic model recovers the classic result on the welfare consequences of linear price discrimination, consider the surplus of segment s

$$V(F_s) = \int_\theta (\theta + \lambda\, \mathbb{E}_{F_s}[\theta])^2 \, \mathrm{d}F_s(\theta),$$

which we can write as

$$V(F_s) = \mathbb{E}_{F_s}[\theta^2] + (2 + \lambda)\lambda \left(\mathbb{E}_{F_s}[\theta]\right)^2.$$

Notice that the first term is linear in probabilities, while the second term is convex. Because $\lambda \in [-1, 1]$, we immediately conclude that $V(\cdot)$ is a concave (convex) function if and only if $\lambda < (>)0$. Therefore, if $\lambda < (>)0$, any mean-preserving spread hurts (helps) consumers. In particular, for the fully-informative segmentation $\mathcal{S}^*$, we have $U(\mathcal{S}^*) < (>)U(\varnothing)$.

In particular, if (as in Figure 2), we have $\lambda = -1/2$, we immediately know the provision

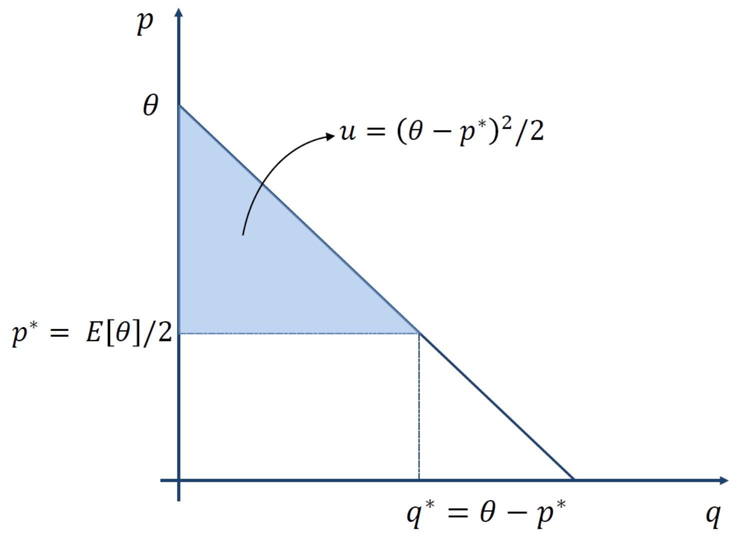

Figure 2: Linear Price Discrimination

of information to a price-discriminating monopolist facing linear demand (and full market coverage) is detrimental to consumer surplus (Robinson, 1933; Schmalensee, 1981).

The model presented in this section is stylized along several dimensions. The general effect of market segmentations and the achievable combinations of consumer and producer surplus are analyzed in the seminar work of Bergemann, Brooks, and Morris (2015) and more recently by Haghpanah and Siegel (2022) and Elliott, Galeotti, and Koh (2022). The consumer's type was assumed one-dimensional, but Ichihashi (2020) and Bonatti and Villas-Boas (2022) illustrate how the main logic of Proposition 1 extends to multidimensional environments such as those, for example, where the consumer has both a vertical willingness to pay attribute and a horizontal product match attribute.

Finally, the consumer was assumed entirely passive, whereas a long literature (summarized in Section 3.2 below) studied the impact of consumer actions on the equilibrium market segmentations. In the remainder of this chapter, we explore the conditions under which a platform can profitably intermediate the exchange of data in markets where consumers like (or dislike) privacy.

3 Data Acquisition

The main message of Section 2 is that we have a language to talk about the availability of information and a consumer's preferences over the amount of data that a platform holds about them. We now focus on the key dimensions of the platform dimension of privacy, namely the collection and the monetization of consumer data, beginning with the former.

Why do consumers allow platforms to collect significant amounts of data? One possibility is that consumers benefit from data collection and that data intermediation is socially efficient. Another possibility is that consumers are unaware of the extent of data collection, or that their stated preferences for privacy differ from their actual preferences—the *privacy paradox* (Athey, Catalini, and Tucker, 2017). In this Section, we specifically ask why platforms are able to intermediate information at little or no cost, why competition does not seem to discipline data acquisition, and whether there are limits to consumer data usage that emerge in a market context—for example, why do we little or no personalized pricing?

3.1 Captive Consumers

Consider a single consumer and a single producer who meet on a monopolist digital platform with no alternative means to contract with each other. Figure 3 simplifies Figure 1 as follows.

FIGURE 3: "Captive" Consumer and Producer

Assume that the consumer makes a one-time participation decision. This decision takes place ex-ante, i.e., before the consumer's type is drawn. If the consumer participates on the platform, which means it uses the platform repeatedly, then it is going to reveal segmentation $\mathcal{S}$ to the platform, which observes a signal realization s and transfers it to the producer. We are going to remain agnostic as to how this transfer occurs—whether the data is effectively sold to the producer, or the producer is merely able to learn something about the consumer when it interacts with them on the digital platform. At this level of abstraction, data intermediation is equivalent to buying data $\mathcal{S}$ from the consumer and reselling it to the producer. With one platform and one producer, it is also immediate to show that the platform will charge the producer their entire willingness to pay to access the consumer's information. Therefore, we now focus on implications for consumers.

If the consumer participates, their ex-ante surplus, aggregating across both signals and types, is going to be given by $U(\mathcal{S})$ as in the previous section. if participating and zero otherwise.[4] Why is the consumer surplus nil if they do not participate? Because in this

[4]The use of the indirect utility function $U(\cdot)$ here underscores that the value of privacy to the consumer depends on the nature of the producer's actions a and on the underlying interaction $u(\theta, a)$. This is an important departure from philosophy and legal approaches to privacy. Unlike in Zuboff (2019), data collection

setting the platform is necessary for the consumer. For example, the platform lowers search costs, offers valuable independent services and matches of higher quality. At this stage if

$$U(\mathcal{S}) \geq 0,$$

the consumer participates. In addition, if

$$U(\mathcal{S}) \geq \max\{0, U(\varnothing)\}, \tag{2}$$

data intermediation yields a Pareto improvement: the consumer gains from interacting with the platform, and so does the producer. However, the more challenging case is one in which

$$U(\varnothing) > U(\mathcal{S}) > 0. \tag{3}$$

The consumer still finds it profitable to join the platform but loses relative to the case of anonymous transactions. This observation has prompted many scholars to refer to privacy loss as an unobserved price of accessing a digital platform. This occurs when the platform's services are nominally free, but consumers pay with their data.

Under these conditions, it was extremely easy for a platform to acquire the consumer's data. Let us now make the platform's problem more realistic (and a little harder) by allowing consumers and producers to meet off-platform.

3.2 Consumer Consent

Suppose now, as in Figure 4, that the consumer can choose whether to "consent" and reveal information to the platform, or to deny consent and remain anonymous. If they do not reveal information, they can still interact with the producer in an anonymous transaction (for example, because they can visit the producer's own website). This is akin to consent requirements in recent legislative efforts aimed at protecting consumer privacy.

In this model, absent any form of compensation, the consumer agrees to reveal their information if and only if they dislike privacy. When consumers have a positive value of privacy, as in (3), the platform must compensate consumers to reveal their information. While direct monetary payments are quite rare, compensation can occur through better quality services and matches.

To quantify those payments, let us maintain the assumption that the platform is a monopolist facing a single producer. Thus, the platform can extract the producer's entire value

makes no first-order difference to a consumer unless of course privacy enters utility function (which may well be the correct behavioral assumption).

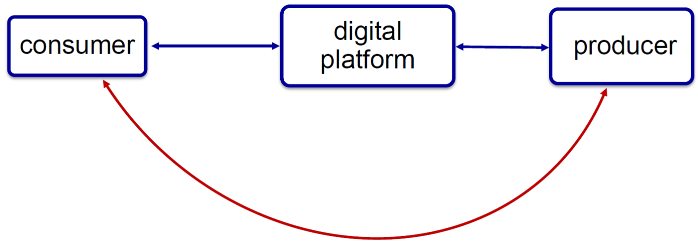

Figure 4: Consumer and Producer with Alternative Channel

of information downstream. This is the setting that has prompted many scholars to appeal to the Coase Theorem (Coase, 1960) and argue that the simple assignment of property rights over data is going to yield the efficient level of information intermediation. The idea is simple and appealing: say the consumer owns the rights to their data and can sell them to the platform. In turn, the platform sells the consumer data to the producer. The three parties will be able to agree on the terms of trade—a price paid by the platform to the consumer and a price paid by the producer to the platform—if and only if the transfer of data from consumer to producer increases total surplus. In other words, if the loss in consumer privacy is worth more than the value of the information for the producer, then the platform will not be able to profitably intermediate this transaction.[5] This suggests that under well-specified property rights, the only trades of data that take place are those of equation (2).

In practice, however, there are at least two problems with the efficiency of the market for consumer information. The first problem is moral hazard: consumers do not reveal their information directly, e.g., by uploading spreadsheets with all their purchase data to an online retail platform. Instead, consumer reveal information through their online (and sometimes offline) behavior. The nature of data usage is critical for the trade of information in this setting. For instance, if consumers know their data will be used to set prices or steer their searches toward more expensive products, they have an incentive to distort their behavior. Such manipulation incentives may both bias and confound the information collected by the platform, thereby reducing its value to the producers.

These forces were first uncovered in the literature on behavior-based price discrimination and ratchet effects. The classic papers by Taylor (2004), Villas-Boas (2004), Acquisti and Varian (2005), and Calzolari and Pavan (2006) allow consumers to take actions (e.g., the level of purchases) at two different times in order to manipulate the second-period firm behavior. More recently, Bonatti and Cisternas (2020) show that the applicability of these models goes

[5]The 2020 California Privacy Rights Act also implicitly appeals to the Coase Theorem: consumers who opt out of data sharing have a *right to equal service and price*, but firms can "offer a different price, rate, level, or quality of goods or services to the consumer if that price or difference is reasonably related to the value provided to the business by the consumer's data."

beyond business to consumer relationships. For example it can be used to shed light on B2B price discrimination.[6] While business privacy is not typically an object of study, many of the same trade-offs face businesses and consumers who are aware of data collection. Argenziano and Bonatti (2021) study how consent regulation and other forms of property rights over data impact the level of trade and welfare in a signaling model.

The second problem is due to externalities, which we explore at length below.

3.3 Social Data

Unlike in the model discussed so far, the consumer is not alone. As many consumers make the decision as to whether to participate in the platform simultaneously, a central dimension of information intermediation is its social aspect. The social aspect of information refers to the correlation in the underlying traits of consumers who join the same platform. Their decisions interact with one another, not directly, but indirectly through the correlation structure of their types. This may lead to a market failure because the social nature of data generates a data externality—the phenomenon that some consumers' data reveal information about other consumers. Data externalities do not have an a priori sign like carbon emissions or vaccinations. For example, if my data is used to offer better products to others, then I impose a positive externality on them; but if others' data is used to steer me towards expensive products instead, others impose a negative externality on me.

A recent and growing literature has shown how data externalities can reduce the cost of acquiring information from consumers—see for example Choi, Jeon, and Kim (2019), Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022), and Ichihashi (2021b). The core idea is the following: when there are many consumers, even if the aggregate effect of revealing all their data might be large and negative for the surplus of any individual, the marginal impact of a single consumer's decision to participate on a digital platform is small. In the language of our basic framework, even if consumer i chooses not to participate on the platform, the producer will have now have access to a potentially very informative segmentation $\mathcal{S}_{-i}$. Figure 3.3 illustrates this scenario.

To formalize this intuition, we follow Bergemann, Bonatti, and Gan (2022), who develop a model of monopolistic data intermediation with $i = 1, \ldots, N$ consumers. In their setting, as in the previous section, a platform can compensate each consumer for their own data, which it then resells to a single producer.

Suppose platform offers t_i to each consumer i for access to (data leading to) a seg-

[6] For example, "Google induced advertisers to bid their true value, only to override pre-set AdX floors and [...] generate unique and custom per-buyer floors depending on what a buyer had bid in the past." (Texas vs. Google).

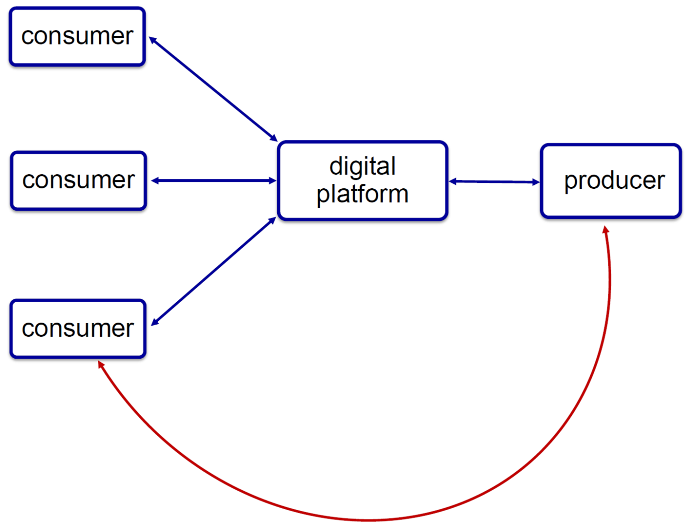

FIGURE 5: Many Consumers with Competing Channel

mentation $\mathcal{S}_i$ of i's type. Denote by $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_N)$ the segmentation induced by every consumer's data. Consumer i makes a participation decision prior to learning their type. This consumer participates if and only if

$$t_i + U_i(\mathcal{S}) \geq U_i(\mathcal{S}_{-i}). \tag{4}$$

The interpretation of this participation constraint is that the transfer t_i must induce the consumer to prefer segmentation $\mathcal{S}$ to the alternative of withholding their data, in which case the platform collects and transmits segmentation $\mathcal{S}_{-i}$. We can then formally define a data externality as follows.

Definition 1 (Data Externality)**.**
The data externality imposed by consumers $-i$ on consumer i is given by

$$DE_i(\mathcal{S}) \triangleq U_i(\mathcal{S}_{-i}) - U_i(\varnothing).$$

Thus, the data externality captures the welfare effect for consumer i of all consumers $j \neq i$ revealing their data while i withholds theirs. We can then immediately put the data externality to work and obtain a characterization of profitable intermediation. Let $W_i(\mathcal{S})$ denote the total surplus (consumer welfare plus producer profits) generated by consumer i when the producer is endowed with segmentation $\mathcal{S}$, and define

$$\Delta W_i(\mathcal{S}) \triangleq W_i(\mathcal{S}) - W_i(\varnothing).$$

Bergemann, Bonatti, and Gan (2022) then show the following result.

Proposition 2 (Profitability of Intermediation)**.**
Intermediation of data $\mathcal{S}$ is profitable if and only if, for all i,

$$\Delta W_i(\mathcal{S}) - DE_i(\mathcal{S}) \geq 0.$$

Intuitively, there are two channels through which a platform can potentially profit from data intermediation. A classic channel is that of surplus creation, which operates when revealing information to the producer helps (or does not excessively hurt) consumers. In particular, the transmission of information may increase total surplus ($\Delta W_i > 0$), in which case data intermediation is both profitable for the platform and socially efficient. A more novel channel operates through the social dimension of the data: if individual consumers' decisions impose negative data externalities on other consumers ($DE_i < 0$), the platform can enlist additional consumers at lower marginal cost, thereby directly increasing its profits.

The latter scenario is more likely as the number of consumers increases. It is not hard to find conditions as in Figure 6 below, where consumer surplus decreases in the number of signals the platform procures, but it does so at a decreasing rate. Thus, a negative data externality combined with a diminishing marginal impact of each consumer's signal allow data intermediation to be both profitable and socially inefficient.

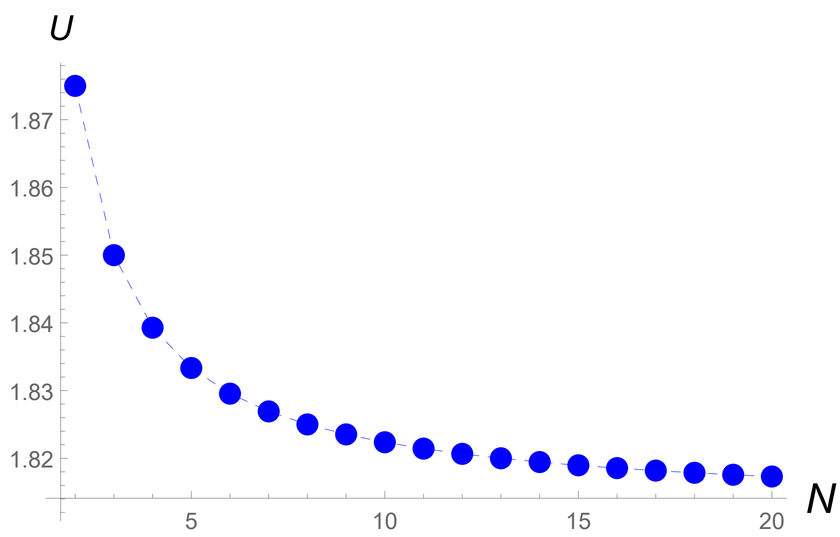

FIGURE 6: Consumer Surplus $U(\mathcal{S}^*(N))$

At this point, it may seem like this model predicts complete and unhinged data sharing. This is not always the case. Indeed, Bergemann, Bonatti, and Gan (2022) also show that the platform-optimal data sharing policy does not necessarily involve complete data sharing. In this sense, the nature of information qualifies the externality effect above and extends insights from the literature on contracting with externalities (Segal, 1999) to the case of social data. In particular, the platform finds it optimal to intermediate individual-level information when the data increases total surplus (e.g., in the case of customized product recommendations). Conversely, when this information is used for (socially inefficient) price discrimination, the platform aggregates the consumers' signals and intermediates market-level information.

To summarize, the platform-optimal data sharing policy involves socially efficient data anonymization decisions. Nonetheless, there are very few guarantees, if any, that the allocation of data is going to be socially efficient. After all, consumers are compensated for their individual harm, but not for the social harm that they create. Finally, as the markets grow large, which is a reasonable approximation for digital platforms, the cost of acquiring the information from consumer vanishes, while the gains persist.

The social aspect of the data relates to the digital privacy paradox, whereby consumers require negligible compensation to reveal their data, in contrast with their stated preferences.[7] These results have prompted several scholars, most notably in psychology, philosophy, and law, to refer to privacy as a collective issue or public good, because the effectiveness of the tools used to monetize and leverage our information depends on our collective choices. Most notably, Zuboff (2019) argues that "Privacy is not private, because the effectiveness of these and other private or public surveillance and control systems depends upon the pieces of ourselves that we give up."

3.4 Regulation and Competition

The potential market failures highlighted in this section naturally pose the question of the effectiveness of regulation. The discussion of data externalities above strongly suggests that individual-level regulation is unlikely to restore efficient outcomes in data collection.[8] A more promising market structure, without the aid of regulatory interventions, would be one where multiple platforms compete as in Rochet and Tirole (2003) for the (ideally exclusive) engagement of every consumer.

However, several recent papers have shown that the effect of competition is not at all straightforward, and that it is not hard to imagine realistic settings where platform competition does not lead to gains in consumer surplus. Most notably, Ichihashi (2021a) develops

[7]This result appears in the randomized control trial of Athey, Catalini, and Tucker (2017), and it was also true in a recent paper on the effects of the GDPR (Aridor, Che, and Salz, 2020). In that paper, a large number of users paid no attention whatsoever to cookies and privacy-enhancing techniques even prior to the regulation. This is consistent with, even though not causally related to, the privacy paradox.

[8]Viljoen (2021) emphasizes the relational aspect of digital markets whereby data creates value by enabling people to connect and the difficulties in regulating the nexus of links created by online data.

a model of competing data intermediaries that can acquire one or more data "units" from a single consumer. The key property of data is that it can be sold to any number of intermediaries at zero cost by the consumer. Furthermore, all copies of the data must be identical—there is no room for selling differentiated data products as in Admati and Pfleiderer (1986). Therefore, if multiple intermediaries hold the consumer's data, they compete away all profits. In this model, when revealing their data has a negative impact on consumer surplus, a single platform is able to make an offer to the consumer that leaves them exactly indifferent. In equilibrium, no other platform can then offer a positive price to the consumer for the data. Hence, the monopoly outcome obtains.

In complementary work, Casadesus-Masanell and Hervas-Drane (2015) offer an explanation for the shortcomings of competition, based on service quality; Loertscher and Marx (2020) provide an explanation for the emergence monopoly platforms based on data aggregation; and Prüfer and Schottmüller (2021) develop a dynamic model of "tipping" in data-rich industries that also supports the near-natural-monopoly theory.

Finally, even if competitive forces were strong, "privacy fixing" has emerged as a new anti competitive concern. The idea is that, instead of fixing prices (because they are constrained to be zero), competing platforms might agree to not preserve their users' privacy. For example, the 2020 Texas v. Google complaint claims that "of course, effective competition is concerned about both price and quality, and the fact that Google coordinates with its competitors on the quality metric of privacy—one might call it privacy fixing—underscores Googles selective promotion of privacy concerns only when doing so facilitates its efforts to exclude competition."[9]

4 Data Monetization

The mechanisms by which data is monetized are critical to understand the privacy implications of data intermediation. In this section, we consider a model where a platform has freely collected a single consumer's information, with the understanding that this is a metaphor for the equilibrium effect of data externalities. We also imagine that the platform can monetize this data by allowing any number of producers in a given industry to access the consumers' attention and target them with personalized offers.

Before turning to the privacy implications of such a market structure, let us think for a moment about how data should *not* be sold.

[9]The U.S. DOJ & FTC 2010 Horizontal Merger Guidelines make it clear that "when the Agencies investigate whether a merger may lead to a substantial lessening of non-price competition, they employ an approach analogous to that used to evaluate price competition."

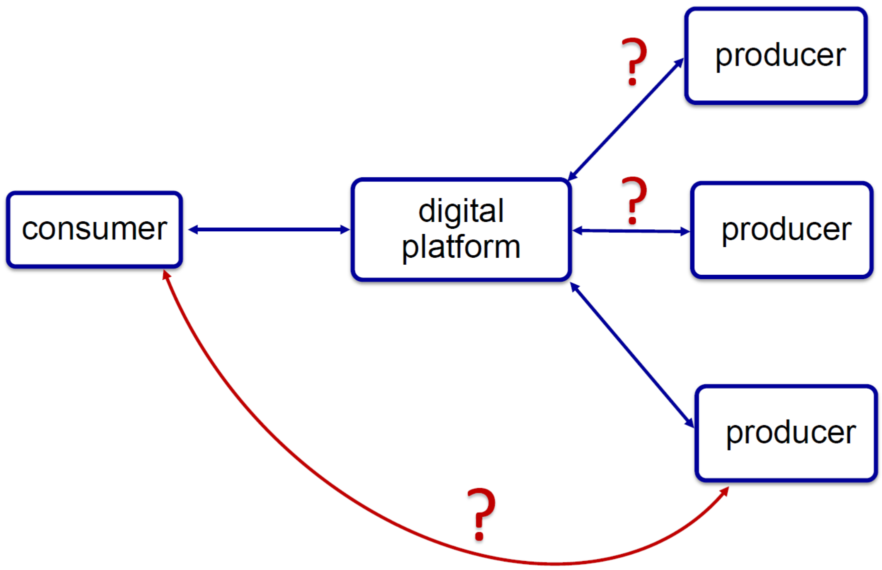

Figure 7: (Potentially) Competing Producers

4.1 Direct Sale of Information

In practice, digital platforms very rarely sell consumer data directly to advertisers and other parties. For one, the reputation backlash and the risk of leakages would be significant, but it is equally important to understand why this would be a suboptimal strategy even absent these concerns.

Indeed, there are at least five reasons why platforms would not want to sell data directly.

1. The first problem a platform would face when selling data directly would be that information about consumers' willingness to pay is likely to create negative externalities downstream: if two or more competitors are informed about the correct product or price level to offer, each one is forced to lower prices. In this world, relative to physical goods, exclusive sales tend to be more profitable as shown in the classic contribution by Admati and Pfleiderer (1986).

2. The second problem relates to data pricing under exclusive sales. Let us entertain the possibility that a digital platform sells individual level data to a single merchant only. The value of this information is a complicated equilibrium object, which depends on the complex game between one informed firm and its uninformed competitors (Bonatti, Dahleh, Horel, and Nouripour, 2022).

3. The third problem is a classic difficulty with selling information. would also arise. "Selling wine without bottles" is a famous metaphor (Barlow, 1994) that refers to the zero marginal cost of data reproduction, which might easily lead to a profitable resale

market for data (Shapiro and Varian, 1999; Jones and Tonetti, 2020). In other words, any data-selling platform creates its own competition by simply letting the data flow out of its own hands.

4. The fourth problem is that data about an individual consumer becomes obsolete over time, but not very quickly. Therefore a data seller is able to charge for the incremental information that they provide over and above the data buyer's initial information (Bergemann, Bonatti, and Smolin, 2018). In other words the platform can charge for the innovation component in the data, and not for the entire value of the dataset.

5. The fifth and fundamental problem relates to how to measure the causal impact of data sales. In practice, it is difficult to prove how much a data product is worth without giving away the information contained in the data itself. This is the famous *information paradox* pointed out by Arrow (1962).

4.2 Indirect Sale of Information

While direct sales of information are problematic, targeted advertising is a superior, more profitable means to monetizing consumer information. Consider for example Google or Amazon search ads (or paid placement on Taobao.com). Advertisers buy a slot on a keyword-results page, which means they can tailor their message, the link they want to show, to the consumer's search query, which is informative of their underlying preferences. Of course, the search engine could sell data about those searches directly, but prefers to leverage the data to sell access to qualified eyeballs instead. This is a far better idea and a far larger market than data direct sales, which is entirely consistent with what economic theory would have predicted (Admati and Pfleiderer, 1990; Bergemann and Bonatti, 2019).

Indeed, selling access to consumers directly solves all five problems we mentioned above. It solves the data exclusivity problem by offering a scarce number of slots. It solves the problem of competition under asymmetric information structures because only a few informed parties access the consumer at one time. It solves the resale and rental problem by never really giving out the data. Finally, it solves the quality measurement problem because advertisers have a number of conversion metrics available to them. Thus, it is only by bundling qualified eyeballs and advertising space that a large digital platform is able to monetize the troves of data at its disposal.[10] With these foundations in mind, we want to understand the implications of selling exclusive access to consumers through targeted advertising space.

[10] Indirect sales of information in digital markets are not limited to search advertising platforms: the same advantages relative to direct sales apply to large display advertising networks such as Google, Meta, Criteo, and Microsoft, as well.

4.3 Mechanisms for Digital Advertising

Consider then a large digital platform that matches heterogeneous buyers and sellers, running individual level auctions for targeted advertising. A first treatment of this topic is in de Cornière and de Nijs (2016), who focus on bidding and unit pricing, and derive conditions under which the platform prefers targeting vs. a random allocation of slots. In what follows, we follow the more recent contribution of Bergemann and Bonatti (2022), who introduce the notion of a "managed campaign." Relative to that paper, we simplify the exposition by considering single-product sellers only.[11]

There are J sellers who offer horizontally differentiated products at no cost and a unit mass of consumers. Each consumer has a multidimensional type $\theta = (\theta_1, \ldots, \theta_j, \ldots, \theta_J) \in \mathbb{R}^J$. Each type component θ_j denotes the consumer's value for the product of firm j.

Independent of their type, a fraction $\lambda \in [0, 1]$ of these consumers use a platform that runs ads in order to find a seller. The remaining $1 - \lambda$ consumers buy directly from sellers, face search costs $\sigma > 0$ after the first free search as in Diamond (1971).

The platform observes all types θ while consumers have arbitrarily precise beliefs m about their valuations. The platform offers a single "sponsored" advertising slot per consumer. In allocating the slot, the consumer's type θ serves as a *targeting category:* the firms' ads can condition on the entire type.

More formally, the platform offers a *managed campaign* mechanism, which consists of the following. The platform charges a fixed fee t to participating sellers. (This can be viewed as a minimum mandatory campaign budget.). The platform specifies which seller j (among those who pay the fee) obtains the slot for which consumers θ. The platform reveals to the consumer—by means of additional information—her θ_j for the advertised product j. Finally, the platform enables each selected seller j to advertise a personalized price $p_j(\theta)$ to the consumer.

Simultaneously to making their participation and personalized pricing decisions on the platform, the sellers also set posted prices $\hat{p}_j$, which are intended for the (anonymous) off-platform consumer. The two sales channels (on- and off-platform) interact because on-platform consumers can also search and (if they find a lower price or better product) buy off-platform. This introduces a "showrooming constraint" as in Wang and Wright (2020) and Teh and Wright (2022) whereby each seller j must provide weakly greater utility to their on-platform consumers than their off-platform consumers. Figure 8 illustrates.

This is a very different model from a Varian (1980) model of sales. In that model, consumers can be distinguished into shoppers and loyal and derive their surplus from price

[11] See also Bergemann, Bonatti, and Wu (2023) for a comparison between the managed campaign model and targeted auctions for digital advertising with manual bidding.

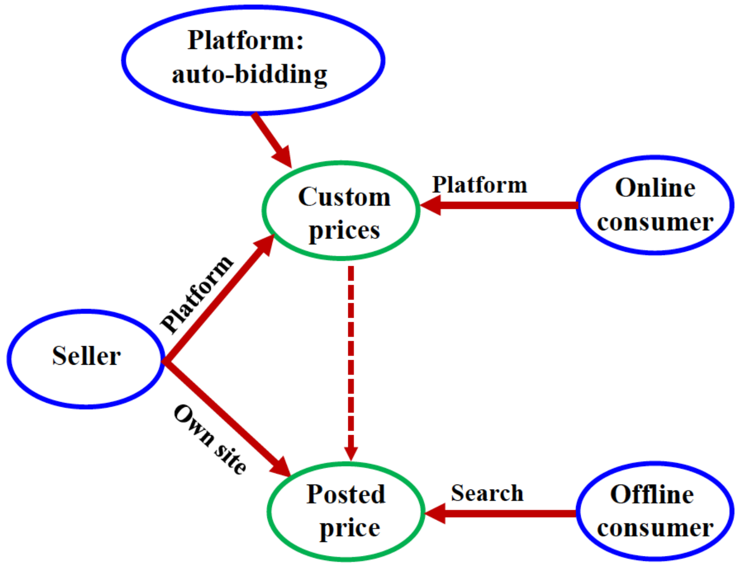

FIGURE 8: Model: Summary

competition for shopping consumers. Here the off platform sales channel provides the consumer's outside option. In equilibrium, we will see that consumers obtain any surplus only because they could act anonymously and leverage their own right to privacy, so to speak, in order to acquire a good from the seller's direct channel. More generally, the on-platform consumer's search behavior depends on the criteria by which the platform assigns a sponsored link. Bergemann and Bonatti (2022) establish the following intuitive result, which has immediate implications for the equilibrium search patterns.

Proposition 3 (Optimal Matching Mechanism).
The platform maximizes revenues by matching each consumer θ to most their favorite seller $j^* = \arg\max_j \theta_j$ among those who participate in the managed campaign mechanism.

Under this matching mechanism, the platform fully exploits its informational advantage: the λ on-platform consumers infer that the displayed seller is $\theta_{j^*} = \max_j \theta_j$, and they cannot detect any deviations by non-participating sellers. Furthermore, by showrooming, these consumers expect symmetric prices off the platform. Consequently, Bergemann and Bonatti (2022) show that these consumers only consider offers by the advertised seller.

Proposition 4 (Consideration Sets). *Every online consumer θ only compares the displayed seller j^*'s personalized (on-platform) and posted (off-platform) prices, $p_{j^*}(\theta)$ and $\hat{p}_{j^*}$.*

Off the platform, consumers act as in the Diamond (1971) model. These $1-\lambda$ consumers with beliefs m face search costs $\sigma > 0$ after the first search; they expect symmetric prices

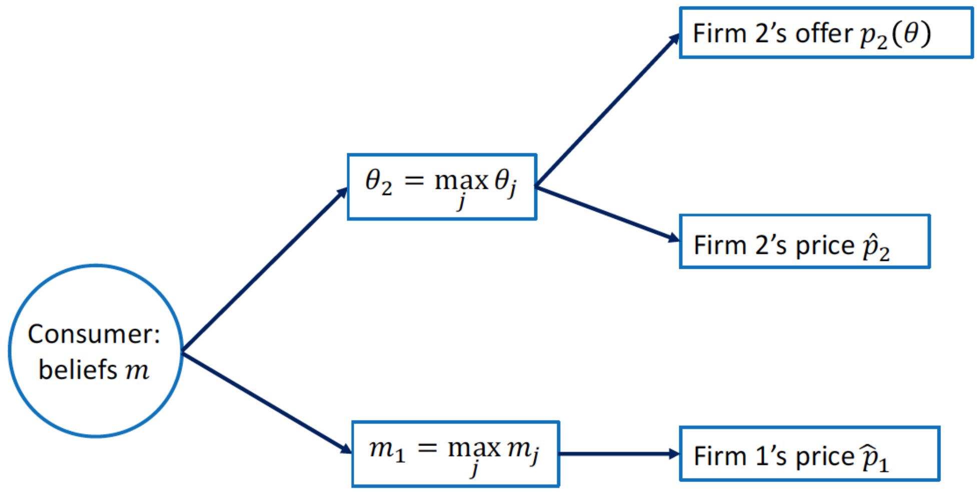

Figure 9: Search Patterns

and hence visit $\hat{j} = \arg\max_j m_j$ only. Figure 9 illustrates the search patterns of a consumer with beliefs m and true type θ both off platform and on platform.

The key result is that the platform is able to completely shield the most efficient producer from competition. After a link by the highest-value firm is shown to the consumer. The consumer infers that is indeed the highest value firm. If this consumer were to showroom, they would only visit that firm's website. Indeed, the model admits an equivalent interpretation wherein each brand has an identical fraction $(1-\lambda)/J$ of loyal, imperfectly informed consumers who are already shopping off of the platform. The remaining λ consumers are not currently shopping, but they can be alerted to the existence of a brand. Once they are alerted by an ad, they contemplate shopping either on or off the platform. The equivalence with this behavioral model requires arbitrarily small amounts of search costs and informational advantages by the platform: without an informational advantage, the platform will not be able to control the consumers' outside options because the consumer's own beliefs will determine where they search first off the platform.

Finally, let us look at the results from a welfare perspective. The platform sells prominence to the highest bidder. This enables trade under symmetric information and induces higher total surplus. In this sense it has a positive social effect.[12] The platform, however, also sells market power. Indeed, the firms never compete in price, which leads to higher prices both on and off the platform. This is mostly due to the platform's informational advantage, which narrows the consumers' search options. The growth of a platform's database

[12]Trading through the platform is inherently more efficient even if consumers know their types. This is because under symmetric information, the platform eliminates any distortions from uniform monopoly pricing—with personalized pricing, all consumers buy. See Hidir and Vellodi (2021) on the price discrimination vs. product matching tradeoff.

(through more consumers λ or, in some cases, better information) reduces outside options and leads to higher prices—a different kind data externality as pointed out in Kirpalani and Philippon (2021).

If, in addition, firms were heterogeneous in their cost function or in the number of on- versus off-platform consumers, the platform would introduce a further element of inefficiency. In particular, lower-quality brands with a smaller off-platform presence might be able to generate higher bids (or be willing to invest larger budgets), and their products might generate lower value for consumers. This scenario is qualitatively consistent with the evidence in Mustri, Adjerid, and Acquisti (2022).

4.4 Privacy and Competition

The results in the managed campaign model make apparent the privacy vs. competition tradeoff. With any indirect sale of data (such as digital advertising auctions and managed campaigns), advertisers learn relatively little about consumers. The key to the success of this intermediation mechanism is that advertisers are able to use the information exactly as if they had it. But in practice, they only learn summary statistics on the return on their investment. With automated bidding, they do not even necessarily know how much they bid for each consumer category, because the platform does so for them. Furthermore, because only the platform ever holds the consumer data, this reduces the risk of leakages and spillovers.[13]

However, because only a few firms (one, in the model) are allowed to use the information at any time, the additional privacy gains can come at the cost of worse terms of trade for the consumer. This is consistent with the concern in Cremèr, de Montjoye, and Schweitzer (2019) that

> One cannot exclude the possibility that a dominant platform could have incentives to sell "monopoly positions" to sellers by showing buyers alternatives which do not meet their needs.

In this sense, the optimal managed campaign mechanism is successful precisely because it restricts competition. Privacy protections sounds anti competitive in the context of this model, but this is not yet a general conclusion—a lot more work is warranted on this topic, especially as it relates to data-driven mergers (Chen, Choe, Cong, and Matsushima, 2022). I outline further critical areas for research below.

[13]See Fainmesser, Galeotti, and Momot (2022), Jullien, Lefouli, and Riordan (2020), Tucker (2018) for a discussion of exogenous and endogenous (equilibrium) risks of data leakages.

5 Conclusions

We have focused on the amount of information that large digital platforms are able to collect from individual users, and on the mechanisms by which they monetize this information with advertisers. Several features of digital markets suggest a new, "two-sided" dimension of user privacy, whereby the behavior of all players on both sides of the platform (i.e., users and advertisers) determines each individual consumer's privacy level.

Let us summarize the main findings. First, a platform's ability to profitably collect an individual's data depends not just on that individual's actions or on the rights awarded to them by the law. The social dimension of the data—by which others' data is sometimes as informative as my own—introduces a data externality that drives a wedge between the profitable and the efficient allocation of information, even under well-specified property rights.

Second, the profitability of selling targeted advertising increases as more firms compete for preferential access to a consumer's attention. This increases the incentives to collect that information, and potentially improves the quality of the match between consumer and producer, thanks to stronger selection effects. At the same time, better quality matches might also mean each advertiser now faces a (smaller) more homogeneous consumer population, which facilitates surplus extraction through prices, even without price discrimination. Finally, the collection and monetization aspects of platforms interact. The growth of a platforms database through the participation of more consumers facilitates data acquisition but also raises advertisers' willingness to pay for preferential placement, reduces the value of their private sales channels, and with that the value of each consumer's outside option.

There is a lot of work left to do. For example, the question of competing data platforms and data sellers is conspicuously understudied—promising initial treatments are in Ichihashi (2021a) and De Corniere and Taylor (2020). Data combination, federated learning and other privacy-preserving initiatives are also worth further study (Bergemann, Bonatti, Demirer, and Vilfort, 2023), as is the evaluation, both theoretical and empirical, of recent regulatory interventions (Ali, Lewis, and Vasserman, 2019; Argenziano and Bonatti, 2021; Chen, 2022). Finally, the information design approach can apply to equally, if not more, important dimensions of consumer privacy, such as the political economy implications of government surveillance. Questions of algorithmic fairness, differential privacy, the tradeoff between the efficacy of industrial policy and individual liberties (Beraja, Kao, Yang, and Yuchtman, 2022), as well as the special status of health data (Miller, 2022) are all areas deserving of further treatment.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022): "Too Much Data: Prices and Inefficiencies in Data Markets," *American Economic Journal: Microeconomics*, forthcoming.

Acquisti, A., C. R. Taylor, and L. Wagman (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54(2), 442–92.

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

Admati, A. R., and P. Pfleiderer (1986): "A monopolistic market for information," *Journal of Economic Theory*, 39(2), 400–438.

——— (1990): "Direct and indirect sale of information," *Econometrica*, 58(4), 901–928.

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

Argenziano, R., and A. Bonatti (2021): "Data Linkages and Privacy Regulation," Discussion paper, Essex and MIT.

Aridor, G., Y.-K. Che, and T. Salz (2020): "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR," Discussion Paper 26900, National Bureau of Economic Research.

Arrow, K. (1962): "Economic Welfare and the Allocation of Resources for Invention," *NBER Chapters*, pp. 609–626.

Athey, S., C. Catalini, and C. Tucker (2017): "The Digital Privacy Paradox: Small Money, Small Costs, Small Talk," Discussion Paper 23488, National Bureau of Economic Research.

Barlow, J. P. (1994): "The Economy of Ideas," *Wired Magazine*.

Beraja, M., A. Kao, D. Y. Yang, and N. Yuchtman (2022): "AI-tocracy," *Review of Economic Studies*, forthcoming.

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

Bergemann, D., and A. Bonatti (2022): “Data, Competition, and Digital Platforms,” Discussion paper, Yale University and MIT.

Bergemann, D., A. Bonatti, M. Demirer, and V. Vilfort (2023): “Privacy, Federated Learning, and the Value of Data,” Discussion paper, Yale University and MIT.

Bergemann, D., A. Bonatti, and T. Gan (2022): “The Economics of Social Data,” *RAND Journal of Economics*, 53(2), 263–296.

Bergemann, D., A. Bonatti, and A. Smolin (2018): “The Design and Price of Information,” *American Economic Review*, 108(1), 1–48.

Bergemann, D., A. Bonatti, and N. Wu (2023): “Managed Campaigns and Data-Augmented Auctions for Digital Advertising,” Discussion paper, Cowles Foundation for Research in Economics.

Bergemann, D., B. Brooks, and S. Morris (2015): “The Limits of Price Discrimination,” *American Economic Review*, 105, 921–957.

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

Bonatti, A., M. Dahleh, T. Horel, and A. Nouripour (2022): “Selling information in Competitive Environments,” Discussion paper, MIT.

Bonatti, A., and J. Villas-Boas (2022): “A Theory of the Effects of Privacy,” Discussion paper, MIT and UC Berkeley.

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

Casadesus-Masanell, R., and A. Hervas-Drane (2015): “Competing with privacy,” *Management Science*, 61(1), 229–246.

Chen, Z. (2022): “Privacy Costs and Consumer Data Acquisition: An Economic Analysis of Data Privacy Regulation,” Discussion paper, Monash University.

Chen, Z., C. Choe, J. Cong, and N. Matsushima (2022): “Data-driven mergers and personalization,” *RAND Journal of Economics*, 53(1), 3–31.

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

Coase, R. H. (1960): "The Problem of Social Cost," *The Journal of Law and Economics*, 3, 1–44.

Cremèr, J., Y.-A. de Montjoye, and H. Schweitzer (2019): "Competition policy for the digital era," Discussion paper, European Commission.

de Cornière, A., and R. de Nijs (2016): "Online Advertising and Privacy," *Rand Journal of Economics*, 47(1), 48–72.

De Corniere, A., and G. Taylor (2020): "Data and competition: a general framework with applications to mergers, market structure, and privacy policy," Discussion paper, Oxford and TSE.

Diamond, P. A. (1971): "A model of price adjustment," *Journal of Economic Theory*, 3(2), 156–168.

Elliott, M., A. Galeotti, and A. Koh (2022): "Market Segmentation through Information," Discussion paper, Cambdrige University.

Fainmesser, I. P., A. Galeotti, and R. Momot (2022): "Digital privacy," *Management Science*, forthcoming.

Haghpanah, N., and R. Siegel (2022): "The Limits of Multiproduct Price Discrimination," *American Economic Review: Insights*, 4(4), 443–58.

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.

——— (2021a): "Competing data intermediaries," *RAND Journal of Economics*, 52(3), 515–537.

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

Johnson, G. (2022): "Economic research on privacy regulation: Lessons from the GDPR and beyond," Discussion paper, National Bureau of Economic Research.

Jones, C. I., and C. Tonetti (2020): "Nonrivalry and the Economics of Data," *American Economic Review*, 110(9), 2819–58.

Jullien, B., Y. Lefouli, and M. Riordan (2020): "Privacy Protection, Security, and Consumer Retention," Discussion paper, Columbia University and TSE.

Kirpalani, R., and T. Philippon (2021): "Data Sharing and Market Power with Two-Sided Platforms," Discussion Paper 28023, NBER.

Loertscher, S., and L. M. Marx (2020): "Digital monopolies: Privacy protection or price regulation?," *International Journal of Industrial Organization*, 71, 1–13.

Miller, A. (2022): "Privacy of digital health information," Discussion paper, National Bureau of Economic Research.

Mustri, E. A. S., I. Adjerid, and A. Acquisti (2022): "Behavioral advertising and consumer welfare: An empirical investigation," Discussion paper, Carnegie Mellon University.

Posner, R. A. (1981): "The economics of privacy," *American Economic Review*, 71(2), 405–409.

Prüfer, J., and C. Schottmüller (2021): "Competing with big data," *The Journal of Industrial Economics*, 69(4), 967–1008.

Robinson, J. (1933): *The Economics of Imperfect Competition*. Macmillan, London.

Rochet, J.-C., and J. Tirole (2003): "Platform competition in two-sided markets," *Journal of the European Economic Associationconomic association*, 1(4), 990–1029.

Schmalensee, R. (1981): "Output and welfare implications of monopolistic third-degree price discrimination," *American Economic Review*, 71(1), 242–247.

Segal, I. (1999): "Contracting with Externalities," *Quarterly Journal of Economics*, 114, 337–388.

Shapiro, C., and H. R. Varian (1999): *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Stigler, G. J. (1980): "An introduction to privacy in economics and politics," *Journal of Legal Studies*, 9(4), 623–644.

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

Teh, T.-H., and J. Wright (2022): "Intermediation and steering: Competition in prices and commissions," *American Economic Journal: Microeconomics*, 14(2), 281–321.

Tucker, C. (2018): "Privacy, algorithms, and artificial intelligence," in *The economics of artificial intelligence: An agenda*, pp. 423–437. University of Chicago Press.

Varian, H. (1980): "A Model of Sales," *American Economic Review*, 70(4), 651–659.

Viljoen, S. (2021): "A relational theory of data governance," *Yale Law Journal*, 131.

Villas-Boas, J. (2004): "Consumer Learning, Brand Loyalty, and Competition," *Marketing Science*, 23(1), 134–145.

Wang, C., and J. Wright (2020): "Search platforms: Showrooming and price parity clauses," *RAND Journal of Economics*, 51(1), 32–58.

Yang, K. H. (2022): "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and Its Consequences," *American Economic Review*, 112(4), 1364–93.

Zuboff, S. (2019): *The Age of Surveillance Capitalism*. Public Affairs, New York.