IFN Working Paper No. 1481, 2023

Paying with Personal Data

Gisle J. Natvik and Thomas P. Tangerås

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Paying with personal data*

Gisle J. Natvik[†] and Thomas P. Tangerås[‡]

December 6, 2023

Abstract

We study commercialization of personal data through personalized advertising by a content platform. Content consumption generates productive data about consumer preferences. The firm invests in artificial intelligence (AI) to improve analytical power and in quality to stimulate content consumption. The profit-maximizing tariff is zero if productive data are highly valuable. Subsidization of usage would generate nonproductive data and be unprofitable. Data provision is efficient when users pay entirely with personal data because then content consumption optimally trades off improvements in user experience against losses in privacy rent. Still, privacy protection is inefficient because of distorted incentives to invest in AI.

Keywords: Artificial intelligence, content platform, personalized advertising, privacy, quality
JEL codes: D82, L12, L15, L81, M37

*This paper was initiated when both authors were visiting Stanford University. Natvik thanks the Department of Economics and Tangerås the Program on Energy and Sustainable Development (PESD) for their hospitality. We thank Malin Arve, Pär Holmberg, Melinda Suveg, Joacim Tåg and participants at the 11th Nordic Workshop on Industrial Organization (NORIO XI) in Stockholm for comments. We are also grateful to the Marianne and Marcus Wallenberg Foundation (grant 2020.0049) for financial support.

†Department of Economics, BI Norwegian Business School, 0442 Oslo. Email: gisle.j.natvik@bi.no. Website: sites.google.com/site/gjnatvik/home.

‡Research Institute of Industrial Economics (IFN), P.O. Box 55665, Grevgatan 34, 102 15 Stockholm. Associate researcher, Energy Policy Research group (EPRG) at University of Cambridge. Faculty affiliate, Program on Energy and Sustainable Development (PESD) at Stanford University. E-mail: thomas.tangeras@ifn.se. Website: ifn.se/thomast.

> Consumers use search engines that produce incredibly accurate results. Social networks let people keep in touch with friends, wherever they are in the world. And they don't pay a single penny for those services. Instead, they pay with their data. That doesn't have to be a problem, as long as people are happy that the data they share is a fair price to pay for the services they get in return. Personal data has become a valuable commodity.
>
> Margrethe Vestager, EU Commissioner for Competition.[1]

1 Introduction

Many of the world's most successful digital platforms, such as search engines and social media, earn their revenue mainly by commercializing data collected from users. Applications of increasingly powerful prediction tools built on artificial intelligence (AI) have been instrumental in driving the development. This business model challenges our valuation of consumer privacy, and has attracted the attention of competition authorities and the concern of policy makers. For instance, the German Competition Authority (GCA) recently conducted an investigation into Facebook's practice of collecting data from third-party websites and decided to restrict external data collection by requiring that Facebook only do so subject to users' individual consent.[2] At a European level, Regulation 2016/679, better known as the General Data Protection Regulation (GDPR), establishes rules relating to the protection of individuals within the European Union (EU) with regard to the processing and movement of personal data. Recent development in generative AI have accelerated these privacy concerns, with calls being made to halt further development and use of such tools. Yet, despite the prominence of these phenomena in daily life and policy debates, the discourse lacks a coherent framework to understand markets where customers buy services with their personal data and firms invest in analysing them.

We develop a model where a monopoly platform invests in content quality to attract consumers and to increase content consumption, and in AI to improve data analysis. The platform profits from information through an improved ability to target users with personalized advertising. On the demand side, agents choose whether to participate on the platform, how much content to consume, and whether to buy products advertised on the platform. Content consumption generates productive user data that reveal information about the individual. A novel and key aspect of our model is that agents can use "bots" to generate additional non-productive user data through insincere platform usage. The inability to charge agents for different types of platform use constrains the firm's ability to stimulate content consumption through subsidies.[3] The platform may then fail to internalize economic effects of data collection and analysis through the platform tariff. Our results deliver insights into four interrelated policy questions.

[1] ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/making-data-work-us_en

[2] See bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02+_2019_Facebook.html

[3] The social media platform *X* (formerly *Twitter*) has recently announced a plan to charge all participants for platform usage. The main purpose is to get rid of bots and fake accounts on the platform, according to the owner, Elon Musk; see bbc.com/news/technology-66850821.

How does investment in AI affect privacy and efficiency? To analyze technology investment in relation to privacy issues, we distinguish between data and information. Information is the output obtained by the firm after feeding user data into a prediction machine (Agrawal et al., 2018). Better AI technology improves the platform's information for any given quantity of user data by increasing the power of the prediction machine. Hence, a user may experience a *loss in information privacy* (which occurs when the firm knows more about the user) as a consequence of AI investment, which is distinct from that individual's *loss in data privacy* (which occurs when the firm collects more user data).[4] The distinction matters because an individual has some control over the own data provision through individual platform usage, but has no influence over how the platform processes this information other than through regulation. A user may benefit from losing some privacy due to improved AI because more precise information about user preferences improves the individual user experience on the platform. However, better information enables the firm to extract more consumer rent through better targeted ads. The platform ignores both effects on infra-marginal consumers under a zero platform tariff. We establish necessary and sufficient conditions for overinvestment in AI to occur. This equilibrium has too little information privacy in the sense that information extraction by the platform is excessive.

Do platforms have distorted incentives to invest in quality? The platform invests in quality to stimulate participation and content consumption. Increased participation is valuable because of the advertising profit additional consumers bring in. Increased content consumption is valuable because it generates additional user data which the platform can analyze to improve advertising. The platform underinvests in quality under a zero platform tariff because it does not account for the positive effect of an increase in quality on infra-marginal users.

Do the platform's pricing policies yield inefficient platform usage and data collection? The platform tariff affects participation on the extensive margin and platform usage on the intensive margin. It would like to subsidize participation and content consumption if users and the information generated by content consumption are very valuable. But subsidies would attract opportunists and trigger insincere platform use, both of which are worthless to the platform.[5] To avoid freeriding, the equilibrium tariff has a particular structure which encompasses many commonly observed tariffs as special cases: Users pay a non-negative subscription fee for accessing the platform in return for a free user allowance. The platform charges an overage fee for excessive platform usage if the allowance is limited. From a consumer perspective, free access and free usage maximize expected utility in the set of non-negative platform tariffs. It follows that participation and content consumption are efficient if the zero tariff maximizes platform profit.[6] Data privacy is efficient because content consumption is efficient. Instead, participation,

[4] An individual concerned with data privacy dislikes sharing of personal data. Somebody who values information privacy is concerned about the consequences of sharing data in terms of the information these data reveal and how such information might affect him or her. The former is sometimes referred to as an intrinsic and the latter as an instrumental preference for privacy; see for instance Lin (2022) and Acemoglu (forthcoming).

[5] Gans (2021) argues that free disposal of goods would generate a mass point of demand at a retail price equal to zero. In this spirit, agents have free disposal of platform participation in our model by the possibility to create a fake user profile at no cost and free disposal of platform usage by the possibility to program a bot at no cost.

[6] Armstrong (2006) provides the standard argument why access would be free for consumers on a two-sided

content consumption and data generation are inefficiently low if the equilibrium platform tariff is positive or the user allowance limited.

Is advertising excessive? The platform engages in advertising if and only if information about consumer preferences is sufficiently precise. The firm is a pure content provider otherwise. Advertising, if it occurs, will be excessive for two reasons. First, the platform disregards the nuisance cost of advertising because advertising is chosen after the consumer has joined the platform. Second, the platform assesses the impact on the marginal instead of the average consumer when choosing advertising intensity, measured by the number of different varieties advertised on the platform. The benefit to the marginal consumer of an increase in product variety is larger than the benefit to the average consumer constituting the appropriate welfare benchmark. This distortion causes too much product variety and thereby too much advertising in equilibrium.

The rest of the paper proceeds as follows. Section 2 places our paper in the literature. We present our model in Section 3. It is solved by backward induction by first analyzing personalized advertising in Section 4, consumers' optimal platform participation and usage in Section 5, and the profit-maximizing platform tariff in Section 6. We consider investments in AI in Section 7 and in quality in Section 8. The paper ends with a policy discussion in Section 9. The appendix contains proofs of some formal statements in the main text.

2 Contributions to the literature

Markets for information have been extensively researched; see Bergemann and Bonatti (2019) for a survey. Many analyze the downstream side of the market where an intermediary, the platform in our case, sells user data to advertisers or data brokers. A main issue is imperfect competition in this market. However, trade of information raises questions about privacy protection which these papers often leave aside.[7] Privacy is instead of key concern to an evolving literature analyzing the upstream relationship between users and the intermediary (e.g. Bergemann and Bonatti, 2019; Jones and Tonetti, 2020; Ichihashi, 2021; Bergemann et al., 2022 and Acemoglu et al., 2022). An interesting question is whether allocation of data property rights can resolve privacy concerns through the pricing mechanism. Results show that outcomes in the data market typically are inefficient, particularly in the presence of data externalities that arise when individual data provide information about other users.

We consider the upstream relationship between users and the platform. Downstream inefficiencies are ignored by an assumption that advertising profit measures the full producer surplus associated

platform. The access fee for users on one side of the platform are smaller in equilibrium if the externalities they exert on users on the other side are more positive. By implication, consumers pay a non-positive fee if they are sufficiently valuable to advertisers. That contribution and most others in the field, see Jullien et al. (2021) for a comprehensive survey, assume that an individual's only decision is whether to join the platform, whereas usage is exogenous from the viewpoint of each individual. In our model, participation and platform usage are endogenous.

[7]An exception is de Corniere and de Nijs (2016) who examine incentives for a platform to disclose individual user data to advertisers who use data to infer demand. Disclosure improves the match between products and consumers, which particularly benefits the industry through higher prices of advertised goods. In our model, more precise information increases advertising profit as well as the total advertising surplus.

with goods advertised on the platform. For instance, advertising could be the outcome of efficient bargaining between the platform and advertisers.[8] This simplification facilitates analysis of a fundamental aspect of information markets generally disregarded by the previous literature. Instead of data being exogenously given quantities, we treat data generation as an endogenous outcome of participation and usage decisions made by agents. In addition, we incorporate an important aspect of content platforms: They provide an infrastructure through which advertisers can reach potential customers. Advertising on the platform affects the perceived benefits and costs of joining and using the platform, which has consequences for the quantity and quality of user data the platform will be able to collect. Our model captures this inter-dependency. It encapsulates the possibility of data trade between users and the platform through the nonlinear platform tariff, which formally allows subsidization of platform usage. Data markets turn out to be nonviable in our framework because the buyer (the platform) cannot distinguish between productive and non-productive data offered by a seller (consumer) just by examination of the data. Buying data will not enable the monopoly to collect more productive data than what it could obtain for free through the user allowance.

A content platform shares common features with a media platform generating business by connecting advertisers and subscribers; see Anderson and Jullien (2015) for a survey. While advertisers value subscribers, the media platform must provide content to attract users because they dislike advertising. A main question is whether advertising levels are efficient, with effects typically being ambiguous. Most studies assume that individuals only decide whether to join the platform. An exception is Reisinger (2012) who establishes circumstances with over- or under-provision of advertising under the assumption of endogenous platform usage and a zero platform tariff. Papers in this literature have not analyzed data issues and privacy.

We analyze both the decision to engage in advertising and how many varieties to advertise. Advertising is excessive when it occurs. The mechanisms are different from those explored elsewhere. In the seminal contribution by Anderson and Coate (2005), for instance, the platform decides on advertising intensity before agents join the platform, under complete information about the willingness to pay for advertised products. In our model, the platform decides on advertising after agents have joined the platform and is incompletely informed about demand. We consider participation and content consumption under a nonlinear tariff.

A small number of papers incorporate privacy issues in the economic analysis of platforms; see Acquisti et al. (2016) for a survey of the economics of privacy. In Kox et al. (2017), competing platforms collect personal information from their subscribers. There may be too much or too little information collection in equilibrium depending on subscribers' valuation of privacy, but any distortion hangs on individuals being unable to detect whether platforms collect information. Under complete information, platforms implement efficient data privacy in equilibrium. In

[8]One could argue whether our model fits the description of a two-sided platform by our assumption of efficient contracting between the platform and advertisers. Hagiu and Wright (2015) provide a characterization and perform an interesting analysis of a platform's choice between vertical integration and "two-sidedness", which describes a platform where contracting is decentralized to buyers and sellers.

Dimakopoulos and Sudaric (2018), platforms impose a minimal data requirement for consumers to gain access to their platform under a zero tariff. Data collection is inefficient in equilibrium because platforms ignore infra-marginal consumers' value of privacy. An access fee would restore efficiency by enabling platforms to fully internalize consumers' value of privacy. In Choi et al. (2019), data collection from platform users reduces the privacy of *non-users* because of a data externality. Data collection is excessive even in an equilibrium with full market coverage because reduced privacy diminishes the value of the non-user outside option and thereby enables the platform to extract more rent through a higher subscription fee.

In our framework, the profit-maximizing platform tariff can lead to inefficient data collection even under complete information and absent any data externalities. We complement the previous literature by incorporating non-price strategies into the analysis, in terms of investment in content quality and AI. Our paper particularly adds to an emerging literature emphasizing consequences of AI for privacy. Acemoglu (forthcoming) builds on Acemoglu et al. (2022) to argue that AI can be harmful because of data externalities across consumers. We derive circumstances when AI is harmful even absent data externalities because of excessive incentives to extract information from consumers by investment in artificial intelligence.

3 The model

This section describes the general properties of our model. The demand side consists of agents who choose whether to subscribe to a content platform, how much content to consume, and whether to purchase advertised products. On the supply side, a monopoly invests in content quality and AI to analyze user data, sets a platform tariff, and decides about advertising on the platform. Section 3.1 presents general expressions for the agents' utility functions, section 3.2 the platform's profit function, and section 3.3 the joint welfare function. The timing of the game is described in section 3.4. Presentation of the specific model of personalized advertising is deferred to section 4.

3.1 Agents

There are two types of agents who might participate on the platform, consumers and opportunists. A consumer is valuable to the platform, but an opportunist is not.

Consumers There is one representative consumer who derives utility from consuming content and therefore may have a willingness to pay for access to and use of the platform. This consumer also has a willingness to pay for goods advertised on the platform. The net surplus from using the platform and purchasing advertised goods is $y + V(q) - T(d)$. Here, y is an exogenous value of participating on the platform regardless of platform usage, $V(q)$ is the consumer's gross utility of content consumption $q \geq 0$, and $T(d)$ is a nonlinear platform tariff that depends on the total quantity $d \geq 0$ of platform usage. Content consumption is not the only way to use the platform, as we discuss below. The outside option of not participating on the platform is zero.

The exogenous valuation y is private information to the consumer, but it is common knowledge that $y \in [-\underline{y}, \bar{y}] = \mathcal{Y}$ with cumulative distribution function $G(y)$ and density function $g(y)$. We impose the monotone hazard rate property that $\frac{g(y)}{1-G(y)}$ is non-decreasing. The hazard rate is a measure of the semi-elasticity of participation demand, in absolute value terms. Participation demand is more elastic the larger is the hazard rate. We assume $\underline{y} > \max_{q \geq 0} V(q)$, so that the consumer opts out if the valuation y is sufficiently low and the platform tariff is non-negative. We also assume $\bar{y} > -\min_{q \geq 0} V(q)$, so that the consumer opts in if the valuation y is sufficiently large and the platform tariff is zero or negative.

The gross utility of content consumption consists of three additively separable terms

$$V(q) \;=\; \underbrace{U(q) - hN(\Phi(q))}_{\text{User experience}} + \underbrace{CS(\Phi(q))}_{\text{Privacy rent}}. \tag{1}$$

$U(q)$ measures the direct utility from consuming platform content in quantity q. This utility is continuous and strictly quasi-concave, with a bliss point $b > 0$. Additional consumption above b reduces utility. This makes sense if, for instance, q measures the time spent on the platform, and there is an opportunity cost of time that may dominate the direct utility of using the platform. Another interpretation is that the consumer experiences a disutility of giving up privacy which is embedded in the utility function $U(q)$. The privacy effect dominates for sufficiently large q, in which case additional content consumption reduces direct utility. $N(\Phi(q))$ is the advertising intensity on the platform, and $h > 0$ is the marginal disutility or nuisance that the consumer experiences from additional advertising. The nuisance cost arises whether or not the consumer buys any advertised product. Advertising intensity depends on content consumption indirectly through the function $\Phi(q)$. As will become clear in section 4, $\Phi(q)$ measures the precision of the platform's information about the consumer obtained by analysing user data derived from content consumption. We assume that the advertising intensity is strictly decreasing in content consumption because the platform can better target the consumer if it has more precise information. We define the *user experience* in (1) as the direct utility of content consumption minus the nuisance cost of advertising. The function $CS(\Phi(q))$ measures the consumer's expected value of purchasing goods advertised on the platform. This consumer surplus depends on content consumption indirectly through the information it reveals about the consumer. We also refer to $CS(\Phi(q))$ as the consumer's *privacy rent*. This privacy rent is decreasing in content consumption because the platform can exploit better information about the consumer to extract rent.

The consumer can also engage in insincere activities in quantity $q_0 \geq 0$ on the platform, in addition to consuming content q. The total platform usage amounts to $d = q + q_0$. Insincere platform usage is neither associated with any intrinsic benefit nor any intrinsic cost from the viewpoint of the consumer. We can think of it as using a bot that randomly browses content on the platform. The consumer might engage in such activities especially if the platform pays participants to increase their usage. Insincere platform usage limits the platform's gains from stimulating content consumption through the platform tariff. We assume that the consumer does not engage in insincere activities unless he or she strictly benefits from doing so.

Opportunists There exists a measure $\rho \geq 0$ of opportunists, indexed by superscript o. An opportunist derives no exogenous value of being on the platform, no direct utility from consuming platform content, has no disutility of receiving advertising, nor any willingness to pay for goods advertised on the platform. Hence, $y^o = 0$ and $V^o(q^o) = 0$ for all $q^o \geq 0$. However, the opportunist will subscribe to the platform if paid to do so, that is, if $T(d) < 0$ for some $d \geq 0$. The opportunist will then generate platform usage by engaging in insincere activities in quantity $q_0^o \geq 0$ to minimize $T(d)$. An example of such behavior would be to construct a fake user profile on the content platform with a bot to randomly engage on the platform. We assume that opportunists opt out if the platform tariff is non-negative for all $d \geq 0$. Opportunists represent an obstacle to attracting consumers through platform subsidies, as will be clear below.

3.2 The monopoly content platform

The monopoly platform supplies content and may engage in advertising directed towards its users. The total profit of attracting the representative consumer (not an opportunist) to the platform equals $\Pi(q) - cq_0 + T(d)$, where

$$\Pi(q) = \underbrace{R(\Phi(q))}_{\text{Advertising revenue}} - \underbrace{fN(\Phi(q))}_{\text{Advertising cost}} - \underbrace{cq}_{\text{Data cost}} \tag{2}$$

represents the profit associated with the consumption of platform content in quantity q. The advertising revenue is what the platform expects to earn on selling goods advertised on the platform. This revenue depends on content consumption indirectly through the information $\Phi(q)$ it reveals about the consumer and is an increasing function of q. Advertising is costly, and we assume that the advertising cost scales linearly with advertising intensity, where the unit cost of advertising equals $f > 0$. This cost represents not only the production of actual ads, but also the cost of obtaining and possibly holding inventories of the different varieties. The parameter $c \geq 0$ measures the constant unit cost of handling data. We assume that the advertising revenue is sufficiently high for $\Pi(q)$ to be non-negative in the relevant domain. The profit function $\Pi(q)$ is strictly concave by assumption. The total profit of attracting the consumer is reduced by an additional data cost cq_0 if the consumer generates non-productive data $q_0 > 0$.

The purpose of collecting user data is to attain information about the consumer. We assume a one-for-one relationship between the consumption of platform content and generation of *productive data*, so that q also measures the quantity of productive data the platform collects from the consumer. Insincere platform usage by the consumer generates additional *non-productive data* in quantity q_0. The platform cannot tell the difference between productive and non-productive data just by observation. Instead, it feeds all the collected user data $d = q + q_0$ into an AI algorithm, a *prediction machine*, that delivers information about the consumer measured by $\Phi(d) \geq 1$. In particular, $\Phi(d) = \Phi(q)$ for all $q_0 \geq 0$ so that non-productive data contain no information about the consumer. We let $\Phi(q)$ be a strictly increasing function so that more content consumption provides more information. For technical reasons, we also assume that $\Phi^{-\frac{1}{2}}(q)$ is convex.[9] An opportunist only engages in insincere activities on the platform, $d^o = q_0^o$, thus generating information $\Phi(q_0^o) = \Phi(0) = 1$. Non-separability of q from q_0 implies that the monopoly charges a nonlinear fee $T(d)$ that only depends on total platform usage.[10]

The monopoly can make two types of investments. It can invest in improved AI to enhance the power of its prediction machine. We parameterize the power of the AI by $\theta \in [0, \bar{\theta}]$, so that the output of the machine is $\Phi(q, \theta)$, which is increasing in θ for all $q > 0$. The platform can also invest in improving the quality of the platform service. This could for instance involve production of content by the platform. We parameterize this component by s, so that the consumer's direct utility of using the platform is $U(q, s)$. In particular, the total and marginal utility of using the platform are both increasing in s. The cost $\Psi(\theta)$ of AI is an increasing function of θ, and the cost $\Upsilon(s)$ of quality is an increasing function of s.

3.3 Welfare of content consumption

Adding up the gross utility of content consumption in (1) and the profit associated with the consumption of platform content in (2) produces the welfare of content consumption

$$W(q) = V(q) + \Pi(q) = U(q) - cq + S(\Phi(q)). \tag{3}$$

It consists of the consumer's direct utility $U(q)$ of content consumption minus the platform's associated data cost cq plus the total advertising surplus

$$S(\Phi(q)) = CS(\Phi(q)) + R(\Phi(q)) - (h + f)N(\Phi(q)) \tag{4}$$

defined as the sum of the consumer surplus associated with purchasing goods on the platform and the advertising revenue minus the sum of the nuisance and advertising cost. We assume that $S(\Phi(q))$ is a strictly concave function of q.

3.4 Timing

The model has four stages with the following timing:

1. The monopoly invests in platform quality and AI.

2. The monopoly commits to a non-linear platform tariff.

[9] This condition is satisfied, for instance, by the constant elasticity function $\Phi(q) = (q+1)^\theta$ and the exponential function $\Phi(q) = e^{\theta q}$, where $\theta > 0$.

[10] Since the consumer is privately informed about its preference y for participating on the platform, the monopoly would more generally offer an incentive compatible mechanism specifying platform usage $d(y)$ and a non-linear tariff $T(y)$ as functions of the possible types of the consumer. The consumer would then self-select among the menu of contracts offered on the platform. The additive separability between the preference y and the gross utility $V(q)$ of using the platform strongly limits the monopoly's ability to screen among different consumer types through contracting. In fact, the monopoly can do no better than to offer one single type-independent contract.

3. The agents decide whether to participate on the platform and how to use the platform in that case. Participants pay the platform tariff as a function of their usage.
4. The monopoly analyzes user data. It then decides how many (if any) varieties of a product to advertise to each individual participant, the characteristics of those varieties, and their prices. Participants make their purchase decisions.

We use backward induction to solve the model for sub-game perfect equilibrium.

4 Personalized advertising

This section analyses stage four of the game in which the platform decides on personalized advertising. We present our model of personalized advertising heuristically. The formal analysis is in an online appendix (Natvik and Tangeras, 2023). A main purpose is to understand how the platform's information affects advertising intensity and profit, the effect on platform users and whether advertising is efficient.

The model The representative consumer is of type i located on a circle $\mathcal{I}$ with unit circumference. The type determines the consumer's preferences over differentiated goods also located on the circle. The consumer derives utility $\bar{v} - p_n - \frac{1}{\sigma}|i - l_n|$ from purchasing one unit of variety n located at $l_n \in \mathcal{I}$ when the price of that variety is p_n. The parameter $\bar{v} > 0$ represents the maximal willingness to pay for any good on the platform. The parameter $\sigma > 0$ is a measure of horizontal product differentiation, a higher σ meaning less differentiation. The consumer buys at most one variety if there are multiple varieties to choose from, and at most one unit of the good. The utility of not buying any item is zero.

The platform has prior knowledge that i is uniformly distributed on $\mathcal{I}$, but the actual type is the consumer's private information. Platform usage generates data the platform can analyse to retrieve information about i. A consumer of type i transmits a signal to the platform that he or she is of type $z \in \mathcal{I}$. The conditional density function $m(z|i)$ of the signal is uniform, $m(z|i) = \phi$, for all signals $z \in [i - \frac{1}{2\phi}, i + \frac{1}{2\phi}]$, and $m(z|i) = 0$ otherwise. The model links content consumption to information through an assumption that $\phi = \Phi(q)$. An increase in q increases the precision of the signal z through an increase in ϕ.

The output $z \in \mathcal{I}$ and $\phi \geq 1$ of the prediction machine returns a posterior density function $m(i|z, \phi)$ of the consumer's type i characterized by $m(i|z, \phi) = \phi$ for all $i \in [z - \frac{1}{2\phi}, z + \frac{1}{2\phi}]$, and $m(i|z, \phi) = 0$ for all types outside this confidence interval. The variable z represents the expected type of the consumer, whereas ϕ measures the precision with which the type is measured.

The monopoly would like to avoid advertising to opportunists on the platform because advertising is costly and opportunists have no willingness to pay for advertised products. By an assumption that opportunists do not send any signal about their type, $z^o \in \emptyset$, the monopoly can always tell a consumer apart from an opportunist based on the information that analysis of user data produces.

The following analysis therefore only concerns personalized advertising directed towards the representative consumer.

Profit-maximizing personalized advertising The platform must decide how many different varieties $N \geq 0$ to advertise to the user, the location $l_n \in \mathcal{I}$ in product space and the price $p_n \geq 0$ of each variety n. We assume that the production cost of advertised goods is zero.[11]

The N varieties offered to the consumer span an interval $[z - \frac{1}{2K}, z + \frac{1}{2K}]$ of possible consumer types. The platform could increase its probability of selling goods by moving some varieties uniformly towards z if the distribution of varieties was non-convex. The distribution is symmetric around z because i is uniformly distributed around z. $K \geq \phi$ because there is no point in advertising goods that are attractive no non-existent user types. Placing all varieties at the same distance from each other minimizes the expected "transportation cost" of the consumer and maximizes the price the platform can charge from the consumer. By linearity, the price is the same for all varieties and set at the point at which a consumer type placed at maximal distance from one variety is indifferent between buying a good or not. Formally,

$$\bar{v} - P - \frac{1}{\sigma}\frac{1}{2NK} = 0 \Leftrightarrow P(NK) = \bar{v} - \frac{1}{2\sigma NK}.$$

The platform then maximizes the expected advertising profit

$$\int_{z-\frac{1}{2K}}^{z+\frac{1}{2K}} P(NK)\phi dx - fN = (\bar{v} - \frac{1}{2\sigma NK})\frac{\phi}{K} - fN$$

over the degree K of product variety and the number N of advertised products. For analytical reasons, it will be convenient to treat the platform's profit-maximizing *advertising intensity* $N(\phi)$ as a continuous function of signal precision. This advertising intensity is nonlinear. For $\phi < \frac{2f}{\sigma\bar{v}^2}$, the signal is so imprecise relative to the advertising cost that it is not worthwhile for the monopoly to advertise any products. The platform is a *pure content provider* in this case. Above the threshold, the monopoly has sufficiently precise information about the consumer to engage in advertising on the platform. The platform then offers enough product variety to ensure that every possible consumer type buys a product, $K = \phi$, and advertises with intensity

$$N(\phi) = \frac{1}{\sqrt{2\sigma\phi f}}. \tag{5}$$

The advertising intensity decreases as precision improves because the platform then can target the user more efficiently with ads.

[11] An equivalent assumption would be that the unit production cost γ is the same for all goods and arises after the consumer has submitted the purchase order. One can then subtract γ from the consumer's gross valuation $\tilde{v}$ to get the net valuation $\bar{v} = \tilde{v} - \gamma$ and then proceed as in the main text.

We can then use $K = \phi$ and the expression $N(\phi)$ to obtain the advertising revenue

$$R(\phi) = \int_{z-\frac{1}{2\phi}}^{z+\frac{1}{2\phi}} P(N(\phi)\phi)\phi dx = \bar{v} - \sqrt{\frac{f}{2\sigma\phi}}. \tag{6}$$

$R(\phi)$ also measures the price of one unit of the good because the platform always sells one variety of the good, all varieties have the same price, and the consumer buys at most one unit. The advertising revenue is an increasing function of ϕ because because increased precision enables the monopoly to offer better targeted products for which it can charge higher prices.

We calculate next the expected consumer surplus $CS(\phi)$ of purchasing goods advertised on the platform. The consumer buys one variety regardless of its type i. The price of all varieties is the same and equal to $R(\phi)$. The consumer surplus of a consumer of type i then depends on the distance to the closest variety offered on the platform. These varieties are uniformly distributed around the consumer's type. The expected consumer surplus thus equals

$$CS(\phi) = 2N(\phi) \int_0^{\frac{1}{2N(\phi)\phi}} [\bar{v} - R(\phi) - \frac{1}{\sigma}(\frac{1}{2N(\phi)\phi} - z)]\phi dz = \frac{1}{2}\sqrt{\frac{f}{2\sigma\phi}} > 0. \tag{7}$$

The consumer surplus is positive, unlike in most other models of advertising which assume full surplus extraction (see for instance, Anderson and Coate, 2005). To some extent, the user benefits from an increase in signal precision because better targeted products reduces the average transportation cost. But the seller charges higher prices for those products, and the overall effect is a reduction in consumer surplus. In the limit as $\phi \to \infty$, the platform extracts the full surplus of advertised products. This is the case of perfect price discrimination.

The consumer surplus plus the advertising revenue minus the nuisance and advertising cost jointly yield the total advertising surplus

$$S(\phi) = CS(\phi) + R(\phi) - (h + f)N(\phi) = \bar{v} - \sqrt{\frac{f}{2\sigma\phi}}\frac{3f + 2h}{2f}.$$

This surplus can be positive of negative, but is strictly increasing in signal precision ϕ.[12]

Inefficiencies of personalized advertising To gauge the welfare implications of advertising, we derive in the online appendix Natvik and Tangeras (2023) the advertising strategy that maximizes welfare. It turns out that all distortions can be attributed to differences between the equilibrium advertising intensity $N(\phi)$ and the efficient advertising intensity $N^*(\phi)$.

Proposition 1. *A monopoly engaging in personalized advertising does so excessively, $N(\phi) > 0$ implies $N(\phi) > N^*(\phi)$.*

[12]Based on the above expressions, the advertising profit $R(\Phi(q)) - fN(\Phi(q))$ and the total advertising surplus $S(\Phi(q))$ are concave functions of q by the assumed convexity of $\Phi^{-\frac{1}{2}}(q)$.

Proof. See the online appendix (Natvik and Tangeras, 2023). □

There are two explanations for this inefficiency. First, the platform fails to internalize the consumer's nuisance cost hN of advertising because advertising intensity is chosen after agents have joined the platform. If advertising instead were set in advance of the participation decision, the nuisance cost would be internalized in the platform tariff. In other words, there is a time-inconsistency problem as the platform cannot commit to (low) advertising before collecting data. Second, the monopoly targets the *marginal* instead of the *average* consumer when deciding on personalized advertising. The profit-maximizing product price extracts the full surplus of the marginal consumer, whose willingness to pay for the product is $\bar{v} - \frac{1}{2N\sigma\phi}$. In contrast, the relevant efficiency benchmark is the expected consumer's willingness to pay for an item, namely $\bar{v} - \frac{1}{4N\sigma\phi}$. The effect on the product price of a marginal increase in advertising intensity N is larger than the increase in the expected consumer's willingness to pay. These two effects both contribute to excessive advertising in equilibrium.

The statement in Proposition 1 is contingent on advertising occurring in equilibrium, $N(\phi) > 0$. Natvik and Tangeras (2023) show that the monopoly under certain circumstances fails to engage in advertising altogether even though advertising is efficient, $N(\phi) = 0 < N^*(\phi)$. This happens over an intermediate range of ϕ if $h < f$. From now on we assume that the platform always advertises.[13] Advertising is always excessive in this case.

5 Participation and usage from a consumer perspective

This section analyses stage three of the game, where agents decide on platform participation and usage. Assume for now that platform usage is free and unlimited. Then, maximization of the gross utility $V(q)$ of content consumption characterized in (1) trades off the marginal improvement in user experience against the marginal loss in privacy rent

$$V'(q) = \underbrace{U'(q) - hN'(\Phi(q))\Phi'(q)}_{\text{Marginal improvement in user experience}} + \underbrace{CS'(\Phi(q))\Phi'(q)}_{\text{Marginal loss in privacy rent}}. \tag{8}$$

Advertising intensity decreases when content consumption increases, and therefore the nuisance cost of being on the platform is lower when the subscriber consumes more content. However, the privacy rent is also lower because better information about the consumer's type enables the platform to extract relatively more of the advertising surplus. We assume that $V(q)$ has a unique and positive optimum and denote this quantity of content consumption by $q^u > 0$.

The platform cannot distinguish between content consumption q and insincere platform usage q_0. Only the total quantity $d = q + q_0$ of platform usage by the consumer is verifiable. The consumer's possibility to generate non-productive data through insincere platform usage limits

[13] The firm advertises on the platform if and only if $\phi \geq \frac{2f}{\bar{v}^2\sigma}$. A sufficient condition for $N(\phi) > 0$ for all $\phi \geq 1$ then is $\bar{v}^2\sigma \geq 2f$.

how much productive data the platform tariff can motivate the user to provide through consumption of platform content. In fact, the monopoly can at most implement content consumption q^u regardless of the platform tariff $T(d)$. For any ambition to implement $\hat{q} > q^u$ and $\hat{q}_0 \geq 0$, the consumer can consume content in quantity q^u and achieve total platform usage $\hat{d} = \hat{q} + \hat{q}_0$ through insincere platform usage $q_0 = \hat{q}_0 + \hat{q} - q^u > \hat{q}_0$.

Lemma 1. *The platform can implement content consumption $\hat{q}$ only if $\hat{q} \in [0, q^u]$ and $V(\hat{q}) \geq V(q)$ for all $q \in [0, \hat{q}]$.*

Proof. Suppose the nonlinear tariff $\hat{T}(d)$ implements $\{\hat{q}; \hat{q}_0\}$, where $\hat{q} > q^u$ and $\hat{q}_0 \geq 0$. The consumer could then consume platform content in quantity $q = q^u$ and engage in insincere platform usage $q_0 = \hat{q} - q^u + \hat{q}_0 > \hat{q}_0$ and achieve strictly higher utility

$$y + V(q) - \hat{T}(q + q_0) = y + V(q^u) - \hat{T}(\hat{q} + \hat{q}_0) > y + V(\hat{q}) - \hat{T}(\hat{q} + \hat{q}_0)$$

under $\{q; q_0\}$ compared to the utility under the intended platform usage $\{\hat{q}; \hat{q}_0\}$. This contradicts the assumed implementability of $\{\hat{q}; \hat{q}_0\}$ by $\hat{T}(d)$. By a similar argument, the monopoly cannot implement $\{\hat{q}; \hat{q}_0\}$, where $\hat{q} \in [0, q^u]$ and $\hat{q}_0 \geq 0$, for any $\hat{T}(d)$ if $V(q) > V(\hat{q})$ for some $q \in [0, \hat{q})$ The consumer could then reduce the consumption of platform content to q and increase insincere platform usage to $q_0 = \hat{q} - q + \hat{q}_0 > \hat{q}_0$ and achieve strictly higher utility than under $\{\hat{q}, \hat{q}_0\}$. □

The gross utility function $V(q)$ is not necessarily quasi-concave, so the monopoly may not be able to implement all content consumption in $[0, q^u]$. We let $\mathcal{Q}$ be the feasible set of content consumption the platform can implement by some platform fee $\hat{T}(d)$.[14]

Turning to the extensive margin, the consumer has total utility $y + V(\hat{q}) - \hat{T}(\hat{q} + \hat{q}_0)$ of participating on the platform if the platform tariff $\hat{T}(d)$ implements platform usage $\{\hat{q}; \hat{q}_0\}$, where y is the exogenous value of participating on the platform as explained in Section 3.1. The value of the consumer's outside option is normalized to zero. Hence, the consumer joins the platform if and only if y exceeds the *participation threshold* $\hat{y} = \hat{T}(\hat{q} + \hat{q}_0) - V(\hat{q})$.

6 The profit-maximizing platform tariff

This section analyses the second stage of the game in which the monopoly chooses its platform tariff to maximize expected profit. A tariff $\hat{T}(d)$ that implements platform usage $\{\hat{q}, \hat{q}_0\}$ and a participation threshold $\hat{y}$ by the consumer, yields expected monopoly profit

$$\underbrace{[1 - G(\hat{y})][\Pi(\hat{q}) - c\hat{q}_0 + \hat{T}(\hat{q} + \hat{q}_0)]}_{\text{Expected profit from consumer}} + \underbrace{\rho[\min\{\hat{T}(\hat{q}_0^o); 0\} - c\hat{q}_0^o]}_{\text{Expected loss from opportunists}}. \quad (9)$$

In this expression, $1 - G(\hat{y})$ measures the probability that the consumer joins the platform.

[14]Formally, $\mathcal{Q} = \{q \in [0, q^u] \mid V(q) \geq V(\tilde{q}) \; \forall \tilde{q} \in [0, q]\}$.

A profit-maximizing tariff structure We previously showed that the possibility to generate non-productive data through insincere platform usage q_0, restricted how much content consumption q the monopoly could implement through its platform tariff. This constraint has strong implications for the profit-maximizing platform tariff structure.

Lemma 2. *In order to maximize profit, it is sufficient for the monopoly to implement the platform tariff structure*

$$T(d) = \begin{cases} F & \forall\ d \in [0, \hat{d}] \\ F + t \times (d - \hat{d}),\ \textit{some}\ t \geq 0, & \forall\ d > \hat{d} \end{cases} \tag{10}$$

The consumption of platform content is no larger than $\hat{d}$, and no participant engages in insincere platform usage under this tariff.

Proof. See Appendix A. □

The platform tariff in Lemma 2 entails a fixed subscription fee F that allows free usage of the platform, possibly up to a limit $\hat{d}$. The monopoly levies an overage charge t for all platform usage in excess of this limit. This tariff structure encompasses many actual platform tariffs as special cases. Search engines and social media platforms typically allow free access to and unlimited free usage of the platform ($F = t = 0$). Media platforms offer digital subscriptions with unlimited free usage in return for a subscription fee ($F > 0 = t$). Streaming platforms often offer subscriptions with free access and user limitations ($F = 0 < t$).

The platform tariff works as follows. The monopoly can never implement content consumption above q^u by Lemma 1. The easiest way to accomplish q^u is to allow free unlimited usage of the platform. If the platform instead wants to limit consumption of platform content to $\hat{q} < q^u$, where $\hat{q} \in \mathcal{Q}$, then it can incentivize the subscriber by setting overage charge t at such a level that it becomes too expensive for the subscriber to consume content above $\hat{q}$. Consumption of platform content in any quantity $q \in [0, \hat{q}]$ is free. Therefore, the consumer optimally chooses $q = \hat{q}$ since $V(\hat{q}) \geq V(q)$ for all $q \in [0, \hat{q}]$; see Lemma 1. The monopoly attracts the desired amount of consumers by varying the subscription fee F.

The optimality of (10) depends crucially on the assumption that the platform only can offer a platform tariff that depends on the consumer's aggregate data usage $d = q + q_0$. If instead q and q_0 were separately verifiable, then the platform could eliminate insincere platform usage by implementing an overage charge on such behavior. The platform could implement any content consumption $\hat{q} \geq 0$ through an appropriate marginal usage price which could be negative.

The platform never collects any overage charges because participants' (consumers and opportunists) optimal platform usage never exceeds $\hat{d}$ in (10). By implication, the monopoly only has two potential sources of revenue, the subscription fee F and the advertising revenue $R(\Phi(\hat{q}))$. If the subscription fee is zero or if the monopoly subsidizes participation on the platform, so that $F \leq 0$, then the advertising revenue represents the monopoly's sole source of income. The consumer pays entirely with personal data in this case. We measure the *implicit price of platform* *usage* by the loss in privacy rent associated with content consumption q relative to not using the platform at all:

$$P^I(q) = CS(\Phi(0)) - CS(\Phi(q)) = \sqrt{\frac{f}{8\sigma}}[1 - \Phi^{-\frac{1}{2}}(q)], \tag{11}$$

where we have used $CS(\phi)$ characterized in (7) to obtain the rightmost expression. The implicit price is increasing in q because higher content consumption generates additional productive data that the platform can use to extract additional rent from the consumer through better targeted advertising. Holding content consumption fixed, the implicit price of platform usage is smaller when the marginal advertising cost f is smaller or when the goods advertised on the platform are less differentiated, meaning that σ is larger.

Having shown that the monopoly maximizes expected profit by implementing the tariff structure described in (10), the next question is how many participants to attract to the platform and how much content consumption to stimulate. The first issue we address is subsidization of platform participation through a negative subscription fee.

Can it be a profit-maximizing strategy to subsidize participation? The platform can implement content consumption $\hat{q} \in \mathcal{Q}$ by an overage charge t for data usage d in excess of $\hat{q}$. It can implement the desired participation $\hat{y} \in \mathcal{Y}$ by setting a subscription fee equal to $F = \hat{y} + V(\hat{q})$. Substituting the simplified platform fee (10) into the expression (9) for platform profit, returns the expected platform profit

$$\Pi^e(\hat{y}, \hat{q}) = [1 - G(\hat{y})][\hat{y} + W(\hat{q})] + \rho \min\{\hat{y} + V(\hat{q}); 0\} \tag{12}$$

purely as a function of the consumer's participation threshold $\hat{y}$ and consumption $\hat{q}$ of platform content, where $W(q)$ measures the welfare of content consumption q defined in (3).

The expected profit from consumers may increase by subsidizing access to the platform through a negative subscription fee. Subsidization is particularly profitable if the welfare of content consumption is large. However, subsidization will also attract opportunists that represent a cost to the monopoly. This cost is larger if there are more opportunists.

Lemma 3. *The platform does not subsidize subscriptions for any content consumption $q \in \mathcal{Q}$ if the share ρ of opportunists is sufficiently high.*

Proof. Holding q fixed, $\hat{y} = -V(q)$ maximizes $\Pi^e(y, q)$ over $y \in [-\underline{y}, -V(q)]$ if and only if

$$[1 - G(y)][y + W(q)] + \rho[y + V(q)] \leq [1 - G(-V(q))]\Pi(q) \ \forall y \in [-\underline{y}, -V(q)),$$

which is equivalent to

$$\rho \geq \tilde{\Omega}(y, q) = G(y) - 1 + \frac{G(y) - G(-V(q))}{y + V(q)}\Pi(q) \ \forall y \in [-\underline{y}, -V(q)).$$

Define $\Omega(q) = \sup_{y \in [-\underline{y}, -V(q))} \tilde{\Omega}(y,q)$. The monopoly then does not want to set $y < -V(q)$ for any $q \in \mathcal{Q}$ if $\rho \geq \max_{q \in [0,q^u]} \Omega(q)$. We conclude that the profit-maximizing subscription fee is non-negative for all $q \in \mathcal{Q}$ if ρ is sufficiently large. □

We assume from now on that the share of opportunists is so large that Lemma 3 holds. The profit-maximizing subscription fee then is non-negative.

When is the profit-maximizing platform tariff zero? The monopoly maximizes the expected profit $\Pi^e(y,q)$ over $y \in \mathcal{Y}$ and $q \in \mathcal{Q}$ subject to the non-negativity constraint $y + V(q) \geq 0$ on the subscription fee. Our next result characterizes the profit-maximizing platform tariff.

Proposition 2. *The platform allows free access to and free unlimited use of the platform, so that the consumer entirely pays with personal data, if and only if:*

1. *Intensive margin content consumption is large:*
$$\Pi'(q^u) \geq 0. \tag{13}$$

2. *Extensive margin participation semi-elasticity is high:*
$$\frac{1 - G(-V(q^u))}{g(-V(q^u))} \leq \Pi(q^u). \tag{14}$$

If the inequality in condition (14) *is strictly reversed while* (13) *holds, then the monopoly charges a positive subscription fee while still permitting free unlimited use of the platform. If inequality* (13) *is strictly reversed, then the platform implements $\hat{q} \in \mathcal{Q}$ by offering free platform usage up to $\hat{q}$, and an overage charge, $t > 0$, on usage thereafter. Any strictly positive profit-maximizing subscription fee $\hat{F}$ is characterized by*

$$\frac{1 - G(\hat{F} - V(\hat{q}))}{g(\hat{F} - V(\hat{q}))} = \Pi(\hat{q}) + \hat{F}. \tag{15}$$

Proof. See Appendix B. □

On the intensive margin, larger content consumption increases the advertising revenue and reduces the advertising cost because the resulting increase in productive data increases the platform's ability to predict the consumer's type with greater precision. However, increased content consumption also increases the platform's data cost. Differentiation of the profit function $\Pi(q)$ characterized in (2) yields the marginal profit

$$\Pi'(q^u) = \underbrace{R'(\Phi(q^u))\Phi'(q^u)}_{\text{Marginal advertising revenue}} - \underbrace{fN'(\Phi(q^u))\Phi'(q^u)}_{\text{Marginal advertising cost}} - \underbrace{c}_{\text{Marginal data cost}} \tag{16}$$

evaluated at the consumer's most preferred level of content consumption. Under condition (13), the marginal increase in advertising revenue and the marginal reduction in advertising cost are so large relative to the marginal increase in the data cost that the platform would prefer content consumption equal to or above q^u. However, it is impossible to induce content consumption beyond q^u because any attempt to do so would only generate non-productive user data through insincere platform usage. The platform instead maximizes the feasible content consumption, which is accomplished by allowing free unlimited platform usage.

The platform wants to reduce platform usage below q^u if the marginal advertising revenue and the marginal reduction in advertising cost are small relative to the marginal data cost. In that case, the platform offers free platform usage up to its preferred level of content consumption $\hat{q}$, after which consumers are penalized for any excess usage.

On the extensive margin, a higher subscription fee increases the profit $\Pi(q^u)+F$ on any consumer the platform manages to attract to the platform. However, a higher subscription fee also deters potential consumers from joining the platform. By how much, depends on the semi-elasticity of participation demand $\frac{g(y)}{1-G(y)}$. If inequality (14) is met, then participation demand is so elastic relative to the profit per user that the monopoly wants to maximize participation. The platform achieves this objective by setting the subscription fee to zero. Instead of paying with money, consumers pay entirely with personal data.

In an interior optimum, the profit-maximizing subscription fee balances the marginal benefit of extracting rent from infra-marginal consumers through a higher subscription fee against the marginal loss associated with having fewer subscribers on the platform. The associated optimality condition (15) is very similar to the condition in Armstrong (2006) for the profit-maximizing user price in a two-sided market. The main difference is that the subscription fee is adjusted by the profit of content consumption in our context, whereas the user price is adjusted by the economic magnitude of the cross-group externality in the two-sided market.

To better understand the underlying characteristics that drive Proposition 2, we can substitute the marginal advertising intensity $N'(\phi)$ derived from (5) and the marginal advertising revenue $R'(\phi)$ derived from (6) into $\Pi'(q^u)$ identified in (16) to obtain the marginal profit expression

$$\Pi'(q^u) + c = \sqrt{\frac{f}{2\sigma\Phi(q^u)}}\frac{\Phi'(q^u)}{\Phi(q^u)} > 0.$$

The consumer's most-preferred content consumption q^u is independent of the platform's marginal data cost c, which implies that condition (13) is satisfied if c is sufficiently small. Other plausible characteristics also generate positive marginal profit. Substitute the marginal advertising intensity and the marginal consumer surplus $CS'(\phi)$ derived from (7) into $V'(q)$ from (8) to get the marginal utility

$$V'(q) = U'(q) - \sqrt{\frac{f}{2\sigma\Phi(q)}}\frac{\Phi'(q)}{\Phi(q)}\frac{f-2h}{4f}$$

of content consumption. Observe in particular that this marginal utility is strictly negative for all content consumption q above the bliss point b if the marginal marketing cost f is sufficiently large relative to the marginal nuisance cost h. In this case, the marginal loss in privacy rent associated with content consumption is so large relative to the reduction in the nuisance cost that the consumer prefers content consumption $q^u \leq b$. By strict concavity of the profit function,

$$\Pi'(q^u) \geq \Pi'(b) = \sqrt{\frac{f}{2\sigma\Phi(b)}}\frac{\Phi'(b)}{\Phi(b)} - c.$$

The right-hand side of this expression is strictly positive if f is large or σ is sufficiently small so that the consumer values product characteristics highly in the decision whether to purchase products advertised on the platform. We conclude that free unlimited platform usage represents a profit-maximizing strategy under robust circumstances where, for instance, the marginal data cost is small, the marginal marketing cost is large, or product differentiation in the advertisement market is large. However, it is also plausibly the case that the platform prefers to limit content consumption below q^u. This occurs, for instance, if the marginal data cost c is large, or if the marginal advertising cost is small, $f < 2h$, and products are relatively homogeneous in the sense that σ is large. As for the subscription fee, q^u is independent of the maximal willingness $\bar{v}$ to pay for advertised products. Any increase in $\bar{v}$ instead goes to the platform as a one-for-one increase in advertising revenue; see (6). Hence, condition (14) is met and the profit-maximizing subscription fee zero if consumers' willingness to pay for advertised products is sufficiently large.

The monopoly may want to subsidize participation if demand on the extensive margin is very elastic. However, a monetary subsidy would be prohibitively costly by the assumption that such a subsidy attracts a large share of opportunists. One way to circumvent this problem could be a contingent subsidy that only has value if used within the context of the platform. For instance, agents who sign up for the platform could receive a coupon which entitles the agent to a discount B on any purchased good advertised on the platform. This coupon would be equivalent to a monetary subsidy in amount B for the consumer, but worthless to an opportunist without any willingness to pay for goods advertised on the platform. In a similar spirit, Amelio and Jullien (2012) analyze how a platform may sell a bundled good at a discount as an implicit subsidy to induce platform participation. We discuss implications of contingent subsidies below.

Does the platform tariff distort data privacy? To evaluate the effects of the platform tariff on data privacy and efficiency, we assume that the monopoly decides on advertising efforts after agents have joined and used the platform, also in the derivation of the welfare maximizing participation threshold $\hat{y}^*$ and content consumption $\hat{q}^*$. Conditional on the marketing intensity $N(\phi)$, a social planner maximizes expected welfare

$$W^e(y, q) = \int_y^{\bar{y}} [\tilde{y} + W(q)] g(\tilde{y}) d\tilde{y}. \tag{17}$$

Even the social planner is limited to selecting q from the set $\mathcal{Q}$ because non-verifiability of insincere platform usage makes it impossible to implement content consumption $q \notin \mathcal{Q}$. Implementation of participation threshold $y < -V(q)$ would imply that a subset of consumer types were strictly better off by not joining the platform. We therefore constrain the social planner to implement $y + V(q) \geq 0$. That is, our efficiency benchmark subjects the social planner to the same incentive compatibility and participation constraints as the monopoly.

The main difference between maximization of the profit function (12) and the welfare function (17) lies in the determination of the level of consumer participation on the extensive margin. Reducing participation through a larger subscription fee can be profitable to the monopoly because of increased rent extraction from infra-marginal consumer types. Such rent extraction only represents pure redistribution between the firm and the consumer from the social planner's perspective who only cares about the marginal effect

$$W_y^e(y,q) = -[y + W(q)]g(y) = -[y + V(q) + \Pi(q)]g(y) < 0 \; \forall y > -V(q) \tag{18}$$

of reduced participation on expected welfare. The welfare created on the platform is so large that the social planner maximizes consumer participation. In terms of a welfare optimal tariff structure, the socially optimal subscription fee would be zero.

The social planner maximizes $W^e(-V(q), q)$ over $q \in \mathcal{Q}$ and thereby uses the allowance both to increase the welfare associated with content consumption on the intensive margin and to increase participation on the extensive margin. The monopoly generally uses the allowance to increase welfare on the intensive margin, but the subscription fee to attract consumers on the extensive margin. Content consumption is downward distorted in this case because the monopoly fails to internalize the welfare benefit on the extensive margin of attracting consumers to the platform through a larger allowance. However, there is too little content consumption in equilibrium even if the subscription fee is zero so that the monopoly allows free access to the platform. In this case, the problem is that the monopoly fails to internalize the welfare benefit of a larger allowance on content consumption on the intensive margin. To see this, observe that we can write expected welfare as

$$W^e(-V(q), q) = \int_{-V(q)}^{\bar{y}} [y + V(q)]g(y)dy + \Pi^e(-V(q), q)$$

if the subscription fee is zero. The welfare effect of a marginal increase in content consumption equals the sum of the effects on the consumer and the monopoly,

$$\frac{d}{dq}W^e(-V(q), q) = [1 - G(-V(q))]V'(q) + \frac{d}{dq}\Pi^e(-V(q), q),$$

but the monopoly only accounts for the latter effect. Since there is a direct relationship between content consumption and generation of productive data, we obtain:

Proposition 3. *The profit-maximizing platform tariff generally reduces participation ($\hat{y} \geq \hat{y}^*$) on the extensive margin and reduces content consumption ($\hat{q} \leq \hat{q}^*$) on the intensive margin* *relative to the social optimum. Platform usage generates too little productive data and therefore excessive data privacy in equilibrium. An exception occurs if free access and free unlimited usage maximize the platform's expected profit. Participation, content consumption and data generation then maximize expected welfare in equilibrium so that data privacy is efficient in equilibrium.*

Proof. See Appendix C. □

Platform participation and content consumption generally are downward distorted in equilibrium, but an exception arises at the corner solution at which the monopoly neither charges anything for access to nor usage of the platform. The zero tariff maximizes the expected utility of the consumer in the class of non-negative platform tariffs. This tariff is therefore welfare optimal if it also maximizes monopoly profit.

The distortion of content consumption stems fundamentally from the assumption that it is prohibitively costly to subsidize participation. Suppose instead the monopoly and the social planner can attract the desired amount of consumers and simultaneously avoid opportunists, for instance by use of a conditional subsidy $B = -(y+V(q))$ as discussed above. By inspection of the marginal expected welfare expression (18), we see that the welfare optimal participation threshold equals $-\max\{W(q); \underline{y}\}$. The social planner then maximizes the expected welfare

$$W^*(q) = \int_{-\max\{W(q);\underline{y}\}}^{\bar{y}} [\tilde{y} + W(q)]g(\tilde{y})d\tilde{y}.$$

over $q \in \mathcal{Q}$. The solution to this problem is to maximize ex post welfare $W(q)$ of content consumption, which is exactly the same as the platform would do. However, participation would generally be downward distorted in equilibrium.[15]

From a general policy viewpoint, competition authorities are correct about being concerned about the tariffs applied by content platforms. They have an incentive to reduce participation by charging excessive subscription fees, and to limit usage of the platform by overly restrictive usage allowances. Such tariffs reduce generation of productive data below the efficient level. However, free access to and free unlimited usage of the platform is not a source of policy concern as far as efficient platform tariffs are concerned. Paying exclusively with personal data in fact maximizes the expected utility of the consumer compared to all other non-negative tariff structures by the platform. The associated data generation is efficient in this case.

7 Too much AI?

We now enter the first stage of the game, at which the platform invests in AI to increase the precision of the prediction machine it employs to extract information about the subscriber from

[15] Assume that $W(\hat{q}) = \max_{q\in\mathcal{Q}} W(q)$. The welfare optimal participation level is then given by $\hat{y}^* = -\max\{W(\hat{q}); \underline{y}\}$. If $W(\hat{q}) < \underline{y}$, then the profit-maximizing threshold $\hat{y}$ directly satisfies $\hat{y} \geq -\underline{y} = \hat{y}^*$. Consider the alternative possibility $W(\hat{q}) \geq \underline{y}$. The platform maximizes $\Pi^e(y, \hat{q})$ over y, where $\rho = 0$. Strict quasi-concavity of $\Pi^e(y, q)$ in y and $\Pi^e_y(\hat{y}^*, \hat{q}) = 1 - G(\hat{y}^*) > 0$ imply $\hat{y} > \hat{y}^*$.

the analysis of user data. We describe AI as a variable θ that increases $\phi = \Phi(q, \theta)$.

We first assume that the platform allows free access to and free usage of the platform. Necessary and sufficient conditions for when a zero tariff maximizes the expected platform profit were established in Proposition 2. We also showed that the zero platform tariff was efficient conditional on the underlying technology. The question is whether this property is sufficient to generate efficient investment incentives.

A subscriber with free unlimited access to the platform chooses content consumption $q \geq 0$ to maximize gross utility

$$V(q, \theta) = U(q) - hN(\Phi(q, \theta)) + CS(\Phi(q, \theta)). \tag{19}$$

We denote the solution to this problem by $q^u(\theta)$, and let $v(\theta) = V(q^u(\theta), \theta)$ be the consumer's indirect utility of AI technology θ. The platform profit contingent on the AI technology θ equals $\pi(\theta) = \Pi(q^u(\theta), \theta)$, where

$$\Pi(q, \theta) = R(\Phi(q, \theta)) - fN(\Phi(q, \theta)) - cq$$

measures the platform profit as a function of the consumption q of platform content and the direct effect of the AI technology.

The platform selects the AI technology $\theta \in [0, \bar{\theta}]$ that maximizes its expected platform profit

$$\pi^e(\theta) = [1 - G(-v(\theta))]\pi(\theta) - \Psi(\theta).$$

AI has a direct effect on the consumer and the platform because of the direct effect on the prediction $\Phi(q, \theta)$ about the consumer's type. It also has an indirect effect through the consumption $q^u(\theta)$ of content and the consumer's participation threshold $-v(\theta)$. The platform's marginal profit of investing in AI therefore equals

$$\pi^e_\theta(\theta) = [1 - G(-v(\theta))][\Pi_\theta + \Pi_q q^u_\theta] + g(-v(\theta))\pi(\theta)V_\theta - \Psi'(\theta).$$

The marginal direct effect Π_θ on platform profit of investing in AI is positive because more precise information about the consumer's type increases the advertising revenue and simultaneously reduces the advertising cost. The marginal indirect effects are positive if better AI increases content consumption and consumer participation, but goes in the opposite direction if better AI reduces content consumption and participation. Assume that the profit-maximizing investment is strictly positive and unique so that the equilibrium power of the AI satisfies $\hat{\theta} > 0$.

We next turn to the welfare optimal investment in AI. The social planner chooses AI to maximize expected welfare $w^e(\theta) = v^e(\theta) + \pi^e(\theta)$, where

$$v^e(\theta) = \int_{-v(\theta)}^{\bar{y}} [y + v(\theta)]g(y)dy$$

measures the consumer's expected utility of the AI technology. The difference between the social planner and the profit-maximizing platform is that the former accounts for the effect

$$v_\theta^e(\theta) = [1 - G(-v(\theta)]V_\theta(q^u(\theta), \theta)$$

of a marginal improvement in AI on the surplus of consumer types on the intensive margin, whereas the platform does not account for this effect. AI investment will generally be distorted because of this externality.

The externality is generally of ambiguous sign because improved AI both has positive and negative effects on the consumer:

$$V_\theta(q, \theta) = -hN'(\phi)\Phi_\theta + CS'(\phi)\Phi_\theta = \sqrt{\frac{f}{2\sigma\phi}}\frac{2h - f}{4f}\frac{\Phi_\theta}{\phi} \gtreqless 0. \tag{20}$$

On the one hand, better AI improves the user experience by exposing the consumer to less advertising when the platform has more precise information. This is the first marginal effect. On the other hand, there is a loss in privacy rent because the platform can better tailor its advertising efforts with more precise information about the consumer's type. This is the second marginal effect. Applying the advertising model in Section 4 enables us to scrutinize this trade off. Substitution of the marginal marketing intensity, $N'(\phi) < 0$, derived from (5) and the marginal consumer surplus, $CS'(\phi) < 0$, derived from (7), delivers the rightmost expression in (20). Either of the two marginal effects in (20) can dominate depending on the magnitudes of underlying parameters. The marginal improvement in user experience dominates the marginal privacy loss if the marginal nuisance cost is sufficiently large relative to the marginal advertising cost, $2h > f$. The consumer prefers more to less AI in this case. The opposite is true if the marginal advertising cost is large relative to the marginal nuisance cost, $f > 2h$. The consumer is indifferent in the knife-edge case $2h = f$. Based on these results about the consumer externality, the following statements summarize the welfare effects of AI investment.

Proposition 4. *A monopoly that allows free access to and free unlimited usage of the platform invests too much in AI from a joint welfare perspective if the marginal nuisance cost of advertising is small relative to the marginal advertising cost, $2h < f$. The consumer's loss of privacy rent then dominates the improved user experience associated with more AI. There is underinvestment in AI in the opposite case, $2h > f$.*

Proof. Let θ^* maximize the expected welfare $w^e(\theta)$. The welfare difference between $\hat{\theta}$ and some arbitrary θ can be written as

$$w^e(\hat{\theta}) - w^e(\theta) = -\int_{\hat{\theta}}^{\theta} v_\theta^e(x)dx + \pi^e(\hat{\theta}) - \pi^e(\theta).$$

If $f > 2h$, then the first term on the right-hand side is strictly positive for all $\theta > \hat{\theta}$ by $v_\theta^e < 0$. Moreover, $\pi^e(\hat{\theta}) \geq \pi^e(\theta)$ for all θ since $\hat{\theta}$ maximizes the expected platform profit. Hence, $f > 2h$

implies $\theta^* \leq \hat{\theta}$. The inequality is strict because $w^e_\theta(\hat{\theta}) = v^e_\theta(\hat{\theta}) + \pi^e_\theta(\hat{\theta}) = v^e_\theta(\hat{\theta}) < 0$. If $2h > f$, then the first term on the right-hand side is strictly positive for all $\theta < \hat{\theta}$ by $v^e_\theta > 0$. Hence, $\theta^* \geq \hat{\theta}$ in this case. The inequality is strict by $w^e_\theta(\hat{\theta}) = v^e_\theta(\hat{\theta}) > 0$. □

The underlying logic is that investment in AI has an effect on the utility of consumers on the intensive margin that the platform fails to internalize because it cannot extract this rent through the platform tariff. The consumer gross utility can increase or decrease in the power of the prediction machine, which implies that AI investment can be too high or too low. Notice also that data privacy is efficient for all levels θ of AI since content consumption $q^u(\theta)$, and therefore data provision, maximizes the consumer's utility. Yet, information privacy, measured as $\Phi(q^u(\theta), \theta)$, is distorted because AI investment is inefficient.

Constrained advertising Underinvestment in AI occurs in our model if the marginal cost of advertising is small relative to the consumer's marginal nuisance cost of advertising. Advertising intensity goes to infinity when f converges to zero; see expression (5). Presumably there is an upper bound on the amount of advertising the platform can expose a consumer to. For instance, the attention span may limit the amount of advertising a consumer can digest. Advertising could also be constrained by a regulatory mandate. A constraint on advertising may substantially affect the incentives to invest in AI.

Impose an advertising constraint $N \leq \bar{N}$, where $\bar{N} \geq \frac{1}{\bar{v}\sigma}$. If this constraint is binding, then the advertising revenue/price of advertised goods becomes

$$R(\phi, \bar{N}) = \int_{z-\frac{1}{2\phi}}^{z+\frac{1}{2\phi}} P(\bar{N}\phi) = \bar{v} - \frac{1}{2\sigma\bar{N}\phi}.$$

Substituting the price $R(\phi, \bar{N})$ and advertising intensity $\bar{N}$ into (7) yields the consumer surplus

$$CS(\phi, \bar{N}) = \frac{1}{4\sigma\bar{N}\phi}$$

under constrained advertising. The total advertising surplus becomes

$$S(\phi, \bar{N}) = CS(\phi, \bar{N}) + R(\phi, \bar{N}) - (h + f)\bar{N} = \bar{v} - \frac{1}{4\sigma\bar{N}\phi} - (h + f)\bar{N}.$$

These expressions have the same qualitative properties as under unconstrained advertising. In particular, advertising revenue and total marketing surplus are both increasing functions of precision ϕ, whereas the consumer surplus is smaller when ϕ is larger. A constraint on advertising has no bearing on the qualitative results in the second and the third stage of the game, but has implications for AI investment.

The gross utility of platform usage under constrained advertising, $N = \bar{N}$, equals

$$V(q, \theta, \bar{N}) = U(q) - h\bar{N} + CS(\Phi(q, \theta), \bar{N}) = U(q) - h\bar{N} + \frac{1}{4\sigma\bar{N}\Phi(q, \theta)}$$

as a function of content consumption q and the power θ of AI. Investment in AI exerts an unambiguously negative externality on consumers, $V_\theta < 0$. The improvement in AI reduces privacy rent, but advertising intensity stays the same. The following result is immediate on the basis of the above discussion and therefore stated without proof.

Proposition 5. *A monopoly that allows free access to and free unlimited usage of the platform invests too much in AI from a joint welfare perspective if advertising on the platform is constrained by $\bar{N} \geq \frac{1}{\bar{v}\sigma}$ and the marginal advertising cost f is sufficiently small.*

This result shows that overinvestment in AI does not have to be conditional on a high monetary advertising cost if advertising is constrained in other dimensions.

Subsidization of platform participation Overinvestment in AI occurs only if subsidization of platform participation is suboptimal, so that subscription fees are non-negative. Suppose the platform and social planner can subsidize consumer participation without simultaneously attracting opportunists to the platform, so that neither is constrained by $y + V(q) \geq 0$. An example could be a contingent bonus to be used for buying goods advertised on the platform. Then the monopoly platform and social planner alike implement the content consumption $\hat{q}(\theta)$ that maximizes welfare $W(q, \theta)$ over the feasible subset $\mathcal{Q}$.

The monopoly invests in AI to maximize expected profit

$$\pi^*(\theta) = [1 - G(\hat{y}(\theta))][\hat{y}(\theta) + W(\hat{q}(\theta), \theta)] - \Psi(\theta).$$

In this expression, $\hat{y}(\theta)$ characterizes the participation threshold that maximizes the expected monopoly profit given the AI technology θ. The marginal effect of AI investment equals

$$\pi^*_\theta(\theta) = [1 - G(\hat{y}(\theta))]W_\theta(\hat{q}(\theta), \theta) - \Psi'(\theta).$$

Investing in AI is beneficial by increasing the total advertising surplus on the intensive margin, $W_\theta = S'(\phi)\Phi_\theta > 0$, which enables the monopoly to extract more rent from the consumer. The indirect effects working through content consumption $\hat{q}(\theta)$ and platform participation $\hat{y}(\theta)$ are of second-order importance to the monopoly in this case.

The social planner allows more participation than the platform and values also the utility of infra-marginal consumer types, so that the expected welfare of AI with power θ becomes

$$w^*(\theta) = \int_{-\max\{W(\hat{q}(\theta),\theta);\underline{y}\}}^{\bar{y}} [y + W(\hat{q}(\theta), \theta)] - \Psi(\theta).$$

The marginal net benefit

$$w^*_\theta(\theta) = [1 - G(-\max\{W_\theta(\hat{q}(\theta), \theta); \underline{y}\})]W_\theta(\hat{q}(\theta), \theta) - \Psi'(\theta) \geq \pi^*_\theta(\theta)$$

on expected welfare of investing in AI is weakly larger than the marginal net benefit for the monopoly because platform participation is weakly higher at the social optimum than in equilibrium, $\hat{y}(\theta) \geq -\max\{W_\theta(\hat{q}(\theta), \theta); \underline{y}\}$. The inequalities are strict if and only if there is incomplete participation in equilibrium, $\hat{y}(\theta) > \underline{y}$. We state the following result without proof.

Proposition 6. *The monopoly underinvests in AI from a joint welfare perspective if unconstrained subsidization of consumer participation is possible without simultaneously attracting opportunists to the platform.*

A comparison of Proposition 4 and Proposition 6 shows that the qualitative properties of the inefficiencies associated with AI investment depend crucially on the platform tariffs applied by the monopoly to incentivize participation and usage.

8 Too little quality?

We now consider quality investment in the first stage of the game. Investing in quality s increases both the consumer's direct utility $U(q, s)$ and marginal utility $U_q(q, s)$ of consuming content. We assume that the consumer has free access to and free unlimited usage of the platform also in this case, although this does not matter for the results.

The gross utility of content consumption is

$$V(q, s) = U(q, s) - hN(\Phi(q)) + CS(\Phi(q))$$

as a function of the quantity of content consumption q and the quality s of the platform service. We let $q^u(s)$ be the content consumption that maximizes consumer gross utility. Content consumption $q^u(s)$ increases in quality by the assumption that the marginal direct utility of content consumption is increasing in quality. We denote by $v(s) = V(q^u(s), s)$ the consumer's indirect utility of quality s. An increase in quality increases also participation by $v'(s) = U_s(q^u(s), s) > 0$.

The expected monopoly profit of offering quality s equals

$$\pi^e(s) = [1 - G(-v(s))]\Pi(q^u(s)) - \Upsilon(s).$$

The monopoly invests in quality to increase content consumption and platform participation:

$$\pi^e_s(s) = [1 - G(-v(s))]\Pi'(q^u)q^u_s + g(-v(s))\Pi(q^u)U_s(q^u, s) - \Upsilon'(s).$$

Assume that the profit-maximizing quality investment $\hat{s}$ is unique and positive.

The social planner chooses quality to maximize expected welfare $w^e(s) = v^e(s)+\pi^e(s)$, where

$$v^e(s) = \int_{-v(s)}^{\bar{y}} [y + v(s)]g(y)dy.$$

measures the consumer's expected utility of quality s. The platform's investment in quality is distorted because it ignores the marginal expected benefit

$$v_s^e(s) = [1 - G(-v(s))]U_s(q^u(s), s) > 0$$

on consumer surplus on the intensive margin of an increase in quality.

Proposition 7. *A monopoly that allows free access to and free unlimited usage of the platform invests too little in quality from a joint welfare perspective.*

Proof. Let s^* maximize the expected welfare $w^e(s)$. The welfare difference between $\hat{s}$ and arbitrary $s < \hat{s}$ can be written as

$$w^e(\hat{s}) - w^e(s) = \int_s^{\hat{s}} v_s^e(x)dx + \pi^e(\hat{s}) - \pi^e(s) > 0$$

by way of $v_s^e(s) > 0$ and $\pi^e(\hat{s}) \geq \pi^e(s)$. Hence, $s^* \geq \hat{s}$. The inequality is strict because $w_s^e(\hat{s}) = v_s^e(\hat{s}) + \pi_s^e(\hat{s}) = v_s^e(\hat{s}) > 0$. □

Investment in quality has a positive effect on infra-marginal consumer types that the platform fails to internalize because it only cares about the effect on the marginal consumer in the choice of quality.

9 Concluding remarks

The dominance of individual search engines, social media and streaming platforms has raised concerns about competition and consumer protection in markets where collection and analysis of large quantities of personal data represent the main source of income. The accelerating development of analytical tools built on artificial intelligence (AI) to utilize these data has accentuated these concerns.

Our study offers a framework within which to examine the potential sources of inefficiency and conflicts of interest that such digital markets give rise to. A distinguishing feature of our framework is the treatment of the main choices of consumers and firms as endogenous and mutually dependent. These choices include entry, content consumption and responses to advertising on the consumer side, and pricing, advertising, AI investments and content quality investments on the platform side. We analyze how margins of choice are jointly determined in equilibrium.

A main finding is that for a given technology, data harvesting will be efficient if free access and free unlimited usage represent a profit-maximizing platform tariff. Under a zero platform tariff, users consume content and generate user data in a quantity that for them strikes an optimal balance between the marginal improvement in user experience and the marginal loss in privacy rent. Thus, a business model in which users pay only with personal data for platform access and content consumption does not entail any direct welfare loss. The result suggests that a zero platform tariff in itself need not cause any economic harm to users that regulators should be concerned about.

However, the degree of consumer privacy will still be inefficient because the platform has distorted incentives to invest in AI to analyze available data. Under plausible assumptions, the monopoly firm over-invests and consequently extracts too much information about its users. Under a zero platform tariff the platform fails to internalize the negative effect of the compiled information on users along the intensive margin. This result indicates that policies to restrict the utilization of AI technology on zero-tariff digital platforms could be warranted.

The platform generally under-invests in quality under a zero platform tariff because the platform fails to internalize the value of improved quality on infra-marginal consumers.

Our analysis builds on the assumption that users rationally foresee the personal consequences of their platform use and associated provision of individual user data. Acemoglu et al. (2023) develop a model of users who are unable to foresee the consequences of providing individual data to the platform. We abstract away from data externalities across users and from platform competition. A string of recent papers (for instance Ichihashi, 2021; Bergemann et al., 2022 and Acemoglu et al., 2022) consider such data externalities. Extending our model in these directions would be fruitful, but we surmise that investments in AI and quality would be distorted also in those circumstances.

References

Acemoglu, Daron, "Harms of AI," *Oxford Handbook of AI Governance*, forthcoming.

_ , Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar, "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, November 2022, *14* (4).

_ , _ , _ , and _ , "A model of behavioral manipulation," Unpublished manuscript MIT, 2023.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman, "The economics of privacy," *Journal of Economic Literature*, 2016, *54* (2), 442–492.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Predicition Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press; Boston, Massachusetts, 2018.

Amelio, Andrea and Bruno Jullien, "Tying and freebies in two-sided markets," *International Journal of Industrial Organization*, 2012, *30*, 436–446.

Anderson, Simon P. and Bruno Jullien, "The advertising-financed business model in two-sided media markets," in Simon P. Anderson, Joel Waldfogel, and David Strmberg, eds., *Handbook of Media Economics*, Vol. 1a, Nort-Holland, 2015, chapter 2, pp. 41–90.

— **and Stephen Coate**, "Market provision of Broadcasting: A welfare analysis," *Review of Economics Studies*, October 2005, *72* (4), 947–972.

Armstrong, Mark, "Competition in two-sided markets," *RAND Journal of Economics*, Autumn 2006, *37* (3), 668–691.

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan, "The economics of social data," *RAND Journal of Economics*, Summer 2022, *53* (2), 263–296.

— **and** — , "Markets for information: An introduction," *Annual Review of Economics*, August 2019, *11*, 1–23.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim, "Privacy and personal data collection with information externalities," *Journal of Public Economics*, 2019, *173*, 113–124.

de Corniere, Alexandre and Romain de Nijs, "Online advertising and privacy," *RAND Journal of Economics*, Spring 2016, *47* (1), 48–72.

Dimakopoulos, Philipp D. and Slobodan Sudaric, "Privacy and platform competition," *International Journal of Industrial Organization*, 2018, *61*, 686–713.

Gans, Joshua S., "The specialness of zero," *Journal of Law and Economics*, February 2021, *64*, 157–176.

Hagiu, Andrei and Julian Wright, "Multi-sided platforms," *International Journal of Industrial Organization*, 2015, *43*, 162–174.

Ichihashi, Shota, "The economics of data externalities," *Journal of Economic Theory*, September 2021, *196*, 105316.

Jones, Charles I. and Christopher Tonetti, "Nonrivalry and the economics of data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Jullien, Bruno, Alessandro Pavan, and Marc Rysman, "Two-sided markets, pricing, and network effects," in Kate Ho, Ali Hortacsu, and Alessandro Lizzeri, eds., *Handbook of Industrial Organization*, Vol. 4, Nort-Holland, 2021, chapter 7, pp. 485–592.

Kox, Henk, Bas Straathof, and Gijsbert Zwart, "Targeted advertising, platform competition, and privacy," *Journal of Economics and Management Strategy*, Fall 2017, *26* (3), 557–570.

Lin, Tesary, "Valuing intrinsic and instrumental preferences for privacy," *Marketing Science*, 2022, *41* (4), 663–681.

Natvik, Gisle J. and Thomas P. Tangeras, "Paying with personal data: Online appendix," www.ifn.se/thomast 2023.

Reisinger, Markus, "Platform competition for advertisers and users in media markets," *International Journal of Industrial Organization*, 2012, *30*, 243–252.

Appendices

A Proof of Lemma 2

We do the proof in reverse order by first showing that no participant can strictly profit from engaging in insincere platform usage under the platform tariff $T(d)$ characterized in eq. (10). Assume that the consumer consumes platform content in quantity q and engages in insincere platform usage in quantity q_0. We obtain

$$y + V(q) - F - t \times \max\{q + q_0 - \hat{q}; 0\} \leq y + V(q) - F - t \times \max\{q - \hat{q}; 0\},$$

where the left-hand side of the inequality is the consumer utility of platform usage $\{q; q_0\}$ and the right-hand side is the consumer utility of $\{q; 0\}$. Hence, the consumer optimally sets $q_0 = 0$. If an opportunist engages in insincere platform usage in quantity q_0^o, then

$$-F - t \times \max\{q_0^o - \hat{d}; 0\} \leq -F$$

where the left-hand side of the inequality is the opportunist's utility of insincere platform usage q_0^o and the right-hand side is the utility of zero platform usage. Hence, the opportunist optimally sets $q_0^o = 0$.

We next show that the consumption of platform content never exceeds $\hat{d}$ subject to an appropriate choice of t. Let

$$t = \max\{\max_{q \in [0, q^u]} V'(q); 0\}.$$

The consumer utility equals $y + V(q) - F - t \times (q - \hat{d})$ for any $q > \hat{d}$. The consumer would never choose $q > \max\{\hat{d}; q^u\}$ because this would strictly reduce $V(q)$ and weakly increase the platform fee relative to content consumption q^u. This part of the proof is done if $\hat{d} \geq q^u$. Assume therefore that $\hat{d} < q^u$ and consider $q \in (\hat{d}, q^u]$. The difference between this strategy and $q = \hat{d}$ can be written as

$$y + V(q) - F - t \times (q - \hat{d}) - [y + V(\hat{d}) - F] = \int_{\hat{d}}^{q} (V'(x) - t)dx \leq 0$$

for all $q \in (\hat{d}, q^u]$ by the definition of t. Hence, $T(d)$ implements content consumption smaller than or equal to $\hat{d}$. The consumer chooses $q \in [0, \hat{d}]$ to maximize $y + V(q) - F$. Let $Q(\hat{d})$ be the maximal solution to this problem. In particular, $Q(\hat{d}) = \hat{d}$ if $\hat{d} \in \mathcal{Q}$ by Lemma 1.

To find an expression for the expected monopoly profit under $T(d)$, denote the participation threshold by $Y(\hat{d}, F)$. Specifically, $Y(\hat{d}, F) = \bar{y}$ if $\bar{y} + V(Q(\hat{d})) \leq F$ because the subscription fee is so high in this case that no consumer type wants to join the platform. At the other polar extreme, $Y(\hat{d}, F) = -\underline{y}$ if $-\underline{y} + V(Q(\hat{d})) \geq F$ because the subscription fee then is so small that all consumer types want to join the platform. For intermediary subscription fees it follows that $Y(\hat{d}, F) = F - V(Q(\hat{d}))$. We can then write the platform's expected profit as

$$[1 - G(Y(\hat{d}, F))][\Pi(Q(\hat{d})) + F] + \rho \min\{F; 0\}. \tag{21}$$

Observe that there are no costs associated with insincere platform usage in this expression.

The final step of the proof is to show that a tariff structure $T(d)$ characterized by (10) maximizes profit. To do so, consider an arbitrary platform tariff $\hat{T}(d)$ that implements consumer platform usage $\{\hat{q}; \hat{q}_0\}$, where $\hat{q} \in \mathcal{Q}$ and $\hat{q}_0 \geq 0$, and a consumer participation threshold $\hat{y} \in \mathcal{Y}$. Assume also that opportunists generate quantity $\hat{q}_0^o \geq$ of non-productive data. The expected platform profit is then given by expression (9). Consider now $T(d)$ of the form (10), where $\hat{d} = \hat{q}$ and $F = \hat{T}(\hat{q} + \hat{q}_0)$. This platform tariff implements content consumption $Q(\hat{q}) = \hat{q}$ since $\hat{q} \in \mathcal{Q}$. The consumer's utility of participating on the platform equals

$$y + V(Q(\hat{q})) - F = y + V(\hat{q}) - \hat{T}(\hat{q} + \hat{q}_0),$$

which is exactly the same as under $\hat{T}(d)$. Hence, $Y(\hat{q}, \hat{T}(\hat{q} + \hat{q}_0)) = \hat{y}$. Inserting this participation threshold and $Q(\hat{q}) = \hat{q}$ into (21) returns the expected profit

$$[1 - G(\hat{y})][\Pi(\hat{q}) + \hat{T}(\hat{q} + \hat{q}_0)] + \rho \min\{\hat{T}(\hat{q} + \hat{q}_0); 0\}$$

of non-linear tariff $T(d)$ where $\hat{d} = \hat{q}$ and $F = \hat{T}(\hat{q} + \hat{q}_0)$. This profit is at least as large as the expected platform profit (9) because there is no insincere platform usage, $q_0 = q_0^o = 0$, under the platform tariff $T(d)$, and the monopoly subsidizes opportunists by a relatively smaller amount, $\min\{\hat{T}(\hat{q} + \hat{q}_0); 0\} \geq \min\{\hat{T}(\hat{q}_0^o); 0\}$.$\square$

B Proof of Proposition 2

We show that $T(d) = 0$ for all $d \geq 0$ is a profit-maximizing tariff if and only if conditions (13) and (14) are both met.

(i) Necessity of (13). Suppose the monopoly charges $T(d) = 0$ for all $d \geq 0$, but $\Pi'(q^u) < 0$. The expected profit of the monopoly equals

$$\Pi^e(-V(q^u), q^u) = [1 - G(-V(q^u))]\Pi(q^u)$$

under the zero tariff. Observe that $V(q) \geq V(q')$ for all $q' \in [0, q]$ and all q in a neighborhood of q^u because q^u would not be a strict maximum of $V(\cdot)$ otherwise. By implication, $q \in \mathcal{Q}$ for all q in a non-degenerate interval $[\tilde{q}, q^u]$. Consider a deviation to the tariff (10) with zero subscription fee, $F = 0$, but a user limit $\hat{d} \in [\tilde{q}, q^u)$ and a prohibitive marginal tariff $t > 0$ for all platform usage $d > \hat{d}$. This alternative tariff yields expected profit

$$\Pi^e(-V(\hat{d}), \hat{d}) = [1 - G(-V(\hat{d}))]\Pi(\hat{d})$$

Seeing as

$$\lim_{\hat{d} \to q^{u-}} \frac{d}{d\hat{d}} \Pi^e(-V(\hat{d}), \hat{d}) = [1 - G(-V(q^u))]\Pi'(q^u) < 0$$

by $V'(q^u) = 0$ and the assumption that $\Pi'(q^u) < 0$, a strictly profitable deviation from the zero tariff then exists.

(*ii*) Necessity of (14). Suppose the monopoly charges $T(d) = 0$ for all $d \geq 0$, but (14) is strictly violated. Consider a deviation to the tariff (10) with positive subscription fee, $F > 0$, and zero marginal tariff, $t = 0$. This tariff yields expected profit

$$\Pi^e(F - V(q^u), q^u) = [1 - G(F - V(q^u))][F + \Pi(q^u)] \tag{22}$$

Seeing as

$$\lim_{F \to 0^+} \frac{d}{dF} \Pi^e(F - V(q^u), q^u) = 1 - G(-V(q^u)) - g(-V(q^u))]\Pi(q^u) < 0$$

by the assumption that (14) is strictly violated, a strictly profitable deviation from the zero tariff then exists.

(*iii*) A zero marginal tariff, $t = 0$, is part of a profit-maximizing tariff under assumption (13). Assume that the profit-maximizing tariff (10) features subscription fee $\hat{F} \geq 0$, implements platform usage $\hat{q} \in \mathcal{Q}$ and participation threshold $\hat{y} = \hat{F} - V(\hat{q}) \in \mathcal{Y}$. This tariff yields expected profit

$$\Pi^e(\hat{y}, \hat{q}) = [1 - G(\hat{y})][\hat{y} + W(\hat{q})]$$

where, recall, $W(q) = V(q) + \Pi(q)$ characterizes ex-post welfare of content consumption. Consider now the alternative tariff (10) with subscription fee $F = V(q^u) - V(\hat{q}) + \hat{F} \geq \hat{F}$ and where $t = 0$ so that all platform usage is free and unlimited. The consumer consequently chooses content consumption q^u. The utility of participating on the platform conditional on optimal content consumption becomes

$$y + V(q^u) - F = y + V(\hat{q}) - \hat{F} = y - \hat{y}.$$

Hence, the alternative tariff implements participation threshold $\hat{y}$. The tariff thus generates expected profit

$$\Pi^e(\hat{y}, q^u) = [1 - G(\hat{y})][\hat{y} + W(q^u)] \geq \Pi^e(\hat{y}, \hat{q}).$$

$W(q^u)$ is at least as large as $W(\hat{q})$ because q^u maximizes $V(q)$ and $\Pi(q^u) \geq \Pi(q)$ for all $q \leq q^u$ by strict concavity of $\Pi(q)$ and the assumption that $\Pi'(q^u) \geq 0$.

(*iv*) Zero subscription fee, $\hat{F} = 0$, is part of a profit-maximizing tariff if $t = 0$ and assumption (13) holds. A tariff (10) with subscription fee $F \in [0, \bar{y} + V(q^u)]$ and marginal tariff $t = 0$ generates expected profit (22). This profit function is strictly quasi-concave in F by the monotone hazard rate property of $G(y)$. Under assumption (13), $\lim_{F \to 0} \frac{d}{dF}\Pi^e(F - V(q^u), q^u) \leq 0$, in which case $\hat{F} = 0$ maximizes (22) in the domain $F \geq 0$.

If the inequality in condition (14) is strictly reversed while (13) holds, then we know from (*iii*) that a zero marginal tariff maximizes profit and from (*ii*) that the profit-maximizing subscription fee is positive. If inequality (13) is strictly reversed, then we know from (*i*) that profit-maximizing content consumption $\hat{q} \in \mathcal{Q}$ satisfies $\hat{q} < q^u$. The platform can implement $\hat{q}$ by offering free platform usage up to $\hat{q}$, and charging a positive marginal fee, $t > 0$, on usage thereafter. Any strictly positive profit-maximizing subscription fee $\hat{F}$ is characterized by the solution (15) to the platform's first-order condition $\Pi^e_F(\hat{F} - V(\hat{q}), \hat{q}) = 0$ for profit maximization with respect to the subscription fee.

We can combine the first-order condition $V'(q^u) = 0$ for the consumer's optimal content consumption, where $V'(q)$ was characterized in (8), with the platform's marginal profit expression $\Pi'(q^u)$ identified in (16) and the marginal effects, $R'(\phi)$, $N'(\phi)$ and $CS'(\phi)$, from the model of personalized advertising in Section 4 to get

$$(\frac{f-2h}{4f})^2(\Pi'(q^u) + c) \ = \frac{f-2h}{4f}U'(q^u) = (\frac{f-2h}{4f})^2\sqrt{\frac{f}{2\sigma\Phi(q^u)}\frac{\Phi'(q^u)}{\Phi(q^u)}} > 0, \ f \neq 2h,$$

after simplification. Content consumption q^u is independent of the marginal data cost, so that condition (13) is satisfied if and only if c is sufficiently small. Content consumption q^u becomes very small or very large, depending on the sign of $f - 2h$, if σ goes to zero so that product differentiation in the advertising market becomes very large. Either way, the left-hand side of the above expression becomes very large. Based on these comparative statics results, we conclude that the profit-maximizing tariff features free unlimited data usage for instance if the marginal data cost is small or product differentiation in the advertisement market is large. □

C Proof of Proposition 3

The expected welfare $W^e(y, q)$ is strictly decreasing in y. Hence, $y = -V(q)$. The social planner then maximizes $W^e(-V(q), q)$ over $q \in \mathcal{Q}$ to derive the ex ante efficient platform usage $\hat{q}^*$. The corresponding efficient participation threshold is $\hat{y}^* = -V(\hat{q}^*)$.

Compare the efficient allocation $\{\hat{y}^*; \hat{q}^*\}$ to the allocation $\{\hat{y}; \hat{q}\}$ that maximizes $\Pi^e(y, q)$ over $y \in \mathcal{Y}$ and $q \in \mathcal{Q}$, subject to $y + V(q) \geq 0$. We first show that $\hat{q} \leq \hat{q}^*$. The chain of inequalities, $\hat{y} \geq -V(\hat{q}) \geq -V(\hat{q}^*) = \hat{y}^*$, then imply $\hat{y} \geq \hat{y}^*$. Assume first that $\hat{y} + V(\hat{q}) > 0$. In this case,

$$W^e(-V(\hat{q}), \hat{q}) - W^e(-V(q), q) = \int_{-V(\hat{q})}^{-V(q)} [y + W(\hat{q})]g(y)dy + [1 - G(-V(q))][W(\hat{q}) - W(q)] \geq 0$$

for all $q \in \mathcal{Q}$ such that $q < \hat{q}$. The first expression is non-negative because $V(\hat{q}) \geq V(q)$ for all $(q, \hat{q}) \in \mathcal{Q}^2$ such that $\hat{q} > q$ and because

$$y + W(\hat{q}) > -V(\hat{q}) + W(\hat{q}) = \Pi(\hat{q}) \geq 0 \; \forall y > -V(\hat{q}).$$

The second expression is non-negative because $\hat{q} \in \arg\max_{q \in \mathcal{Q}} W(q)$ if $\hat{y} > -V(\hat{q})$. We conclude that $\hat{q}^* \geq \hat{q}$ if $\hat{y} + V(\hat{q}) > 0$.

Assume next that $\hat{y} + V(\hat{q}) = 0$. The difference

$$W^e(-V(\hat{q}), \hat{q}) - W^e(-V(q), q) = V^e(-V(\hat{q}), \hat{q}) - V^e(-V(q), q) + \Pi^e(-V(\hat{q}), \hat{q}) - \Pi^e(-V(q), q)$$

in expected welfare between $\hat{q}$ and $q < \hat{q}$, where $q \in \mathcal{Q}$, is non-negative if $V^e(-V(\hat{q}), \hat{q}) \geq V^e(-V(q), q)$ because $\Pi^e(-V(\hat{q}), \hat{q}) \geq \Pi^e(-V(q), q)$ by the assumption that $\hat{y} = -V(\hat{q})$ and $\hat{q}$ maximize the platform's expected profit. Seeing as

$$V^e(-V(\hat{q}), \hat{q}) - V^e(-V(q), q) = \int_{-V(\hat{q})}^{-V(q)} [y + V(\hat{q})] g(y) dy + [1 - G(-V(q))][V(\hat{q}) - V(q)] \geq 0$$

by $V(\hat{q}) \geq V(q)$ for all $(q, \hat{q}) \in \mathcal{Q}^2$ such that $\hat{q} > q$, we conclude that $\hat{q}^* \geq \hat{q}$ even if $\hat{y} + V(\hat{q}) = 0$.

If free access and free unlimited usage maximize the platform's expected profit, then $\hat{q} = \hat{q}^* = q^u$ by $\hat{q} \leq \hat{q}^* \leq q^u$ and $\hat{q} = q^u$. Moreover, $\hat{y} = -V(\hat{q}) = -V(\hat{q}^*) = \hat{y}^*$.

The consumer would like to set the subscription fee as low as possible for any q to obtain expected utility $V^e(-V(q), q)$. Seeing as the expected utility is non-decreasing in $q \in \mathcal{Q}$, the consumer would like to implement q^u. Unsurprisingly, the consumer prefers zero subscription fee and unlimited free access in the class of non-negative platform tariffs.□