WORKING PAPERS

N° 1404

January 2023

“Data and Competition: A Simple Framework”

Alexandre de Cornière and Greg Taylor

Toulouse School of Economics

Data and Competition: A Simple Framework*

Alexandre de Cornière† and Greg Taylor‡

January 27, 2023

Abstract

Does enhanced access to data foster or hinder competition among firms? Using a competition-in-utility framework that encompasses many situations where firms use data, we model data as a revenue-shifter and identify two opposite effects: a mark-up effect according to which data induces firms to compete harder, and a surplus-extraction effect. We provide conditions for data to be pro- or anti-competitive, requiring neither knowledge of demand nor computation of equilibrium. We apply our results to situations where data is used to recommend products, monitor insuree behavior, price-discriminate, or target advertising. We also revisit the issue of data and market structure.

Keywords: competition, data, price discrimination, targeted advertising, market structure.
JEL Classification: L1, L4, L5.

1 Introduction

Data has become one of the most important issues in the debate about competition and regulation in the digital economy.[1] But does the use of data by firms make markets more or less competitive?

*We are grateful to Mark Armstrong, Paul Belleflamme, Daniele Condorelli, Dianne Coyle, Vincenzo Denicolò, George Georgiadis , Bruno Jullien, Volker Nocke, Martin Peitz and Yossi Spiegel for useful conversations and suggestions. Thanks are also due to participants at various seminars and conferences for useful comments and discussions. De Cornière acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program), the TSE Digital Center, and the Artificial and Natural Intelligence Toulouse Institute (ANITI). Taylor acknowledges the financial support of the Digital Economics Research Network.

†Toulouse School of Economics, University of Toulouse Capitole; alexandre.de-corniere@tse-fr.eu; https://sites.google.com/site/adecorniere

‡Oxford Internet Institute, University of Oxford; greg.taylor@oii.ox.ac.uk; http://www.greg-taylor.co.uk

[1] For reports dealing with this issue, see Crémer et al. (e.g., 2019), Furman et al. (2019), and Scott Morton et al. (2019). An example hearing on the topic is the FTC's recent Hearing on Privacy, Big Data, and Competition, see `https://www.ftc.gov/news-events/events-calendar/ftc-hearing-6-competition-consumer-protection-21st-century`, accessed 1 May 2019.

On the one hand, data is a source of efficiencies. It enables firms to offer new or better products, to make personalized recommendations to consumers, or to improve monetization opportunities. On the other hand, observers have raised many concerns. One class of concerns reflects fears of exploitative behavior such as privacy violations, price-discrimination, and more generally excessive surplus extraction.[2] A second set of concerns encompass adverse implications for market structure, such as raising barriers to entry or creating winner-take-all situations (see, e.g., Furman et al., 2019, 1.71 to 1.79).

One challenge in studying the competitive effects of data lies in the variety of its uses, from targeted advertising to customized product recommendations to personalized pricing. Surprisingly, while many recent papers study markets in which firms can collect, trade, or use consumer data in various ways (see our literature review below), we are not aware of any attempt at systematically categorizing situations depending on whether data plays a pro- or an anti-competitive role.[3] Our first contribution in this paper is to provide such a characterization. To do so, we use a simple model of competition-in-utility à la Armstrong and Vickers (2001), where each firm chooses the mean utility u it provides to consumers. This approach is flexible enough to encompass various business models, such as price competition (with uniform or personalized prices), ad-supported business models, or competition in quality. We model data as a factor δ that generates more revenues for a given level of utility provided, a natural property across many uses of data (we provide several microfoundations in Section 4). This might be because data can be used to increase the surplus created by a product (e.g., through better personalization) or because the data can be used to extract a bigger share of the surplus (e.g., through price discrimination) or both. Formally, we assume that the mark-up of a firm takes the form $r(u, \delta)$, increasing in δ (we later show how this reduced form can be given an informational microfoundation).

Our first main result characterizes environments where data is *unilaterally pro-competitive*, in the sense that a better dataset induces a firm to offer more utility to consumers, keeping its rivals' offers fixed (i.e. the firm's best-response in the utility space shifts upwards). Data is *unilaterally anti-competitive* when it shifts the best response downwards. We highlight a potential trade-off between two effects. The first is the *mark-up effect*: because data increases firms' mark-ups, it also induces them to compete more fiercely to attract consumers. The second effect, which we call the *surplus extraction effect*, is more ambiguous: depending on the way it is used, data may enable firms to extract or on the contrary to provide consumer surplus more efficiently. We then show that, in many cases, the overall competitive effect of data can be determined without

[2]E.g., Scott Morton et al. (2019), p.37: "[Big Data] enables firms to charge higher prices (for goods purchased and for advertising) and engage in behavioral discrimination, allowing platforms to extract more value from users where they are weak."

[3]This statement does not apply to the literature on competitive price-discrimination, as reviewed for instance by Stole (2007).

having to compute the equilibrium or to make functional form assumptions about demand, as it depends only on the shape of the mark-up function $r(u, \delta)$, and in particular on whether it is super-modular.

One attractive property of the competition-in-utility model is that it can accommodate both strategic complementarity and substitutability, depending (as we show) on whether firms' revenues come at the expense of consumers or not (i.e. the sign of $\partial r/\partial u$). Since this strategic effect and the earlier unilateral effect can both be characterized from the per-consumer revenue function r, we obtain sufficient statistics for the equilibrium competitive effects of data directly from model primitives without the need to compute equilibrium. We highlight the implications for policies such as mandated data sharing for an incumbent or more stringent constraints on data collection.

To encompass many different business models and uses of data, our model is initially somewhat abstract, especially in its treatment of data and revenues. We therefore make the application to data more concrete by giving it a microfoundation in four models of markets where data plays an important role. Besides showing how various uses of data can be cast into our framework, this allows us to show how the trade-offs described above play out differently when data is used in different ways.

In the first application, data is used to improve the quality of firms' products, for instance by enabling better product recommendations. We show that our framework nests a model where Bayesian recommenders learn about consumers' tastes from past customers' feedback. In such situations, only the mark-up effect operates, and data is unilaterally pro-competitive. Secondly, we study a model of moral hazard where insurance providers have access to data on the agent's effort. Data mitigates the hidden information problem, reducing the opportunity cost of providing utility. The surplus extraction effect therefore runs in consumers' favour, which means data is pro-competitive. Thirdly, we consider a price discrimination model where multi-product retailers can use data to learn about consumers' willingness to pay for some products. By reducing the deadweight loss, data improves the efficiency of surplus extraction, and this effect tends to dominate the mark-up effect, making data anti-competitive. Lastly, in a model of targeted advertising on platforms, we show that whether data is pro- or anti-competitive is determined by how it affects the elasticity of advertisers' demand. This, in turn, depends on whether the data mostly contains information about consumers' match to a broad product category, or to specific advertisers. The overall takeaway is that different uses of data produce starkly different predictions about its competitive effects. Nevertheless, in each of these cases those effects can be decomposed into mark-up and surplus extraction effects, and are easily characterized using the simple conditions from our baseline analysis.

We next consider the endogenous process by which data is collected. To do so, we embed the static model into a dynamic framework where data generated by a sale in one period can be used in the next. We address the question of whether data is a barrier to entry and can form part of an entry deterrence strategy. We show that there is a data barrier to entry if and only if data is unilaterally pro-competitive, allowing us to apply the supermodularity condition derived earlier. Similarly, in a simple model of competition over an infinite horizon with very impatient firms, we find that data can lead to long-run concentration only if it is unilaterally pro-competitive. These results highlight a tension between static (exploitative) and dynamic (exclusionary) concerns. Dynamic concerns arise precisely when data is not used in a statically exploitative way and vice-versa. Our model therefore provides a guide on when each theory of harm is most relevant.

The organization of the paper is as follows: after discussing the related literature, we present the basic framework in Section 2. In Section 3 we derive conditions for data to be unilaterally pro- or anti-competitive. We apply these conditions to four microfounded models of markets with data use in Section 4 to show how the unilateral effects of data can be determined. We extend the unilateral analysis to study the equilibrium effects of data in Section 5, which also allows us to study some dynamic issues in Section 6. Section 7 discusses the model and shows how the analysis can be extended to incorporate consumer privacy concerns and data externalities. We conclude in Section 8.

Related Literature

Data takes many forms and has many different users and uses (Acquisti et al., 2016). Much of the literature has therefore focused on the study of particular applications of data (see Pino, forthcoming, for a survey). For example, one active literature considers the consequences of allowing firms to use data for personalized pricing (e.g., Thisse and Vives, 1988; Fudenberg and Tirole, 2000; Taylor, 2004; Acquisti and Varian, 2005; Calzolari and Pavan, 2006; Anderson et al., 2016; Belleflamme and Vergote, 2016; Kim et al., 2018; Montes et al., 2018; Bonatti and Cisternas, 2019; Gu et al., 2019; Chen et al., 2020; Ichihashi, 2020; Bounie et al., 2021). Another literature studies targeted advertising (e.g., Roy, 2000; Iyer et al., 2005; Galeotti and Moraga-González, 2008; Athey and Gans, 2010; Bergemann and Bonatti, 2011; Rutt, 2012; Johnson, 2013; Bergemann and Bonatti, 2015; de Cornière and de Nijs, 2016). These papers provide a rich picture of how data affects market outcomes in particular institutional environments. However, that picture is complex, with data sometimes being pro-competitive, but reducing consumer surplus on other occasions. Our contribution is to develop a framework that allows us to systematically characterize the competitive effects of data while remaining agnostic about how the data is used. We stress that we do not aim to nest all extant models—the variety of modelling approaches is too great—but we do offer a model that reflects some of the most important trade-offs and shows how they play out in different contexts.

One important theme in the policy debate concerns the relationship between data use or accumulation and market structure. Recent papers such as Farboodi et al. (2019), Prüfer and Schottmüller (2021) and Hagiu and Wright (forthcoming) study long-run market dynamics when data-enabled learning helps firms improve their products, and emphasize the potential for data to lead to increased concentration (this is related to earlier work on learning-by-doing, e.g., Dasgupta and Stiglitz, 1988; Cabral and Riordan, 1994).[4] In Section 6 we apply our framework to this question, and show that the way in which data is used can have a significant effect on its implications for market dynamics. On a related note, some commentators have argued that data may create a barrier to entry (e.g., Grunes and Stucke, 2016). Building on the classic analysis of Fudenberg and Tirole (1984) (see also Bulow et al., 1985), we use our framework to show that the viability of an entry-deterrence strategy also depends on how the data is used.

2 Model

Demand We consider a market with $n \geq 1$ firms. As in Armstrong and Vickers (2001), each firm chooses a mean utility level $u_i \in \mathbb{R}$, resulting in demand $D_i(u_i, \mathbf{u}_{-i})$, where $\mathbf{u}_{-i}$ are the mean utilities available from other firms and the outside option. Depending on the context, u_i may depend on firm i's price, on its quality, or on any of its strategic choices, such as the "ad load" that a media firm imposes on viewers for instance. We provide several illustrative examples in Section 4. Demand is assumed to be continuously differentiable, and such that $\frac{\partial D_i(u_i,\mathbf{u}_{-i})}{\partial u_i} \geq 0$ and $\frac{\partial D_i(u_i,\mathbf{u}_{-i})}{\partial u_j} \leq 0$ for $j \neq i$. [5]

Mark-up and fixed costs Firms' marginal cost is constant and normalized to zero. The choice of a mean utility u_i determines firm i's per-consumer revenue (which is also the mark-up), $r(u_i)$, which we assume is continuously differentiable.

The fixed cost of choosing u_i is $C(u_i)$, with $C'(u_i) \geq 0$ and $C''(u_i) \geq 0$.[6]

Data Each firm has access to data containing strategically relevant information about the market. Data need not be quantitative, and may include qualitative or unstructured elements. The quality of the data may vary with the number of variables or observations it contains, or with the relevance, accuracy or recency of those observations. To reflect the differing qualities of datasets, we assume that they can be ranked such that a better (e.g., more informative) dataset allows the firm to generate more revenue per-consumer

[4] See, also, Campbell et al. (2015), Lam and Liu (2020) for theoretical studies of how data regulations may affect market structure, and Johnson et al. (2021) for a related empirical study on the effects of European privacy regulations.

[5] Such a formulation is consistent with discrete choice models such that the utility that consumer l obtains from firm i is of the form $u_{il} = u_i + \epsilon_{il}$, where ϵ_{il} is a random taste shock. In the nested logit model, for instance, we have $u_i = x_i\beta - \alpha p_i + \xi_i$ where x_i is a vector of product characteristics, p_i is the price, and ξ_i an unobservable (to the econometrician) shock. Such a model can also be interpreted as one with a representative consumer with taste for diversity (Anderson et al., 1988).

[6] In Armstrong and Vickers (2001), $C(u_i) = 0$, which holds when u_i depends on firm i's price only. With investments in quality, one may have $C'(u_i) > 0$.

for any level of utility. This guarantees that it is possible to represent each dataset by a score, $\delta_i \in \mathbb{R}$, such that the associated mark-up, $r(u_i, \delta_i)$, is increasing in δ_i.[7] Given this representation, we take r as a primitive and perform our analysis using δ_i rather than the dataset it represents.

Assumption 1. A firm with a better dataset (i.e., a higher δ_i) achieves a higher mark-up for any given utility level provided to consumers: $\frac{\partial r(u_i,\delta_i)}{\partial \delta_i} > 0$.

We often say a firm with a higher δ_i has 'more' data, even though a larger δ_i might actually correspond to a more informative dataset of equal size. We can think of r as capturing the technology of data use, and we will see that different ways of using data—such as targeted advertising or price discrimination—generate quite different rs, and thus imply different effects of data.[8]

To give a simple example, suppose that the mean utility has the form $u_i = V(\delta_i) - p_i$, where $V(\delta_i)$ is consumers' valuation for product i, which we assume is increasing in the quality of firm i's data, and where p_i is product i's price. Then we have $r(u_i, \delta_i) = p_i = V(\delta_i) - u_i$. We provide a microfoundation for $V'(\delta_i) > 0$ in Section 4, along with other examples that satisfy Assumption 1.

There are two ways to interpret δ_i. Firstly, it might measure the aggregate data held by i about the overall population of consumers. Having such data might enable the firm to provide a better offer to all consumers as, for example, when a search engine provides better results for queries it has seen before. Alternatively, δ_i might measure the amount of data the firm has about a single specific consumer, in which case u_i is interpreted as a personalized offer to that consumer and each consumer is treated as a separate market, buying from i with probability $D_i(u_i, \mathbf{u}_{-i})$.

Of course, the data used by firms is often personal data, raising potential concerns around privacy or data externalities between consumers. We abstract away from intrinsic privacy concerns in the main model, but show how these issues can be incorporated into the analysis in Section 7.

Firms simultaneously choose their u_i to maximize profit

$$\pi_i(u_i, \mathbf{u}_{-i}, \delta_i) = r(u_i, \delta_i) D_i(u_i, \mathbf{u}_{-i}) - C(u_i), \tag{1}$$

which we assume to be quasi-concave in u_i for any $\mathbf{u}_{-i}$, δ_i. Sufficient conditions for this are (i) that C is sufficiently convex, or (ii) that both r and D_i are log-concave in u_i.

[7]Let the set of all datasets be Ω and the per-consumer revenue associated with $\omega \in \Omega$ be $\tilde{r}(u, \omega)$. Then, so long as better datasets are associated with higher revenue, one example of a valid representation, $\delta : \Omega \to \mathbb{R}$, is $\delta(\omega) = \tilde{r}(0, \omega)$. For any valid representation we have $r(u_i, \delta_i) := \tilde{r}(u_i, \delta^{-1}(\delta_i))$.

[8]Data might also lower the fixed cost. If data reduces the incremental fixed cost of providing utility, $\frac{\partial^2 C_i}{\partial u_i \partial \delta_i} \leq 0$, then this effect in isolation unambiguously leads the firm to offer higher utility so data would more often be pro-competitive. The statement of Proposition 1 below, though, would remain unchanged.

Occasionally, and where no confusion results, we write $D_i \equiv D_i(u_i, \mathbf{u}_{-i})$ and $r_i \equiv r(u_i, \delta_i)$ for conciseness.

3 Unilateral effects of data and monopolists' incentives

We begin by studying how data affects firms' unilateral incentives to offer utility. Let $\hat{u}_i(\mathbf{u}_{-i}, \delta_i)$ be firm i's best-response function. We use the following definition.

Definition 1. We say that data is *unilaterally pro-competitive* (UPC) for firm i for a given $\mathbf{u}_{-i}$ if $\frac{\partial \hat{u}_i(\mathbf{u}_{-i},\delta_i)}{\partial \delta_i} > 0$. We say that data is *unilaterally anti-competitive* (UAC) when the inequality is reversed.

This notion of pro- or anti-competitiveness of data captures the "unilateral" effect of data: data is UPC if better data induces a firm to offer more utility to consumers, keeping any rivals' utility offers constant. It therefore fully characterises how a monopolist responds to a change in the data available, as well as being an important ingredient in the competitive equilibrium analysis to follow. While Definition 1 is a local property, we will see in section 4 that in many applications data is UPC or UAC more globally.

Given the expression for firm i's profit, (1), its best response function, $\hat{u}_i(\mathbf{u}_{-i}, \delta_i)$, is found as the solution to its first-order condition:

$$\frac{\partial \pi_i(u_i, \mathbf{u}_{-i}, \delta_i)}{\partial u_i} = \frac{\partial r(u_i, \delta_i)}{\partial u_i} D_i(u_i, \mathbf{u}_{-i}) + \frac{\partial D_i(u_i, \mathbf{u}_{-i})}{\partial u_i} r(u_i, \delta_i) - \frac{\partial C(u_i)}{\partial u_i} = 0. \quad (2)$$

By standard arguments, firm i's best-response is increasing in δ_i if and only if $\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} > 0$. Differentiating (2) with respect to δ_i, the condition $\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} > 0$ can be rewritten as:

$$\frac{\partial D_i(u_i, \mathbf{u}_{-i})}{\partial u_i} \frac{\partial r(u_i, \delta_i)}{\partial \delta_i} + \frac{\partial^2 r(u_i, \delta_i)}{\partial u_i \partial \delta_i} D_i(u_i, \mathbf{u}_{-i}) > 0. \quad (3)$$

Data affects the incentive to provide utility in two ways. Firstly, an extra unit of data increases the mark-up earned from an additional consumer and therefore the incentive to attract consumers with high utility offers. This *mark-up effect* corresponds to the first term in (3), which is always positive. Secondly, data may affect the opportunity cost (or benefit) of providing utility to a consumer. For example, the opportunity cost of showing consumers fewer ads is higher the more precisely targeted the foregone ads would have been. This gives rise to the second term in (3), whose sign is ambiguous. This second term can also be interpreted as a *surplus extraction effect*: when $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i}$ is negative, data makes the firm more efficient at extracting surplus from consumers. Equation (3) thus reveals that a sufficient condition for data to be UPC is that r be supermodular, $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} \geq 0$.

When $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} < 0$, data simultaneously increases the value of each extra consumer and makes surplus extraction more attractive, so that its overall effect may be UPC or UAC. One way to make further progress is to consider the case where the fixed cost is constant, i.e. $C'(u_i) = 0$ (see Section 4 for several natural examples). Then we can substitute the first-order condition, $r_i \frac{\partial D_i}{\partial u_i} + \frac{\partial r_i}{\partial u_i} D_i = 0$, into (3) and obtain that data is UPC if and only if $r_i \frac{\partial^2 r_i}{\partial u_i \partial \delta_i} > \frac{\partial r_i}{\partial u_i}\frac{\partial r_i}{\partial \delta_i}$, which is equivalent to $\frac{\partial^2 \ln(r_i)}{\partial u_i \partial \delta_i} > 0$. We summarize this discussion in the following proposition (whose proof is in Appendix A):

Proposition 1. *(i) If r is supermodular ($\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} \geq 0$) then data is unilaterally pro-competitive for firm i for all $\mathbf{u}_{-i}$.*

(ii) When fixed costs are constant, data is unilaterally pro-competitive for firm i for all $\mathbf{u}_{-i}$ if and only if r is log-supermodular ($\frac{\partial^2 \ln(r(u_i,\delta_i))}{\partial u_i \partial \delta_i} > 0$).

An interesting feature of Proposition 1 is that the conditions do not depend on the demand function D_i. Moreover, because these primitive conditions hold for all $\mathbf{u}_{-i}$, one does not have to compute the equilibrium to be able to determine whether data is UPC or UAC.[9] This is particularly valuable in setups with more than two potentially asymmetric firms, where explicitly computing the equilibrium might prove impossible. Instead, what is most important is the economic technology, $r(u_i, \delta_i)$, that connects data, utility, and revenue.

So far we have treated the technology in a deliberately abstract fashion in order to accommodate as many business models and uses of data as possible. Consequently, Proposition 1 is a relatively general statement that can be applied to factors other than data that increase mark-ups (such as the stock of cost-reducing innovations). One thing that makes the application to data economically interesting is just how naturally it generates both pro- and anti-competitive effects because different but plausible uses of data resolve the trade-off between mark-up and surplus extraction effects quite differently.[10] To see this, it will now be necessary to make the application to data more concrete and consider several microfounded models of situations where data is used.

4 Applications

In this section we discuss four applications that build on established models of product improvement, moral hazard, price discrimination, and targeted advertising by media platforms. In each case we can use Proposition 1 to quickly characterize the effects of data.

[9] This property is somewhat reminiscent of the sufficient statistics approach in public economics (Chetty, 2009).

[10] In contrast, a textbook model of cost-reducing innovations would normally be pro-competitive. Indeed, in a competition-in-utility framework, Shelegia and Wilson (forthcoming) provide several examples of revenue-shifting technologies, all of which would be pro-competitive in our framework.

More substantively, these applications also allow us to decompose the markup and surplus extraction effects for different uses of data, and thereby provide economic intuition for why data has different competitive implications in these different uses. To show the spectrum of possibilities, we consider in turn a case where the surplus extraction effect is absent, positive, negative (and dominant), and ambiguous. Additionally, this section also shows how the reduced form treatment of data can be given an informational microfoundation.

In each application firms serve $D_i(u_i, \mathbf{u}_{-i})$ consumers. But, because the effects of data can be characterized independently of demand or equilibrium, we focus here on the properties of the mark-up function $r(u_i, \delta_i)$.

4.1 Product improvement

An important use of data is to improve the quality of the products or services offered by firms based on the feedback or choices of past customers. For instance, search engine algorithms use data about past queries to improve their results. This improvement can also take the form of more personalized recommendations without affecting the quality of the underlying products: a movie streaming service suggesting shows to its users based on the viewing history of others like them, or an online retailer suggesting products to consumers based on past purchases.

Suppose that consumers have unit demand, buying from i with probability $D_i(u_i, \mathbf{u}_{-i})$. Firm i chooses price p_i, resulting in utility $u_i = V + f(\delta_i) - p_i$. Here, V is the standalone value of the product and $f'(\delta_i) > 0$ captures the idea that data allows the firm to offer a better product. We provide a Bayesian microfoundation immediately below, but it is instructive to first consider the implications of such a technology. This structure, for example, was used in a recent paper by Hagiu and Wright (forthcoming),[11] where f is an increasing function of past sales (interpreted as product improvement via customer feedback).

The per-consumer revenue is equal to p_i, meaning we can invert the utility function to write $r(u_i, \delta_i) = V + f(\delta_i) - u_i$. It is clear that $\frac{\partial^2 r_i}{\partial u_i \partial \delta_i} = 0$; the surplus extraction effect is inactive here because the firm can extract surplus via the price, independent of δ_i. Since only the markup effect remains, data is UPC by Proposition 1.

Proposition 2. *In the model of product improvement, data is UPC.*

Intuitively, data increases the quality of the product, allowing the firm to hold u_i constant while charging a higher price. This makes the marginal consumer more valuable at any given u_i so the firm wants to increase utility to attract more consumers.

[11] Guembel and Hege (2021) also study a related model where consumers observe the realisation of the firm's signal before purchasing. One could also cast that model in a competition in utility framework, with a small extra notational burden. Note that one substantial difference between our model and Hagiu and Wright (forthcoming) and Guembel and Hege (2021) is that they do not have horizontal differentiation so that the equilibrium is not always given by the first-order condition.

To microfound $f'(\delta_i) > 0$, consider a situation where multiproduct firms offer to recommend an experience good (e.g., movie) to a consumer. The set of products that the firm can recommend is represented by the real line. The consumer has latent taste $\theta_0 \in \mathbb{R}$ and enjoys gross utility $V - (\theta_0 - x)^2$ from product x. The firm has a dataset, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$, about n past customers' tastes. Each $\hat{\theta}_l = \theta_l + \epsilon_l$ is a signal with noise $\epsilon_l \sim N(0, \sigma_\epsilon^2)$. The true tastes, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_n)$, are jointly normally distributed with means zero and variances σ^2, while the covariance between any two consumers' tastes is $\chi > 0$.[12] For experience goods the consumer can't observe $|\theta_0 - x|$ before purchasing. But the firm has an incentive to develop a reputation for good recommendations because this increases consumers' expected match quality, and consumers anticipate this.

Using standard results from probability theory, the firm's posterior belief about θ_0, given data $\hat{\boldsymbol{\theta}}$, follows a normal distribution $N(\mu, \frac{1}{\delta})$, where

$$\mu = \frac{\chi \sum_{l=1}^{n} \hat{\theta}_l}{(n-1)\chi + \sigma^2 + \sigma_\epsilon^2}, \qquad \frac{1}{\delta} = \sigma^2 - \frac{n\chi^2}{(n-1)\chi + \sigma^2 + \sigma_\epsilon^2}.$$

Given this posterior, the strategy that maximizes the expected value of the product is to recommend product μ. The consumer's expected mismatch is then $E((\theta_0 - \mu)^2) = \frac{1}{\delta}$. Letting $f(\delta) = -1/\delta$, we therefore have $u = V + f(\delta) - p$, with $f'(\delta) > 0$ as required.

In this example the dataset's value depends on three properties, namely its size (n), accuracy (σ_ϵ^2) and relevance (χ). But the impact of these three attributes can be conveniently summarized by a single parameter, δ.

As a final remark, we can extend the model of product improvement to the case where consumers are one-stop shoppers with downward-sloping demand by using the framework of Cowan (2004). In Appendix B.1 we show that data is again UPC in such a setting.

4.2 Moral hazard

Data can also be used to alleviate problems of asymmetric information in insurance markets and other situations of moral hazard. For example, insurers like Geico and UnitedHealth Group have turned to technologies such as vehicle telematics or personal fitness trackers to log customers' behavior and condition insurance contracts on the data recorded. This novel form of data helps to mitigate the hidden action problem that has plagued insurers for centuries.

Consider a model of insurance under moral hazard in the tradition of Holmström (1979), with binary effort. A risk-averse consumer who exerts no protection effort incurs a

[12] We could easily incorporate the case where consumers are heterogeneous (with a general covariance matrix for $\boldsymbol{\theta}$) at the cost of additional notational complexity, provided we treat each consumer as a separate market. An alternative interpretation is that $\boldsymbol{\theta}$ are realizations of a single consumer's tastes at different points of time.

loss L with probability 1. If he exerts effort, he avoids the loss with probability α. The utility function is separable in money and effort: if the final wealth is W then utility is $V(W) - ke$, where $e \in \{0, 1\}$ is the level of effort and $k > 0$ the cost of effort. V is increasing and concave, and we normalize consumers' initial wealth to zero. Insurers collect data about consumers' behavior over the relevant period. More precisely, when a consumer suffers a loss even though he exerted effort, his insurer i observes a signal that proves that the consumer exerted effort with probability δ_i. With probability $1 - \delta_i$ the data is inconclusive and the insurer learns nothing from it.[13]

Each risk-neutral insurer offers a contract $\{p_i, X_{Hi}, X_{Li}\}$, where p_i is the insurance premium that consumers pay irrespective of whether they incur the loss, X_{Hi} is the amount to be reimbursed in case of a loss if the insurer's data proves the consumer exerted effort, and X_{Li} is the amount to be reimbursed in case of a loss if the data is inconclusive.[14] A contract induces a mean utility u_i, which leads to a demand $D_i(u_i, u_{-i})$ for insurer i. The model is thus equivalent to one with a single agent and "random" participation.[15] From now on we focus on the strategy of one insurer, and drop the index i. We assume that it is optimal for insurers to design contracts that induce effort.

Suppose that an insurer wishes to offer a level of expected utility equal to u. The optimal way to provide such a utility level is the solution to the following program:

$$\max_{p, X_H, X_L} p - (1 - \alpha)\left(\delta X_H + (1 - \delta) X_L\right) \tag{4}$$

subject to incentive compatibility (consumers find it optimal to exert effort), and the requirement that $\{p, X_H, X_L\}$ yields expected utility u. We solve this problem in Appendix B.2. Any given (u, δ) pair implies an optimal contract, $\{p, X_H, X_L\}$. Substituting this contract into (4) then yields an expression for revenue of the form $r(u, \delta)$. We can therefore exploit Proposition 1 to establish:

Proposition 3. *In the model of insurance with moral hazard, the surplus extraction effect is positive* ($\frac{\partial^2 r(u,\delta)}{\partial u_i \partial \delta_i} > 0$) *if consumers have constant absolute risk aversion or a constant relative risk-aversion above* $1/2$*. Data is then UPC.*[16]

Like the previous application, data is UPC. What's new in this case is that the surplus extraction effect is active. Data mitigates the hidden action problem, allowing higher

[13]We choose such a stylized technology for analytical tractability, but the main insights do not depend on it. For instance, a technology where the insurer receives a signal when the consumer does not exert the effort would deliver similar results. The important point is that a more precise signal will lead the insurer to offer more insurance, as we discuss below.

[14]We assume that the insurer cannot pretend not to have received a signal.

[15]By opposition to the standard principal-agent model where the principal knows the agent's outside option. See Roger (2016) for more on moral hazard with random participation.

[16]When the constant relative risk aversion is below $1/2$, the surplus extraction effect is negative, and must be compared with the mark-up effect. Numerical methods have not delivered a single example where data is UAC.

levels of insurance to be offered. More insurance means less risk for the consumer who, because he is risk averse, therefore requires less wealth to reach the same utility u. Lower wealth in turn means the consumer's marginal utility of wealth is higher and it is cheaper to give the consumer additional utility: $\frac{\partial r(u,\delta)}{\partial u \partial \delta} > 0$. In other words, the surplus extraction effect here works in consumers' favor and, along with the markup effect, makes data UPC.

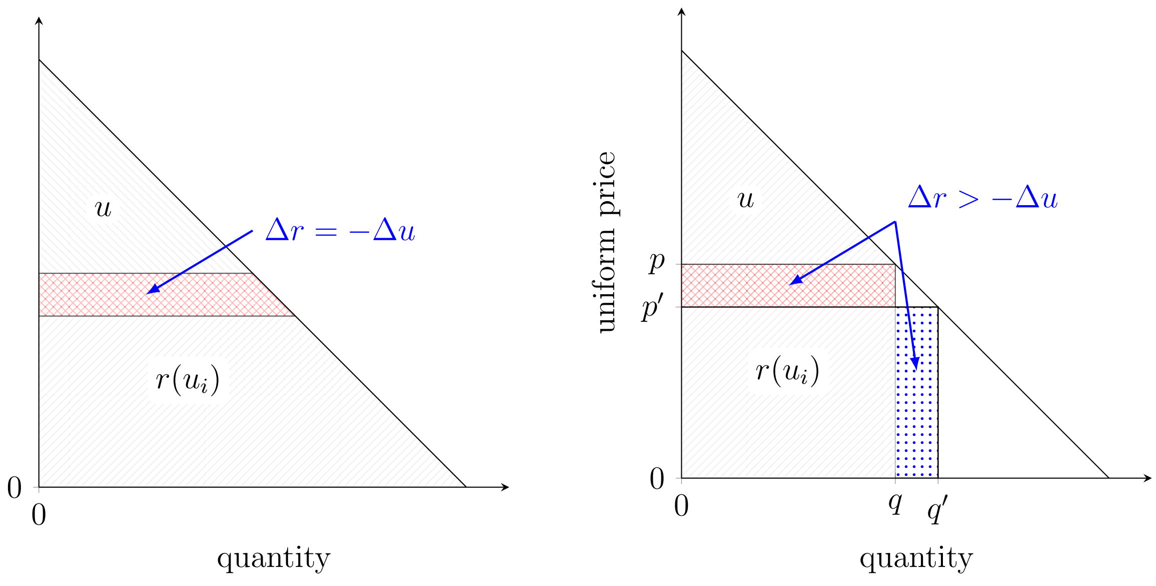
4.3 Price-discrimination

Armstrong and Vickers (2001) use the competition-in-utility framework to study competitive price-discrimination.[17] We adapt their framework to study the competitive effects of data when firms price-discriminate. We consider a model with multi-product retailers and one-stop shoppers who have an idiosyncratic willingness to pay for each product. Data allows retailers to identify consumers' willingness to pay for a fraction δ_i of products and charge a personalized price. For the remaining $1 - \delta_i$ products, the firm can't identify consumers' willingness to pay and sets a uniform price. Consumers observe all the prices, and choose the retailer that provides the largest surplus. We provide the formal analysis of such a model in Appendix B, where we show the following result:

Proposition 4. *(i) In the game of competitive price-discrimination à la Armstrong and Vickers (2001), the surplus extraction effect is negative ($\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} < 0$). (ii) If the demand for individual products is linear or has a constant elasticity, then data is UAC.*

To get some intuition, consider a firm's marginal incentive to provide utility in the extreme cases where the firm has either perfect or no data. If the firm knows the consumer's willingness to pay for all products then there is no deadweight loss and offering one additional unit of utility corresponds to a profit decrease of 1 (left panel of Figure 1). If the firm does not know the willingness to pay, it sells with a probability lower than 1 (deadweight loss). The same increase in utility is achieved through a price decrease from p to p', and is associated with an increase in the quantity, so that the cost in terms of reduced profit is smaller than 1 (right panel of Figure 1). In other words, unlike the product improvement and moral hazard applications, the firm becomes more efficient at extracting surplus as it gathers more data. This negative surplus extraction effect must be weighed against the markup effect that favors higher utility offers. For linear or constant-elasticity demand we find that the surplus extraction effect is dominant, so data is UAC.

[17] While most of the analysis in Armstrong and Vickers (2001) takes place in an environment of intense competition (so that the equilibrium is close to marginal cost-pricing), they provide a condition analogous to $\frac{\partial^2 \ln[r_i(u_i,\delta_i)]}{\partial u_i \partial \delta_i} > 0$ for discrimination to benefit consumers (their Lemma 3), and apply it to compare uniform pricing and two-part tariffs (Corollary 1). By explicitly incorporating data in the model we are able to study marginal improvements in the ability to price-discriminate, as well as asymmetric situations.



(a) $\delta = 1$

(b) $\delta = 0$

Figure 1: (a) If the firm can perfectly discriminate, offering one unit of utility reduces revenues by 1. (b) If the firm cannot observe consumers' willingness to pay, offering one unit of utility reduces revenues by less than 1.

4.4 Targeted advertising

One major use of data is to facilitate the targeting of advertising. We build upon the seminal model of media market competition in Anderson and Coate (2005), to which we add a role for targeted ads.

Consider media platforms that sell advertising space to a unit mass of advertisers and compete for consumers to whom these ads are shown. For our purposes, there are two important features of the Anderson and Coate (2005) framework. The first is that ads impose a (linear) nuisance cost on viewers: if the platform shows n ads then the utility of consuming its content is $u = V - \gamma n$, where V is the baseline value of the content and γ measures the nuisance cost of an advertisement. The second is that each advertiser has an idiosyncratic probability of matching with a consumer and captures all of the surplus from trade.[18] This implies an inverse demand for advertising slots, $P(n)$. We introduce targeting into the model by assuming that advertisers' demand for slots also depends on the accuracy of targeting, parameterized by δ. We thus rewrite advertisers' demand for slots as $P(n, \delta)$, which we assume is increasing in δ in the relevant range (we provide a Bayesian microfoundation at the end of this subsection).[19]

The platform's revenue from ads is $nP(n, \delta)$. Inverting the utility function, we can

[18]This assumption could be relaxed, provided that the nuisance cost of being exposed to an ad is higher than consumers' gain from trade.

[19]In the literature on advertising, targeting is often measured as a clockwise rotation of the inverse demand (Johnson and Myatt, 2006). Here we assume that the equilibrium ad supply is always below the rotation point.

express $n = (V - u)/\gamma$ and therefore write revenue as $r(u, \delta) = \frac{V-u}{\gamma} P(\frac{V-u}{\gamma}, \delta)$. The model is thus expressed in the form needed to apply Proposition 1. In particular, this is an environment with constant fixed cost so it suffices to check the log-supermodularity of r.

Proposition 5. *In the Anderson and Coate (2005) model with ad targeting, data is UPC if and only if* $\frac{\partial^2 \ln(P(n,\delta))}{\partial n \partial \delta} < 0$*, i.e. if an only if demand for ad slots becomes more elastic as* δ *increases.*

Because Proposition 5 gives a condition on how the elasticity of demand for ad slots changes when more data is available, it could potentially be used in empirical work. To better understand the forces at play, let us analyze the sign of the surplus extraction effect $\frac{\partial^2 r(u,\delta)}{\partial u \partial \delta}$. Using $n(u) = (V - u)/\gamma$, we have

$$\frac{\partial^2 r(u, \delta)}{\partial u \partial \delta} = n'(u) \left(\frac{\partial P(n(u), \delta)}{\partial \delta} + n(u) \frac{\partial^2 P(n(u), \delta)}{\partial n \partial \delta} \right)$$

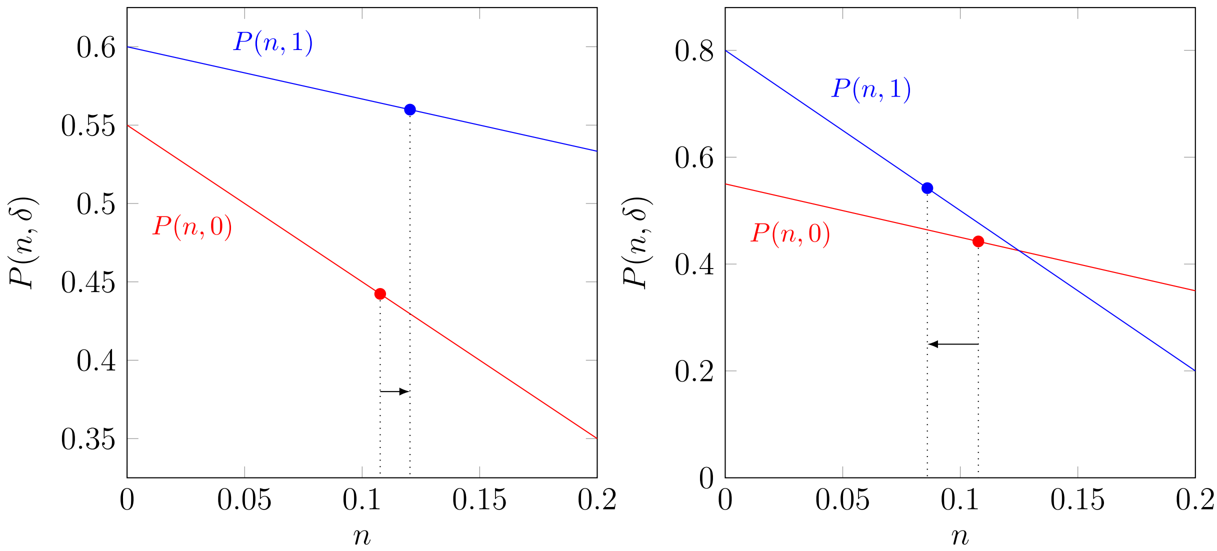
The first term between the brackets captures the idea that, as δ increases, the price for ad slots increases, so that the opportunity cost of not showing an ad is higher. This pushes towards a lower utility. The second term shows that an increase in δ may change the slope of demand for ad slots. When this term is negative, this pushes towards fewer ads, i.e. a higher utility. This is especially likely when targeting is useful to identify a small number of relevant ads, in which case the platform will reduce n to focus on extracting a high price from the few matched advertisers.

Microfoundation Because the function $P(n, \delta)$ is itself reduced-form, let us provide a more concrete example building on Anderson and Coate (2005) and Johnson and Myatt (2006). Suppose that there are an infinite number of product categories, each with a continuum of advertisers, and that each consumer is interested in a finite number K of categories. If a consumer is interested in a category, he is prepared to buy the product of advertiser a with latent probability θ_a, uniformly distributed on $[0, 1]$.

Platform i uses its data to target the ads it sells. With probability $\lambda(\delta_i)$, each relevant category is identified as such. Conditional on observing such a "relevance" signal, each advertiser a also receives a signal s about θ_a, which, following Lewis and Sappington (1994) and Johnson (2013), is equal to θ_a with probability $\mu(\delta_i)$ and is a random shock from a uniform distribution with probability $1 - \mu(\delta_i)$. The functions λ and μ respectively capture how informative data is about category- and brand match, and both are non-decreasing.

The willingness to pay of an advertiser who receives a signal s about θ_a is $\mu(\delta_i)s + \frac{(1-\mu(\delta_i))}{2}$. Therefore, the inverse demand for advertising slots is such that

$$K\lambda(\delta_i) Pr \left[\mu(\delta_i)s + \frac{(1 - \mu(\delta_i))}{2} \geq P(n_i, \delta_i) \right] = n_i.$$



(a) Data about relevant categories

(b) Data about preferred brands

Figure 2: On the left, the number of ads goes up when data increases: data is UAC. On the right, the opposite is true. The equilibrium point is marked in each case. See the appendix for more details.

This implies

$$P(n_i, \delta_i) = \frac{1 + \mu(\delta_i)}{2} - \frac{\mu(\delta_i)}{K\lambda(\delta_i)} n_i. \tag{5}$$

We can then write the per-consumer revenue, $n_i P(n_i, \delta_i)$, as a function of u_i and δ_i as above. In Appendix B.4.1 we show data is UPC if it mostly facilitates brand-specific matches (if λ' is small relative to μ'), but UAC if it mostly identifies relevant categories. When targeting identifies brand preferences it causes demand to rotate à la Johnson and Myatt (2006). More data leads the platform to lower ad volumes and focuses on charging a high price to the few advertisers with the best match. If targeting identifies relevant categories then the platform shows more ads to extract the increased willingness to pay of all sellers in the category. An example of both cases can be found in Figure 2 (details in Appendix B.4.1).

Additional considerations First, suppose that consumers can multihome across media platforms. A common theme in the literature is that advertisers have a lower willingness to pay to reach a given consumer a second time (e.g. Ambrus et al., 2016; Anderson et al., 2016). While the unilateral effect of δ_i on i's best response is still given by Proposition 1, some new effects emerge. Suppose that an increase in δ_i implies a higher probability for platform i to match its consumers with relevant ads. All else equal, the price at which platform j sells its advertising slots then goes down, because of the increased competition from i. This in turn affects firm j's best response, unlike in our baseline model where a firm's best response only depends on the quality of its own data. In order to obtain precise results on the direction in the shift of j's best response, we need to put more structure to the model. In some cases we have found that the shift in j's best-response is of the opposite sign as the shift in i's best response, though we leave this analysis out for the sake of brevity.

Second, another possible effect of targeting is the reduction in the nuisance experienced by consumers (see Johnson, 2013, for instance). One could easily add this component by changing the utility function to $u_i = V - \gamma(1 - \nu\delta_i)n_i$, where ν would measure the extent to which consumers prefer targeted ads. Such an analysis would make it more likely that data is UPC.

Third, one can also enrich the model to study the situation where firms can also directly charge consumers (see Kawaguchi et al., 2020, for a structural model of the mobile applications market using competition in utility and mixed business models). In Appendix B we show that data is always UPC when firms can charge consumers as well as showing them targeted ads. The intuition is that ad levels are chosen efficiently while firms use prices to adjust their utility offers, and that data does not affect the efficiency of surplus extraction through price, so only the mark-up effect applies. This example also illustrates that our framework, where firms choose u_i, can accommodate various situations in which the underlying decision problem (e.g., price and ad load) is multi-dimensional.

5 Equilibrium competitive effects of data

We now turn from the unilateral effects of data to its equilibrium effects under duopoly. Let the market be composed of two firms, each located at opposite ends of a Hotelling line. Demand has the usual Hotelling form, $D_i(u_i, u_j) = \frac{1}{2} + \frac{u_i - u_j}{2t}$, where t measures the level of differentiation. We assume that the game has a unique stable equilibrium.[20]

Giving firm i more data has both a direct (unilateral) effect and an indirect (strategic) effect. The direct effect comes from the unilateral shift in i's best response. This is exactly the effect we saw in Section 3 and its sign is characterized in Proposition 1 (e.g., is given by the log-supermodularity of r if fixed costs are constant).

The indirect effect comes as both firms strategically adjust their utility offers to restore equilibrium, given i's new best-response function. The direction of this strategic effect depends on whether firms' actions are strategic complements or substitutes. One advantage of the competition-in-utilities approach is that it can readily accommodate both possibilities. But this leaves open the question of how to determine which is the relevant case in any given market. Here, we can usefully invoke the concepts of congruence and conflict from de Cornière and Taylor (2019).

[20]Formally, a standard sufficient condition for this is that $\frac{\partial^2 \pi_i}{\partial u_i^2} + \left|\frac{\partial^2 \pi_i}{\partial u_i \partial u_j}\right| < 0$, i.e. $\frac{\partial^2 r}{\partial u_i^2} D_i + \frac{\partial r}{\partial u_i} \frac{3}{2t} - C''(u_i) < 0$ (see Vives, 2001).

Definition 2. Payoffs are *congruent* whenever $\frac{\partial r(u_i,\delta_i)}{\partial u_i} > 0$. When the inequality is reversed, we say that payoffs are *conflicting*.

While the examples of Section 4 all feature conflicting payoffs, a simple model with congruent payoffs would be one where media firms' per-consumer advertising revenue increases with the quality of their content, either because consumers consume more content or because advertisers are willing to pay a premium to be associated with quality content. See Appendix B.5 for an example.

From now on, let us assume that $\frac{\partial r_i}{\partial u_i}$ is of constant sign in the relevant domain.[21] Then the congruence/conflict property suffices to characterize the strategic effect that is the missing ingredient in our equilibrium analysis:

Proposition 6. *(i) With Hotelling demand, u_i and u_j are strategic complements if payoffs are conflicting and strategic substitutes if payoffs are congruent.*

(ii) The effect of an increase in δ_i, on u_i^* and u_j^* is given in the following table:

	Data	
Payoffs	*UAC*	*UPC*
Conflicting	$\downarrow u_i^*, \downarrow u_j^*$	$\uparrow u_i^*, \uparrow u_j^*$
Congruent	$\downarrow u_i^*, \uparrow u_j^*$	$\uparrow u_i^*, \downarrow u_j^*$

The proof of Proposition 6 is in Appendix A. Propositions 1 and 6 together allow us to reduce the problem of signing the unilateral and equilibrium effects of data to the much simpler one of signing two derivatives of r_i. This obviates, in particular, the need to fully compute equilibrium in order to obtain comparative statics. Instead, one need only identify enough parameters of r_i to sign the two derivatives of interest. Although we have assumed Hotelling demand, Proposition 6 continues to hold for other demand specifications so long as either (i) $\frac{\partial^2 D_i}{\partial u_i \partial u_j}$ is small enough or (ii) the congruence or conflict property is sufficiently strong (i.e., $|\frac{\partial r_i}{\partial u_i}|$ is large).[22]

It will often be possible to determine whether payoffs are congruent or conflicting from a simple inspection of firms' business model. For example, each of the applications in Section 4 exhibits conflict because firms increase per-consumer revenue purely instruments (prices or ad loads) that reduce utility. Thus, applying Proposition 6, we see that data has a consistent industry-wide impact that depends only on its unilateral effect. Giving any one firm more data would lead to an intensification of competition in the product improvement application, but leads to worse outcomes for all consumers under price discrimination with linear or constant-elasticity demand.

[21]In the applications of Section 4, $\frac{\partial r_i}{\partial u_i}$ is of constant sign, except for the targeted advertising one. In that application the relevant domain is the values of u for which $\frac{\partial r_i}{\partial u_i} \leq 0$ because outside of this range it would always be profitable for the firm to offer more utility.

[22]In particular, we have strategic complementarity if $\frac{\partial^2 \pi_i}{\partial u_i \partial u_j} = \frac{\partial r_i}{\partial u_i}\frac{\partial D_i}{\partial u_j} + r_i \frac{\partial^2 D_i}{\partial u_i \partial u_j} > 0$, and strategic substitutability if the inequality is reversed.

A remark on data sharing policies One family of oft-mooted policy proposals aims to improve firms' access to data by, for example, forcing a dominant firm to share its data with smaller rivals. For example, Article 6(11) of the EU's Digital Markets Act imposes obligations for incumbent "gatekeeper" platforms to share search query and other types of data with rival firms. Formally, this amounts to an increase in δ_i, starting from $\delta_i < \delta_j$. Our results provide guidance on when such a policy would be effective, and sounds a note of warning about cases where it might be counter-productive. If data is UPC and payoffs are conflicting then Proposition 6 tells us that such a data-sharing mandate would be unambiguously pro-competitive. But if data is UAC or payoffs are congruent then data sharing would lead to at least one firm reducing its utility offer.

6 Endogenous data collection and dynamic implications

One distinguishing property of data is that it is often collected as a byproduct of firms' interactions with consumers. This implies a dynamic structure where a firm's economic decisions today determine the data available to it tomorrow. While our model is static, one can embed it into a dynamic framework to shed light on further policy issues: when does data constitute a barrier to entry? When does it favor concentration? Here we discuss a few insights that emerge from simple dynamic extensions of the model.[23]

6.1 Data as a barrier to entry

A recurring and contentious theme of the policy dabate around data is whether data per se constitutes a barrier to entry (e.g., Grunes and Stucke, 2016; Sokol and Comerford, 2016). The EU's Digital Markets Act takes the clear position that data advantages have limited the contestability of core platform services.[24] Consider a two-period entry game, where an incumbent initially operates alone on a market, before a potential entrant decides whether to enter and compete in the second period. Entry will occur only if the entrant expects a profit sufficient to cover its entry cost. Suppose that data is a by-product of firms' economic activity, so that the quantity of data available to the incumbent in the

[23] Another important situation is one where a firm endogenously collects data in its primary market that might be useful to firms in an adjacent market. This raises interesting questions about how a merger would affect the collection and use of data. In de Cornière and Taylor (2022) we use the framework developed here to study such questions.

[24] See, for example, paragraph 32 of the preamble "The features of core platform services in the digital sector, such as [...] benefits from data have limited the contestability of those services", or paragraph 36 of the preamble "The processing, for the purpose of providing online advertising services, of personal data from third parties using core platform services gives gatekeepers potential advantages in terms of accumulation of data, thereby raising barriers to entry. This is because gatekeepers process personal data from a significantly larger number of third parties than other undertakings."

second period is an increasing function of its first-period sales (and thus of the first period utility offer). A first remark is that, using the Fudenberg and Tirole (1984) terminology, data makes the incumbent look *tough* when it is UPC: an incumbent with more data will offer a larger utility, which reduces the entrant's profit. Conversely, more data makes the incumbent look *soft* when it is UAC. We can therefore use Proposition 1 to characterize the effect of data on entry, as shown in the following result.

Proposition 7. *(1) Data acts as a barrier to entry if and only if it is UPC. (2) If data is UPC then the incumbent can deter entry by over-collecting data in the first period, which benefits first-period consumers. (3) If data is UAC then deterrence is achieved by under-collecting data in the first period, which harms first-period consumers.*

Our characterization of strategic substitutability/complementarity depending on whether payoffs are congruent/conflicting is also useful here, as it allows us to discuss the nature of an accommodation strategy, again following Fudenberg and Tirole (1984). Table 1 accordingly uses Propositions 1 and 6 to summarize the optimal entry deterrence and accommodation strategies.

Table 1: Should an incumbent firm over- or under-collect data? A: optimal accommodation strategy, D: optimal entry deterrence strategy.

	UPC	UAC
Conflict	A: under-collection D: over-collection	A: over-collection D: under-collection
Congruence	A: over-collection D: over-collection	A: under-collection D: under-collection

6.2 Data and concentration

Consider now a dynamic game, where two firms repeatedly compete over an infinite horizon, and where data accumulates as a function of a firm's past sales, potentially with some depreciation. Several recent papers study the implications for the long-run evolution of market concentration (Prüfer and Schottmüller, 2021; Farboodi et al., 2019; Hagiu and Wright, forthcoming).[25]

Because the analysis of this kind of game with forward-looking agents is very complex, these papers all assume a specific functional form for profits to make some progress. Indeed, the papers begin with environments that imply data is pro-competitive[26] and proceed to

[25] An earlier version of this idea can also be found in Argenton and Prüfer (2012).

[26] As discussed in Section 4, the baseline model of Hagiu and Wright (forthcoming) fits our competition-in-utility framework and data is UPC. For Prüfer and Schottmüller (2021), casting the model in a

show that data can generate a winner-takes-all dynamic. To be able to accommodate both UPC and UAC cases, it is necessary for us to relax these functional form assumptions and simplify in another dimension to ensure the analysis remains tractable. We focus on the case where firms are myopic. Of course, the myopia assumption is strong, but it allows us to show that relaxing the literature's assumption that data is pro-competitive can lead to quite different long-run effects of data.

Consider a market where two firms, A and B, are myopic and compete over an infinite horizon. At the start of period t, firm i's stock of data is δ_i^t. The initial stocks of data may differ, but firms are otherwise symmetric. Denote $\Delta_i^t = \delta_i^t - \delta_j^t$ for i's data advantage at time t. In every period, each firm chooses a utility offer u_i^t, resulting in demand $D_i^t = D(u_i^t, u_j^t)$ and a mark-up $r_i^t = r(u_i^t, \delta_i^t)$. We assume that firms accumulate data by serving consumers, but that data also depreciates at rate $1 - \Upsilon \in [0, 1]$. Thus, $\delta_i^{t+1} = \Upsilon \delta_i^t + D_i^t$.

Firms with more data offer higher utility if and only if data is UPC. Moreover, because a firm that offers higher utility serves more consumers, it accumulates more new data than its rival. The following proposition is immediate:

Proposition 8. *Suppose that i is the current leader ($\Delta_i^t > 0$) and firms are myopic. Then:*

1. *Data being UPC is a necessary condition for i's data advantage to increase ($\Delta_i^{t+1} > \Delta_i^t$).*
2. *Data being UPC is a necessary and sufficient condition for i's data advantage to increase if data does not depreciate ($\Upsilon = 1$).*

Thus, the log-supermodularity condition from Proposition 1 can be used to characterize the evolution of the data advantage. If data is UPC and the existing stock of data is long-lived ($\Upsilon = 1$) then the leader's advantage grows over time because it accumulates more new data each period. This is the dynamic logic to be found in Prüfer and Schottmüller (2021) and Hagiu and Wright (forthcoming), and can lead to market tipping in environments where data has pro-competitive uses such as product improvement.[27] If data decays over time ($\Upsilon < 1$) then the data advantage can still increase, but only if the leader accumulates enough new data to offset the depreciation of its existing lead. If data is UAC then an initial data advantage cannot increase because the firm that enjoys it offers a lower utility than its rival and collects less data.

competition in utility framework would lead to $\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} > 0$, so data is also UPC. Farboodi et al. (2019)'s model is one with competition in quantities and cannot be expressed in terms of competition-in-utility, but more data leads to higher quantities, and therefore more consumer surplus..

[27]Note that the increase in the data advantage might not imply an increase in market concentration if the marginal value of data decreases quickly.

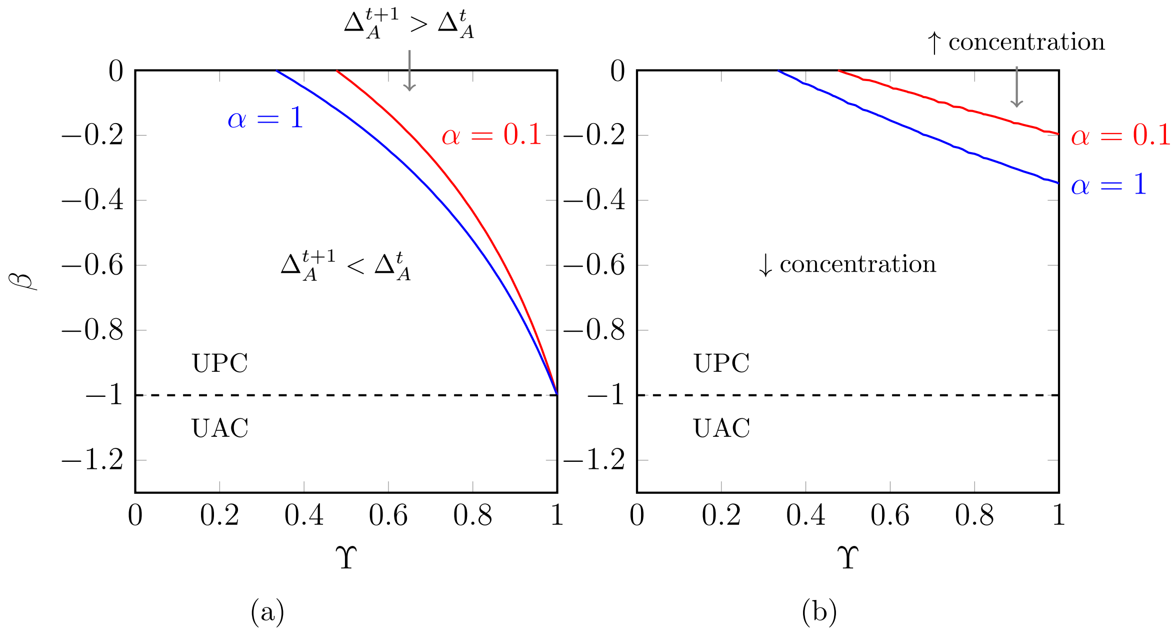


Figure 3: Data is UPC above the dashed line and UAC below it. (a) The leader's data advantage grows between t and $t+1$ above the solid curve corresponding to the relevant value of α. (b) Market concentration grows between t and $t+1$ above the solid curve corresponding to the relevant value of α. The plot is drawn for $\delta_A^t = 0.6$ and $\delta_B^t = 0.4$.

We can illustrate Proposition 8 with a parameterized example. Suppose $D(u_i^t, u_j^t) = \frac{1}{2}+u_i^t-\alpha u_j^t$, where α measures the intensity of competition, and $r(u_i^t, \delta_i^t) = 1+\delta_i^t-u_i^t+\beta\delta_i^t u_i^t$. We take $\beta < 0$ to ensure strategies are strategic complements (cf. Proposition 6). We also observe from Proposition 1 that data is UPC if and only if $\frac{\partial^2 \ln(r_i^t)}{\partial u_i^t \partial \delta_i^t} \geq 0$, that is $\beta \geq -1$. Calculating the equilibrium (u_A^t, u_B^t) for a given $\delta_A^t > \delta_B^t$, we can infer whether the leader's data-advantage increases or decreases between periods, and also whether market concentration increases or decreases. Figures 3a and 3b respectively show the region in which data leads to an increasing advantage and an increase in concentration. Three conditions must be satisfied for data and market concentration to increase: (i) data must not depreciate too quickly (Υ large enough), so that the leader's advantage persists over time; (ii) data must be 'UPC enough' (β large), so that a data advantage translates into a sufficiently higher utility offer; (iii) competition must be strong enough (α large), so that a utility advantage translates into a large enough market share advantage.

6.3 Discussion

The results on dynamics and barriers to entry point to a tension between exploitative and exclusionary theories of harm. When a data advantage leads a firm to extract more of consumers' surplus it will tend to serve fewer consumers and accumulate less data in the future. So short-run exploitation implies less concern about long-run market structure. Conversely, when data is used to consumers' benefit there is less to worry about from a static consumer welfare perspective, but consumers will tend to gravitate towards firms with lots of data. This can harm consumers if it blockades entry or induces laggard firms to exit the market. Note that in each of the applications in Section 4 data would remain UPC after a competitor's exit—it's not the use of data per se that harms consumers, but rather the change in market structure that it induces.

7 Discussion and extensions

Consumer heterogeneity The simplicity of our approach naturally entails some costs. The most significant one, in our view, is the way it restricts consumer heterogeneity. First, the competition-in-utility framework requires that actions that increase or decrease the mean utility u_i affect all of i's customers equally. While standard discrete choice models such as the logit or nested logit are consistent with this specification, models with random coefficients (Berry et al., 1995) are not. Second, our way of modelling data implies that consumers are also homogenous with respect to how data affects their (expected) utility. The framework is thus ill-suited to study issues related to adverse selection or price-discrimination with spatial differentiation (e.g. models à la Thisse and Vives, 1988), where different types of consumers might be made better-off or worse-off by an increase in the quality of data.[28] Note that this is a feature shared by many papers on the economics of data (e.g., Prüfer and Schottmüller, 2021; Hagiu and Wright, forthcoming; Choi et al., 2019; Acemoglu et al., forthcoming, to name a few). Consumer heterogeneity therefore only applies to horizontal brand preferences.

Privacy concerns An important theme in the policy debate around data is the potential for harm to consumers through exploitative data collection and the associated loss of privacy.[29] While situations where data is UAC themselves provide a justification for privacy concerns, our model can be adapted to accommodate intrinsic taste for privacy. Suppose that consumers incur a harm, $h(\delta)$, where h is increasing and convex. If u is the (mean) gross utility offered by the firm (with corresponding revenue $r(u, \delta)$), the net utility is then $U \equiv u - h(\delta)$. Define $R(U, \delta) \equiv r(U + h(\delta), \delta)$, i.e. R is the firm's mark-up as a function of the net utility it offers. We can then use R instead of r in the subsequent analysis. The only substantial change is that R may be decreasing in δ when privacy concerns are sufficiently strong.[30] In that case, Proposition 1(i) changes: supermodularity of R becomes a necessary (but no longer sufficient) condition for data to be UPC. Proposition 1(ii), applied to R instead of r, is unchanged.

[28]See the discussion in Armstrong and Vickers (2001), p.584.

[29]See, for example, Bundeskartellamt (2019) .

[30]For instance, in the product improvement example, if the willingness to pay is $V + f(\delta)$, we have $U = V + f(\delta) - h(\delta) - p$, and, with a marginal cost normalized to zero, $R(U, \delta) = V + f(\delta) - h(\delta) - U$. Whenever $h'(\delta) > f'(\delta)$, R is decreasing in δ.

Data externalities A recent literature (Choi et al., 2019; Ichihashi, 2021; Bergemann et al., 2022; Acemoglu et al., forthcoming; Markovich and Yehezkel, 2021) explores the idea of data externalities, whereby individuals do not internalize the fact that their data can be used to make inferences about others. For instance, a firm may be able to predict a consumer's future behavior based on how consumers with similar characteristics have behaved. We can incorporate this feature by supposing that a firm generates *insights* about consumer l by combining data about l with data about other consumers. Focusing on the monopoly case for brevity, let $I_l \equiv g(\delta_l, \delta_{-l})$ be the quality of the firm's insights about consumer l if δ_l is the data it has about the consumer and δ_{-l} is the data about other consumers (where we assume symmetry among the other consumers). We assume that g is increasing in both arguments. For a consumer over whom the firm has insight I_l and to whom it offers gross utility u_l, the firm's revenue is $r(u_l, I_l)$, increasing in I_l. A consumer buys from the firm if $u_l - h(\delta_l) + \epsilon_l > 0$, where $h(\delta_l)$ is the privacy cost and ϵ_l is an idiosyncratic random taste shock. The firm gets a unit of data about each consumer who buys from it.[31]

In this environment we can immediately see that there is a negative data externality between consumers if data is UAC and a positive externality if it is UPC. Indeed, if data is UAC then each consumer that buys from the firm shares data that leads to other consumers receiving lower utility offers. This is true in aggregate, even if each individual consumer is atomistic. This case, which has been the primary focus of the literature, leads to too much data sharing in equilibrium. Conversely, if data is UPC then consumers benefit from the insights generated from others' data (as in Section 4.1, where a more informed firm makes better product recommendations). The sharing of data then becomes a public good. Proposition 1 can therefore be used to sign the externality in a given market.

8 Conclusion

The wide variety of business models, purposes, and technologies under which data is used make it hard to develop a clear overall picture of its role in competition. One objective of this paper is to suggest a simple yet flexible framework through which to analyze the competitive role of data and potential policy interventions. While we do not claim to nest all possible situations where firms use data, we show how key trade-offs are resolved across a wide range of different scenarios. Understanding these general trade-offs is important as policy makers are working to implement economy-wide regulations for data.

[31]We could alternatively assume that the privacy cost depends on I_l without changing the basic message. Additionally, we could replace the binary decision of whether to buy or not with a richer model where the consumer chooses how much to interact with the firm and suppose that δ_l is increasing in the level of interaction.

We study a model where firms compete in utility levels, and where data allows a firm to generate more revenues for a given level of utility. Considering unilateral effects of data, we identify a key trade-off between a *mark-up* and a *surplus extraction* effect. Data makes each consumer more valuable, thus leading firms to compete harder to attract more of them (mark-up effect). It can also make surplus extraction more efficient, potentially leading to lower utility provision. In many cases, whether data is unilaterally pro- or anti-competitive (UPC or UAC) can be inferred from a simple super- or sub-modularity property of the per-consumer revenue function, independently of market demand and without need to compute the equilibrium.

We illustrate the usefulness of this approach through four applications illustrating varied competitive effects of data. When it is used to improve products, data increases the potential gains from trade and leads firms to compete harder for consumers. In other markets, data helps to resolve moral hazard problems, reducing the cost of providing utility. These are both cases where data is pro-competitive because the surplus extraction effect is inactive or runs in consumers' favor. By contrast, data makes a price discriminator more efficient at extracting surplus, which is anti-competitive. Lastly, ad targeting can have pro- or anti-competitive effects depending on the kinds of ad matches it induces. When data is informative about category matches then many firms within a matched category will want to advertise and data leads to an expansion of ad supply (causing more nuisance for consumers). Conversely, if data identifies a small number of matched brands within a category then ad supply will contract to focus on extracting those advertisers' willingness to pay.

The competition-in-utility framework also accommodates situations of strategic complementarity or substitutability. Restricting attention to a Hotelling duopoly, we provide a simple characterization, based on the relationship between utility and revenue. Coupled with the conditions determining whether data is UPC or UAC, this allows us to obtain a more complete picture of the competitive effects of data, and to discuss policies such as mandated data sharing or overall restrictions of data collection.

Our simple model can also be embedded in a dynamic framework. We highlight that whether data is UPC or UAC determines whether exclusionary or exploitative theories of harm are more likely to apply, but that there is an important tension between the two. Lastly, our analysis can also be applied to situations with data externalities, whose sign depends on whether data is UPC or UAC.

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (forthcoming). "Too Much Data: Prices and Inefficiencies in Data Markets". *American Economic Journal: Microeconomics.*

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). "The economics of privacy". *Journal of Economic Literature* 54.2, pp. 442–92.
Acquisti, Alessandro and Hal R. Varian (2005). "Conditioning Prices on Purchase History". *Marketing Science* 24.3, pp. 305–523.
Ambrus, Attila, Emilio Calvano, and Markus Reisinger (2016). "Either or both competition: A" two-sided" theory of advertising with overlapping viewerships". *American Economic Journal: Microeconomics* 8.3, pp. 189–222.
Anderson, Simon P. and Stephen Coate (2005). "Market Provision of Broadcasting: A Welfare Analysis". *The Review of Economic Studies* 72.4, pp. 947–972.
Anderson, Simon P, André De Palma, and J-F Thisse (1988). "A representative consumer theory of the logit model". *International Economic Review*, pp. 461–466.
Anderson, Simon, Alicia Baik, and Nathan Larson (2016). "The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing". URL: `https://www.tse-fr.eu/sites/default/files/TSE/documents/ChaireJJL/Digital-Economics-Conference/Conference/anderson_simon.pdf`.
Argenton, Cédric and Jens Prüfer (2012). "Search Engine Competition With Network Externalities". *Jornal of Competition Law & Economics* 8.1, pp. 73–105.
Armstrong, Mark and John Vickers (2001). "Competitive price discrimination". *RAND Journal of Economics* 4, pp. 1–27.
Athey, Susan and Joshua S. Gans (2010). "The Impact of Targeting Technology on Advertising Markets and Media Competition". *American Economic Review* 100.2, pp. 608–13. DOI: `10.1257/aer.100.2.608`. URL: `http://www.aeaweb.org/articles?id=10.1257/aer.100.2.608`.
Belleflamme, Paul and Wouter Vergote (2016). "Monopoly price discrimination and privacy: The hidden cost of hiding". *Economics Letters* 149, pp. 141–144.
Bergemann, Dirk and Alessandro Bonatti (2011). "Targeting in advertising markets: implications for offline versus online media". *RAND Journal of Economics* 42.3, pp. 417–443.
— (2015). "Selling cookies". *American Economic Journal: Microeconomics* 7.3, pp. 259–294.
Bergemann, Dirk, Alessandro Bonatti, and Tan Gan (2022). "The Economics of Social Data". *RAND Journal of Economics* 53.2, pp. 263–296.
Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile prices in market equilibrium". *Econometrica: Journal of the Econometric Society*, pp. 841–890.
Bonatti, Alessandro and Gonzalo Cisternas (Sept. 2019). "Consumer Scores and Price Discrimination". *The Review of Economic Studies.*
Bounie, David, Antoine Dubus, and Patrick Waelbroeck (2021). "Selling Strategic Information in Digital Competitive Markets". *RAND Journal of Economics* 52.2, pp. 283–313.

Bulow, Jeremy I, John D Geanakoplos, and Paul D Klemperer (1985). "Multimarket oligopoly: Strategic substitutes and complements". *Journal of Political economy* 93.3, pp. 488–511.

Bundeskartellamt (2019). *Case Summary: Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing*. Germany: Bundeskartellamt.

Cabral, Luis M. B. and Michael H. Riordan (1994). "The Learning Curve, Market Dominance, and Predatory Pricing". *Econometrica* 62.5, p. 1115. URL: https://www.jstor.org/stable/2951509?origin=crossref.

Calzolari, Giacomo and Alessandro Pavan (2006). "On the optimality of privacy in sequential contracting". *Journal of Economic theory* 130.1, pp. 168–204.

Campbell, James, Avi Goldfarb, and Catherine Tucker (2015). "Privacy regulation and market structure". *Journal of Economics and Management Strategy* 24.1, pp. 47–73.

Chen, Zhijun, Chongwoo Choe, and Noriaki Matsushima (2020). "Competitive Personalized Pricing". *Management Science* 66, pp. 4003–4023.

Chetty, Raj (2009). "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods". *Annu. Rev. Econ.* 1.1, pp. 451–488.

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

Cowan, Simon (2004). "Demand shifts and imperfect competition". *Working Paper.*

Crémer, Jacques, Yves-Alexandre de Montjoye, and Heike Schweitzer (2019). *Competition Policy for the Digital Era*. Report. European Commission.

Dasgupta, Partha and Joseph Stiglitz (1988). "Learning-by-doing, market structure and industrial and trade policies". *Oxford Economic Papers* 40.2, pp. 246–268.

De Cornière, Alexandre and Romain de Nijs (2016). "Online Advertising and Privacy". *RAND Journal of Economics* 47.1, pp. 48–72.

De Cornière, Alexandre and Greg Taylor (2019). "A Model of Biased Intermediation". *RAND Journal of Economics* 50.4, pp. 854–882.

— (2022). "Data and Competition: Mergers". Working Paper.

Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp (2019). "Big Data and Firm Dynamics". *AEA Papers and Proceedings* 109, pp. 38–42.

Fudenberg, Drew and Jean Tirole (1984). "The fat-cat effect, the puppy-dog ploy, and the lean and hungry look". *American Economic Review* 74.2, pp. 361–366.

— (2000). "Customer poaching and brand switching". *RAND Journal of Economics*, pp. 634–657.

Furman, Jason, Dianne Coyle, Amelia Fletcher, Derek McAuley, and Philip Marsden (2019). *Unlocking Digital Competition*. Report of the Digital Competition Expert Panel. HM Government.

Galeotti, Andrea and José Luis Moraga-González (2008). "Segmentation, advertising and prices". *International Journal of Industrial Organization* 26.5, pp. 1106 –1119.

Grunes, Alan and Maurice Stucke (2016). *Big data and competition policy*. Oxford University Press.

Gu, Yiquan, Leonardo Madio, and Carlo Reggiani (2019). "Exclusive Data, Price Manipulation and Market Leadership". CESifo Working Paper No. 7853.

Guembel, Alexander and Ulrich Hege (2021). "Data, Product Targeting and Competition". *Working Paper*.

Hagiu, Andrei and Julian Wright (forthcoming). "Data-enabled learning, network effects and competitive advantage". *RAND Journal of Economics*.

Holmström, Bengt (1979). "Moral Hazard and Observability". *The Bell Journal of Economics* 10.1, pp. 74–91. URL: http://www.jstor.org/stable/3003320.

Ichihashi, Shota (2020). "Online Privacy and Information Disclosure by Consumers". *American Economic Review*.

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

Iyer, Ganesh, David Soberman, and J. Miguel Villas-Boas (2005). "The Targeting of Advertising". *Marketing Science* 24.3. URL: http://groups.haas.berkeley.edu/marketing/PAPERS/VILLAS/ms05.pdf.

Johnson, Garrett, Scott Shriver, and Samuel Goldberg (2021). "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR". Working Paper.

Johnson, Justin P. (2013). "Targeted advertising and advertising avoidance". *The RAND Journal of Economics* 44.1, pp. 128–144. URL: http://doi.wiley.com/10.1111/1756-2171.12014.

Johnson, Justin P and David P Myatt (2006). "On the simple economics of advertising, marketing, and product design". *American Economic Review* 96.3, pp. 756–784.

Kawaguchi, Kohei, Toshifumi Kuroda, and Susumu Sato (2020). "Merger Analysis in the App Economy: An Empirical Model of Ad-Sponsored Media". *Available at SSRN 3746830*.

Kim, Jin-Hyuk, Liad Wagman, and Abraham L. Wickelgren (2018). "The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing". *Journal of Economics and Management Strategy*. URL: https://doi.org/10.1111/jems.12285.

Lam, Wing Man Wynne and Xingyi Liu (2020). "Does data portability facilitate entry?" *International Journal of Industrial Organization* 69, p. 102564.

Lewis, Tracy R and David EM Sappington (1994). "Supplying information to facilitate price discrimination". *International Economic Review*, pp. 309–327.

Markovich, Sarit and Yaron Yehezkel (2021). "Data regulation: who should control our data?" *Available at SSRN 3801314*.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2018). “The Value of Personal Information in Online Markets with Endogenous Privacy”. *Management Science*.

Pino, Flavio (forthcoming). “The Microeconomics of Data – A Survey”. *Journal of Industrial and Business Economics*. DOI: 10.1007/s40812-022-00220-6.

Prüfer, Jens and Christoph Schottmüller (2021). “Competing with big data”. *Jornal of Industrial Economics* 69.4, pp. 967–1008.

Roger, Guillaume (2016). “Participation in moral hazard problems”. *Games and Economic Behavior* 95, pp. 10–24.

Roy, S (2000). “Strategic segmentation of a market”. *International Journal of Industrial Organization* 18.8, pp. 1279–1290. URL: http://dx.doi.org/10.1016/S0167-7187(98)00052-6.

Rutt, James (2012). “Targeted Advertising and Media Market Competition”. Working Paper. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103061.

Scott Morton, Fiona, Theodore Nierenberg, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A Douglas Melamed, and Jamie Morgenstern (2019). “Report: Committee for the Study of Digital Platforms-Market Structure and Antitrust Subcommittee”. *George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business*.

Shelegia, Sandro and Chris Wilson (forthcoming). “A Generalized Model of Advertised Sales”. *American Economic Journal: Microeconomics*.

Sokol, D. Daniel and Roisin E. Comerford (2016). “Antitrust and Regulating Big Data”. *George Mason Law Review* 23.5, pp. 1129–1162.

Stole, Lars A (2007). “Price discrimination and competition”. *Handbook of industrial organization* 3, pp. 2221–2299.

Taylor, Curtis R. (2004). “Consumer Privacy and the Market for Customer Information”. *RAND Journal of Economics* 35.4, pp. 631 –650. URL: http://www.jstor.org/stable/1593765.

Thisse, Jacques-Francois and Xavier Vives (1988). “On The Strategic Choice of Spatial Price Policy”. *The American Economic Review* 78.1, pp. 122–137.

Vives, Xavier (2001). *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press.

A Omitted proofs

Proof of Proposition 1. Part i: The first two terms on the right-hand side of (3) are positive: the demand for firm i is increasing in u_i, and its revenue is increasing in δ_i. The sign of $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i}$ is ambiguous but when it is non-negative, we have $\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} > 0$, i.e. data is pro-competitive.

Part ii: When $C'(u) = 0$, we have $\frac{\partial D_i}{\partial u_i}/D_i = -\frac{\partial r_i}{\partial u_i}/r_i$ by (2). We thus have

$$\frac{\partial D_i}{\partial u_i}\frac{\partial r_i}{\partial \delta_i} + \frac{\partial^2 r_i}{\partial u_i \partial \delta_i} D_i > 0 \Leftrightarrow -\frac{\partial r_i}{\partial u_i}\frac{\partial r_i}{\partial \delta_i} + \frac{\partial^2 r_i}{\partial u_i \partial \delta_i} r_i > 0$$

$$\Leftrightarrow \frac{1}{r_i^2}\left(-\frac{\partial r_i}{\partial u_i}\frac{\partial r_i}{\partial \delta_i} + \frac{\partial^2 r_i}{\partial u_i \partial \delta_i} r_i\right) > 0 \Leftrightarrow \frac{\partial}{\partial \delta_i}\left(\frac{\frac{\partial r_i}{\partial u_i}}{r_i}\right) > 0$$

$$\Leftrightarrow \frac{\partial^2 \ln(r_i)}{\partial u_i \partial \delta_i} > 0. \qquad \blacksquare$$

Proof of Proposition 6. Part (i): By definition, payoffs are strategic complements if $\frac{\partial^2 \pi_i}{\partial u_i \partial u_j} > 0$, i.e. if $\frac{\partial D_i(u_i,u_j)}{\partial u_j}\frac{\partial r(u_i,\delta_i)}{\partial u_i} + r(u_i,\delta_i)\frac{\partial^2 D_i(u_i,u_j)}{\partial u_i \partial u_j} > 0$. In the Hotelling model, $D_i(u_i,u_j) = \frac{\tau+u_i-u_j}{2\tau}$, so that $\frac{\partial^2 D_i(u_i,u_j)}{\partial u_i \partial u_j} = 0$, meaning that $\frac{\partial^2 \pi_i}{\partial u_i \partial u_j}$ has the opposite sign to $\frac{\partial r(u_i,\delta_i)}{\partial u_i}$.

Part (ii): We find u_i^* as the solution to

$$\hat{u}_i(\hat{u}_j(u_i^*),\delta_i) - u_i^* = 0 \tag{6}$$

(recalling that $\hat{u}_i$ is i's best response function). The left-hand side of (6) is decreasing in u_i^* when $\frac{\partial^2 \pi_i}{\partial u_i^2} + \left|\frac{\partial^2 \pi_i}{\partial u_i \partial u_j}\right| < 0$, which must be true at a stable equilibrium.

Suppose data is UPC. Then the left hand side of (6) is increasing in δ_i so u_i^* must increase with δ_i. The effect on u_j^* follows immediately from the definition of strategic complements and substitutes along with part (i). A symmetric argument holds for the UAC case. $\blacksquare$

B Proofs and supplementary material for the applications of Sections 4 and 5

B.1 Product improvement model with downward-sloping demand

An alternative approach to modelling product improvement is to think of consumers as one-stop shoppers who choose firm i with probability $D_i(u_i, \mathbf{u}_{-i})$ and buy $Q(p_i,\delta_i)$ products from their chosen firm whose price is p_i. This allows us to apply the demandshifting framework of Cowan (2004). Let the inverse demand be $P(q_i, \delta_i)$. Data improves the product and causes demand to shift up: $\frac{\partial Q(p_i,\delta_i)}{\partial \delta_i} > 0$. This can be microfounded in a similar fashion to the unit demand case.

Utility when the price is p_i is given by the standard consumer surplus measure,

$$u_i = \int_{p_i}^{\infty} Q(x, \delta_i)\, dx, \tag{7}$$

while per-consumer revenue is $Q(p_i, \delta_i)p_i$. We can rewrite this revenue in the form $r(u_i, \delta_i)$ by inverting (7) to get p_i as a function of u_i and δ_i, and therefore apply our supermodularity conditions. We consider specifications where δ_i shifts the inverse (or direct) demand additively ($P(q_i, \delta_i) = P(q_i) + \delta_i$) or multiplicatively ($P(q_i, \delta_i) = (1 + \delta_i)P(q_i)$).

We start with the following lemma.

Lemma 1. *Consider a decreasing and twice-differentiable demand function $Q(p)$, and its inverse, $P(q)$. If P is log-concave, then $Q'(p) + pQ''(p) \leq 0$. Similarly, if Q is log-concave, $P'(q) + qP''(q) \leq 0$.*

Proof. We have $Q'(p) = \frac{1}{P'(Q(p))}$. Differentiating once more, we obtain $Q''(p) = -\frac{P''(Q(p))}{P'(Q(p))^3}$. Then,

$$Q'(p) + pQ''(p) \leq 0 \Leftrightarrow (P'(Q(p)))^2 - P(Q(p))\, P''(Q(p)) \geq 0$$

which is true if P is log-concave. ∎

Applying Proposition 1, we now have:

Corollary 1. *Suppose $Q(p_i, \delta_i)$, is log-concave in p_i. Then if data shifts the corresponding inverse demand, $P(q_i, \delta_i)$, additively or multiplicatively, it is UPC. The same results apply if $P(q_i, \delta_i)$ is log-concave in q_i and data shifts $Q(p_i, \delta_i)$ additively or multiplicatively.*

Proof. First, $\hat{p}(u_i, \delta_i)$, the price that generates utility u_i, is implicitly defined by

$$u_i = \int_{\hat{p}(u_i,\delta_i)}^{\infty} Q(x, \delta_i) dx.$$

We have $\frac{\partial \hat{p}}{\partial \delta_i} \geq 0$. Firm i's per-consumer profit is $r(u_i, \delta_i) = \hat{p}(u_i, \delta_i)Q(\hat{p}(u_i, \delta_i), \delta_i)$. Using the property that $\frac{\partial \hat{p}}{\partial u_i} = -\frac{1}{Q(\hat{p}(u_i,\delta_i),\delta_i)}$ (by the implicit function theorem), we can write

$$\frac{\partial r(u_i, \delta_i)}{\partial u_i} = -1 + \eta(u_i, \delta_i),$$

where $\eta(u_i, \delta_i) = -\frac{\partial Q}{\partial p}\frac{p}{Q}$ is the price elasticity of demand. The cross-derivative of the per-consumer profit is then

$$\frac{\partial^2 r(u_i, \delta_i)}{\partial u_i \partial \delta_i} = \frac{\partial \eta(u_i, \delta_i)}{\partial \delta_i}$$

By Proposition 1 (1), we know that $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} \geq 0$ is a sufficient condition for data to be pro-competitive. Let us now show that $\frac{\partial \eta(u,\delta)}{\partial \delta} \geq 0$ in the four examples mentioned.

(i) If $Q(p_i, \delta_i) = \delta_i + Q(p_i)$, $\eta(u_i, \delta_i) = -\frac{\hat{p}(u_i,\delta_i)Q'(\hat{p}(u_i,\delta_i))}{\delta_i + Q(\hat{p}(u_i,\delta_i))}$. Then, $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i}$ is of the same sign as

$$-\frac{\partial \hat{p}(u_i, \delta_i)}{\partial \delta_i}\Big\{ \big[Q'(\hat{p}(u_i, \delta_i)) + \hat{p}(u_i, \delta_i)Q''(\hat{p}(u_i, \delta_i)\big]\big(\phi(\hat{p}(u_i, \delta_i)) + \delta_i\big) - \hat{p}(u_i, \delta_i)(Q'(\hat{p}(u_i, \delta_i)))^2\Big\}.$$

This is positive if $Q'(p) + pQ''(p) \leq 0$, which, by Lemma 1, is true if P is log-concave.

(ii) If $Q(p_i, \delta_i) = \delta_i Q(p_i)$, then $\eta(u_i, \delta_i) = -\frac{\hat{p}(u_i,\delta_i)Q'(\hat{p}(u_i,\delta_i))}{Q(\hat{p}(u_i,\delta_i))}$ and a similar calculation to case (1) applies.

For cases (iii) $(P(q_i, \delta_i) = \delta_i + P(q_i))$ and (iv) $(P(q_i, \delta_i) = \delta_i P(q_i))$, write $\eta(u_i, \delta_i) = -\frac{P(\hat{q}(u_i,\delta_i))}{\hat{q}(u_i,\delta_i)\frac{\partial P(\hat{q}(u_i,\delta_i),\delta_i)}{\partial q_i}}$. Then, $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i}$ is of the same sign as

$$-\Big\{\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial \delta_i}\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i}\hat{q}(u_i, \delta_i) - P(\hat{q}(u_i, \delta_i)\Big[\frac{\partial \hat{q}(u_i, \delta_i)}{\partial \delta_i}\Big(\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i} + \hat{q}(u_i, \delta_i)\frac{\partial^2 P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i^2}\Big) + \hat{q}(u_i, \delta_i)\frac{\partial^2 P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i \partial \delta_i}\Big]\Big\}.$$

The term $\frac{\partial P(\hat{q}(u_i,\delta_i),\delta_i)}{\partial q_i} + \hat{q}(u_i, \delta_i)\frac{\partial^2 P(\hat{q}(u_i,\delta_i),\delta_i)}{\partial q_i^2}$ is non-positive when Q is log-concave, and $\frac{\partial^2 P(\hat{q}(u_i,\delta_i),\delta_i)}{\partial q_i \partial \delta_i}$ is equal to zero in case (3), and to $P'(\hat{q}(u_i, \delta_i)) < 0$ in case (4), so that $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i} > 0$ in both cases. ■

In this case the surplus extraction effect is active and runs in consumers' favor, reinforcing the UPC effect of data. To see why, suppose the firm increases utility by lowering its price. Since the area beneath the demand curve is fixed, the extra utility must come from a mix of lower revenue and reduced deadweight loss. Under the conditions specified in Corollary 1, less revenue must be sacrificed to provide a utility increase after data shifts demand.

B.2 Moral hazard

Here we prove Proposition 3.

The incentive compatibility and target utility constraints are respectively

$$\alpha V(-p) + (1 - \alpha)\left[\delta V(-p + X_H - L) + (1 - \delta)V(-p + X_L - L)\right] - k \geq V(-p + X_L - L) \quad (8)$$

and

$$\alpha V(-p) + (1-\alpha)\left[\delta V(-p+X_H-L) + (1-\delta)V(-p+X_L-L)\right] - k = u. \tag{9}$$

It is fairly easy to prove that (8) must bind in equilibrium: the insurer could improve upon a non-binding constraint by offering slightly more insurance in exchange for a higher premium, until the constraint binds. Combining the two constraints therefore implies that $V(-p+X_L-L) = u$, i.e. $X_L = L + p + V^{-1}(u)$. We then substitute X_L in the objective (4) and in (9), and write the Lagrangian

$$\begin{aligned}\mathcal{L} = p - (1-\alpha)\left[\delta X_H + (1-\delta)(L+p+V^{-1}(u))\right] \\ + \lambda\left\{\alpha V(-p) + (1-\alpha)\left[\delta V(-p+X_H-L) + (1-\delta)u\right] - k - u\right\}. \end{aligned} \tag{10}$$

By combining the first-order conditions with respect to p and X_H, we obtain $V'(-p) = V'(-p-L+X_H)$, i.e. $X_H = L$: it is optimal for the insurer to fully compensate consumers when it can prove that they exerted effort. Replacing X_H by L in the constraint, we obtain $V(-p) = u + k/(\alpha+\delta-\alpha\delta)$, i.e $p(u,\delta) = -V^{-1}\left(u + k/(\alpha+\delta-\alpha\delta)\right)$. This also implies that $X_L(u,\delta) = L + p(u,\delta) + V^{-1}(u) = L - V^{-1}\left(u + k/(\alpha+\delta-\alpha\delta)\right) + V^{-1}(u)$.

We can now rewrite the per-consumer profit as a function of u:

$$\begin{aligned} r(u,\delta) &= p(u,\delta) - (1-\alpha)\left[\delta X_H + (1-\delta)X_L(u,\delta)\right] \\ &= (\alpha+\delta-\alpha\delta)\left[V^{-1}(u) - V^{-1}\left(u + \frac{k}{\alpha+\delta-\alpha\delta}\right)\right] - V^{-1}(u) - (1-\alpha)L. \end{aligned} \tag{11}$$

The cross-derivative of $r(u,\delta)$ is

$$\begin{aligned}\frac{\partial r(u,\delta)}{\partial u \partial \delta} = (1-\alpha)\Bigg[(V^{-1})'(u) - (V^{-1})'\left(u + \frac{k}{\alpha+\delta-\alpha\delta}\right) \\ + \frac{k}{\alpha+\delta-\alpha\delta}(V^{-1})''\left(u + \frac{k}{\alpha+\delta-\alpha\delta}\right)\Bigg]. \end{aligned} \tag{12}$$

This is positive whenever $(V^{-1})'$ is convex. To see this, notice that (12) is positive if

$$(V^{-1})''(u+x) > \frac{(V^{-1})'(x+u) - (V^{-1})'(u)}{(x+u) - u},$$

where $x = k/(\alpha+\delta-\alpha\delta)$.

For constant absolute risk aversion we have

$$V(W) = c - e^{-\beta W} \iff (V^{-1})'(v) = \frac{1}{(c-v)\beta} \iff (V^{-1})'''(v) = \frac{2}{(c-v)^3\beta} > 0.$$

With constant relative risk aversion, the utility takes the form $V(W) = \frac{W^{1-\theta}-1}{1-\theta}$

(and the initial wealth is high enough that wealth is never negative), so that $V^{-1}(v) = ((1-\theta)v+1)^{\frac{1}{1-\theta}}$. We then have $\left(V^{-1}\right)'(v) = ((1-\theta)v+1)^{\frac{\theta}{1-\theta}}$. $\left(V^{-1}\right)'$ is convex if $\theta > 1/2$.

B.3 Price discrimination

Consider a market with several retailers, each offering a continuum of products $z \in [0,1]$. Each product has a constant marginal cost c (we momentarily relax the assumption that $c = 0$ since $c > 0$ allows us to consider the example of constant-elasticity per-consumer demand). Consumer's valuations are independently and identically distributed across products, with a cumulative distribution function $F(v_z)$. Let $Q(p) = 1 - F(p)$ be the demand for each product if it is sold at a uniform price p, such that $p \mapsto (p-c)Q(p)$ is concave.

Each retailer i has a dataset that allows it to perfectly infer consumer l's valuations for a share δ_{il} of the products. Retailers can offer fully personalized prices: for each product z and consumer l, retailer i sets a price p_{izl}. Let us order retailer i's products so that i observes the consumer's valuations for products $z \in [0, \delta_{il}]$ (we call them the *identified* products).

Consumers are one-stop shoppers, and their utility has a retailer-specific shock ϵ_{il}, of zero mean. They observe all valuations and prices, and visit the retailer that offers them the highest surplus. Because firms can offer fully personalized prices, we can consider each consumer as a separate market. We thus drop the index l from now on.

The per-consumer profit is given by

$$r_i = \int_0^{\delta_i} (p_{iz} - c)\mathbb{1}_{p_{iz} \leq v_{iz}} dz + \int_{\delta_i}^1 (p_{iz} - c)Q(p_{iz})dz$$

Our objective is to write r_i as a function of u_i and δ_i so that we can exploit Proposition 1 to study the effects of data. It will be useful to write the mean utility as

$$u_i = \underbrace{\int_0^{\delta_i} (v_z - p_{iz})\mathbb{1}_{p_{iz} \leq v_{iz}} dz}_{\equiv U^I} + \underbrace{\int_{\delta_i}^1 (v_z - p_{iz})\mathbb{1}_{p_{iz} \leq v_{iz}} dz}_{\equiv U^{NI}}$$

where U^I (resp. U^{NI}) is the utility provided through identified (resp. non identified) products.

For each identified product the firm observes the consumer's valuation and can therefore avoid any deadweight loss. Providing a level of utility U^I through these products allows the firm to generate a profit of $r^I(U^I, \delta_i) \equiv \delta_i \bar{w} - U^I$, where $\bar{w} = \int_c^\infty Q(x)dx$ is the maximal total surplus.

We define $\rho(s)$ to be such that

$$s = \int_{\rho(s)}^{\infty} Q(x)dx.$$

In words, $\rho(s)$ is the price that delivers a consumer surplus s for a single product. If retailer i wants to provide a utility U^{NI} through a mass $(1-\delta_i)$ of non-identified products, it cannot do better than set a consumer-specific uniform price $\rho\left(\frac{U^{NI}}{1-\delta_i}\right)$. The associated profit is $r^{NI}(U^{NI},\delta_i) \equiv (1-\delta_i)\left[\rho\left(\frac{U^{NI}}{1-\delta_i}\right) - c\right] Q\left(\rho\left(\frac{U^{NI}}{1-\delta_i}\right)\right)$.

Putting things together, if firm i wants to provide a mean utility u_i to a given consumer, the maximal profit it can achieve if this consumer chooses i is

$$r(u_i,\delta_i) = \max_{U^{NI}\geq 0} \left\{ r^I(u_i - U^{NI},\delta_i) + r^{NI}(U^{NI},\delta_i) \right\}. \tag{13}$$

We restrict attention to the parameter region such that Assumption 1 holds, that is where $u_i < (1-\delta_i)\bar{w}$.[32] We are now ready to prove Proposition 4.

B.3.1 Proof of Proposition 4

As a first step we prove the following Lemma, which characterizes the optimal way to provide a given target utility.

Lemma 2. *If firm i wishes to offer utility $u_i \leq (1-\delta_i)\overline{w}$, then it optimally extracts all the value from identified products: $U_i^I = 0$.*

Proof. Suppose that $u_i \leq (1-\delta_i)\overline{w}$, and that $U_i^I > 0$. Let us show that it would be optimal for firm i to increase $U_i^{NI} > 0$ (and thus to decrease $U_i^I > 0$ since u_i is kept constant). Recall that the per-consumer profit is $r(u_i,\delta_i) = \max_{U_i^{NI}\geq 0}\left\{r^I(u_i - U_i^{NI},\delta_i) + r^{NI}(U_i^{NI},\delta_i)\right\}$. Let us define $\tilde{r}(u_i,U_i^{NI},\delta_i) \equiv r^I(u_i - U_i^{NI},\delta_i) + r^{NI}(U_i^{NI},\delta_i)$. Using the expressions in the text for r^I and r^{NI}, as well as $\rho'(s) = -1/Q(\rho(s))$, we find that

$$\frac{\partial \tilde{r}(u_i,U_i^{NI},\delta_i)}{\partial U_i^{NI}} = -\frac{\left[\rho\left(\frac{U_i^{NI}}{1-\delta_i}\right) - c\right] Q'\left(\rho\left(\frac{U_i^{NI}}{1-\delta_i}\right)\right)}{Q\left(\rho\left(\frac{U_i^{NI}}{1-\delta_i}\right)\right)} > 0$$

Therefore $U_i^I > 0$ cannot be optimal.

■

Proof of part (i) of Proposition 4. Since we restrict attention to situations where

[32] This in particular implies that competition among firms is not too strong. When the inequality is reversed firm i has enough data to eliminate all the deadweight loss, and $\partial r/\partial \delta_i = 0$. In that region one can show that data is competitively neutral.

$u_i \leq (1-\delta_i)\overline{w}$, by Lemma 2, we have $r^I = \delta\bar{w}$. Thus,

$$r(u_i, \delta_i) = \delta\bar{w} + (1-\delta)\left[\left(\rho\left(\frac{u_i}{1-\delta_i}\right) - c\right) Q\left(\rho\left(\frac{u_i}{1-\delta_i}\right)\right)\right]. \tag{14}$$

We have $\frac{\partial r_i}{\partial u_i \partial \delta_i} < 0$ because ρ is decreasing and $p \mapsto (p-c)Q(p)$ is concave. ■

Proof of part (ii) of Proposition 4: *Linear demand.* Suppose demand has the form $Q(p) = a - bp$ for $a, b > 0$. We can compute

$$\rho(s) = \frac{a - \sqrt{2}\sqrt{b}\sqrt{s}}{b}, \quad \bar{w} = \frac{(a-bc)^2}{2b}.$$

Substituting these expressions into $r(u_i, \delta_i)$ (equation 14), we find that $\frac{\partial \ln(r)}{\partial u_i, \delta_i}$ has the same sign as

$$a\left(2\sqrt{2}bc(2-\delta_i) + \frac{8\sqrt{b}u_i}{\sqrt{\frac{u_i}{1-\delta_i}}}\right) - b\left(4\sqrt{2}u_i + \sqrt{2}bc^2(2-\delta_i) + \frac{8\sqrt{b}cu_i}{\sqrt{\frac{u_i}{1-\delta_i}}}\right) - \sqrt{2}a^2(2-\delta_i).$$

This is concave in u_i and maximized at $u_i = (1-\delta_i)\bar{w}$. Making this substitution, the expression's maximal value is $-\sqrt{2}(a-bc)^2\delta \leq 0$.

Constant elasticity of demand. Suppose demand has the form $Q(p) = ap^{-\eta}$ for $a > 0$ and constant elasticity $\eta > 1$. We can compute

$$\rho(s) = \left(\frac{s(\eta-1)}{a}\right)^{-\frac{1}{\eta-1}}, \quad \bar{w} = \frac{ac^{1-\eta}}{\eta-1}$$

Substituting these expressions into $r(u_i, \delta_i)$ (equation 14), we can compute (writing $\rho = \rho\left(\frac{u_i}{1-\delta_i}\right)$)

$$\frac{\partial \ln(r)}{\partial u_i, \delta_i} = \frac{c\rho^{-\eta}\left\{c^{-\eta}\left[c\eta((1-\delta)\eta - 1) - (1-\delta)(\eta-1)^2\rho\right] - (1-\delta)(\eta-1)\rho^{1-\eta}\right\}}{u(\eta-1)\left[c^{1-\eta}\delta + (1-\delta)(\eta-1)(\rho-c)\rho^{-\eta}\right]^2}.$$

This has the same sign as the term in curly brackets, which is decreasing in ρ. Letting $\rho \to c$, the term in curly brackets becomes

$$-c^{1-\eta}\delta\eta < 0.$$

■

B.4 Targeted advertising

B.4.1 Example

Per-consumer revenue is $n_i P(n_i, \delta_i)$, i.e. $r(u_i, \delta_i) = \frac{V-u_i}{\gamma} P\left(\frac{V-u_i}{\gamma}, \delta_i\right)$, where $P(n_i, \delta_i)$ is given in (5). We then have

$$\frac{\partial^2 \ln\left(r(u_i, \delta_i)\right)}{\partial u_i \partial \delta_i} = \frac{2k\left(\lambda(\delta_i)\mu'(\delta_i) - \mu(\delta_i)(1+\mu(\delta_i))\lambda'(\delta_i)\right)}{k\lambda(\delta_i) + \mu(\delta_i)\left(k\lambda(\delta_i) - 2(V-u)\right)^2}.$$

While ambiguous in general, the sign of $\frac{\partial^2 \ln(r(u_i,\delta_i))}{\partial u_i \partial \delta_i}$ is unambiguous when either λ or μ is constant. When $\lambda'(\delta_i) = 0$, we have $\frac{\partial^2 \ln(r(u_i,\delta_i))}{\partial u_i \partial \delta_i} > 0$: when data only brings information about the horizontal taste of consumers within a category, data is UPC, as more data leads firms to show fewer ads to consumers. On the other hand, when $\mu'(\delta_i) = 0$, $\frac{\partial^2 \ln(r(u_i,\delta_i))}{\partial u_i \partial \delta_i} < 0$: when data is informative about relevant categories, it is UAC, as it leads firms to increase the ad load.

To construct the example in Figure 2 we suppose the platform is a monopolist facing consumer demand $D(u_i) = 4u - 3$. We also let $\lambda(\delta) = \frac{1}{10} + \lambda\delta$ and $\mu(\delta) = \frac{1}{10} + \mu\delta$. In the left panel of Figure 2 we set $\lambda = \frac{1}{2}$ and $\mu = \frac{1}{10}$ so data mostly identifies relevant categories. In the right panel $\lambda = \frac{1}{10}$ and $\mu = \frac{1}{2}$, meaning data mostly identifies relevant brands.

B.5 An example with congruent payoffs

Consider a media market in which firms compete for attention by investing $C(u_i)$ in providing free content that generates average utility u_i. Firms' revenue comes from selling n targeted ads with inverse demand from advertisers $P(n, \delta_i)$, decreasing in n and increasing in targeting accuracy, δ_i over the relevant range. The firm can show at most one ad for each unit of attention its content attracts. One can construct a model of consumers' time use in which the amount of attention, a, is increasing in content quality, $a'(u_i) > 0$.[33] Per-consumer revenue is therefore $r(u_i, \delta_i) = \max_{n \le a(u_i)} nP(n, \delta_i)$. As long as $\frac{\partial[nP(n,\delta_i)]}{\partial n}|_{n=a(u_i)} > 0$ the attention constraint $n \le a(u_i)$ is binding, and payoffs are congruent: $\frac{\partial r_i}{\partial u_i} > 0$.[34] We also have $\frac{\partial^2 r_i}{\partial u_i \partial \delta_i} > 0$, so data is UPC by Proposition 1. Thus, Proposition 6 allows us to characterise the competitive effects of data in this market: an increase in δ_i leads to an increase in u_i^* and to a decrease in u_j^*.

[33] For example, suppose the firm chooses quality q_i at cost $C(q_i)$. Consumers get utility $\sqrt{4aq}$ from spending a units of attention consuming content of quality q, and one unit of utility for each unit of attention spent on the outside option. Then the indirect utility is $u(q_i) = \max_a\{\sqrt{4aq_i} - a\}$. We find $u(q_i) = q_i$ with the optimal a being $a(q_i) = q_i$. We can therefore use a change of variables and write $C(u_i)$ and $a(u_i) = u_i$. We see that $a(u_i)$ is indeed increasing.

[34] The attention constraint will bind, for example, if $C'(u_i)$ is large enough.