Data and Competition: a Simple Framework, with Applications to Mergers and Market Structure*

Alexandre de Cornière† and Greg Taylor‡

December 3, 2021

Abstract

What role does data play in competition? This question has been at the center of a fierce debate around competition policy in the digital economy. We provide a simple framework for studying the competitive effects of data, encompassing a wide range of applications (product improvement, targeted advertising, price-discrimination) using a competition-in-utilities approach. We model data as a revenue-shifter, and identify conditions for data to be pro- or anti-competitive. The conditions are simple and often do not require knowledge of market demand or calculation of equilibrium. We use this framework to address policy-relevant questions related to market structure and data-driven mergers. We show that the effects of a data-driven merger between firms operating on adjacent markets depend both on whether data is pro- or anti-competitive and on firms' ability to trade data absent the merger.

Keywords: competition, big data, data-driven mergers, 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

*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

or less competitive? It is a difficult question because firms use data in many different ways, be it targeted advertising, price-discrimination, or product improvement (e.g. better search results, more personalized recommendations). Moreover, large-scale use of data enables various efficiencies (such as making new kinds of products possible), but also raises many concerns. A first concern is that data may facilitate exploitative behavior, allowing firms to extract more surplus from consumers.[2] Secondly, there are concerns that data may have adverse implications for market structure, raising barriers to entry or creating winner-take-all situations (see, e.g., Furman et al., 2019, 1.71 to 1.79). Thirdly, an increasing number of mergers in the digital sector involve data,[3] and there is still a debate as to how such data-driven mergers should be regulated (Grunes and Stucke, 2016).

A defining feature of data in the digital economy is 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.[4] 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 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. Inspired by Armstrong and Vickers' work on price-discrimination, we model data as a factor that generates more revenues for a given level of utility provided, a natural property across many uses of data. 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.

Our first main result characterizes environments where data is *unilaterally pro-competitive*, in the sense that a better dataset shifts a firm's best-response in the utility space 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 more consumers. The second is the *surplus extraction effect*: data

Data, and Competition, see `https://www.ftc.gov/news-events/events-calendar/ftc-hearing-6-competition-consumer-protection-21st-century`, accessed 1 May 2019.

[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] See Argentesi et al. (2019) and Motta and Peitz (2020) for an overview of mergers involving large technology firms.

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

sometimes enables firms to extract consumer surplus in a more efficient manner, thereby increasing the opportunity cost of providing more utility. We then show that, in many cases, this trade-off can be resolved without having to compute the equilibrium or to make functional form assumptions about demand, but instead depends only on the mapping between utility and per-consumer revenue, more precisely on the semi-elasticity of the latter with respect to the former. In particular, we can often pin-down the unilateral effects of data by checking a simple super-modularity condition on the per-consumer revenue function.

We apply the result to several examples inspired by standard models of data usage. In the first, data is used to improve the quality of the firm's product, allowing it to charge a higher mark-up. This gives the firm an incentive to offer higher utility to attract more consumers (i.e., data is pro-competitive). Secondly, we consider a firm that can use data to price discriminate. Having more data increases the opportunity cost of providing utility because any extra utility increasingly comes out of the firm's revenues rather than the (diminishing) deadweight loss. This tends to make data anti-competitive. Lastly, in a model of targeted advertising, data increases the number of ads that a media platform chooses to show because well-targeted ads are more valuable. This can be either pro- or anti-competitive, depending on whether data increases or reduces the elasticity of advertisers' demand. These three applications show that the trade-off described above can play out quite differently, depending on how exactly data is used. However, in all three cases we are able to use our simple conditions to characterize data's effects.

One attractive property of the competition-in-utility model is that it can accommodate cases of strategic complementarity or substitutability, depending (as we show) on whether firms' revenues come at the expense of consumers or not. Since this strategic effect and the earlier unilateral effect can both be characterized from the per-consumer revenue function, 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.

Embedding the static model into a dynamic framework, we next turn to some important wider policy issues. First, we study the link between data and market structure by considering a dynamic model 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 leads 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.

Another issue where our model can be usefully applied is the study of data-driven mergers. We consider two adjacent markets: the data generated on the (monopolized) market A can be used by firms that compete on market B. Here, data is an endogenous byproduct of activity on market A, and thus depends positively on the utility offered to consumers on that market. We look at a merger between the monopolist on market A and one of the B competitors, and study in particular how the merger may affect the incentives of firm A to collect data by providing utility to consumers. In this context, a specificity of data is that it may not be possible for firm A to license its data to a B firm absent the merger, either because of regulatory constraints or contractual frictions. We show that whether data trade is possible without the merger is an important factor, along with the pro- or anti-competitive nature of data, in determining whether the merger benefits consumers in each market. We discuss how this theory of harm differs from standard vertical foreclosure theories, and how it could be applied in practice.

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 some classic 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 5.2. Section 6 uses the framework to study data-driven mergers. We conclude in Section 7.

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. For example, active literatures consider 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; Chen et al., 2020; Gu et al., 2019; Ichihashi, forthcoming) or 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. Our contribution is to develop a framework that allows us to systematically analyze the competitive effects of data while remaining agnostic about how the data is used, and provide new results on how some of the key trade-offs involved play out in different contexts.

One important contemporary question concerns the control of mergers involving the exchange of data. A few papers (Chen et al., forthcoming; Kim and Choi, 2010; Esteves and Vasconcelos, 2015; Kim et al., 2018) study this question in models where data is used for price-discrimination purposes. Prat and Valletti (2019) consider mergers between media platforms offering slots for targeted advertising. By contrast, our framework allows us to study the effects of a data-driven merger across different business models and link those effects back to the underlying technology of data use. While related to the literature on vertical integration (Riordan, 2005), a data-driven merger differs from a standard vertical one: both the "upstream" and the "downstream" firms in our model may face the same set of consumers, and in some cases a merger is the only way to transfer the input (data) among firms.[5]

Another important theme in the policy debate concerns the relationship between data use or accumulation and market structure. Recent papers such as Prüfer and Schottmüller (Forthcoming), Farboodi et al. (2019) and Hagiu and Wright (2020) 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).[6] In Section 5.2 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 speculated 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 and unilateral effects of data

2.1 Model description

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, 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 $\frac{\partial D_i(u_i,\mathbf{u}_{-i})}{\partial u_i} \geq 0$

[5] Condorelli and Padilla (2020) also look at cross-market data use, but study a different question, namely "predatory entry".

[6] 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 an empirical study on the GDPR.

and $\frac{\partial D_i(u_i, \mathbf{u}_{-i})}{\partial u_j} \leq 0$ for $j \neq i$. [7]

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 $C''(u_i) \geq 0$.[8]

Data Each firm has access to data containing strategically relevant information about the market. The quality of the data may vary with the number of variables or observations it contains, or with the relevancy, accuracy or recency of those observations. The key assumption in our model is that datasets can be indexed by $\delta_i \in \mathbb{R}$ such that a firm's mark-up, $r(u_i, \delta_i)$, is increasing in δ_i.

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: $\frac{\partial r(u_i, \delta_i)}{\partial \delta_i} \geq 0$.

In other words, the quality of a dataset is measured by its potential to increase the mark-up.[9] This assumption follows a long tradition in which the informativeness of a signal can be measured by its value to a decision-maker (e.g., Blackwell, 1951; Blackwell, 1953; Athey and Levin, 2018). 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. This way of introducing data in a competition-in-utilities approach allows us to flexibly analyze a variety of different business models and technologies for using data—each corresponding to a different relationship between u, δ, and r.

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.[10] Then we have $r(u_i, \delta_i) = p_i = V(\delta_i) - u_i$.

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

[7]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).

[8]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 $C'(u_i) > 0$.

[9]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, though, would remain unchanged.

[10]In Section 4.1 we provide a microfoundation for this formulation based on firms making product recommendations.

provides better results for queries it has seen before. Alternatively, the 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 $D_i(u_i, \mathbf{u}_{-i})$.

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.[11]

3 Unilateral effects of data and monopolists' incentives

We begin by studying how data affects firms' incentives to offer utility, treating δ_i as an exogenous parameter. We will later endogenize δ_i by considering various ways that data is obtained as a by-product of economic activity, starting in Section 5.2.

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 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.

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:

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

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:

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

[11] Because competition in utilities can encompass a wide variety of strategic environments, it will typically be necessary to check quasi-concavity for any given application of the framework. But some sufficient conditions are (i) that C is sufficiently convex, or (ii) that C is non-concave and that both r and D_i are log-concave.

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 some natural examples). Then we can substitute the first-order condition, $r\frac{\partial D_i}{\partial u_i} + \frac{\partial r}{\partial u_i} D_i = 0$, into (3) and obtain that data is UPC if and only if $r\frac{\partial^2 r}{\partial u_i \partial \delta_i} > \frac{\partial r}{\partial u_i}\frac{\partial r}{\partial \delta_i}$, which is equivalent to $\frac{\partial^2 \ln(r)}{\partial u_i \partial \delta_i} > 0$. We summarize this discussion in the following proposition (whose proof is in Appendix A):

Proposition 1. 1. *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 $\mathbf{u}_{-i}$.*

2. *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.[12] 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. We next turn to some examples of such technologies.

4 Applications

One advantage of our framework is that it is simple but flexible enough to accommodate various uses of data. But, because competition in utility may seem somewhat abstract, it is probably helpful to outline some concrete applications. In this section we discuss

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

three applications that build on classic models of product improvement (Cowan, 2004), price discrimination (Armstrong and Vickers, 2001), and targeted advertising by media platforms (Anderson and Coate, 2005). We show how the competition in utility framework can be used to study these issues, and how Proposition 1 can inform us regarding the pro or anti-competitive effects of data. At the end of the section we also discuss some limitations of the model. Formal details can be found in Appendix B.

4.1 Product improvement

One important use of data is to improve the quality of the products or services offered by firms. 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 their viewing history, or an online retailer suggesting products to consumers based on past purchases.

As a concrete example, consider a situation where multiproduct firms use data to recommend experience goods (e.g., movies) to their users. The product space is the real line, and firms may recommend any product to their customers. A consumer of type $\theta \in \mathbb{R}$ places value $v - (\theta - x)^2$ on product $x \in \mathbb{R}$. Analysis of the firm's data yields a noisy signal, s_i, about the consumer's tastes, normally distributed with mean θ and variance $\frac{1}{\delta_i}$. The firm recommends product s_i at price p_i. Consumers do not observe the realisation of s_i prior to the purchasing decision, but know the precision of the recommendation. Letting $V(\delta_i) \equiv E[v - (\theta - s_i)^2] = v - \frac{1}{\delta_i}$, the mean utility of accepting this recommendation is $u_i = V(\delta_i) - p_i$. Thus, firm i's per-user revenue is $r_i(u_i, \delta_i) = V(\delta_i) - u_i$. We have $\frac{\partial^2 r_i}{\partial u_i \partial \delta_i} = 0$, so that the surplus extraction effect is inactive and data is UPC by Proposition 1(1). 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 u_i to attract more consumers.

In a recent paper on data-enabled learning, Hagiu and Wright (2020) study a model with a similar structure:[13] there the utility of a consumer is $u_i = s_i + \delta_i - p_i$ where s_i is the standalone value of i, and δ_i is a function of past sales.[14] Data is therefore also UPC in their model.

These examples belong to a more general class of models where data is a demand shifter (Cowan, 2004). For instance, consider a model where consumers can buy multiple products from firm i. Each consumer demands $Q(p_i, \delta_i)$ units at price p_i, with corresponding inverse

[13] 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 (2020) 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.

[14] We discuss some dynamic aspects in Section 5.

demand $P(q_i, \delta_i)$. Data improves the product and causes demand to shift up: $\frac{\partial Q(p_i,\delta_i)}{\partial \delta_i} > 0$. Utility when the price is p_i is given by the standard consumer surplus measure,

$$u_i = \int_{p_i}^{\infty} Q(x, \delta_i)\, dx, \tag{4}$$

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 (4) to get p_i as a function of u_i and δ_i, and therefore apply our supermodularity conditions. In Appendix B 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 ($P(q_i, \delta_i) = (1 + \delta_i)P(q_i)$). Applying Proposition 1 we prove the following result there:

Application 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.*

4.2 Price-discrimination

Armstrong and Vickers (2001) use the competition-in-utility framework to study competitive price-discrimination.[15] 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, implying a downward-sloping per-consumer demand curve, $Q(p)$, which we assume is such that $p \mapsto (p - c)Q(p)$ is concave (c being the marginal cost).[16] 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. The utility of choosing a firm is given by the expected consumer surplus it provides. We provide the formal analysis of such a model in Appendix B, where we show the following result:

Application 2. *In the game of competitive price-discrimination à la Armstrong and Vickers (2001), $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} < 0$.*

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).

[15] 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.

[16] We momentarily relax the assumption that $c = 0$ since $c > 0$ allows us to consider the example of constant-elasticity per-consumer demand.

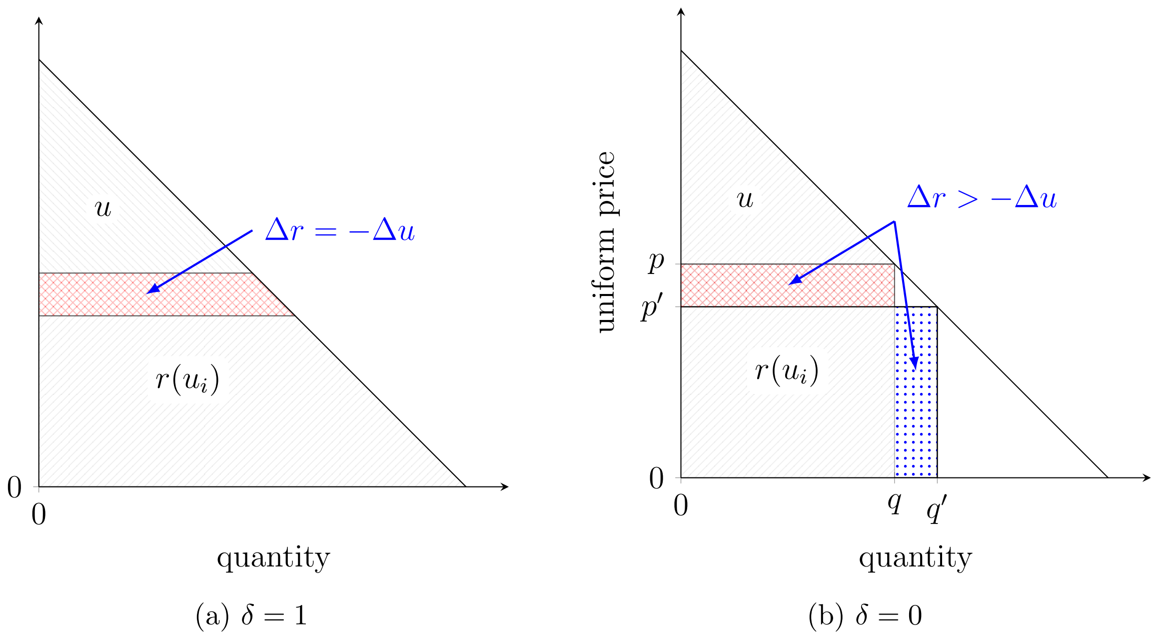


Figure 1: (a) If the firm can perfectly discriminate, offering one unit of utility reduces revenues by 1. (b) If the firm cannot observe consumers' willingness to pay, offering one unit of utility reduces revenues by less than 1.

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). This makes firms less willing to offer high levels of utility when their data improves.

The result above does not directly imply that data is always unilaterally anti-competitive, but does mean that data makes surplus extraction more efficient, which potentially offsets the increased value of marginal consumers. In Appendix B we derive sufficient conditions for data to be UAC in the case of linear or constant-elasticity demand functions, using the log-submodularity condition of Proposition 1.

4.3 Targeted advertising

Another 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 targeted advertising.

Consider horizontally differentiated media platforms that show content to consumers and sell advertising space to advertisers. Data held by platform i allows it to offer personalized advertising. As in Anderson and Coate (2005), ads impose a linear nuisance cost on viewers: if the firm shows n_i ads then utility is

$$u_i = v - \phi n_i, \tag{5}$$

where v is the baseline value of the content and ϕ measures the nuisance cost of an advertisement.[17]

The platform chooses how many ad slots to sell for each consumer visit, and runs a uniform-price auction among advertisers. The data held by the platform can be used by advertisers to compute their willingness to bid, by generating an informative signal about the likelihood that the consumer is interested in the advertiser's product. This implies an inverse demand for advertising slots $P(n_i, \delta_i)$. We assume that, in the relevant range, $\frac{\partial P(n_i,\delta_i)}{\partial \delta_i} \geq 0$.[18] The platform's per-consumer revenue is $n_i P(n_i, \delta_i)$ which, using (5), can be rewritten as $r(u_i, \delta_i) = \frac{v-u_i}{\phi} P\left(\frac{v-u_i}{\phi}, \delta_i\right)$.

This model is one where fixed costs are constant ($C'(u) = 0$). By Proposition 1 (2), we know that data is UPC if and only if $\frac{\partial^2 \ln(r(u_i,\delta_i))}{\partial u_i \partial \delta_i} > 0$. Let $\epsilon(n_i, \delta_i)$ be the elasticity of advertisers' demand, i.e. $\epsilon(n_i, \delta_i) \equiv \frac{-P(n_i,\delta_i)}{n_i \frac{\partial P(n_i,\delta_i)}{\partial n_i}}$. We have the following:

Application 3. *In the Anderson and Coate (2005) model with targeted advertising, data is UPC if and only if the demand for ads becomes less elastic with better targeting , i.e.* $\frac{\partial \epsilon(n_i,\delta_i)}{\partial \delta_i} < 0$.

An increase in the price of ads due to better targeting has two opposite effects. First, it makes each consumer worth more to the platform. In order to attract more consumers, the platform has an incentive to lower the ad load. This corresponds to what we call the mark-up effect. Second, the higher ad price increases the opportunity cost of providing utility to consumers, since doing so entails showing fewer ads. In other words, more data makes the platform more efficient at extracting surplus from consumers: the same ad nuisance corresponds to higher revenues.

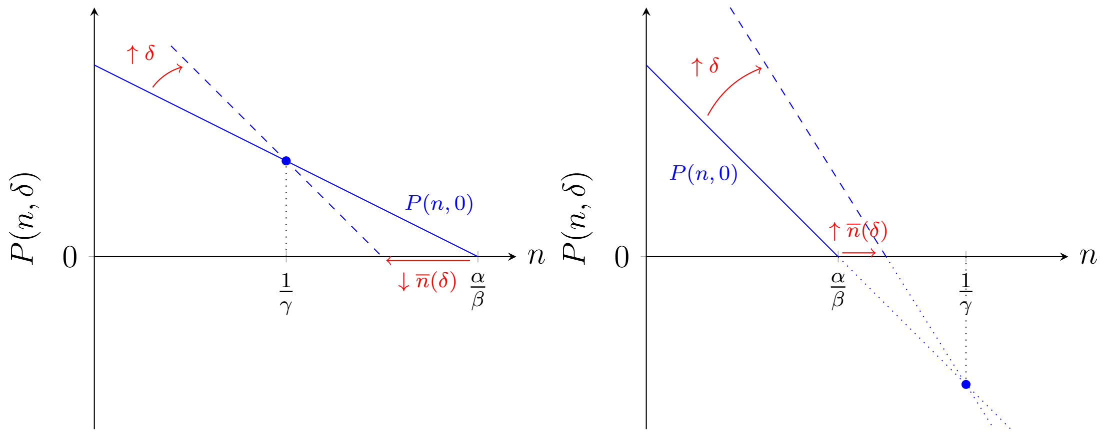
Which effect dominates depends on how the elasticity of the demand for ads changes with data. If demand for ads becomes less elastic as the platform collects more data, a small decrease in the number of ads generates a large increase in their price. The cost of showing fewer ads is then offset by the increase in the price, and the mark-up effect dominates: data is UPC.

Empirically, estimating advertisers' demand elasticity for various levels of targeting may be challenging given limited resources. To go further, one could assume that the demand for advertising slots takes the following form:

$$P(n_i, \delta_i) = \max\{\alpha + \delta_i - (\beta + \gamma\delta_i)n_i, 0\}.$$

[17] At the end of this sub-section we discuss the case where targeting lowers the nuisance cost of an ad.

[18] See Appendix B for a more detailed explanation. If targeting induces a demand rotation as in Johnson and Myatt (2006), this assumption means that the quantity of ads is below the rotation point. In other words, the marginal advertiser's willingness to pay increases with targeting (see also Rutt (2012) for a model of platform competition with targeted advertising using demand rotations).



(a) $\frac{1}{\gamma} < \frac{\alpha}{\beta}$: data is UPC (b) $\frac{1}{\gamma} > \frac{\alpha}{\beta}$: data is UAC

Figure 2: Rotation in linear ad demand caused by ad targeting. When the rotation point is below α/β data is UPC and leads fewer advertisers to bid. If the rotation point is above α/β the opposite is true.

This is a demand rotation with a rotation point at $n_i = \frac{1}{\gamma}$.[19]

In that case the condition on the elasticity of ad demand for data to be UPC is $\frac{1}{\gamma} < \frac{\alpha}{\beta}$. Let $\overline{n}(\delta_i)$ be the number of advertisers willing to submit a positive bid, i.e. the smallest n such that $P(\overline{n}(\delta_i), \delta_i) = 0$. Note that $\overline{n}(0) = \frac{\alpha}{\beta}$. This means that data is UPC if and only if the rotation point $\frac{1}{\gamma}$ is smaller than $\overline{n}(0)$. Because demand further rotates as δ_i increases, $\overline{n}(\delta_i)$ is then decreasing: when data is UPC the number of advertisers willing to submit a positive bid is smaller as targeting becomes more accurate (see Figure 2a). The reverse logic applies when data is UAC, so that $\overline{n}(\delta_i)$ is increasing (Figure 2b).

The implication of this analysis is that one could use the number of bidders as a proxy for the pro- or anti-competitive nature of data: if targeting attracts more bidders, data is UAC. Conversely, if targeting attracts fewer bidders, data is UPC.

Another possible effect of targeting is the reduction in the nuisance experienced by consumers. One could easily add this component by changing the utility function in (5) into $u_i = v - \phi(1 - \lambda\delta_i)n_i$, where λ would measure the extent to which consumers prefer targeted ads. Such an analysis would make it more likely that data is UPC.

Finally, one can 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

[19] And so, by footnote 18, we restrict attention to situations where the equilibrium is $n_i^* < \frac{1}{\gamma}$.

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.

4.4 Limitations and discussion

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 the focus is on the quality of data, and not on what we learn from it. More specifically, the utility offered to a consumer does not depend on the content of the information, only about its quality. 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).[20] But this is a features shared by many papers on the economics of data (e.g., Farboodi et al., 2019; Prüfer and Schottmüller, Forthcoming; Hagiu and Wright, 2020; Choi et al., 2019; Acemoglu et al., 2019, to name a few).

In our framework, δ_i can be interpreted as a parameter other than data that affects the mark-up of a firm. For instance, we could interpret δ_i as i's stock of cost-reducing innovations. However, data is particularly interesting because, as the applications above demonstrate, it naturally gives rise to both pro- and anti-competitive effects. In contrast, a textbook model of cost-reducing innovations would normally be UPC.[21] To further enrich the model in a way that is more particular to data, Sections 5.2 and 6 introduce the feature that data is often the endogenous by-product of past interactions with consumers.

To facilitate a clean exposition, we have omitted from the model some features common to many data-rich environments, among them network effects and consumer privacy concerns.[22] We show in the online appendix how these features can be incorporated into the model. Network effects can be accommodated in the unilateral analysis considered above. The main difference is that firm i's per-consumer revenue now also depends on u_j, which makes it harder to characterize the nature of the strategic interaction (strategic complements or substitutes, as we discuss in the next section). Strong privacy

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

[21] In a competition-in-utility framework, Shelegia and Wilson (forthcoming) provide several examples of revenue-shifting technologies, all of which would satisfy our UPC condition.

[22] A recent literature has emerged to study various issues related to the economics of privacy. Examples include Hermalin and Katz (2006), Casadesus-Masanell and Hervas-Drane (2015), Kim and Wagman (2015), Acemoglu et al. (2019), Bergemann et al. (2019), Choi et al. (2019), Dosis and Sand-Zantman (2019), Jann and Schottmüller (2019), Ichihashi (2020), Fainmesser et al. (2020), and Jones and Tonetti (forthcoming).

concerns may lead to the property that $\frac{\partial r_i}{\partial \delta_i} \leq 0$ because the firm has to compensate users for invasion of their privacy. This is readily accommodated after a straightforward adjustment to Proposition 1. In particular, we obtain an analogous necessary and sufficient log-supermodularity condition for data to be UPC.[23]

5 Equilibrium competitive effects of data

5.1 Static duopoly

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.[24]

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 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).

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.4 for an example.

From now on, let us assume that $\frac{\partial r}{\partial u}$ is of constant sign in the relevant domain.[25] Then

[23] The main difference is that submodularity of (privacy-adjusted) revenues is now a sufficient condition for data to be UAC (rather than supermodularity being sufficient for UPC).

[24] 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).

[25] In the applications of Section 4, $\frac{\partial r}{\partial u}$ is of constant sign, except for the targeted advertising one (Section 4.3). In that application the relevant domain is the values of u for which $\frac{\partial r}{\partial u} \leq 0$ because outside of this range it would always be profitable for the firm to offer more utility.

the congruence/conflict property suffices to characterize the strategic effect that is the missing ingredient in our equilibrium analysis:

Proposition 2. *(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 2 is in Appendix A. Propositions 1 and 2 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 2 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).[26]

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 using instruments (prices or ad loads) that reduce utility. Thus, applying Proposition 2, 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 ad targeting when data increases the elasticity of ad demand.

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.[27] 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 2 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.

[26]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.

[27]For example, Article 6(1) of the EU's proposed Digital Markets Act imposes obligations for incumbent "gatekeeper" platforms to share search query and other types of data with rival firms.

5.2 Some dynamic implications

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.

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). 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 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. The following remark is immediate.

Remark 1. *Data acts as a barrier to entry if and only if it is UPC.*

One can then apply Fudenberg and Tirole (1984)'s results, and conclude that an entry deterrence strategy will involve over-collection of data in the first period (to the benefit of early consumers) if data is UPC, while it will involve under-collection of data in the first period if it is UAC.

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 summarizes the results.

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

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, Forthcoming; Farboodi et al., 2019; Hagiu and Wright, 2020).

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[28] and proceed to 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. 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 starkly different long-run effects of data.

The formal analysis is in Appendix C, but the intuition is simple to describe. Suppose that we start from a situation where firm A has a data advantage. If data is UPC, firm A will offer a larger utility than firm B, thereby collecting more data, and will start the following period still with an advantage. UPC data therefore creates some degree of persistence of the data advantage, and *may* lead to increased concentration when the utility difference is very sensitive to differences in the quantity of data. On the other hand, when data is UAC, an initial data advantage can never result in increased concentration because the firm that enjoys it offers a lower utility than its rival and collects less data.

This points to a tension between exploitative and exclusionary theories of harm. When a data advantage leads a firm to act exploitatively 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 an immediate consumer welfare perspective, but consumers will tend to gravitate towards firms with lots of data, giving rise to entrenched market leadership.

6 Data-driven mergers

As noted in the previous section's discussion of dynamics, many firms collect data as a by-product of their interactions with consumers. Interestingly, customer data collected in one market is often of value to firms in another. This creates incentives to share data between

[28] As discussed in Section 4, the baseline model of Hagiu and Wright (2020) fits our competition-in-utility framework and data is UPC. For Prüfer and Schottmüller (Forthcoming), casting the model in a 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..

firms. Sometimes this is achieved through data sharing agreements (i.e., by trading data between independent firms). On other occasions data trade is impossible or impracticable, leading firms to merge in order to acquire data. Barriers to data trade might include data protection regulation,[29] moral hazard (the recipient might not protect the data, as when Facebook shared data with Cambridge Analytica), or fears that the shared data may enable the recipient to enter and compete in the the sharer's primary market. Such barriers are particularly likely to be important when the data in question is personal data or commercially sensitive. Significant mergers for which the acquisition of data was widely viewed as a key motivating factor include Microsoft–LinkedIn,[30] Facebook–WhatsApp,[31] and Google–FitBit.[32]

In this section we build upon our baseline framework to study data-driven mergers when pre-merger data trade is and is not possible. We model data as a byproduct of economic activity: the quantity (and quality) of data generated by a firm is an increasing function of the usage of its product. In order to focus on the data-related aspects of the merger, we assume that the merging firms operate on separate markets and are therefore not direct competitors.[33] We label the two markets A and B, and assume that data generated on market A can be used in market B.[34]

Such a structure shares some similarities with a vertical merger,[35] in the sense that a firm in a "downstream" market (B) obtains an input (data) from a firm in an "upstream" market (A).[36] A first difference is that the question of whether the input can be sold absent the merger plays an important role in the case of data-driven mergers but does not feature in analyses of standard vertical mergers. The second main difference is that selling data is not necessarily the primary purpose of the firm in the A market, which has

[29] The UK Competition and Markets Authority observes that mergers can be used to circumvent data protection rules: "The GDPR makes gaining and managing consent [...] within an undertaking, or group of undertakings in common control, an easier exercise than sharing data between undertakings to deliver the same purpose." See `https://assets.publishing.service.gov.uk/media/5dfa0580ed915d0933009761/Interim_report.pdf`, accessed 12 February 2020.

[30] In the Microsoft and LinkedIn case, LinkedIn's data could be used by Microsoft to customize its Customer Relationship Management (CRM) software, Dynamics 360. See `https://www.reuters.com/article/us-microsoft-linkedin-idUSKBN17Q1FW`, accessed 13 December 2019. Salesforce, the leader in the CRM market, was also reportedly interested in acquiring LinkedIn.

[31] Facebook has been in a position to use the data from WhatsApp to offer more personalized advertisements, even though it initially claimed this would not be technically feasible. See `http://europa.eu/rapid/press-release_IP-17-1369_en.htm`, accessed 13 December 2019.

[32] The Australian Competition and Consumer Commission (ACCC) recently offered a preliminary view that the aggregation of FitBit's health and fitness data by Google may substantially lessen competition in the markets for data-dependent health services and targeted advertising (ACCC, 2020).

[33] While this assumption seems plausible in the Microsoft/LinkedIn merger, it is more controversial in the Facebook/WhatsApp case, as both firms could be viewed as competitors in the market for social network services. The European Commission considered that the two companies are distant competitors, due to distinguishing features and consumers' ability to multi-home.

[34] For simplicity we ignore the possibility that data generated on B could be used on A as well.

[35] See (Riordan, 2005) on that topic.

[36] Also related, Condorelli and Padilla (2020) look at the related issue of cross-market data use, but study a different question, namely "predatory entry".

its own consumers (unlike an upstream manufacturer in a standard supply chain). An important aspect, which so far has been relatively neglected, is therefore how the merger will affect the consumer surplus in the A market.

6.1 The model

Firm A is a monopolist on market A and offers a mean utility u_A, leading to a profit $\tilde{\pi}_A(u_A)$. Serving consumers on its primary market allows firm A to collect a quantity of data $\delta_A \equiv \delta(u_A)$, with $\delta'(u_A) > 0$, either because more consumers use product A (extensive margin) or because consumers use product A more, which allows the firm to collect more data (intensive margin).[37] It will be convenient to operate a change of variables and say that firm A directly chooses a quantity of data δ_A, corresponding to a utility level $u_A(\delta_A)$, with $u'_A(\delta_A) > 0$.[38] Firm A's profit on its primary market is $\pi_A(\delta_A) \equiv \tilde{\pi}_A(u_A(\delta_A))$, which we assume is quasi-concave and maximized for $\widehat{\delta}$ such that $\pi'_A(\widehat{\delta}) = 0$.

The data can be used on a secondary market, B, where it can either be UPC or UAC. Two firms (B_1 and B_2) compete on market B, along the lines of the model described in Section 2: firms offer utility level u_i, $i \in \{1,2\}$, resulting in a demand $D_i(u_i, u_j)$. We assume that A is the unique source of data so that, if A transfers a quantity δ_i to firm B_i, the latter's per-consumer revenue is $r(u_i, \delta_i)$.[39] In a slight departure from the notation used above, we adopt the reduced-form profit expressions $\pi_i(\delta_i, \delta_j) \equiv r_i\left(u_i^*(\delta_i, \delta_j), \delta_i\right) D_i\left(u_i^*(\delta_i, \delta_j), u_j^*(\delta_j, \delta_i)\right) - C_i\left(u_i^*(\delta_i, \delta_j)\right)$, where u_i^* denotes the equilibrium utility level provided by B_i in the subgame where the data levels are δ_i and δ_j.

On the B market, we assume that the u_i's are strategic complements,[40] and that a firm's profit is increasing in the amount of data it has: $\frac{\partial \pi_i(\delta_i, \delta_j)}{\partial \delta_i} > 0$.[41]

We will consider two scenarios, depending on whether data trade between two independent firms can happen. As we will show, this is a critical determinant of whether the merger is likely to benefit consumers. The game proceeds as follows: At $t = 1$, firm A chooses δ_A. At $t = 2$ data trade takes place when possible. At $t = 3$ the firms in market B observe δ_1 and δ_2 and choose their utility offers.

[37] Alternatively, A might collect more data through more invasive data collection practices, in which case we could have $u'(\delta_A) < 0$. This case is easily incorporated into the analysis, as we discuss at the end of this section.

[38] This is different from the UPC condition. Here, data is collected as a byproduct of the economic activity: the choice of u_A determines the amount of data collected.

[39] Another equivalent interpretation is that the B firms start with the same level of data, and δ_i measures the additional data provided by A.

[40] Recall that, with additively separable demand, u_1 and u_2 are strategic complements if and only if $\frac{\partial r_i}{\partial u_i} < 0$.

[41] This last assumption rules-out situations where an increase in δ_i would lead firm B_j to compete so much more fiercely that B_i would prefer to commit not to use the data.

6.2 Merger when data trade is not possible

Suppose that pre-merger data trade between A and the B firms is impossible. We assume that the merger allows the new firm to transfer the data between A and B_1. We compare the equilibrium outcome when firms are independent to the case where A and B_1 merge. We use a superscript I for the case of independent firms, and a superscript M for the case where A and B_1 merge.

Independent firms Given that trade is impossible, firm A focuses solely on maximization of its A-market profit. It therefore chooses to collect $\delta_A^I = \widehat{\delta}$ by offering utility $\widehat{u}_A = u_A(\widehat{\delta})$. Since the B firms have no access to data, they offer utilities $u_i^*(0,0)$.

Merger At $t = 1$, firm AB_1 maximizes the joint profit of the integrated unit, $\pi_A(\delta_A) + \pi_1(\delta_A, 0)$. Given that $\frac{\partial \pi_1}{\partial \delta_1} > 0$, in equilibrium we must have $\delta_A^M > \widehat{\delta}$.

Comparison Since $\delta_A^M > \delta_A^I$ and $u'_A(\delta_A) > 0$, we have $u_A^M > u_A^I$. In words: consumer surplus on market A increases after the merger. In market B, the merger results in firm 1 having access to an additional δ_A^M data. The effect of the merger on consumer surplus on market B therefore depends on whether data is unilaterally pro- or anti-competitive.

Proposition 3. *When data trade between independent firms is not possible:*

1. *If data is UPC on market B the merger increases consumer surplus on both markets.*
2. *If data is UAC on market B the merger increases consumer surplus on market A but reduces it on market B.*

The merger allows data to find a new use in market B. This makes data more profitable, leading A to collect more data, requiring it to increase u_A. If data is UPC then the use of data also induces B-firms to increase their equilibrium utility offer, resulting in an unambiguous gain for all consumers. Such circumstances therefore favour a more lenient merger policy. If data is UAC, on the other hand, the use of data reduces utility offers in market B and consumers' gain in market A must be weighed against this loss.

6.3 Merger when data trade is possible

We now turn to the case where data can be traded without the merger. In practice, this situation could be characterized either by the existence of data trade prior to the merger, or by a high probability that such trade would happen in the near future. For instance, in the Google–Fitbit merger, the ACCC considered that, absent the acquisition, Fitbit would be likely to enter the market for health data supply.[42]

Data is a non-rival but excludable good, so that in theory firm A could sell it to one or both B firms, and indeed both options can be profitable depending on the parameters. For ease of exposition we focus on the case in which an exclusive sale is optimal. We present the condition for optimality and analyze the non-exclusive case in Appendix D, where we show that firm A optimally makes the same exclusivity decisions both pre- and post-merger.[43]

Given exclusivity, B_1 and B_2 compete in an auction to obtain the data. B_i is willing to pay up to the difference in its profit between the case in which it gets the data and the case in which B_j gets it. The price of data is thus $\pi_i(\delta_A, 0) - \pi_i(0, \delta_A)$.

Independent firms Firm A's profit is given by the combination of the profit in its primary market and the revenue from selling data: $\pi_A(\delta) + [\pi_i(\delta, 0) - \pi_i(0, \delta)]$. The first-order condition for firm A is therefore

$$\pi'_A(\delta) + \left[\frac{\partial \pi_i(\delta, 0)}{\partial \delta_i} - \frac{\partial \pi_i(0, \delta)}{\partial \delta_j}\right] = 0. \tag{6}$$

The amount of data collected affects the price of data through two channels: first, collecting more data increases the profit of the data holder by assumption. Second, it also affects the outside option by changing the profit of the firm that does not obtain the data. The sign of this second effect depends on whether data is unilaterally pro- or anti-competitive. Indeed, if data is UPC then a higher δ_i implies a higher u_i, which is bad for B_j's profit: $\frac{\partial \pi_j(0,\delta)}{\partial \delta_i} < 0$. This gives an extra incentive to collect data in order to degrade the outside option of not buying the data. The reverse holds when data is UAC.

Merger If A and B_1 merge, A still has the option to sell the data to B_2. However, if exclusivity is preferred when A is independent, such a strategy is never profitable (see Appendix D). Therefore, the profit of the integrated firm is comprised of the profit in market A and the profit that can be made by exclusively using the data itself in market B: $\pi_A(\delta_A) + \pi_1(\delta_A, 0)$. The corresponding first-order condition is

$$\pi'_A(\delta) + \frac{\partial \pi_i(\delta, 0)}{\partial \delta_i} = 0. \tag{7}$$

[42] See ACCC (2020), para. 18: " In the absence of the proposed acquisition, there is a greater likelihood that Fitbit—either under current or alternative ownership—will enter partnerships to make its data available to alternative suppliers of ad tech services (subject to privacy laws)".

[43] Thus, input foreclosure does not arise here. We stress that this modelling choice reflects our desire to emphasize the novel features of our approach (foreclosure having been widely studied elsewhere, e.g., Rey and Tirole, 2007) rather than a view that foreclosure is not an important consideration.

After the merger, firm AB_1 fully internalizes B_1's profit and no longer needs to manipulate its outside option, so the last term in (6) disappears.[44]

Comparing the first-order conditions (6) and (7), we therefore obtain the following:

Proposition 4. *When data trade among independent firms is possible:*

1. *If data is UPC on market B the merger leads to less data collection, reducing consumer surplus on both markets A and B.*

2. *If data is UAC on market B the merger leads to more data collection, increasing surplus on market A but reducing it on market B.*

6.4 Summary and discussion

A data-driven merger can affect consumers through two channels: by changing the distribution of data (and intensity of competition) in market B, and by changing incentives to collect data in the primary market A. Combining Propositions 3 and 4: If data is unilaterally pro-competitive, we find that surplus in markets A and B is aligned: the merger benefits consumers in both markets if data trade is impossible prior to the merger and harms them otherwise. If data is unilaterally anti-competitive then the surplus effects of the merger differ across markets: consumers benefit in market A from better offers designed to generate more data, but are harmed in market B where the extra data softens competition. The effect of the merger on utility in the two markets is summarized in Table 2.

Table 2: Effect of a data-driven merger on the utility offered on market A (u_A), and the utility offered on market B (u_B).

	data is UPC	data is UAC
Pre-merger data trade	$\downarrow u_A, \downarrow u_B$	$\uparrow u_A, \downarrow u_B$
No pre-merger data trade	$\uparrow u_A, \uparrow u_B$	$\uparrow u_A, \downarrow u_B$

Policy implications If we focus on the case where data is unilaterally pro-competitive, our analysis offers both an efficiency argument in favor of a data-driven merger (it enables data uses in adjacent markets) and a new theory of harm (the merger reduces incentives to collect data, leading to a lower utility in the primary market). The key condition is whether data trade is possible absent the merger.

[44] After (but not before) the merger, δ is chosen to maximize AB_1's joint profit. The merger is therefore profitable.

Absent pre-merger trade, and if there are no indications that such trade might take place in the near future, it is important to identify the source of the friction: a merger allowing firms to bypass regulations may undermine other public objectives, and the efficiency argument should be given less weight. One could even interpret the existence of regulations as an indication that the use of data does not increase consumers' utility, an argument in favour of blocking the merger. If, on the other hand, trade of data is hindered by other types of (e.g., contracting) frictions, our analysis suggests that the merger is more likely to benefit consumers.

To what extent should authorities and courts use the existence of pre-merger trade as a legal test for evaluating these arguments? Suppose first that market investigations reveal the existence of such trade. Authorities should then obviously lend less credence to the above efficiency argument. However, before accepting the theory of harm that we have proposed, several conditions should be checked: (i) Firm A has market power on the data market. Absent this condition, firm A would have no incentive to manipulate B firms' willingness to pay; (ii) Data trade is an important part of firm A's activity. Indeed, the main driving force of our result is that the incentive to manipulate the price of data is strong enough to affect A's behaviour in its primary market; (iii) The value of the dataset of firm A depends positively on the utility it offers to its primary customers. The idea is that a firm offering higher utility attracts more consumers and therefore gathers more data. But if consumers have strong privacy concerns, and if data is collected with privacy-invading technologies, we could have $u'_A(\delta_A) < 0$. In such a case, the implications, of the merger on market B are unchanged, but those on market A are reversed, so that the effects of the merger on each market are of opposite sign when data is UPC, while the merger is always harmful when data is UAC.

7 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.

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 also makes 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 three applications, where data is used respectively to improve product quality, to target advertising, and to price-discriminate.

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.

Finally, we turn our attention to data-driven mergers, where data collected in the market of one merging party can be used in the other's. We show that one key determinant of the effect of the merger on consumer surplus, beyond the UPC or UAC nature of data, is whether data can be traded absent the merger. Our analysis offers both a theory of harm and an efficiency argument, and we discuss the conditions for each to apply.

References

ACCC (2020). *Statement of Issues - Google LLC Proposed Acquisition of Fitbit*. Tech. rep. Australian Competition and Consumer Commission. URL: `https://www.accc.gov.au/public-registers/mergers-registers/public-informal-merger-reviews/google-llc-proposed-acquisition-of-fitbit-inc`.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar (2019). "Too Much Data: Prices and Inefficiencies in Data Markets". NBER Working Paper 26296. URL: `https://www.nber.org/papers/w26296.pdf`.

Acquisti, Alessandro and Hal R. Varian (2005). "Conditioning Prices on Purchase History". *Marketing Science* 24.3, pp. 305–523.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (2016). "The economics of privacy". *Journal of Economic Literature* 54.2, pp. 442–92.

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`.

Argentesi, Elena, Paolo Buccirossi, Emilio Calvano, Tomaso Duso, Alessia Marrazzio, and Salvatore Nava (2019). "Ex-post Assessment of Merger Control Decisions in Digital Markets." *Document prepared by Lear for the Competition and Markets Authority.*

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`.

Athey, Susan and Jonathan Levin (2018). "The value of information in monotone decision problems". *Research in Economics* 72.1, pp. 101–116. DOI: `https://doi.org/10.1016/j.rie.2017.01.001`. URL: `https://www.sciencedirect.com/science/article/pii/S1090944316303064`.

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 (2019). "The Economics of Social Data". Working Paper. URL: `http://www.mit.edu/~bonatti/social.pdf`.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile prices in market equilibrium". *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Blackwell, David (1951). "Comparison of Experiments". *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 99–102.

— (1953). "Equivalent Comparisons of Experiments". *The Annals of Mathematical Statistics* 24.2, pp. 265–272. URL: `http://www.jstor.org/stable/2236332`.

Bonatti, Alessandro and Gonzalo Cisternas (2019). "Consumer Scores and Price Discrimination". *The Review of Economic Studies*.

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.

Casadesus-Masanell, Ramon and Andres Hervas-Drane (2015). "Competing with Privacy". *Management Science* 61.1, pp. 229–246. DOI: 10.1287/mnsc.2014.2023.

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

Chen, Zhijun, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima (forthcoming). "Data-Driven Mergers and Personalization". *RAND Journal of Economics.*

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.

Condorelli, Danielle and Jorge Padilla (2020). "Data-driven Predatory Entry with Privacy-Policy Tying". Working Paper.

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.

Dosis, Anastasios and Wilfried Sand-Zantman (2019). "The ownership of data". *TSE Working Paper.*

Esteves, Rosa-Branca and Helder Vasconcelos (2015). "Price discrimination under customer recognition and mergers". *Journal of Economics & Management Strategy* 24.3, pp. 523–549.

Fainmesser, Itay P., Andrea Galeotti, and Ruslan Momot (2020). "Digital Privacy". *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 (2000). "Customer poaching and brand switching". *RAND Journal of Economics*, pp. 634–657.

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

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.

Ganuza, Juan-José and Jose S Penalva (2010). "Signal orderings based on dispersion and the supply of private information in auctions". *Econometrica* 78.3, pp. 1007–1030.

Gilbert, Richard J and David MG Newbery (1982). "Preemptive patenting and the persistence of monopoly". *The American Economic Review*, pp. 514–526.

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 (2020). "Data-enabled learning, network effects and competitive advantage". Working Paper.

Hermalin, Benjamin E. and Michael L. Katz (2006). "Privacy, property rights and efficiency: The economics of privacy as secrecy". *Quantitative Marketing and Economics* 4.3, pp. 209–239. URL: `http://www.springerlink.com/index/10.1007/s11129-005-9004-7`.

Ichihashi, Shota (2020). "Non-competing Data Intermediaries". Working Paper. URL: `https://shota2.github.io/research/data.pdf`.

— (forthcoming). "Online Privacy and Information Disclosure by Consumers". *American Economic Review*.

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`.

Jann, Ole and Christoph Schottmüller (2019). "An Informational Theory of Privacy". *The Economic Journal*. URL: `https://doi.org/10.1093/ej/uez045`.

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.

Jones, Charles I and Christopher Tonetti (forthcoming). "Nonrivalry and the Economics of Data". *American Economic Review*.

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, Byung-Cheol and Jay Pil Choi (2010). “Customer information sharing: Strategic incentives and new implications”. *Journal of Economics & Management Strategy* 19.2, pp. 403–433.

Kim, Jin-Hyuk and Liad Wagman (2015). “Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis”. *The RAND Journal of EConomics* 46.1, pp. 1–22.

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.

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

Motta, Massimo and Martin Peitz (2020). “Big Tech Mergers”. CEPR Discussion Paper No. DP14353.

Prat, Andrea and Tommaso M Valletti (2019). “Attention oligopoly”. *Available at SSRN 3197930*.

Prüfer, Jens and Christoph Schottmüller (Forthcoming). “Competing with big data”. *Jornal of Industrial Economics*.

Rey, Patrick and Jean Tirole (2007). “A primer on foreclosure”. *Handbook of industrial organization* 3, pp. 2145–2220.

Riordan, Michael H (2005). “Competitive effects of vertical integration”.

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 1: 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 2: When $C'(u) = 0$, we have $\frac{\partial D_i}{\partial u_i}/D_i = -\frac{\partial r}{\partial u_i}/r$ by (2). We thus have

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

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

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

Proof of Proposition 2. 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

$$\hat{u}_i(\hat{u}_j(u_i^*), \delta_i) - u_i^* = 0 \tag{8}$$

(recalling that $\hat{u}_i$ is i's best response function). The left-hand side of (8) 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 (8) 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 (Application 1)

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,

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

which is true if P is log-concave. ■

Proof of Application 1. 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 $r(u_i, \delta_i) = \hat{p}(u_i, \delta_i)Q\left(\hat{p}(u_i, \delta_i), \delta_i\right)$. 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

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

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

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

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

$$\begin{aligned}-\Big\{&\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial \delta_i}\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i}\hat{q}(u_i, \delta_i) \\ &- P(\hat{q}(u_i, \delta_i)\Big[\frac{\partial \hat{q}(u_i, \delta_i)}{\partial \delta_i}\left(\frac{\partial P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i} + \hat{q}(u_i, \delta_i)\frac{\partial^2 P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i^2}\right) \\ &+ \hat{q}(u_i, \delta_i)\frac{\partial^2 P(\hat{q}(u_i, \delta_i), \delta_i)}{\partial q_i \partial \delta_i}\Big]\Big\}.\end{aligned}$$

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. ■

B.2 Price discrimination (Application 2)

Consider a model in which a consumer has an idiosyncratic willingness to pay for each of a continuum of goods drawn independently from distribution F, implying demand $Q(p) = 1 - F(p)$. Each firm i sells its own version of every good, produced at marginal cost c, and has data that allows it to determine consumers' willingness to pay for a fraction δ_i of them (we call these goods *identified*) and thus extract as much of the surplus as it wants. For the remaining $1 - \delta_i$ *unidentified* goods, the firm only knows that consumers have demand $Q(p)$ and can do no better than setting a uniform price. Consumers one-stop shop, and the utility of choosing firm i is given by the standard consumer surplus measure.

Let $\mathcal{I}_{j,l}$ be the set of products for which j observes l's willingness to pay (j's identified products). For $z \in \mathcal{I}_{j,l}$, let $v_{z,l}$ and $p_{j,z,l}$ denote respectively the consumer's willingness to pay for product z and the price at which firm j sells it to her. The mean utility offered by firm j is then

$$u_{j,l} = \int_{z\in\mathcal{I}_{j,l}} (v_z - p_{j,z,l})dz + (1-\delta_{j,l})\int_{p_{j,l}^{NI}}^{\infty} Q(x)dx.$$

Because firms can set personalized offers $u_{j,l}$ to each consumer l, we can consider each consumer as a separate market, and we now drop the l index for notational convenience.

We decompose the utility u_j in two: $u_j = U_j^I + (1-\delta_j)u_j^{NI}$. The first term, U_j^I, is the utility offered through identified products, while the second, $(1-\delta_j)u_j^{NI}$, is the utility offered through non-identified products.

Let $r^I(U,\delta)$ be the revenue generated by the share δ of identified products if the associated utility is U. If we denote the maximal surplus generated by a product as $\overline{u}$ (i.e. $\overline{u} = \int_0^\infty Q(x)dx$), we have

$$r^I(U,\delta) = \delta(\overline{u} - c) - U. \tag{9}$$

Let $r^{NI}(u,\delta)$ be the profit generated by non-identified products if the expected surplus for each one is u. We have

$$r^{NI}(u,\delta) = (1-\delta)(p^{NI}(u) - c)q\left(p^{NI}(u)\right). \tag{10}$$

where $p^{NI}(u)$ satisfies $u = \int_{p^{NI}(u)}^{\infty} Q(x)dx$. Letting $\eta(u) \equiv -\frac{(p^{NI}(u)-c)Q'(p^{NI}(u))}{Q(p^{NI}(u))}$ denote the mark-up elasticity of demand, observe that

$$\frac{\partial r^{NI}(u,\delta)}{\partial u} = (1-\delta)\frac{\partial p^{NI}(u)}{\partial u}\left[Q\left(p^{NI}(u)\right) + \frac{\partial Q\left(p^{NI}(u)\right)}{\partial u}\left(p^{NI}(u)-c\right)\right] = (1-\delta)[\eta(u)-1],$$

where the last equality follows from

$$u = \int_{p^{NI}(u)}^{\infty} Q(x)dx \implies \frac{\partial p^{NI}(u)}{\partial u} = -\frac{1}{Q\left(p^{NI}(u)\right)}.$$

As a preliminary step, we have the following result which says, to the extent possible, the firm prefers to provide utility by lowering the price of the unidentified products:

Lemma 2. *Suppose that firm j wishes to offer utility u_j.*

- *If $u_j \le (1-\delta_j)\overline{u}$, j optimally extracts all the value from identified products: $U_j^I = 0$.*
- *If $u_j > (1-\delta_j)\overline{u}$, all non-identified products are sold at marginal cost: $u_j^{NI} = \overline{u}$.*

Proof of Lemma 2. Suppose first that $u_j \le (1-\delta_j)\overline{u}$, and that $U_j^I > 0$. Consider the following reallocation of utility provision: firm i reduces the utility offered through identified products by $dU_j^I = -\epsilon$, and increases the utility provided by each non-identified product by $du_j^{NI} = \epsilon/(1-\delta_j)$, so that the overall utility u_j remains the same. The change in profit is $\left(\frac{1}{1-\delta_j}\frac{\partial r^{NI}(u_j^{NI},\delta_j)}{\partial u} - \frac{\partial r^I(U_j^I,\delta_j)}{\partial U}\right)\epsilon = \eta(u_j^{NI})\epsilon > 0$, so that $U_j^I > 0$ cannot be optimal.

If $u_j > (1-\delta_j)\overline{u}$, having $U^I = 0$ is no longer possible: selling all non-identified products at marginal cost would not be enough to provide utility u_j. But a similar logic to that above implies that the first step is indeed to lower the price of non-identified products to marginal cost, before starting to lower the prices of identified products. ∎

Intuitively, providing utility is cheaper by lowering the price of non-identified products, because some of the extra utility comes out of deadweight loss rather than the firm's revenue. When the firm already offers as much surplus as it can from the non-identified products, it has to start lowering the price of identified products.

Since non-identified products are sold at marginal cost when $u_j \ge (1-\delta_j)\overline{u}$, we then have $r_j(u_j,\delta_j) = \delta_j(\bar{u}-u_j)$, implying $\frac{\partial^2 r_j}{\partial u_j \partial \delta_j} < 0$. However note that $\frac{\partial^2 \ln[r_j]}{\partial u_j \partial \delta_j} = 0$ so data is neither pro- nor anticompetitive.

We now focus on the case with $u_j < (1-\delta_j)\overline{u}$ where data is non-neutral. By Lemma 2, we then have

$$r(u_j,\delta_j) = \delta_j\overline{u} + r^{NI}\left(\frac{u_j}{1-\delta_j},\delta_j\right) = \delta_j\overline{u} + (1-\delta_j)p^{NI}\left(\frac{u_j}{1-\delta_j}\right)Q\left[p^{NI}\left(\frac{u_j}{1-\delta_j}\right)\right]. \quad (11)$$

Note that, if $p \mapsto pQ(p)$ is concave, we have $\frac{\partial r_j}{\partial u_j \partial \delta_j} < 0$.

Linear demand Suppose demand is given by $Q(p) = 1 - bp$, and marginal cost $c \to 0$. We have $p^{NI}(u) = \frac{1-\sqrt{2}\sqrt{b}\sqrt{u}}{b}$ and $\overline{u} = \frac{1}{2b}$. Substituting $p^{NI}(u)$ into (11) we compute

$$\frac{\partial^2 \ln r}{\partial u_j \partial \delta_j} = -\frac{\left[4\sqrt{2}bu_j + \sqrt{2}(2-\delta_j) - 8\sqrt{b}\sqrt{u_j}\sqrt{1-\delta_j}\right]\sqrt{\frac{u_j}{1-\delta_j}}\sqrt{b}}{2u\left(\sqrt{2}\sqrt{b}\sqrt{u_j}\sqrt{1-\delta_j} + \delta_j - 4bu_j\right)^2},$$

which is negative if $-[4\sqrt{2}bu_j + \sqrt{2}(2-\delta_j) - 8\sqrt{b}\sqrt{u_j}\sqrt{1-\delta_j}] < 0$. The left-hand side is concave in u and maximized at $u_j = \bar{u}(1-\delta_j) = \frac{1-\delta_j}{2b}$. Making this substitution, r is log-submodular if $-\delta_j < 0$, which is true so data is UAC.

Constant elasticity Suppose demand is given by $Q(p) = p^{-\sigma}$, where $\sigma > 1$ is the price-elasticity of demand. We have $p^{NI}(u) = (u(\sigma-1))^{\frac{1}{1-\sigma}}$ and $\overline{u} = \frac{c^{1-\sigma}}{\sigma-1}$. Using (11), we find that r_j is log-submodular (and thus data is UAC) if and only if $\sigma < 1/(1-\delta_j)$.

B.3 Targeted advertising (Application 3)

In this appendix subsection, we make explicit how targeted advertising can generate a family of demand rotations $P(n, \delta_i)$. This presentation borrows from Rutt (2012), itself based upon Ganuza and Penalva (2010).

Suppose that each advertiser is a monopolist on its product market. For each product, consumers' willingness to pay is $v \in \{0, V\}$. Suppose, for notational simplicity, that advertisers are ex ante identical. Each advertiser receives an informative signal $x \sim \mathcal{U}[0,1]$. Its updated belief that $v = V$ is denoted $\tilde{\sigma}(x, \delta_i)$, with $\frac{\partial \tilde{\sigma}(x,\delta_i)}{\partial x} \geq 0$. The signal is more informative as δ_i increases: there exists $\tilde{x} \in (0,1)$ such that $\frac{\partial \tilde{\sigma}(x,\delta_i)}{\partial \delta_i} > 0$ if and only if $x > \tilde{x}$.

The willingness to pay of an advertiser who has received signal x is thus $V\tilde{\sigma}(x, \delta_i)$. If the platform decides to sell n_i slots, the n_i advertisers with the highest signals win a slot, and the uniform price is then $P(n_i, \delta_i) = V\tilde{\sigma}(1 - n_i, \delta_i)$. As long as $n_i < 1 - \tilde{x}$, we have $\frac{\partial P(n_i,\delta_i)}{\partial \delta_i} > 0$: the willingness to pay of advertisers who receive a "good" signal ($x > \tilde{x}$) increases when signals are more accurate.

Targeted advertising and pricing Suppose firms can use two instruments: a price and a quantity of ads. The utility of a consumer is $u_i = v - p_i - \phi n_i$, while the per-consumer revenue is $p_i + n_i P(n_i, \delta_i)$. In order to compute $r(u_i, \delta_i)$, let us first solve the following problem:

$$\max_{p_i, n_i} p_i + n_i P(n_i, \delta_i) \quad \text{s.t. } v - p_i - \phi n_i = u_i$$

Substituting p_i by $v - u_i - \phi n_i$ into the objective, we find that the optimal number of ads is given by $P(n_i^*, \delta_i) + n_i^* \frac{\partial P}{\partial n_i} - \phi = 0$: firm i equalizes the marginal revenue and the advertising nuisance. Indeed, in order to maintain the utility level u_i, an additional ad must be accompanied by a price decrease of ϕ, and the latter is thus the effective marginal cost of advertising. Firm i's per-consumer revenue is then

$$r(u_i, \delta_i) = v - u_i - \phi n_i^* + n_i^* P(n_i^*, \delta_i)$$

We have $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} = 0$. By Proposition 1, we conclude that data is UPC.

Note that the previous analysis ignores the possible non-negativity constraint on prices. Indeed, if competition is very strong, firms might want to subsidize participation by setting $p_i < 0$. If we restrict p_i and n_i to be non-negative, then, whenever the constraint $p_i \geq 0$ binds, firm i generates all its revenue through advertising and the UPC/UAC condition is that given in the main text in the case without prices.

B.4 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, $a'(u_i) > 0$.[45] Per-consumer revenue is therefore $r_i(u_i, \delta_i) = \max_{n \leq a(u_i)} nP(n, \delta_i)$. As long as $\frac{\partial [nP(n,\delta_i)]}{\partial n}|_{n=a(u_i)} > 0$ the attention constraint $n \leq a(u_i)$ is binding, and payoffs are congruent: $\frac{\partial r_i}{\partial u_i} > 0$.[46] We also have $\frac{\partial^2 r_i}{\partial u_i \partial \delta_i} > 0$, so data is UPC by Proposition 1. Thus, Proposition 2 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^*.

C Dynamics and market concentration

Consider a market where two firms, A and B, 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 a market share $D_i^t = D(u_i^t, u_j^t)$ 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 $\gamma \in [0, 1]$. Thus, $\delta_i^{t+1} = \gamma \delta_i^t + D_i^t$. Given discounting

[45] 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.

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

rate β, firms' value for the problem is

$$V_i^t(\delta_i^t, \delta_j^t) = \max_{u_i^t} \left[\pi_i(u_i^t, u_j^t, \delta_i^t) + \beta V_i^{t+1}(\delta_i^{t+1}, \delta_j^{t+1}) \right],$$

subject to the law of motion for data and taking rival's equilibrium play as given. The main purpose of this appendix is to study how the leader's data advantage evolves over time. In order to obtain analytical results, we focus on the case where firms are myopic, i.e., $\beta = 0$. Then 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 5. *Suppose data does not depreciate ($\gamma = 1$) and that i is the current leader ($\Delta_i^t > 0$). Then $\Delta_i^{t+1} > \Delta_i^t$ if and only if data is UPC.*

Thus, the log-supermodularity condition from Proposition 1 can be used to characterize the evolution of data concentration when data is long-lived. Note that the increase in data concentration does not imply an increase in market concentration, as the marginal value of data may be decreasing in some examples.

If data decays over time ($\gamma < 1$) then a similar force is at work but the leader must accumulate enough new data each period to offset the depreciation of its existing data advantage in order for its advantage to be increasing. It follows that data being UPC is a necessary (but not sufficient) condition for data concentration to increase. We can illustrate this with a parameterized example. Suppose $D(u_i^t, u_j^t) = \frac{1}{2} + u_i^t - \sigma u_j^t$ and $r(u_i^t, \delta_i^t) = 1 + \delta_i^t - u_i^t + \theta \delta_i^t u_i^t$. We take $\theta < 0$ to ensure strategies are strategic complements (cf. Proposition 2). We also observe from Proposition 1 that data is UPC if and only if $\frac{\partial^2 \ln(r_i^t)}{\partial u_i^t \partial \delta_i^t} \geq 0 \iff \theta \geq -1$. 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 (the latter being less likely than the former): (i) data must not depreciate too quickly (γ large enough), so that the leader's advantage grows over time; (ii) data must be 'UPC enough' (θ large), so that a data advantage translates into a sufficiently higher utility offer; (iii) competition must be strong enough (σ large), so that a utility advantage translates into a large enough market share advantage.

D Mergers with non-exclusive data trade

In this section we relax the assumption that only one B-firm can use the data collected by A.

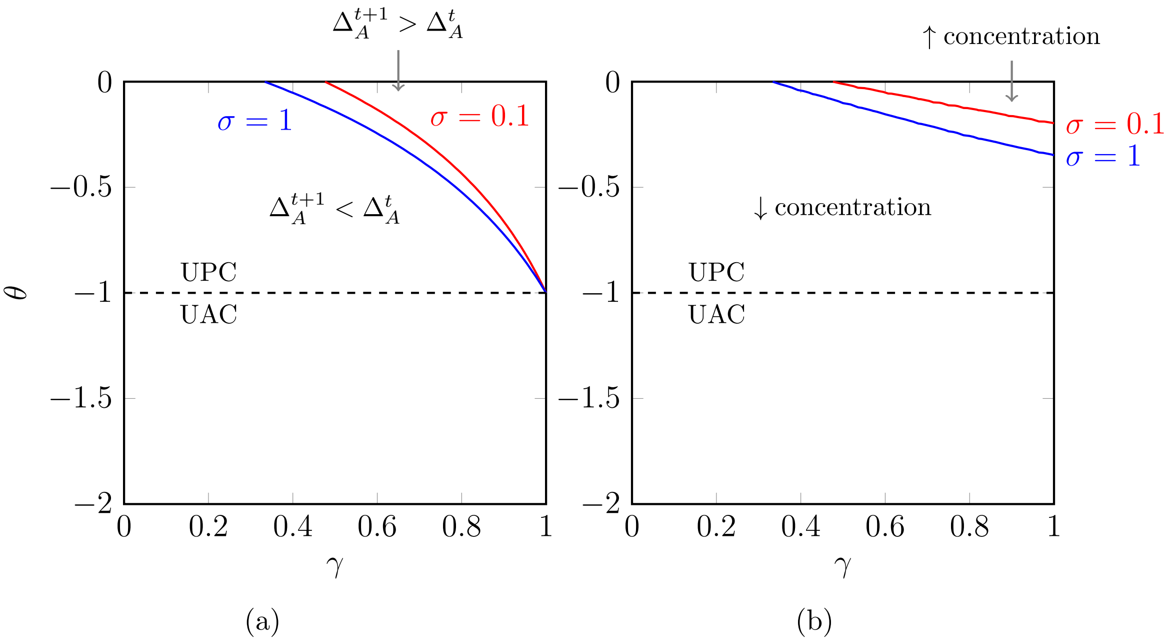


Figure 3: Data is UPC above the dashed line and UAC below it. (a) The leader's data advantage grows between t and $t+1$ above the solid curve corresponding to the relevant value of σ. (b) Market concentration grows between t and $t+1$ above the solid curve corresponding to the relevant value of σ. The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

Independent firms Suppose that firm A can choose whether to offer an exclusive deal for data or not. At $t = 1$, firm A chooses u_A and collects the data as a byproduct of its activity. At $t = 2$, it chooses between exclusive and non-exclusive sale. If the sale is exclusive, A runs an auction as in the main text. In case of a non-exclusive deal, it simultaneous offers to sell data to B_i at a price T_i. B firms simultaneously accept or reject the offers. At $t = 3$, the B firms observe the outcome of data trade and compete on the B-market.

As in the main text, we proceed to a change of variables and let firm A directly choose how much data to collect (so that u_A is a function of δ_A). We know that the outcome of stage 3 is given by the profit functions $\pi_i(\delta_i, \delta_j)$. At stage 2, we know that A's profit in case of an exclusive deal is $\pi_i(\delta_A, 0) - \pi_i(0, \delta_A)$ (see the main text). In case of a non-exclusive deal, firm A can charge $T_i = \pi_i(\delta_A, \delta_A) - \pi_i(0, \delta_A)$ to B_i, corresponding to the value of data to B_i given the (correct) expectation that B_j will also obtain the data. A's profit is then $2\left(\pi_i(\delta_A, \delta_A) - \pi_i(0, \delta_A)\right)$.

Exclusivity is thus preferred when

$$\pi_i(\delta_A, 0) + \pi_i(0, \delta_A) > 2\pi_i(\delta_A, \delta_A). \tag{12}$$

The data is sold exclusively if and only if exclusivity maximizes industry profit, which is a version of the well-known "efficiency effect" (Gilbert and Newbery, 1982). For simplicity, let us assume that either (12) holds for all values of δ_A or that it does not hold for any value of δ_A, so that the decision of exclusivity is independent of δ_A.

In the case where (12) holds, the first-order condition governing the choice of δ_A is (6), which we rewrite below:

$$\pi_A'(\delta) + \left[\frac{\partial \pi_i(\delta, 0)}{\partial \delta_i} - \frac{\partial \pi_i(0, \delta)}{\partial \delta_i}\right] = 0.$$

When (12) does not hold, the first-order condition is

$$\pi_A'(\delta) + 2\left[\frac{\partial \pi_i(\delta, \delta)}{\partial \delta_i} - \frac{\partial \pi_i(0, \delta)}{\partial \delta_i}\right] = 0. \tag{13}$$

Merger Even if A merges with B_1, it has the option to sell data to B_2. The maximal price it can charge is $\pi_2(\delta_A, \delta_A) - \pi_2(0, \delta_A)$, in which case AB_1's joint profit is

$$\pi_A(\delta_A) + \pi_1(\delta_A, \delta_A) + \pi_2(\delta_A, \delta_A) - \pi_2(0, \delta_A)$$

If instead it chooses not to sell the data to B_2, AB_1's joint profit is

$$\pi_A(\delta_A) + \pi_1(\delta_A, 0)$$

Comparing the two expressions, we find that AB_1 prefers not to sell the data if and only if

$$\pi_1(\delta_A, 0) + \pi_2(0, \delta_A) > \pi_1(\delta_A, \delta_A) + \pi_2(\delta_A, \delta_A). \tag{14}$$

If (14) holds, the first-order condition is the same as that in the main text, namely

$$\pi_A'(\delta) + \frac{\partial \pi_1(\delta, 0)}{\partial \delta_1} = 0 \tag{15}$$

If (14) does not hold so that $A - B_1$ sells the data to B_2, the first-order condition is

$$\pi_A'(\delta) + \frac{\partial \pi_1(\delta_A, \delta_A)}{\partial \delta_1} + \frac{\partial \pi_2(\delta_A, \delta_A)}{\partial \delta_1} + \frac{\partial \pi_2(\delta_A, \delta_A)}{\partial \delta_2} - \frac{\partial \pi_2(0, \delta_A)}{\partial \delta_1} = 0 \tag{16}$$

Comparison First, notice that (14) is the same condition as (12). This means that whether one or both B firms obtain the data is independent of the merger. In other words, the merger does not raise foreclosure concerns because an independent firm A would make the same trading decision.

Second, suppose that (12) does not hold (the other case is already in the main text). Comparing the first-order conditions (13) and (16), we find that the merger reduces the incentives to collect if and only if $\frac{\partial \pi_i(\delta,\delta)}{\partial \delta_j} + \frac{\partial \pi_i(0,\delta)}{\partial \delta_j} < 0$, which is true if data is UPC and false if data is UAC. We thus obtain the same results as under exclusive offers, presented in Table 2.

W Web Appendix

W.1 Model with network effects

Suppose that the mean utility that a consumer obtains from firm i depends on how many consumers also buy from i. Let us assume that the value of these network effects is $\alpha(q_i, \delta_i)$, where q_i is the number of consumers who buy from i. The stand-alone value of product i is V_i, and its price is p_i, so that

$$u_i = V_i - p_i + \alpha(q_i, \delta_i)$$

We know that $q_i = D_i(u_i, u_j)$, and can use this fact to write

$$r_i = p_i = V_i - u_i + \alpha(D_i(u_i, u_j), \delta_i)$$

In this model, data is pro-competitive if $\frac{\partial^2 \alpha}{\partial q_i \partial \delta_i} > 0$. For instance, we might expect this to be the case if the network effects arise because consumers value a larger pool of potential matches and data allows the firm to match consumers more effectively. One difference with the baseline model is that the per consumer revenue depends on u_{-i}, through its effect on the choice of quality. Although this has little impact on the unilateral analysis, it does mean that, even if we specify the model to a Hotelling duopoly, the nature of the strategic relation between u_1 and u_2 is no longer given by whether we have congruent or conflicting payoffs.

W.2 Model with consumer 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.[47] Suppose that δ_i measures how much information i collects about each consumer and introduce privacy concerns by assuming that consumers incur a disutility $\gamma(\delta)$, where γ 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 $U \equiv u - \gamma(\delta)$. We write $R(U, \delta)$ for the per-consumer revenue as a function of the *net* utility U and of the amount of data δ, i.e. $R(U, \delta) = r(U + \gamma(\delta), \delta)$

The main difference with the baseline model is that privacy concerns may make $R(U, \delta)$ decreasing in δ. For instance, in the product improvement example suppose that consumer have unit demand for the product, and that the willingness to pay (ignoring privacy concerns) is $v(\delta)$. We then have $U = v(\delta) - \gamma(\delta) - p_i$, and, with a marginal cost normalized to zero, $R(U, \delta) = v(\delta) - \gamma(\delta) - U$. Whenever $\gamma'(\delta) > v'(\delta)$, R is decreasing in δ.

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

For simplicity we assume that the fixed cost is constant, and that $u \mapsto r(u,\delta)$ is decreasing and log-concave. The firm's profit can then be written as

$$\pi(U,\delta) = R(U,\delta)D\left(U\right) - C \tag{17}$$

Let $\widehat{U}(\delta)$ be the profit-maximizing net utility if the firm collects an amount of data δ. By analogy to Definition 1, we say that data is pro-competitive if $\widehat{U}'(\delta) > 0$, and anti-competitive if $\widehat{U}'(\delta) < 0$. We then find the following result, which parallels 1 (and is obtained in the same manner):

Proposition W1. *In the model with privacy concerns, data is pro-competitive if and only if* $\frac{\partial^2 \ln(R(U,\delta))}{\partial U \partial \delta} > 0$.

The necessary and sufficient condition for data to be procompetitive is similar to the the baseline model , i.e. given by the sign of $\frac{\partial^2 \ln(R(U,\delta))}{\partial U \partial \delta}$. However, the presence of privacy concerns makes it "more likely" that data is anticompetitive in the following sense: if data is anticompetitive absent privacy concerns, it will remain so with privacy concerns, whereas data can be anticompetitive with privacy concerns but procompetitive without. To see this, note that

$$\frac{\partial^2 \ln\left(R(U,\delta)\right)}{\partial U \partial \delta} = \frac{\partial^2 \ln\left(r(u,\delta)\right)}{\partial u \partial \delta} + \gamma'(\delta)\frac{\partial^2 \ln\left(r(u,\delta)\right)}{\partial u^2}.$$

By log-concavity of $u \mapsto r(u,\delta)$, the term multiplying $\gamma'(\delta)$ is negative, meaning that larger privacy concerns make it more likely that $\frac{\partial^2 \ln(R(U,\delta))}{\partial U \partial \delta} < 0$.