# Data and Competition: a General Framework with Applications to Mergers, Market Structure, and Privacy Policy*

Alexandre de Cornière[†] and Greg Taylor[‡]

February 27, 2020

**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 use a competition-in-utilities approach to provide a general framework for studying the competitive effects of data, encompassing a wide range of markets where data has many different uses. We identify conditions for data to be unilaterally pro- or anti-competitive (UPC or UAC). The conditions are simple and often require no information about market demand. We apply our framework to study various applications of data, including training algorithms, targeting advertisements, and personalizing prices. We also show that whether data is UPC or UAC has important implications for policy issues such as data-driven mergers, market structure, and privacy policy.

**Keywords**: competition, big data, data-driven mergers, privacy.
**JEL Classification**: L1, L4, L5.

## 1 Introduction

Data has become one of the most important issues in the ongoing vivid debate about competition and regulation in the digital economy. This is illustrated by recent policy

*We are grateful to Paul Belleflamme, Vincenzo Denicolò, 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).

[†]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

reports (e.g., Crémer et al., 2019; Furman et al., 2019; Scott Morton et al., 2019), policy hearings (such as the FTC's recent Hearing on Privacy, Big Data, and Competition[1]), and newly established specialist policy teams (such as the UK CMA's Data, Technology, and Analytics unit). The idea that firms would seek to gather information about their consumers and market environment is not new, but today's situation stands out by the scale and scope of the data collected, along with its importance to many of the most successful technology firms' business models.

Firms have found many uses for the data they collect or acquire, be it targeted advertising, price-discrimination, or product improvement (e.g. better search results, more personalized product recommendations), often through the help of machine learning algorithms. While observers acknowledge the various efficiencies Big Data brings about, many concerns remain. A first concern is that data may hamper effective competition, by raising barriers to entry or by creating winner-take-all situations (see e.g. Furman et al., 2019, 1.71 to 1.79). A second, related, concern is that dominant firms may also engage in exclusionary conduct related to data, by refusing to provide access to data to other firms, by signing exclusive contracts or by employing tying and cross-usage agreements (Autorité de la Concurrence and Bundeskartellamt, 2016, pp 17-20). A third broad concern is exploitative behavior, when a firm either uses its dominant position to collect excessive amounts of data (see the recent Facebook case by the German Bundeskartellamt) or uses its data to extract surplus from consumers (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"). Finally, an increasing number of mergers in the digital sector involve data (see Argentesi et al., 2019, for recent cases), and there is still a debate as to how such data-driven mergers should be tackled by competition authorities (Grunes and Stucke, 2016).

The importance of data to the digital economy has led to a rapidly growing economics literature (see below for a discussion). Most papers in that literature focus on one kind of data use (e.g. price-discrimination, targeted advertising) and on a narrow set of issues (e.g. exclusive deals, mergers, evolution of market structure). While the correspondingly detailed modelling has allowed researchers to uncover and understand some novel economic mechanisms that apply to some specific situations, one drawback of this approach is that the connection between the various models and issues is not always clear.

In this paper we propose a framework that allows a unified approach to the various usages of data, and we derive a number of results related to the policy issues mentioned above. We consider a model where firms compete in the utility-space. This approach is flexible enough to encompass various business models, such as price competition (with

[1]See `https://www.ftc.gov/news-events/events-calendar/ftc-hearing-6-competition-consumer-protection-21st-century`, accessed 1 May 2019

uniform or personalized prices), ad-supported business models, or competition in quality. Inspired by Armstrong and Vickers (2001)'s work on price-discrimination, we model data as a revenue-shifting input: for a given utility provided, a better dataset enables a firm to generate more revenue from each consumer, a natural property across many uses of data. Our first main result consists in characterizing the environments where data is *unilaterally pro-competitive* (or unilaterally anti-competitive), in the sense that a better dataset induces a firm to offer more (or less) utility to consumers. We show that in many cases the pro or anticompetitive nature of data can be assessed without making specific assumptions about the shape of the demand function,[2] but instead depends only on the mapping between utility and revenue (Proposition 1). We apply the result to various examples inspired by standard models of data usage.

This preliminary static analysis, which only relies on the revenue-shifting property of data, serves as a building block for the rest of the paper. We then consider other properties of data to study various issues.

First, we study data-driven mergers. We consider two adjacent markets: the data generated on the (monopolized) market A can be used by the firms who compete on market B. Here, data is a 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 if the merger benefits consumers.

Next, we turn to the study of 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 later periods. We show that a necessary condition for data to lead to market dominance or to deter entry is that it is unilaterally pro-competitive. While fairly intuitive, this point — which to the best of our knowledge had not been explicitly made — indicates a tension between the static and the dynamic effects of data on market outcomes, which could constitute a guide for practitioners.

Finally, we introduce consumer privacy concerns in a model of data collection by a monopolist. Our baseline model can accommodate such a situation, with the potential tweak that collecting more data may reduce the firm's revenue for a given utility provided. We show that the firm may collect too little or too much data depending on whether data is unilaterally pro- or anti-competitive. In this context, a potential friction may be that consumers cannot observe how much data is collected or sold to third parties

[2] Apart from standard regularity assumptions.

(resulting in privacy costs). Another source of inefficiency lies in the data externalities among consumers: data about a consumer may help a firm learn something about others. We discuss various policy interventions: restrictions on the amount of data collected, increased consumer control of data collection, increased transparency. While the first two policies work well when data is unilaterally anti-competitive, they are ineffective and can even backfire when data is pro-competitive. Transparency offers more flexibility when data is unilaterally pro-competitive, and may achieve the second-best optimum.

**Contribution** In summary, our contribution is two-fold. Firstly, by casting data as a revenue-shifter into a competition-in-utility model, we provide an analysis that is not closely tied to a specific use of data, answering calls for a more general understanding of the competitive effects of data (e.g., Economist, 2017; Furman et al., 2019). In this model, we give conditions for data to be unilaterally pro- or anti-competitive, that hold irrespective of the chosen discrete-choice model specification. By applying this analysis to various "off-the-shelf" models in which data is used in a specific way, we illustrate the versatility and usefulness of the approach, which connect firms' business model to the competitive effects of data.

Secondly, we rely on this framework to generate new insights about several important policy issues related to competition in the presence of data (data-driven mergers, evolution of market structure, privacy), contributing to an ongoing policy debate in this area.

**Roadmap** After a brief discussion of the related literature, we present the basic framework in Section 2. In Section 3 we discuss various applications of the basic framework. We then turn to the issues of data-driven mergers in Section 4, of dynamic market structure in Section 5, and of privacy in Section 6.

## Related Literature

The economic literature has not yet developed a coherent general framework for the analysis of data and competition. One reason is that data takes many forms and has many different users and uses (see Acquisti et al., 2016, for a discussion of this point). 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 to price discriminate (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., forthcoming; Gu et al., 2019; Ichihashi, forthcoming) or target ads (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). In both cases, the competitive effects of data appear ambiguous—depending on both how data is used and the specific modelling assumptions made. Thus, while these papers shed light on the relationship between data and competitive outcomes in some particular situations, it is hard to distil from the literature a single overall message about data's competitive effects. In this paper we adopt a competition-in-utilities approach à la Armstrong and Vickers (2001)[3] to build a model that is agnostic about how data is used. This approach allows us to provide more general results on the economic and competitive effects of data across a range of contexts. It also helps us highlight how specific assumptions can drive important market outcomes. For example, in a model of ad targeting we show that a simple change to the way ad platforms compete can drastically alter the competitive effects of data.

Armed with this framework, we turn our attention to various policy issues related to data. One important contemporary question concerns the control of mergers involving the exchange of data. A few papers (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 discuss the effects of a data-driven merger depending on the way data is used. 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.

Another important theme in the policy debate concerns the relationship between data use or accumulation and market structure. Building on the literature on learning by doing and dynamic network effects (e.g., Cabral and Riordan, 1994; Mitchell and Skrzypacz, 2005), recent papers such as Prüfer and Schottmüller (2017) and Hagiu and Wright (2020) study long-run market dynamics when data-enabled learning helps firms improve their products. These papers take the use of data (product improvement) as given and focus on how firms' learning process can generate data-driven network effects. We also study the relationship between data and market evolution, but instead focus on the role of different business models or different uses of data in driving different dynamic outcomes.

Lastly, a literature has emerged to study issues related to privacy regulation and property rights over personal data. Examples include Hermalin and Katz (2006), Campbell et al. (2015), 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) and Ichihashi (2020). Common questions addressed in this literature include whether the market under-provides privacy or whether naive consumers are vulnerable to the exploitation of their data. Taking a different

[3] See Bliss (1988) for an earlier use of this approach.

approach, we use our framework to study the relationship between data and competition in the presence of consumer privacy concerns. This allows us to analyze how various common proposals for data protection policies affect not only consumer privacy outcomes, but also wider competition in the market.

# 2 The competitive effects of data

## 2.1 Model description

**Competition in utility** Consider a discrete choice model where $n$ firms, indexed by $i \in \{1, \ldots, n\}$, compete in utility (à la Armstrong and Vickers, 2001): each firm chooses an average utility level $u_i$, and the utility that consumer $l$ obtains from firm $i$ is $u_{il} = u_i + \epsilon_{il}$. 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). The fixed cost of choosing $u_i$ is $C(u_i)$.[4] The shocks $\epsilon$ follow a joint distribution $G$, which we assume is continuously differentiable and such that $E[\epsilon_{il}] = 0$. The corresponding demand for firm $i$ is denoted $D_i(\mathbf{u})$, where $\mathbf{u} \equiv (u_1, \ldots, u_n)$, such that $\frac{\partial D_i(\mathbf{u})}{\partial u_i} > 0$ and $\frac{\partial D_i(\mathbf{u})}{\partial u_j} < 0$ for $j \neq i$.

While this competition in utility framework restricts consumer heterogeneity, it is consistent with some standard discrete-choice models, such as the multinomial or nested logit models.

**Data** Each firm has access to a dataset containing information about (actual or potential) consumers' characteristics, market demand, etc. The data available to each firm may differ in terms of scale, scope, accuracy or recency. We assume that all these characteristics determine the "quality" of the dataset available to firm $i$, and denote such quality by $\delta_i \in \mathbb{R}^+$.

There are two ways to interpret $\delta_i$. 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 by, for example, making product recommendations based on the choices or feedback of past customers. Alternatively, the $\delta_i$ 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 $D_i(\mathbf{u})$.

[4]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 $C'(u_i) > 0$. When strategies are multi-dimensional, $C$ might depend on the way in which utility is provided. For instance, an increase in quality entails a cost, unlike a decrease in price. We return to this issue below, but for the moment assume that the strategic choice is uni-dimensional.

The per-consumer revenue of firm $i$ is a function of its mean utility $u_i$ as well as of the quality of its data $\delta_i$. We denote such function by $r_i(u_i, \delta_i)$.

The key assumption of our model is the following:

**Assumption 1.** A firm with a better dataset (i.e. a higher $\delta_i$) generates more revenue for any given utility level provided to consumers: $\frac{\partial r_i(u_i,\delta_i)}{\partial \delta_i} \geq 0$.

In other words, the quality of a dataset is measured by its potential to generate revenue. 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 $u_i$, $\delta_i$, and $r_i$. For illustrative purposes, let us briefly and informally sketch one example application (this and several others are developed more completely in Section 3). Suppose firms set prices $p_i$ for personalized products. The more data the firm has, the better can it personalize the product so that a consumer's value for the product, $v(\delta_i)$, is increasing. We might then have $u_i = v(\delta_i) - p_i$. This implies the firm's per-consumer revenue is $r(u_i, \delta_i) \equiv p_i = v(\delta_i) - u_i$.

Firms simultaneously choose their $u_i$ to maximize profit

$$\pi(\mathbf{u}, \delta_i) = r_i(u_i, \delta_i) D_i(\mathbf{u}) - C(u_i), \tag{1}$$

which we assume to be quasi-concave in $u_i$ for any $\mathbf{u}_{-i}$, $\delta_i$.

We also introduce the following definition:

**Definition 1.** Firms have *constant fixed costs* if $C'(u_i) = 0$ for all $u_i$.

In the previous example the per-user revenue $r_i$ is a decreasing function of $u_i$. While this may seem intuitive when firms compete in prices for instance, there are natural environments in which higher utilities are associated with higher per-consumer revenues for firms. de Cornière and Taylor (2019) define environments with "congruent payoffs" such that $\frac{\partial r_i}{\partial u_i} > 0$, in contrast to "conflicting payoffs" where $\frac{\partial r_i}{\partial u_i} < 0$. A useful result from that paper can be stated as follows:

**Lemma 1.** *Suppose that $\frac{\partial^2 D_i}{\partial u_i \partial u_j} = 0$. Then $u_i$ and $u_j$ are strategic complements if $\frac{\partial r_i}{\partial u_i} < 0$, and strategic substitutes if $\frac{\partial r_i}{\partial u_i} > 0$.*

The assumption that $\frac{\partial^2 D_i}{\partial u_i \partial u_j} = 0$ is of course a restriction, but it applies to the Hotelling-Salop discrete-choice models.

## 2.2 Unilateral effects of data

We begin by studying how data affects firms' incentives to offer utility, treating $\delta_i$ as an exogenous parameter. We will later endogenize $\delta_i$ by considering various ways that data is obtained as a by-product of economic activity, starting in Section 5.

Let $\widehat{u_i}(\mathbf{u}_{-i}, \delta_i)$ be firm $i$'s best-response function. We use the following definition.

**Definition 2.** We say that data is unilaterally pro-competitive (UPC) for firm $i$ for a given $\mathbf{u}_{-i}$ if $\frac{\partial \widehat{u_i}(\mathbf{u}_{-i},\delta_i)}{\partial \delta_i} > 0$.

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 *ceteris paribus*. It is incomplete in two ways: first, it does not capture the equilibrium effects of data. Second, it is a static notion. However, we will see that this unilateral property plays a key role in determining the effects of data, both on equilibrium (Propositions 2.A and 2.B) and on market dynamics (Section 5).

Firm $i$'s best-response is increasing in $\delta_i$ if and only if $\frac{\partial^2 \pi_i(\mathbf{u},\delta_i)}{\partial u_i \partial \delta_i} > 0$. 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:

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

Differentiating with respect to $\delta_i$, we find that data is unilaterally pro-competitive if

$$\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} = \frac{\partial D_i(\mathbf{u})}{\partial u_i} \frac{\partial r(u_i, \delta_i)}{\partial \delta_i} + \frac{\partial^2 r(u_i, \delta_i)}{\partial u_i \partial \delta_i} D_i(\mathbf{u}) \geq 0 \tag{3}$$

Our first result gives a sufficient and a necessary-and-sufficient condition for data to be unilaterally pro-competitive:

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

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

Data affects the incentive to provide utility in two ways. Firstly, an extra unit of data increases the marginal revenue earned from an additional consumer and therefore the incentive to attract consumers with high utility offers. This 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.

Part 1 of Proposition 1 follows immediately from the fact that only the second term of (3) has an ambiguous sign. Part 2 is found by using the first-order condition, (2), to eliminate $D$ from (3). In both cases we obtain a condition that does not depend on $D$ so that precise knowledge of the shape of demand is often not necessary in order to assess the competitive effects of data. Instead, what is most important is the economic technology, $r$, that connects data, utility, and revenue. This technology will be driven by the particular way in which data is being used.

Proposition 1 leads quite naturally to results on equilibrium utility offers. Under monopoly ($n = 1$), the sign of $\frac{\partial u^*}{\partial \delta}$ is given directly by the sign of (3). For $n > 1$, the equilibrium $\mathbf{u}^*$ is given by the intersection of firms' best-response functions. Suppose we assume

$$\frac{\partial^2 \pi_i}{\partial u_i^2} + \sum_{j \neq i} \left| \frac{\partial^2 \pi_i}{\partial u_i \partial u_j} \right| < 0, \tag{4}$$

which guarantees the equilibrium is unique.[5] Then the following proposition describes how data affects the utility that is provided in equilibrium.

**Proposition 2.A.** *Suppose firms are symmetric ($\delta_i \equiv \delta$) and* (4) *holds. In the unique (symmetric) equilibrium, utility offers are increasing in $\delta$ if and only if data is unilaterally pro-competitive.*

**Proposition 2.B.** *Suppose $n = 2$ and* (4) *holds. In the unique equilibrium, if data is unilaterally pro-competitive (anti-competitive) then*

1. *$u_i^* \geq u_j^*$ ($u_i^* \leq u_j^*$) when $\delta_i \geq \delta_j$;*
2. *an increase in $\delta_i$ causes $u_i^*$ to increase (decrease);*
3. *an increase in $\delta_i$ causes $u_j^*$ to increase (decrease) if utilities are strategic complements, and decrease (increase) if they are strategic substitutes.*

Under symmetry, an increase in $\delta$ causes the equilibrium point to shift along the 45° line in the same direction as firms' best responses (up if data is pro-competitive and down otherwise)—see Figure 1. When only one firm's $\delta_i$ increases, that $i$'s utility offer moves in the same direction as its best response shift (up if data is pro-competitive and down otherwise). The equilibrium point then moves *along* the rival's best response function so that the two firms' offers move in the same direction if the situation is one of strategic complements, and in opposite directions when utilities are strategic substitutes. Figure 2 illustrates.

# 3 Applications to product personalization, ad targeting, and price discrimination

The analysis has so far been conducted in terms of an abstract parameter $\delta_i$, which could be interpreted as any factor that shifts $r$. One thing that makes the application to data

[5] Formally, this ensures each firm's best response function is a contraction (see Vives, 2001).

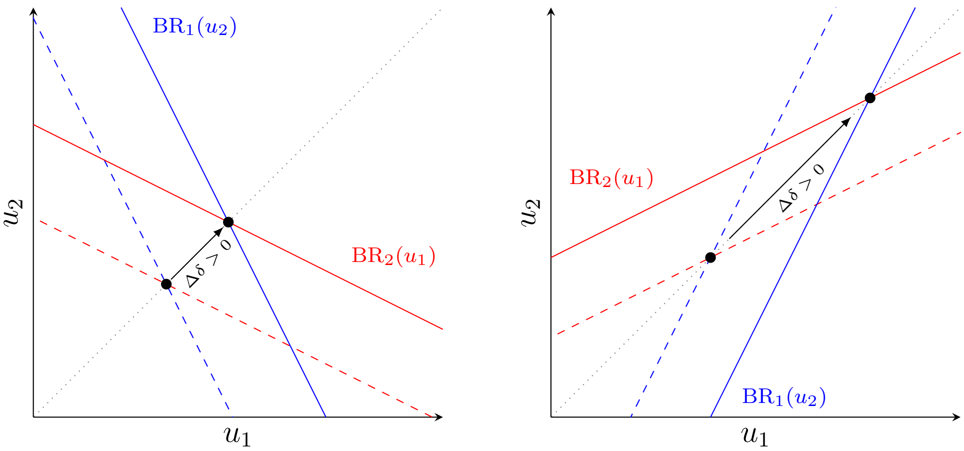


Figure 1: An increase in $\delta_1 \equiv \delta_2 \equiv \delta$ cause equilibrium utility offers to increase when data is pro-competitive. The left panel shows the case of strategic substitutes, the right strategic complements.

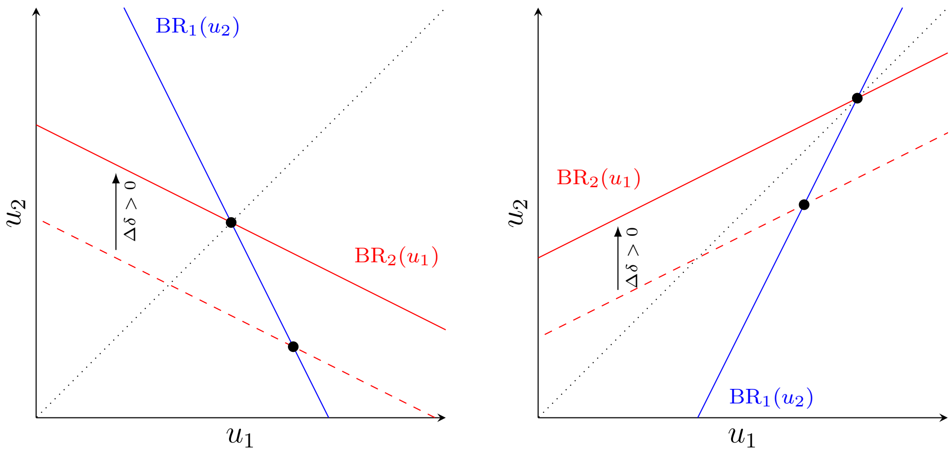


Figure 2: An increase in $\delta_2$ causes the equilibrium $u_2^*$ to increase when data is pro-competitive. The effect on $u_1^*$ depends on whether the game has strategic substitutes (left panel), or complements (right panel).

particularly interesting is that relatively standard models of data use naturally generate both pro- and anti-competitive effects. Here we apply the framework to some of the main ways data is used (product improvement or personalization, targeted advertising, and price discrimination) and show how the model can be used to determine the competitive effects of data in each case.

## 3.1 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.

We already discussed a reduced-form model of product improvement in Section 2.1. To recap: suppose a firm charges $p_i$ for a product that consumers value at $v(\delta_i)$, which is increasing.[6] We then have $u_i = v(\delta_i) - p_i$ and hence $r(u_i, \delta_i) \equiv p_i = v(\delta_i) - u_i$. Because $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} = 0$, Proposition 1 tells us that data is unilaterally pro-competitive in this case. Intuitively, if data is used to improve the product then the firm can charge a higher price while providing a given level of utility. This high price increases the firm's marginal incentive to attract consumers, inducing it to increase $u_i$.

We can extend this result using the more general framework of demand-shifting (Cowan, 2004).[7] In this broader class of models, each consumer who picks firm $i$ buys $q(p_i, \delta_i)$ units of its product, $q$ being non-decreasing in $\delta_i$ (equivalently, we could formulate the model starting from the inverse demand function $P(q_i, \delta_i)$). The utility that a consumer obtains from choosing $i$ is the standard consumer surplus: $u_i = \int_{p_i}^{\infty} q(x, \delta_i)dx$. Inverting this equation we obtain the price (and hence $r(u_i, \delta_i)$) associated with a given $u_i$. In Appendix B.1 we show that data is unilaterally pro-competitive when it is used to improve products in such a way as to induce an additive or multiplicative shift in the per-consumer demand. More precisely:

**Result 1.** *Suppose $\phi$ is a non-increasing function satisfying $\phi'(\cdot) + x\phi''(\cdot) \leq 0$. Then data is unilaterally pro-competitive if any of the following conditions hold: (i) $q(p, \delta) = \delta + \phi(p)$, (ii) $q(p, \delta) = \delta\phi(p)$, (iii) $P(q, \delta) = \delta + \phi(q)$, or (iv) $P(q, \delta) = \delta\phi(q)$.*

Moreover, in this class of applications we have $\frac{\partial r_i}{\partial u_i} < 0$, meaning that, by Lemma 1, utilities are strategic complements whenever $\frac{\partial^2 D_i}{\partial u_i \partial u_j} = 0$.

## 3.2 Targeted advertising

Another major use of data is to facilitate the targeting of advertising. Suppose that the firms are media platforms that face an inverse demand for advertising slots $P(n_i, \delta_i)$, decreasing in the number of ad slots, $n_i$. Having more data allows firms to better target ads, so $P$ is increasing in $\delta_i$. For example, according to an industry report, the price of advertising to a user of the Safari web browser has fallen 60% since it started blocking access to users' data.[8] Suppose that there is a one-to-one mapping between the number of ads shown by firm $i$ and the utility $u_i$, so that the number of ads corresponding to

[6] This reduced-form can be given a microfoundation: suppose each firm is a multi-product retailer of experience goods (e.g., movies or books). Each consumer has an ideal product, $\theta$, and experiences a mismatch from product $x$ such that she values $x$ at $V - (x - \theta)^2$. Each firm obtains a signal, $s_i$ that is distributed according to $\mathcal{N}(\theta, \frac{1}{\delta_i})$. After observing $s_i$, the best that firm $i$ can do is to recommend product $s$. The value when choosing firm $i$ is then $v(\delta_i) = V - E[(s - \theta)^2] = V - \frac{1}{\delta_i^2}$.

[7] Cowan (2004) only considers symmetric demand shifters, and does not look at the equilibrium utility provision.

[8] See `https://www.theinformation.com/articles/apples-ad-targeting-crackdown-shakes-up-ad-market`, accessed 10 December 2019.

$u_i$ is $n(u_i)$. Then the firm's revenue is $r(u_i, \delta_i) = n(u_i)P(n(u_i), \delta_i)$. We consider two formulations of this model corresponding to different assumptions about $n(u_i)$.

Firstly, suppose that consumers dislike seeing ads—a common assumption in the literature (e.g., Anderson and Coate, 2005). Then $n'(u_i) < 0$ in this case. Assuming that fixed costs are constant and using Proposition 1 (2) yields:

**Result 2.** *In the targeted advertising application with $n'(u_i) < 0$, data is unilaterally anti-competitive if and only if $\frac{\partial^2 \ln[P(n_i,\delta_i)]}{\partial n_i \partial \delta_i} > 0$. This is true, in particular, if $\frac{\partial^2 P(n_i,\delta_i)}{\partial n_i \partial \delta_i}$ is non-negative.*

As in the preceding application we have $\frac{\partial r_i}{\partial u_i} < 0$, at least in the relevant region where the number of ads shown by each firm is below the monopoly level (i.e. maximizing $n_i P(n_i, \delta_i)$. We therefore have strategic complements whenever $\frac{\partial^2 D_i}{\partial u_i \partial u_j} = 0$.

Secondly, in the so-called "attention economy" the key advertising bottleneck is often not the number of ads that can be shown to a consumer, but the consumer's willingness to pay attention to them. Suppose that $u_i$ measures the quality of a firm's content (chosen at cost $C(u_i)$). Consumers spend more time (or attention) on a platform with better content and the platform can show one ad per unit of time. Thus, $n'(u) > 0$. Using Proposition 1 (1) yields:

**Result 3.** *In the targeted advertising application with with $n'(u_i) > 0$, data is unilaterally pro-competitive if $\frac{\partial P(n_i,\delta_i)}{\partial \delta_i} + n_i \frac{\partial^2 P(n_i,\delta_i)}{\partial n_i \partial \delta_i} \geq 0$. This is true, in particular, if $\frac{\partial^2 P(n_i,\delta_i)}{\partial n_i \partial \delta_i}$ is non-negative.*

Notice that because $n'(u_i) > 0$, we also have $\frac{\partial r_i}{\partial u_i} > 0$, meaning that utilities are strategic substitutes when $\frac{\partial^2 D_i}{\partial u_i \partial u_j} = 0$.

In Appendix B.2 we provide a microfoundation based on using noisy signals about consumers' preferences to target ads. We derive the implied $P(n_i, \delta_i)$ and show that $\frac{\partial^2 P(n_i,\delta_i)}{\partial n_i \partial \delta_i} = 0$, meaning data is unilaterally anti-competitive in the model with ad nuisance but pro-competitive in the model of competition for attention. Thus, the same inverse demand for advertising can imply quite different effects of data depending on the exact mode of competition. Our point in this paper is not to argue that data is UPC or UAC when used to target ads, but simply to stress the modelling assumptions that drive such a result. When ads create nuisance, increasing $u$ means showing fewer ads. This is more costly if $\delta_i$ is large (because the marginal ad is more valuable). Conversely, when firms compete for attention a higher $u$ means consumers spend more time on the platform. This is more beneficial the larger is $\delta_i$ (because each unit of attention can be more precisely matched with an ad).

## 3.3 Price-discrimination

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

Consider a model in which a consumer has a value for each of a continuum of goods drawn independently from some distribution $F$. Each firm $i$ sells its own version of every good and has data that allows it to determine consumers' willingness to pay for a fraction $\delta_i$ 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 $Q(p) = 1 - F(p)$ 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.

To develop some intuition, first consider the polar cases with $\delta_i \in \{0,1\}$. If $\delta_i = 0$ then, given a target $u_i$, the firm cannot do better than a uniform price for all goods. To increase consumer surplus from area $a$ in Figure 3 to areas $a+b+d$ would require the firm to cut price from $p^*$ to $p^{*\prime}$, causing a change in revenue of $e-b$. If, on the other hand, $\delta = 1$ then the firm is able to perfectly price discriminate and the sum of consumer and firm surplus must equal the total area under the demand curve. Thus, increasing utility from $a$ to $a+b+d$ would result in a change in revenue of $-b-d$. The opportunity cost of providing utility is lower when the firm has no data. This suggests that data might be unilaterally anti-competitive.

We develop this reasoning more formally in Appendix B.3. Denote the maximal social surplus generated by a product as $\overline{u}$ (i.e. $\overline{u} = \int_0^\infty q(x)dx$). We show that a firm wishing to offer utility $u_i$ optimally provides as much of that utility as possible by lowering the uniform price of the non-identified products rather than extracting less surplus from identified products (if $u_i \leq (1-\delta_i)\overline{u}$ then it can provide all of its utility in this way). This implies an optimal set of prices (and thus an $r$) associated with any target utility level. We prove the following result:

**Result 4.** *Consider the equilibrium of the price-discrimination game outlined above.*

1. *If* $u_j^* < (1-\delta_j)\overline{u}$, *we have* $\frac{\partial^2 r(u_j^*,\delta_j)}{\partial u_j \partial \delta_j} < 0$.

2. *If* $u_j^* > (1-\delta_j)\overline{u}$, $\frac{\partial r(u_j^*,\delta_j)}{\partial \delta_j} = 0$*: more data does not affect firm j's equilibrium behavior.*

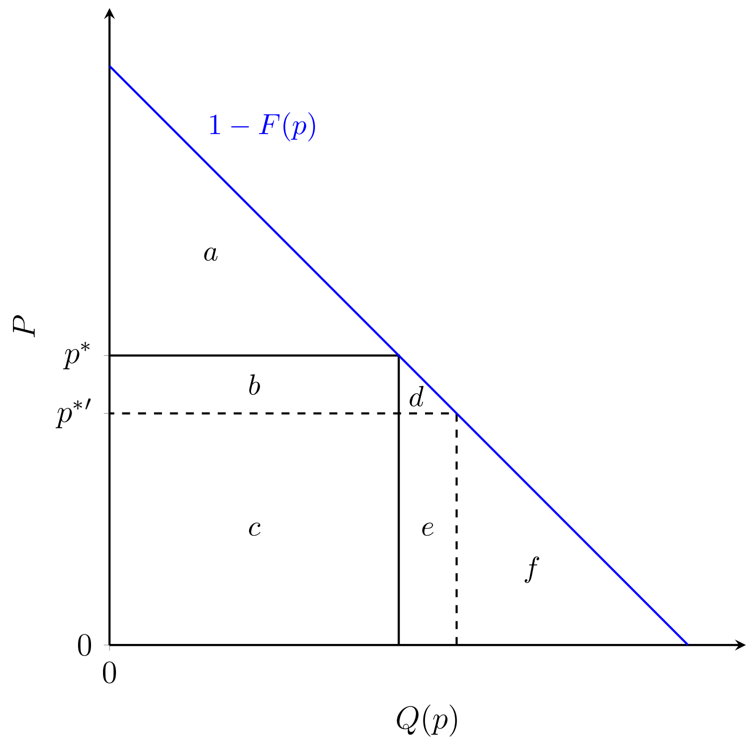


Figure 3: Price discrimination.

While part 1 of Result 4 does not imply that data is unilaterally anti-competitive in the corresponding region (the sufficient (and necessary) condition is $\frac{\partial^2 \ln[r(u_j^*,\delta_j)]}{\partial u_j \partial \delta_j} < 0$), one can check that this is indeed the case when $q(p)$ is linear or has a constant price-elasticity.

Result 4 suggests that data is more likely to be anti-competitive when the initial level of data is small. Indeed, for $\delta_j$ close to zero we necessarily have $u_j^* < (1-\delta_j)\overline{u}$ and therefore $\frac{\partial^2 r(u_j,\delta_j)}{\partial u_j \partial \delta_j} < 0$, while for $\delta_j$ close to 1 $u_j^* > (1-\delta_j)\overline{u}$, which means that data ceases to be anti-competitive.

## 3.4 Other applications

We can also apply competition-in-utility framework to other situations, including where firms' decisions are multi-dimensional (e.g. choice of a price and quality), where there are network effects, or where consumers can consume from more than one firm (multi-homing). We analyze such situations in Appendix B. The model works relatively well in environments with multidimensional decisions or network effects, even though it sometimes loses the attractive property that the pro or anti-competitive nature of data is independent of the choice of the demand function. Regarding multi-homing, we provide an example of targeted advertising with $n = 2$. Firm $j$'s revenue depends on $\delta_i$ and $u_i$ (because advertising on $j$ is less valuable if consumers can be reliably reached through $i$). We show that the effect of an increase in $\delta_i$ on $\widehat{u}_i$ is as above, while $\widehat{u}_j$ shifts in the opposite direction to $\widehat{u}_i$.

So far in the paper, we have mostly treated data as an exogenous parameter. In order to address some of the policy relevant issues surrounding data, we now embed the model in richer frameworks where the quality of data available to each firm is an equilibrium object.

# 4 Data-driven mergers

Many firms collect data as a by-product of their activity. At a basic level, customer lists can be a valuable asset as they allow a firm to send personalized offers to its previous customers. Interestingly, the customer list of a firm in a market can also be of value to firms in another market, for instance if the products are complements, paving the way for data trade opportunities. Alternatively, because the trade of data may be hindered by various frictions (see below for a discussion), acquiring a company can be a way to put a hand on its data. Several recent high-profile mergers, such as that between Microsoft and LinkedIn or Facebook and WhatsApp, have indeed been partially motivated by the acquisition of data. In the Microsoft and LinkedIn case, LinkedIn's data could be used by Microsoft to customize its Customer Relationship Management (CRM) software, Dynamics 360.[9] One should note that Salesforce, the leader in the CRM market, was also reportedly interested in acquiring LinkedIn. Following the Facebook and WhatsApp merger, 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.[10]

In this section we build upon our baseline framework to study data-driven mergers. We enrich the model by incorporating several key features of the relevant cases. 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 (on both the extensive and intensive margins). 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.[11] We label the two markets $A$ and $B$, and assume that data generated on market $A$ can be used in market $B$.[12]

Such a structure shares some similarities with a vertical merger case, in the sense that a firm in a "downstream" market ($B$) obtains an input (data) from a firm in an "upstream"

[9]See `https://www.reuters.com/article/us-microsoft-linkedin-idUSKBN17Q1FW`, accessed 13 December 2019.

[10]See `http://europa.eu/rapid/press-release_IP-17-1369_en.htm`, accessed 13 December 2019.

[11]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. We discuss horizontal data-driven mergers at the end of this section.

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

market ($A$), and competition authorities have paid close attention to theories of harm related to input foreclosure by the integrated firm (Ocello and Sjödin, 2017). There are two main versions of such theories of harm: (i) after the merger, the firm with the data will stop supplying it to its rivals in the $B$ market; (ii) after the merger, the integrated firm will gain exclusive access to the data, which will harm its rivals. While the two theories rely on data being kept internally after the merger, they differ as to whether data is shared before the merger. This point underlines a first difference between a data-driven merger as we model it and a more standard vertical merger: as we argue below, in some cases, regulatory or contractual frictions make data sharing between independent firms impossible or impractical, so that the merger is the only way to share data.

The second main difference between our framework and a vertical one lies in the fact that selling data is not necessarily the primary purpose of the firm in the $A$ market, and that the $A$ and $B$ firms may face the same consumers. We therefore argue that an important aspect, which so far has been relatively neglected, is how the merger will affect the behavior of the $A$ firm (and therefore consumer surplus) in its primary market.

## 4.1 The model

**Market structure** Firm $A$ is a monopolist on market $A$, and offers a mean utility $u_A$, leading to a demand $D_A(u_A)$ and a per-consumer revenue $r_A(u_A)$. Let $\pi_A(u_A) \equiv r_A(u_A)D_A(u_A) - C_A(u_A)$. Serving consumers on its primary market allows firm $A$ to collect a quantity of data $\delta_A \equiv \delta(u_A)$, with $\delta'(u_A) > 0$. Thereafter we operate a change of variables and say that firm $A$ directly chooses a quantity of data $\delta_A$, corresponding to a utility level $u_A(\delta_A)$, with $u'_A(\delta_A) > 0$. Firm $A$'s profit on its primary market is $\pi_A(\delta_A) \equiv \pi_A(u_A(\delta_A))$, which we assume is quasi-concave and maximized for $\widehat{\delta}$ such that $\pi'_A(\widehat{\delta}) = 0$.

The data can also be used on a secondary market $B$, where two firms ($B_1$ and $B_2$) compete. Competition on the $B$ market takes the form 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 $\delta_i$ to firm $B_i$, the latter's per-consumer revenue is $r(u_i, \delta_i)$.[13] We will mostly use the reduced-form profit expressions $\pi_i(\delta_i, \delta_j) \equiv r\left(u_i^*(\delta_i, \delta_j), \delta_i\right) D_i\left(u_i^*(\delta_i, \delta_j), u_j^*(\delta_j, \delta_i)\right) - C\left(u_i^*(\delta_i, \delta_j)\right)$, where $u_i^*$ denotes the utility level provided in the subgame where the data levels are $\delta_i$ and $\delta_j$.

**Data trade** We will consider two scenarios, depending on whether data trade between two independent firms can happen. As we will show, this is actually a critical determinant of whether the merger is likely to benefit consumers.

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

Data is a non-rival but excludable good: when data trade is possible, firm $A$ can choose to sell any vector $(\delta_1, \delta_2) \in [0, \delta_A] \times [0, \delta_A]$,[14] in exchange for payments $T_1$ and $T_2$. The trade mechanism consists in simultaneous take-it-or-leave-it public offers $(\delta_1, T_1)$ and $(\delta_2, T_2)$ made by firm $A$,[15] followed by simultaneous public acceptance decisions by the $B$-firms.

**Extra assumptions and notations** On the $B$ market, we assume that the $u_i$'s are strategic complements,[16] and that a firm's profit is increasing in the amount of data it has: $\frac{\partial \pi_i(\delta_i, \delta_j)}{\partial \delta_i} > 0$. These assumptions entail a loss of generality. In particular, the second one rules out situations where an increase in $\delta_i$ would lead firm $B_j$ to compete so much more fiercely that $B_i$ would prefer to commit not to use the data.

**Timing** The game proceeds as follows: At $t = 1$, firm $A$ chooses $\delta_A$. At $t = 2$ data trade takes place when possible. At $t = 3$ the firms in market $B$ observe $\delta_1$ and $\delta_2$ and choose their utility offers.

## 4.2 Merger when data trade is not possible

Several factors may make data trade between independent firms impractical. For instance, privacy regulations may prevent firms from sharing personal data with third parties. By contrast, a merger may facilitate information sharing within two divisions of a firm, as the following quote from the UK Competition and Markets Authority suggests: "The GDPR makes gaining and managing consent within a 'walled garden' to deliver a particular purpose, either within an undertaking, or group of undertakings in common control, an easier exercise than sharing data between undertakings to deliver the same purpose."[17]

Another possible friction has to do with moral hazard regarding data protection. Suppose that company $A$ licenses its customer data to another firm, $B$. If $B$ does not undertake the appropriate level of investment in cyber-security, or if it uses the data in a fraudulent way itself, consumers may blame company $A$ in case of a breach, which would deter $A$ from licensing the data.[18] With a merger, $B$ would internalize the value of $A$'s reputation and invest accordingly, making data sharing possible.

[14] Whether the data is sold, temporarily licensed, or whether firm $A$ merely allows $B$ firms to send queries to the database without providing the data itself is of no consequence in the model.

[15] The results would hold with Nash bargaining provided $A$ has enough market power over the data.

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

[17] `https://assets.publishing.service.gov.uk/media/5dfa0580ed915d0933009761/Interim_report.pdf`, accessed 12 February 2020. Note that in some jurisdictions mergers may not always allow firms to transfer data internally. Recent calls for "data Chinese walls" within companies would also complicate matters.

[18] The Facebook - Cambridge Analytica scandal is a good illustration of the drawbacks of licensing data to independent third parties. See `https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram`, accessed 13 December 2019.

In our analysis, we thus consider the two situations: data trade without a merger can either be possible or not.

Suppose therefore that data trade between $A$ and the $B$ firms is impossible absent the merger. 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 $A - B_1$ 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 $\delta_A^M > \widehat{\delta}$.

**Comparison** The fact that $\delta_A^M > \delta_A^I$ means that $u_A^M > u_A^I$, i.e. that consumer surplus on market $A$ increases after the merger. In market $B$, the merger results in firm 1 having access to an additional $\delta_A^M$ 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 unilaterally pro-competitive, the merger increases consumer surplus on both markets.*

2. *If data is unilaterally anti-competitive, 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 induce consumers to share more data by increasing $u_A$. If data is pro-competitive, the use of data also induces $B$-firms to increase their utility offer (Proposition 2.B), resulting in an unambiguous gain for all consumers. Such circumstances therefore favour a more lenient merger policy. If data is anti-competitive, 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.

Of course, in the first case the merger will reduce firm $B_2$'s profit and could potentially lead to its exit. While undesirable in itself, this possible "efficiency offence" should probably not be used as an argument against the merger.

### 4.3 Merger when data trade is possible

We now turn to the case where data can be traded even without the merger. A first point to look at is whether firm $A$ finds it more profitable to offer an exclusive deal to one of the $B$ firms or to sell data to both. In the former case, its revenue from the sale of data is $\pi_i(\delta,0)-\pi_i(0,\delta)$, while in the latter it is $2\left(\pi_i(\delta,\delta)-\pi_i(0,\delta)\right)$. Exclusivity is therefore preferred when

$$\pi_i(\delta,0)+\pi_i(0,\delta)>2\pi_i(\delta,\delta), \tag{5}$$

which is a version of the well-know "efficiency effect" (Gilbert and Newbery, 1982). For the sake of brevity we only present the results corresponding to this case (but discuss the other case towards the end of this Section).

The first-order condition for firm $A$ is

$$\pi_A'(\delta)+\frac{\partial\pi_i(\delta,0)}{\partial\delta_i}-\frac{\partial\pi_j(0,\delta)}{\partial\delta_i}=0 \tag{6}$$

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 profit of the firm which does not obtain the data (firms' outside option). The sign of this effect depends on whether data is unilaterally pro- or anticompetitive. Indeed, if data is pro-competitive, a higher $\delta_i$ implies a higher $u_i$, which is bad for $B_j$'s profit: $\frac{\partial\pi_j(0,\delta)}{\partial\delta_i}<0$. The reverse holds when data is anti-competitive.

If $A$ and $B_1$ merge, $A$ still has the option to sell the data to $B_2$. However, such a strategy is never profitable if exclusivity is preferred when $A$ is independent. Indeed, the price at which $A-B_1$ would sell to $B_2$ is $\pi_2(\delta,\delta)-\pi_2(0,\delta)$, but selling implies a loss in profit for $B_1$ of $\pi_1(\delta,0)-\pi_1(\delta,\delta)$. Checking when the loss exceeds the gain yields exactly (5). Therefore, the profit of the integrated firm is $\pi_A(\delta_A)+\pi_1(\delta_A,0)$, and its first-order condition is

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

After the merger, firm $A-B_1$fully internalises $B_1$'s profit and no longer needs to manipulate its outside option, so the last term in (6) disappears.[19]

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

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

1. *If data is unilaterally pro-competitive, the merger leads to less data collection, reducing consumer surplus on both markets $A$ and $B$.*

[19] After (but not before) the merger, $\delta$ is chosen to maximize $A-B_1$'s joint profit. The merger is therefore profitable.

*2. If data is unilaterally anti-competitive, the merger leads to more data collection, increasing surplus on market $A$ but reducing it on market $B$.*

### 4.4 Non-exclusive deals

We have presented results for the case where (5) holds, i.e. where $A$, before and after the merger, finds it optimal to not sell data to $B_2$. But we obtain similar results when $A$'s profit is maximized by selling data to the two $B$ firms before the merger. In that case the merger does not affect which firms get the data, because $A - B_1$ still finds it profitable to sell the data to $B_2$. But the merger still changes the incentives to collect data through its effect on the $A - B_1$ negotiation: before the merger, $A$ wants to reduce the profit that $B_1$ would make if it did not buy the data, whereas such a force disappears after the merger.

We emphasize that, whether (5) holds or not, the number of $B$-firms supplied with data is not affected by the merger. Thus, harms experienced by consumers are not due to a foreclosure effect. Instead, the mechanism is that the merger changes $A$'s incentive to collect data, resulting in different equilibrium utility offers in both market $A$ and (via the effect of data on competition) in market $B$.

### 4.5 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 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 and only if data trade is impossible prior to the merger. If data is 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.

**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). In the model, the key condition is whether data trade is possible absent the merger.

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. There are two dimensions to this statement. The first, less controversial one is that a better product should attract more consumers and therefore allow the collection of more data. Second, collecting data should not be perceived by consumers as a major privacy violation (see Section 6 for a discussion of privacy concerns in a model with endogenous data collection).

In the absence of 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 frictions, our analysis suggests that the merger is more likely to benefit consumers.

**Input foreclosure** By studying a model with public contracts, we have shut down input foreclosure concerns, whereby firm $A$ would stop supplying data to $B_2$ after the merger. In practice, these concerns have been and should continue to be given some attention. Our modelling choice reflects our desire to emphasize the novelty of our approach. Moving to a secret contracting game à la Hart and Tirole (1990) would add a foreclosure dimension to our model, without fundamentally affecting the other insights.

Notice also that in this setup, the one-monopoly-profit theory does not hold: the merger is strictly profitable even with public offers. This is due to our assumption that the choice of firm $A$ takes place before the negotiation stage. Inverting the timing would allow firms to replicate the merger with a contract.

## 4.6 Horizontal data-driven mergers

Some horizontal mergers can also result in competitors merging their databases, thereby gaining further insights about consumers for instance. Our general model can be readily applied to such a situation. Suppose that firms 1 and 2 merge in a market with $n$ firms. First, as is standard in horizontal mergers, unilateral "price" effects exert a downward push on the merging firms' best-response. Let $\delta_i'$ be the post-merger quality of firm $i$'s data. We presumably have $\delta_i' \geq \delta_i$ for $i = 1, 2$. When data is unilaterally anti-competitive, this further reduces the utility offered by merging firms, and the overall effect of the merger is negative. When data is unilaterally pro-competitive, however, the increase in $\delta_1$ and $\delta_2$ may offset the unilateral price effect, in the same manner that reductions in marginal costs due to synergies may make a merger desirable for consumers. The key is of course that the increase in $\delta_1$ and $\delta_2$ should be large enough, i.e. that combining the datasets of merging firms is very valuable. This will be true if data generates enough economies of scale and/or economies of scope, an empirical matter subject to much debate (See Chiou and Tucker, 2017; Schaefer et al., 2018; Claussen et al., 2019, for empirical studies in various contexts).

# 5 Data and market dynamics

While the previous section analyzed a model where data generated in one market is used in another one, in many situations data can be used in the same market in later periods. Taking this into account introduces dynamic considerations that have been a recurring feature of the policy debate around data. For instance, there have been discussions around the potential for data-driven network effects to create monopoly situations, either by inducing exit or by deterring entry (Furman et al., 2019).

In this section we present a simple model of dynamic competition involving data. Two firms compete over multiple periods, and data generated by sales at date $t$ can be used in later periods. We look at properties of the Markov-perfect equilibria when firms are impatient enough (or when the value of data is ephemeral enough). In particular, we are interested in the question of market tipping: under which conditions can the use of data by firms lead to a market structure dominated by a single firm?

Suppose $n = 2$ firms compete over many periods, $t = 1, 2, \ldots$. Firms discount the future at rate $\beta$. At the start of period $t$, firm $i$ has access to data $\delta_i^t$. If the utility offered in period $t$ is $\mathbf{u}^t = (u_i^t, u_j^t)$ then the per-period profit is

$$\pi_i(\mathbf{u}^t, \delta_i^t) = r(u_i^t, \delta_i^t) D_i(\mathbf{u}^t) - C(u_i^t).$$

We assume that (4) holds[20] and that $\pi_i$ is concave in $u_i^t$.

Firm $i$ accumulates new data from interacting with each of the $D(u_i^t, u_j^t)$ consumers it attracts in period $t$, and may also be able to carry-over some data from period to period. The quality of its dataset in period $t+1$ is therefore $\delta_i^{t+1} \equiv f(D_i(\mathbf{u}^t), \delta_i^t)$, non-decreasing in both arguments. Let $\frac{\partial \delta_i^{t+1}}{\partial \delta_i^t} \leq 1$: data at least weakly decays over time (e.g., because it becomes outdated). We focus on Markov-perfect equilibria in which firms offer utility $u_i^*(\delta_i^t, \delta_j^t)$ in each period. Thus, a firm's value function has the form

$$v_i(\mathbf{u}^t, \delta_i^t, \delta_j^t) = \pi_i(\mathbf{u}^t, \delta_i^t) + \beta V_i[\delta_i^{t+1}, \delta_j^{t+1}],$$

where $V_i$ is $i$'s continuation value given the datasets it expects firms to start the next

[20] i.e. that best-responses cross only once in the static game.

period with. We assume $\beta$ is small enough that $v_i(\mathbf{u}^t, \delta_i^t)$ is maximized where the first-order condition $\frac{\partial v_i}{\partial u_i^t} = 0$ holds.

We say that data is transient if it quickly becomes obsolete (formally, if $\frac{\partial \delta_i^{t+1}}{\partial \delta_i^t} = 0$). Firms are myopic if $\beta = 0$. We say that the market tips in favor of firm $i$ (in the long run) if $\lim_{t\to\infty} D_j(\mathbf{u}^t) = 0$.

**Proposition 5.** *Suppose either (i) data is transient, or (ii) firms are myopic. In a Markov-perfect equilibrium:*

1. *if data is unilaterally pro-competitive, a firm with an initial advantage stays ahead forever:* $\delta_i^t \geq \delta_j^t \implies \delta_i^{t+1} \geq \delta_j^{t+1}$.
2. *if data is unilaterally anti-competitive, any initial advantage shrinks over time:* $\delta_j^t \leq \delta_i^t \implies \delta_i^{t+1} - \delta_j^{t+1} \leq \delta_i^t - \delta_j^t$.

If data is unilaterally pro-competitive then a firm with more data will choose to offer higher utility and thus accumulate more new data than its rival each period, leading to an entrenched market-leader. This is a necessary, but not sufficient condition for data to cause market tipping. Whether data actually causes the market to tip also depends on if the extra utility provided by the market leader causes its advantage to grow from period to period. Suppose, for example, we take take our simplest product improvement application ($r(u_i^t, \delta_i^t) = v\delta_i^t - u_i^t$) with $n = 2$ and let demand be à la Hotelling: $D_i(\mathbf{u^t}) = \frac{1}{2} + \frac{u_i^t - u_j^t}{2\tau}$. If data is transient, such that $\delta_i^{t+1} = D_i(\mathbf{u}^t)$, then the system has a single state variable, $\delta_1^t$.[21] We can compute the equilibrium utility offers as $u_1^*(\delta_1^t) = \frac{1}{3}(v + v\delta_1^t - 3\tau)$, $u_2^*(\delta_1^t) = \frac{1}{3}(2v - v\delta_1^t - 3\tau)$. Thus, the transition rule is

$$\delta_1^{t+1} = D[u_1^*(\delta_1^t), u_2^*(\delta_1^t)] = \frac{1}{2} + \frac{(2\delta_1^t - 1)v}{6\tau}.$$

The market tips if and only if $3\tau < v$. If $\tau$ is small then demand is quite elastic and the leader exploits its advantage to capture many more consumers, leading to tipping. Likewise, a large $v$ means that a little extra data gives a big advantage, again leading to tipping. If, on the other hand, $3\tau > v$, a firm with an initial advantage will stay ahead but see its advantage shrink over time, eventually converging to equal market shares.

When data is unilaterally anti-competitive, on the other hand, Proposition 5 is unambiguous: data cannot lead to market tipping. Intuitively, when data is anti-competitive a firm with a current advantage in data offers lower utility and therefore serves fewer consumers than its rival. This leads to the leader's data advantage decreasing over time. We therefore observe an interesting tension between static and dynamic competition concerns related to data: it is when data is unilaterally pro-competitive (a static concept) that long-run dynamics are most likely to foster competitive concerns.

[21]This is because $\delta_2^t = D_2(\mathbf{u}^{t-1}) = 1 - D_1(\mathbf{u}^{t-1}) = 1 - \delta_1^t$.

**Discussion** Among the potential concerns that have been raised surrounding the competitive effects of data, some are of an exploitative nature: exploitation of consumers' cognitive biases (Scott Morton et al., 2019), price-discrimination (Autorité de la Concurrence and Bundeskartellamt, 2016), or broader privacy violations (Bundeskartellamt, 2019)[22]. Others are of an exclusionary nature: risk of market tipping, data as barrier to entry, refusal to supply, limitations to portability (Crémer et al., 2019; Furman et al., 2019). In our framework, one way to conceptualize exploitative concerns is to model data as being unilaterally anti-competitive: more data induces firms to offer less utility to consumers. One message of the simple dynamic model presented here is that exploitative and exclusionary concerns related to data cannot simultaneously apply to any given situation. For instance, for data to lead to market tipping, it has to be unilaterally pro-competitive.

This logic could also be applied to situations where an incumbent faces potential entry. In such situations, data constitutes a barrier to entry only to the extent that it is unilaterally pro-competitive. Indeed, a firm in possession of a large quantity of data that it uses for exploitative purposes would be a *soft* competitor (Fudenberg and Tirole, 1984) and would invite entry. When data is unilaterally pro-competitive, on the other hand, the incumbent could strategically over-invest in data collection in order to deter entry. Such investment would benefit consumers in the early periods (because collecting data requires offering a large utility) while its later periods effect would be ambiguous (better product by the incumbent, but less competition).

Notice that we are not claiming that market-tipping is incompatible with data being used in exploitative ways, but merely that data *cannot be the cause of* market-tipping when it is exploitative. For instance, in situations where data is unilaterally anti-competitive, tipping might occur if the market also exhibits non data-related positive network effects.

Our results can be related to the existing literature. Hagiu and Wright (2020) show that network effects arise when data-enabled learning can be applied across the entire consumer population and can be translated into product improvements within the users' consumption lifetime, or when across-user learning is combined with learning that is consumer specific. Likewise, in Prüfer and Schottmüller (2017), a firm's cost of innovation is decreasing in its past demand. This generates an endogenous network effect through increased investment incentives. In both cases, the positive network effects mean that data is unilaterally pro-competitive and, consistent with our results, these papers show that tipping is possible. Conversely, models with switching costs bear more similarity to environments that are unilaterally anti-competitive. Indeed, such models tend to have negative feedback effects because firms with high past demand charge higher prices. For this reason, new consumers will often choose smaller entrants over larger incumbents,

[22] See Section 6 for an explicit modelling of privacy concerns. The points we make here would be valid there as well.

consistent with the lack of tipping observed in our model. See Farrell and Klemperer (2007) for a discussion of such dynamic switching cost models.

# 6 Data collection with privacy concerns

We have so far focused on the competitive implications of data. However, a significant part of the policy debate around data has revolved around consumer privacy concerns and potentially exploitative practices related to the collection of individuals' data—see, for example, Bundeskartellamt (2019). Several regulatory options have been considered or implemented (through the EU's General Data Protection Regulation for instance): restrictions on the type of data that can be collected, increased transparency regarding data collection and usage (firms' privacy policies are often quite opaque to consumers), or a transfer of control to individual consumers (through consent mechanisms) or to consumer unions.[23]

Privacy concerns enrich our baseline model through two conceptual novelties, which we introduce successively. First, they can lead the revenue function $r$ to be (locally) decreasing in $\delta_i$. In this context we study the effects of limiting data collection and of imposing transparency to otherwise opaque policies. Second, a consumer's privacy concerns arguably only apply to what data *about her* a firm has, even though data about other consumers might be used in the firm's interaction with this consumer, leading to the presence of externalities among consumers.

Because the effects that we discuss here are orthogonal to strategic interactions among firms, we assume that the firm is a monopolist.

## 6.1 A model with privacy concerns

Suppose that $\delta$ measures the share of consumer characteristics that the firm collects (e.g. gender, age, taste in movies, etc.). We introduce privacy concerns by assuming that consumers incur a disutility $\gamma(\delta)$, where $\gamma$ 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 $\delta$, 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 $\delta$, as the following examples illustrate.

**Example: product improvement** Suppose that consumer have unit demand for the product, and that the willingness to pay (ignoring privacy concerns) is $v(\delta)$. We then have

[23] Arrieta-Ibarra et al. (2018) and Posner and Weyl (2018) advocate for the notion of data as labor, which we do not consider here.

$U = v(\delta) - \gamma(\delta) - p_i$, 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 $\delta$.

**Example: targeted advertising with nuisance** If the firm shows $n$ ads (sold at price $P(n,\delta)$), suppose that consumers' net utility is $U = v - kn - \gamma(\delta)$. The per-consumer revenue is $nP(n,\delta)$, which means we can write $R(U,\delta) = \frac{v-U-\gamma(\delta)}{k} P\left(\frac{v-U-\gamma(\delta)}{k}, \delta\right)$, which can be decreasing in $\delta$ if $\gamma'(\delta)$ is large enough compared to $\frac{\partial P}{\partial \delta}$.

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

$$\pi(U,\delta) = R(U,\delta) D(U) - C \tag{8}$$

Let $\widehat{U}(\delta)$ be the profit-maximizing net utility if the firm collects an amount of data $\delta$. We say that data is pro-competitive if $\widehat{U}'(\delta) > 0$, and anticompetitive if $\widehat{U}'(\delta) < 0$.

We have the following characterization:

**Proposition 6.** *In the model with privacy concerns, data is pro-competitive if and only if* $\frac{\partial^2 \ln(R(U,\delta))}{\partial U \partial \delta} > 0$.

Proposition 6 is the analogue of Proposition 1 with privacy concerns (and constant fixed cost). 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

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

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 $\frac{\partial^2 \ln(R(U,\delta))}{\partial U \partial \delta} < 0$.

We now consider a game where the firm chooses both how much data to collect and how much utility to provide. A typical complaint of consumers related to data is the opacity of the system, meaning that it is difficult for consumers to know how much data about them firms actually collect (and what they do with it).[24] We make this our starting point, and consider several policy interventions.

[24] For instance, Scott Morton et al. (2019) argue it would be "absurdly impractical" for consumers to read the privacy policies of the many firms they interact with.

## 6.2 Restrictions on data collection, transparency

Suppose that consumers cannot observe the choice of $\delta$, but can form rational expectations $\delta^e$. This opacity does not prevent the firm from generating revenue thanks to the data, but it means that consumers' participation decision only depends on their expected privacy cost $\gamma(\delta^e)$. If we fix privacy concerns at $\gamma(\delta^e)$, we are back to the model without privacy concerns and utility $u - \gamma(\delta^e)$, in which case the firm's profit is increasing in $\delta$.

Therefore the equilibrium is such that $\delta^* = 1$, which consumers correctly anticipate. The equilibrium level of utility is given by the first-order condition $\frac{\partial \pi(U,1)}{\partial U} = 0$.

Let us study two possible regulatory approaches: restrictions on data collection, and improved transparency.

**Data cap** Suppose that a regulator restricts the amount of data that firms can collect. Formally, we model this with a cap $\bar{\delta} < 1$ that the firm cannot exceed. If we maintain the assumption of opacity, the cap is binding in equilibrium: $\delta^* = \bar{\delta}$. By definition we have the following result:

**Proposition 7.** *A cap on data collection improves consumer welfare if and only if data is anticompetitive.*

While trivial, Proposition 7 highlights that a cap is a fairly crude instrument, that works well when data is anticompetitive, but that can backfire when it is procompetitive.

**Transparency** We now consider a different approach, focused on enabling consumers to observe the firm's data policy (how much data is collected, how it is used, etc.). Suppose that the regulator can ensure that consumers observe and understand the firm's data policy (see our discussion below).

Under such a transparency policy, the firm solves $\max_{U,\delta} \pi(U, \delta)$.

Let $(U_T^*, \delta_T^*)$ be the profit-maximizing strategy. In a second-best world, where the regulator cannot choose $U$, how does the equilibrium level of data collection compare to the one that would maximize consumer surplus? The next result identifies a class of models such that transparency achieves the second-best.

**Proposition 8.** *If* $\frac{\partial^2 R(U_T^*, \delta_T^*)}{\partial U \partial \delta} = 0$*, transparency achieves the second-best consumer surplus.*

**Proof.** Starting from $(U_T^*, \delta_T^*)$, suppose that we force the firm to change the amount of data it collects by some $\epsilon$. The firm will then choose to provide a different utility level. The sign of the change in the net utility provided to consumers is given by the sign of

$$\frac{\partial^2 \pi(U_T^*, \delta_T^*)}{\partial U \partial \delta}\epsilon = \left( \frac{\partial^2 R(U_T^*, \delta_T^*)}{\partial U \partial \delta} D(U_T^*) + \frac{\partial R(U_T^*, \delta_T^*)}{\partial \delta} D'(U_T^*) \right) \epsilon = \frac{\partial R(U_T^*, \delta_T^*)}{\partial \delta} D'(U_T^*)\epsilon \tag{9}$$

There are three possible cases: (i) $\delta_T^* \in (0,1)$, in which case it satisfies $\frac{\partial R(U_T^*,\delta_T^*)}{\partial \delta} = 0$, and therefore utility does not increase. (ii) $\delta_T^* = 1$, which implies that $\frac{\partial R(U_T^*,\delta_T^*)}{\partial \delta} \geq 0$, and a decrease in $\delta$ (weakly) reduces the utility provided. (iii) $\delta_T^* = 0$, which implies that $\frac{\partial R(U_T^*,\delta_T^*)}{\partial \delta} \leq 0$, and an increase in $\delta$ (weakly) reduces the utility provided. ■

An example where $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} = 0$ is the model with product improvement and unit demand (see above): $R(U,\delta) = v(\delta) - \gamma(\delta) - U$. Notice that this example can be given another interpretation: data does not improve the quality (which we denote $s$) but is sold by the firm to a third party for a price $v(\delta) - s$.

In such environments, transparency provides the firm with the right incentives: it collects the amount of information that maximizes the surplus created by each transaction, which then gives it an incentive to generate many such transactions by providing a large net utility. Compared to a cap on data collection, transparency is "information light", in the sense that the regulator does not need to have detailed information about the environment (i.e. whether data is pro- or anticompetitive) for the policy to be effective.

Transparency can also sometimes achieve the second best if it results in a corner solution. If $\delta_T^* = 0$ but $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} < 0$ then the firm's choice to collect as little data as possible is also (second-)best for consumers. Likewise, if $\delta_T^* = 1$ and $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} > 0$ over the relevant range then consumers benefit from the firm's decision to collect as much data as possible.

Unfortunately, the alignment of incentives between consumers and the firm is not always perfect.

**Proposition 9.** *Suppose that $\delta_T^* \in (0,1)$. If $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} > 0$ (resp.$< 0$), the firm collects too little (resp. too much) data from consumers' point of view.*

**Proof.** $\delta_T^* \in (0,1)$ implies $\frac{\partial R(U_T^*,\delta_T^*)}{\partial \delta} = 0$. From (9), we then see that increasing $\delta$ by $\epsilon > 0$ would increase (resp. decrease) the incentive to provide utility if $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} > 0$ (resp. $< 0$). ■

Notice however that, compared to the default opacity regime, transparency is an improvement even when it leads to too much data collection as long as $\delta_T^* < 1$. The only case where opacity is preferred to transparency is when the latter leads to a large drop in data collection even though $\frac{\partial^2 R(U_T^*,\delta_T^*)}{\partial U \partial \delta} > 0$.

**Discussion** The model highlights a few messages that can be helpful in deciding which of these two approaches (data cap or transparency) to prioritize. First, imposing limitations on what data firms can collect (or use) can only be desirable from an efficiency standpoint in the presence of a strong prior belief that data is unilaterally anti-competitive. The important task would then be to identify which type of data is particularly likely to be used against consumers, which may be difficult.

Increased transparency, on the other hand, does not place such a burden on regulators. Indeed the logic is to allow firms to effectively communicate the extent of their data collection and use, allowing consumers to make a more informed choice. It provides more flexibility than a cap: if data is UAC, transparency will induce the firm to collect less data than under the benchmark, but not if data is UPC. Of course, this approach also faces its own challenges. The main one is to make sure that communication is effective. Consumers cannot be relied upon to read hundreds of pages of privacy policies full of legalese. Requiring consumers to give their consent to cookie collection for each of their visit to a website, as is currently the case in Europe, also seems to impose too much of a burden on them. Our favored approach is one where firms' policies would be graded on a standardized scale (see e.g. Ramadorai et al., 2019, for a promising approach, documenting in particular the heterogeneity of firms' privacy policies), and where consumers could specify an ex ante threshold on their browser. Websites whose policy clear the threshold could be accessed without further action, but consumers would need to give their ex post approval for websites with a less protective policy.

## 6.3 Data externalities and control

Let us now explicitly take into account the idea that, in order to generate "insights" about consumer $l$, a firm needs data about this consumer but also data about other consumers, for instance to predict her future behavior based on how consumers with similar characteristics have behaved. Let $\Delta_l \equiv g(\delta_l, \delta_{-l})$ be the quality of the "insights" about consumer $l$ if $\delta_l$ is the quality of the data it has about the consumer and $\delta_{-l}$ is the quality of data about other consumers (where we assume symmetry among the other consumers). We assume that $g$ is increasing in both arguments.

For a consumer over whom the firm has insight $\Delta_l$ and to whom it offers utility $u_l$, the firm's revenue is $r(u_l, \Delta_l)$, increasing in $\Delta_l$. The privacy loss for consumer $l$ is $\gamma(\delta_l)$.[25]

We consider a regime of individual control of data, where each consumer decides how much data to share with the firm.

**Proposition 10.** *If data is pro-competitive ($\widehat{U}'(\Delta) > 0$), individual control results in too little data being collected by the firm compared to the second-best consumer-optimal solution. If data is anticompetitive, individual control results in too much data being collected.*

The result is a straightforward implication of the observation that, when data is pro-competitive (resp. anticompetitive), individual control of data results in a game of

[25]Under this specification, consumers dislike sharing their data, independently of how much data the firm has about other consumers. One could also assume that consumers dislike the firm knowing or inferring things about them, in which case the disutility would be $\gamma(\Delta_l)$. Our results would not change under this alternative specification.

contribution to a public good (resp. public bad): individual consumers do not internalize the effect of their data sharing decision on others.

Compared to a situation where the firm collects the data opaquely, however, it is clear that individual control is an improvement in the anticompetitive case but not in the procompetitive one.

**Discussion** Other recent papers in the literature have identified some of the externalities generated by disclosure of data. In particular, data about a consumer may be informative about other consumers. Choi et al. (2019) show that this externality results in excessive data collection in a setup with privacy-concerned consumers. Acemoglu et al. (2019) and Bergemann et al. (2019) explore the consequences of this externality on the market for data, and show in particular, in two different models, that it can result in a low equilibrium price for data.

Our focus is different here: we do not look at the informational linkage between users, but simply assume that data about users other than $l$ allows the firm to generate more insight about $l$. This could be because of a substitutability property: the firm can learn something about consumer $l$ simply by looking at other consumers, as in the aforementioned papers; or because of a form of complementarity: data about $l$ is more useful if the firm also has some data about other consumers, for instance because it can make recommendations based on the past behavior of consumers with similar characteristics. In this environment, we argue that the externality may be negative (as in Choi et al., 2019), resulting in excessive disclosure, but it can also be positive, the implication being that giving consumers control over their data may result in too little disclosure.

# 7 Conclusion

In a time of intense regulatory scrutiny around the digital economy, it is important that debates take place under a certain conceptual clarity. When it comes to data and its competitive implications, the multiplicity of models developed by economists might run in the way of such clarity.

In this paper we propose another, complementary approach based on a competition-in-utility framework. The flexibility induced by this approach allows us to analyze a variety of policy issues (e.g. market structure, data-driven mergers, data collection regulations) under a variety of scenarios regarding data uses (e.g. product improvement, targeted advertising, price-discrimination). The key property of data in our framework is that it allows firms to generate more revenue per-consumer for a given utility provided.

We show that in many cases the (unilaterally) pro- or anticompetitive nature of data can be inferred from the properties of the per-consumer revenue function independently of the demand function or of the competitive intensity, and illustrate the usefulness of the approach using various examples. Whether data is pro- or anticompetitive in turn has major implications, for instance regarding the evolution of market structure, an issue for which we show that data cannot be anticompetitive in a static sense and simultaneously lead to tipping or barriers to entry. When it comes to data-driven mergers, when data is pro-competitive, a key condition for a merger to be desirable is that trade of data among independent firms be severely constrained. In the domain of privacy regulation, policies restricting data collection or granting individual control to consumers are desirable when data is anticompetitive, but perform poorly when data is pro-competitive. Measures fostering (effective) transparency provide more flexibility and may achieve the second-best in setups where data can be either pro- or anticompetitive.

# 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 $\delta_i$. Intuitively, having more data raises the incentive to provide utility, as the firm generates more revenue from each extra consumer. The sign of $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i}$ is ambiguous (see below for examples). What we can say is that 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 $\frac{\partial D_i}{\partial u_i}/D_i = -\frac{\partial r_i}{\partial u_i}/r_i$ by (2). We thus have

$$\frac{\partial D_i(\mathbf{u})}{\partial u_i}\frac{\partial r(u_i,\delta_i)}{\partial \delta_i}+\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i}D_i(\mathbf{u}) > 0 \Leftrightarrow -\frac{\partial r(u_i,\delta_i))}{\partial u_i}\frac{\partial r(u_i,\delta_i)}{\partial \delta_i}+\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i}r(u_i,\delta_i)) > 0$$

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

$$\Leftrightarrow \frac{\partial^2 \ln\left(r(u_i,\delta_i)\right)}{\partial u_i \partial \delta_i} > 0$$

■

**Proof of Proposition 2.A.** Starting from the first-order condition, (2), and totally differentiating yields

$$\frac{\partial^2 \pi_i}{\partial u_i^2}du_i + \left[\frac{\partial \pi_i^2}{\partial u_i \partial \delta} + \sum_{j \neq i}\frac{\partial^2 \pi_i}{\partial u_i \partial u_j}\frac{du_j}{d\delta}\right]d\delta = 0.$$

Symmetry implies $\frac{du_j}{d\delta} = \frac{du_i}{d\delta}$. Making this substitution and rearranging, we have

$$\frac{du_i}{d\delta} = -\frac{\frac{\partial \pi_i^2}{\partial u_i \partial \delta}}{\frac{\partial^2 \pi_i}{\partial u_i^2} + \sum_{j \neq i}\frac{\partial^2 \pi_i}{\partial u_i \partial u_j}}.$$

The denominator is negative by (4) and the numerator is equal to (3), which is positive if and only if data is pro-competitive. ■

**Proof of Proposition 2.B.** We find $u_1^*$ as the solution to

$$\hat{u}_i(\hat{u}_j(u_i^*),\delta_i) - u_i^* = 0 \tag{10}$$

(recalling that $\hat{u}_i$ is $i$'s best response function). The left-hand side of (10) is decreasing in $u_i^*$ under (4). Suppose data is pro-competitive. The left hand side of (10) is increasing in $\delta_i$ so $u_i^*$ must increase with $\delta_i$ (part 2). Part 3 then follows immediately from the definition of strategic complements and substitutes. A symmetric argument holds for the anti-competitive case.

Parts 2 and 3 along with (4) imply that, starting from a symmetric equilibrium with $\delta_i = \delta_j$, we have $\left|\frac{\partial u_j^*}{\partial \delta_i}\right| < \left|\frac{\partial u_i^*}{\partial \delta_i}\right|$. Part 1 follows immediately. ■

**Proof of Lemma 1.** Totally differentiating (2) yields $\frac{d\hat{u}_i}{du_j} = -\frac{\partial^2 \pi_i}{\partial u_i \partial u_j} / \frac{\partial^2 \pi_i}{\partial u_i^2}$, which has the same sign as $\frac{\partial^2 u_i}{\partial u_i \partial u_j}$. Differentiating (2) yields

$$\frac{\partial^2 \pi_i}{\partial u_i \partial u_j} = \frac{\partial r(u_i, \delta_i)}{\partial u_i} \frac{\partial D_i(\mathbf{u})}{\partial u_j} + \underbrace{\frac{\partial^2 D_i(\mathbf{u})}{\partial u_i \partial u_j}}_{=0} r(u_i, \delta_i),$$

which has the opposite sign to $\frac{\partial r(u_i, \delta_i)}{\partial u_i}$. ■

**Proof of Proposition 5.** Firms choose $u_i^t$ to satisfy $\frac{\partial v_i(u_i^t, u_j^t, \delta_i^t, \delta_j^t)}{\partial u_i^t} = 0$. Thus, firm $i$'s best response, $\hat{u}_i^t$, is increasing in $\delta_i^t$ if

$$\frac{\partial^2 v_i(u_i^t, u_j^t, \delta_i^t)}{\partial u_i^t \partial \delta_i^t} = \frac{\partial^2 \pi_i(u_i^t, u_j^t, \delta_i^t)}{\partial u_i^t \partial \delta_i^t} + \beta \left[ \frac{\partial^2 V_i}{\partial (\delta_i^{t+1})^2} \frac{\partial \delta_i^{t+1}}{\partial \delta_i^t} \frac{\partial \delta_i^{t+1}}{\partial u_i^t} + \frac{\partial^2 V_i}{\partial \delta_i^{t+1} \partial \delta_j^{t+1}} \frac{\partial \delta_i^{t+1}}{\partial \delta_i^t} \frac{\partial \delta_j^{t+1}}{\partial u_i^t} + \frac{\partial V_i}{\partial \delta_i^{t+1}} \frac{\partial^2 \delta_i^{t+1}}{\partial u_i^t \partial \delta_i^t} \right] \tag{11}$$

is positive, and decreasing if (11) is negative. If firms are myopic then $\beta = 0$ so (11) has the same sign as $\frac{\partial^2 \pi_i(u_i^t, u_j^t, \delta_i^t)}{\partial u_i^t \partial \delta_i^t}$. If data is transient then $\frac{\partial \delta_i^{t+1}}{\partial \delta_i^t}, \frac{\partial^2 \delta_i^{t+1}}{\partial u_i^t \partial \delta_i^t} = 0$ so (11) again has the same sign as $\frac{\partial^2 \pi_i(u_i^t, u_j^t, \delta_i^t)}{\partial u_i^t \partial \delta_i^t}$. In both cases, $\hat{u}_i^t$ is increasing in $\delta_i^t$ if and only if data is pro-competitive. By a similar logic, if firms are myopic or data is transient then $\frac{\partial^2 v_i(u_i^t, u_j^t, \delta_i^t)}{\partial u_i^t \partial \delta_j^t} = 0$, meaning $\hat{u}_i^t$ is independent of $\delta_j^t$.

Summarizing, suppose that $\delta_i > \delta_j$ and data is pro-competitive (anti-competitive). Then $i$'s best response is shifted up (down) compared to $j$'s i and we can therefore use the same logic as in Proposition 2.B to show that $u_i^t > u_j^t$ ($u_i^t < u_j^t$) in equilibrium.

To prove part 1 note that if data is pro-competitive and $\delta_1 \geq \delta_2$ then both arguments of $f(D_i(\mathbf{u}^t), \delta_i^t)$ are larger in the case that $i = 1$ than when $i = 2$.

To prove part 2 note that if data is anti-competitive and $\delta_1 \geq \delta_2$ then we must have $D_1(\mathbf{u}^t) < D_2(\mathbf{u}^t)$. Combined with the assumption that $\frac{\partial \delta_i^{t+1}}{\partial \delta_i^t} \leq 1$, this implies that $\delta_i^{t+1} - \delta_i^t = f(D_i(\mathbf{u}^t), \delta_i^t) - \delta_i^t$ is smaller when $i = 1$ than when $i = 2$. ■

# B Proofs and supplementary material for Section 3 applications

## B.1 Product improvement with general demand shifting

Suppose that, if firm $i$ sets a price $p_i$, each consumer who picks firm $i$ buys $q(p_i, \delta_i)$ units of its product, $q$ being non-decreasing in $\delta_i$. The utility that a consumer obtains from choosing $i$ is then

$$u(p_i, \delta_i) = \int_{p_i}^{\infty} q(x, \delta_i) dx.$$

Inverting this equation we obtain $\hat{p}(u_i, \delta_i)$, the unit price that delivers a utility $u_i$. The associated per-consumer profit, assuming a constant marginal cost $c_i$, is

$$r(u_i, \delta_i) = (\hat{p}(u_i, \delta_i) - c_i)\, q\left(\hat{p}(u_i, \delta_i), \delta_i\right),$$

which is easily shown to be non-decreasing in $\delta_i$ for any utility level above the monopoly one.

Equivalently, we could assume that data shifts each consumer's inverse demand, so that a consumer's willingness to pay for a $q$th unit of product $i$ is $P(q, \delta_i)$. We then define $\hat{q}(u_i, \delta_i)$ as the solution to

$$u_i = \int_0^q \left(P(x, \delta_i) - P(q, \delta_i)\right) dx$$

and $r(u_i, \delta_i) = \hat{q}(u_i, \delta_i)\left(P(\hat{q}(u_i, \delta_i), \delta_i) - c_i\right)$.

Let us define the mark-up elasticity of demand:

$$\eta(u, \delta) \equiv -\frac{\frac{\partial q(\hat{p}(u,\delta),\delta)}{\partial p}(\hat{p}(u, \delta) - c)}{q(\hat{p}(u, \delta), \delta)} = -\frac{P(\hat{q}(u, \delta), \delta) - c}{\hat{q}(u, \delta)\frac{\partial P(\hat{q}(u,\delta),\delta)}{\partial q}}$$

The proof of Result 1 consists in showing that $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i} \geq 0$ if and only if $\frac{\partial^2 r(u_i,\delta_i)}{\partial u_i \partial \delta_i} \geq 0$, which is a sufficient condition for data to be pro-competitive for firm $i$ (by Proposition 1 (1)).

**Proof of Result 1.** $\hat{p}(u_i, \delta_i)$, 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 $\frac{\partial \hat{p}}{\partial \delta_i} \geq 0$ and, by the implicit function theorem, $\frac{\partial \hat{p}(u_i,\delta_i)}{\partial u_i} = -\frac{1}{q(\hat{p}(u_i,\delta_i),\delta_i)}$. Firm $i$'s per-consumer profit is $r(u_i, \delta_i) = (\hat{p}(u_i, \delta_i) - c)\, q\left(\hat{p}(u_i, \delta_i), \delta_i\right)$. Using the property that

$\frac{\partial \hat{p}}{\partial u_i} = -\frac{1}{q(\hat{p}(u_i,\delta_i),\delta_i)}$, 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 $q(p_i,\delta_i) = \delta_i + \phi(p_i)$, $\eta(u_i,\delta_i) = -\frac{(\hat{p}(u_i,\delta_i)-c_i)\phi'(\hat{p}(u_i,\delta_i))}{\delta_i+\phi(\hat{p}(u_i,\delta_i))}$. Then, $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i}$ is of the same sign as

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

which is positive if $\phi'(p) + p\phi''(p) \leq 0$.

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

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

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

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 by the assumption that $\phi'(x) + x\phi''(x) \leq 0$, 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 $\phi'(\hat{q}(u_i,\delta_i)) < 0$ in case (4), so that $\frac{\partial \eta(u_i,\delta_i)}{\partial \delta_i} > 0$ in both cases. ■

## B.2 Microfoundation for ad targeting with noisy signals

We can microfound the ad targeting technology described in Section 3.2 by supposing the product space is a circle of circumference 2 with consumers and advertisers uniformly distributed around its perimeter. If a consumer is shown an ad for a product located at distance $0 \leq \epsilon \leq 1$ from their location, the consumer's willingness to pay is $w$ with probability $1-\epsilon^2$, and 0 otherwise. For each consumer, platform $i$ receives a noisy signal about the consumer's location and discloses it to advertisers. If, without loss of generality, the consumer's true location is indexed as zero then the signal is distributed on $[-1,1]$ according to the truncated normal distribution with mean zero and standard deviation $\frac{1}{\delta_i}$. That is, the PDF of signal $s_i$ is $f(s_i,\delta_i)=\frac{\phi(s_i,\frac{1}{\delta_i})}{\Phi(1,\frac{1}{\delta_i})-\Phi(-1,\frac{1}{\delta_i})}$, where $\phi(\cdot,\frac{1}{\delta_i})$ is the normal PDF with mean zero and standard deviation $\frac{1}{\delta_i}$ and $\Phi$ is the corresponding normal CDF. An advertiser located at $x$ is willing to pay up to $w\,(1-E[\epsilon^2|s_i,x])$. If the platform decides to sell $n_i$ advertising slots, the resulting price of an individual slot is given by the marginal advertiser's willingness to pay. That advertiser is located a distance $n_i/2$ from $s_i$, implying a price of $P(n_i,\delta_i)=w\left(1-\frac{1-2f(1,\delta_i)}{\delta_i^2}-(\frac{n_i}{2})^2\right)$. Notice that $\frac{\partial^2 P(n_i,\delta_i)}{\partial n_i \partial \delta_i}=0$.

## B.3 Price discrimination analysis

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 both firms can set personalized offers $u_{j,l}$ and $u_{i,l}$ 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 profit generated by the share $\delta$ of identified products if the associated utility is $U$. If we denote the maximal social surplus generated by a product as $\overline{u}$ (i.e. $\overline{u}=\int_0^\infty q(x)dx$), we have

$$r^I(U,\delta)=\delta\overline{u}-U-\delta c, \quad \text{and} \quad \frac{\partial r^I(U,\delta)}{\partial U}=-1 \tag{12}$$

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

$$r^{NI}(u,\delta)=(1-\delta)(p^{NI}(u)-c)q\left(p^{NI}(u)\right), \quad \text{and} \quad \frac{\partial r^{NI}(u,\delta)}{\partial u}=(1-\delta)(\eta(u)-1) \tag{13}$$

where $p^{NI}(u)$ satisfies $u=\int_{p^{NI}(u)}^{\infty}q(x)dx$ and $\eta(u)\equiv-\frac{(p^{NI}(u)-c)q'(p^{NI}(u))}{q(p^{NI}(u))}$ is the "mark-up elasticity" of demand.

As a preliminary step, we have the following result:

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

- *If $u_j \leq (1-\delta_j)\overline{u}$, $j$ 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: $u_j^{NI} = \overline{u} - c$.*

**Proof of Lemma 2.** Suppose first that $u_j \leq (1-\delta_j)\overline{u}$, 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 of the demand effect that is absent on identified products (their price is always below the consumer's willingness to pay). As the desired level of utility grows, the firm has to start lowering the price of identified products.

Armed with Lemma 2, we can now prove Result 4:

**Proof of Result 4.** Part 1 : When $u_j^* < (1-\delta_j)\overline{u}$, we know by Lemma 2 that $U_j^I = 0$ so that $r(u_j^*, \delta_j) = r^{NI}(\frac{u_j^*}{1-\delta_j}, \delta_j)$. By (13), $\frac{\partial^2 r(u_j^*,\delta_j)}{\partial u_j \partial \delta_j}$ is of the same sign as $\eta'(\frac{u_j^*}{1-\delta_j})$, which is negative by the log-concavity of $q$. Part 2: When $u_j^* > (1-\delta_j)\overline{u}$, we know by Lemma 2 that $u_j^{NI} = \overline{u} - c$ and $U_j^I = u_j^* - (1-\delta_j)(\overline{u} - c)$ so that $r(u_j^*, \delta_j) = r^I(u_j^*, \delta_j^*) = \delta_j(\overline{u} - c) - U_j^I = \overline{u} - c - u_j^*$, which is independent from $\delta_j$. ■

## B.4 Multi-dimensional choice

Suppose that firms choose both price, $p_i$, and quality, $q_i$, at a fixed cost of $K(q_i)$. The utility experienced by a consumer is $U(q_i, p_i, \delta_i)$, increasing in $q_i$ and decreasing in $p_i$. Inverting $U$, we get $p_i$ as a function: $p_i = P(u_i, q_i, \delta_i)$. Profit is then

$$\pi_i(\mathbf{u}, q_i, \delta_i) = P(u_i, q_i, \delta_i)D_i(\mathbf{u}) - K(q_i).$$

We assume that $K$ is sufficiently convex to make $\pi$ concave in $q_i$. The firm will choose $q_i$ to solve

$$\frac{\partial P(u_i, q_i, \delta_i)}{\partial q_i} D_i(\mathbf{u}) = K'(q_i), \tag{14}$$

implying an optimal $q^*(\mathbf{u}, \delta_i)$.

The firm's first-order condition is now

$$\frac{\partial \pi_i(\mathbf{u}, \delta_i)}{\partial u_i} = \left[\frac{\partial P(u_i, q_i, \delta_i)}{\partial u_i} + \frac{\partial P(u_i, q_i, \delta_i)}{\partial q_i}\frac{\partial q^*(\mathbf{u}, \delta_i)}{\partial u_i}\right] D(\mathbf{u}) + P(u_i, q^*, \delta_i)\frac{\partial D(\mathbf{u})}{\partial u_i} - \frac{\partial K(q_i)}{\partial q_i}\frac{\partial q^*(\mathbf{u}, \delta_i)}{\partial u_i},$$

which, using (14), simplifies to

$$\frac{\partial \pi_i(\mathbf{u}, \delta_i)}{\partial u_i} = \frac{\partial P(u_i, q^*, \delta_i)}{\partial u_i} D(\mathbf{u}) + P(u_i, q^*, \delta_i)\frac{\partial D(\mathbf{u})}{\partial u_i}.$$

Data is pro-competitive if

$$\frac{\partial^2 \pi_i(\mathbf{u}, \delta_i)}{\partial u_i \partial \delta_i} = \frac{\partial^2 P(u_i, q^*, \delta_i)}{\partial u_i \partial \delta_i} D(\mathbf{u}) + \frac{dP(u_i, q^*, \delta_i)}{d\delta_i}\frac{\partial D(\mathbf{u})}{\partial u_i} > 0. \tag{15}$$

This is analogous to (3).

By way of example, suppose that $U(q_i, p_i, \delta_i) = \delta_i q_i - p_i$, implying $P(u_i, q_i, \delta_i) = \delta_i q_i - u_i$. Then (15) becomes

$$\underbrace{\left[q^*(\mathbf{u}, \delta_i) + \delta_i \frac{D(\mathbf{u})}{K''(q^*(\mathbf{u}, \delta_i))}\right]}_{\frac{dP(u_i, q^*, \delta_i)}{d\delta_i}} \frac{\partial D(\mathbf{u})}{\partial u_i} > 0,$$

which is satisfied so data is pro-competitive.

## B.5 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.

### B.6 Multi-homing

Suppose that two firms are advertising supported websites, and that consumers can multi-home. Following Ambrus et al. (2016), let us assume that participation decisions are independent, i.e. that consumer $l$ visits website $i$ if $u_i + \epsilon_{il} \geq 0$, where $\epsilon_{il}$ is the consumer-specific taste shock.

We use the ad nuisance model from subsection 3.2 in which $n'(u_i) < 0$, with the difference that advertisers only value the first impression on any given consumer. Let $X_i(\delta_i, \delta_j, n(u_j))$ be the probability that a randomly chosen ad at $i$ is exclusive in the sense that the ad matches the consumers tastes and $j$ does not successully show the same consumer a matching ad. $X$ is increasing in its first argument and decreasing in the second two. We assume that we can write $r_i(\mathbf{u}, \delta) = n(u_i)P_i(n(u_i), X_i)$, where $P_i$ is the price of an ad.

We know from Proposition 1 that an increase in $\delta_i$ causes $i$'s best response function to shift in a direction given by the sign of

$$\frac{\partial \ln(r_i)}{\partial u_i \partial \delta_i} = \underbrace{\frac{\partial X_i}{\partial \delta_i}}_{>0} n'(u_i) \frac{\frac{\partial^2 P_i}{\partial u_i \partial X_i} P_i - \frac{\partial P_i}{\partial u_i}\frac{\partial P_i}{\partial X_i}}{P_i^2}.$$

By a similar reasoning, the effect of $\delta_i$ on $j$'s best response function is given by the sign of

$$\frac{\partial \ln(r_j)}{\partial u_j \partial \delta_i} = \underbrace{\frac{\partial X_j}{\partial \delta_i}}_{<0} n'(u_j) \frac{\frac{\partial^2 P_j}{\partial u_j \partial X_j} P_j - \frac{\partial P_j}{\partial u_j}\frac{\partial P_j}{\partial X_j}}{P_j^2}.$$

We therefore see that $i$'s and $j$'s best responses shift in opposite directions. In particular, if $\frac{\partial^2 P_j}{\partial u_j \partial X_j} \geq 0$ then an increase in $\delta_i$ causes a unilaterally anti-competitive response from $i$ and a unilaterally pro-competitive response from $j$.

# References


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.

Ambrus, Attila, Emilio Calvano, and Markus Reisinger (2016). "Either or both competition: A" two-sided" theory of advertising with overlapping viewerships". *American Economic Journal: Microeconomics* 8.3, pp. 189–222.

Anderson, Simon P. and Stephen Coate (2005). "Market Provision of Broadcasting: A Welfare Analysis". *The Review of Economic Studies* 72.4, pp. 947–972.

Anderson, Simon, 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.

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E. Glen Weyl (2018). "Should We Treat Data as Labor? Moving beyond "Free"". *AEA Papers and Proceedings* 108, pp. 38–42.

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.

Autorité de la Concurrence and Bundeskartellamt (2016). "Competition Law and Data".

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.

Bliss, Christopher (1988). "A Theory of Retail Pricing". *Journal of Inustrial Economics* 36.4, pp. 375–391.

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

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 (forthcoming). "Competitive Personalized Pricing". *Management Science*.

Chiou, Lesley and Catherine Tucker (2017). *Search engines and data retention: Implications for privacy and antitrust*. Tech. rep. National Bureau of Economic Research.

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.

Claussen, Jörg, Christian Peukert, and Ananya Sen (2019). "The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News". *Data and Externalities in Online News (June 5, 2019)*.

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.

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*.

Economist, The (2017). "The world's most valuable resource is no longer oil, but data". May 6th 2017. URL: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

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

Farrell, Joseph and Paul Klemperer (2007). "Coordination and Lock-In: Competition with Switching Costs and Network Effects". Ed. by M. Armstrong and R. Porter. Vol. 3. Handbook of Industrial Organization. Elsevier, pp. 1967–2072. URL: http://www.sciencedirect.com/science/article/pii/S1573448X06030317.

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.

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.

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

Hart, Oliver and Jean Tirole (1990). "Vertical Integration and Market Foreclosure". *Brookings Papers on Economic Activity: Microeconomics* 1990, pp. 205–286.

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, 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.

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`.

Mitchell, Matthew F and Andrzej Skrzypacz (2005). "Network externalities and long-run market shares". *Economic Theory* 29.3, pp. 621–648. URL: `https://www-2.rotman.utoronto.ca/facbios/file/etfinal.pdf`.

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

Ocello, Eleonora and Cristina Sjödin (2017). "Microsoft/LinkedIn: Big data and conglomerate effects in tech markets". *Competition Merger Brief* 1, pp. 1–6.

Posner, Eric A. and E. Glen Weyl (2018). *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

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

Prüfer, Jens and Christoph Schottmüller (2017). "Competing with big data". *Working Paper*.

Ramadorai, Tarun, Ansgar Walther, and Antoine Uettwiller (2019). "The market for data privacy".

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`.

Schaefer, Maximilian, Geza Sapi, and Szabolcs Lorincz (2018). *The effect of big data on recommendation quality: The example of internet search*. DICE Discussion Papers 284. University of DÃ¼sseldorf, DÃ¼sseldorf Institute for Competition Economics (DICE).

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*.

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.