University of Reading

Data brokers co-opetition

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Gu, Y. ORCID: https://orcid.org/0000-0002-4594-4852, Madio, L. and Reggiani, C. (2022) Data brokers co-opetition. Oxford Economic Papers, 74 (3). pp. 820-839. ISSN 0030-7653 doi: 10.1093/oep/gpab042 Available at https://centaur.reading.ac.uk/108166/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1093/oep/gpab042

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Oxford Economic Papers, 74(3), 2022, 820–839
doi: 10.1093/oep/gpab042
Advance Access Publication Date: 12 September 2021

OXFORD

Data brokers co-opetition

Yiquan Gu[a], Leonardo Madio[b,c], and Carlo Reggiani[d]

[a]University of Liverpool Management School, Liverpool L69 7ZH, UK; E-mail: yiquan.gu@liv.ac.uk
[b]Department of Economics and Management, University of Padova, Via del Santo, 33, 35123, Padova, Italy
[c]CESifo
[d]School of Social Sciences-Economics, University of Manchester, Manchester M13 9PL, UK

Abstract

Data brokers share consumer data with rivals and, at the same time, compete with them for selling. We propose a 'co-opetition' game of data brokers and characterize their optimal strategies. When data are 'sub-additive' with the merged value net of the merging cost being lower than the sum of the values of individual datasets, data brokers are more likely to share their data and sell them jointly. When data are 'super-additive', with the merged value being greater than the sum of the individual datasets, competition emerges more often. Finally, data sharing is more likely when data brokers are more efficient at merging datasets than data buyers.

JEL classifications: D43, L13, L86, M31

1. Introduction

In today's highly digitized economy, data have become particularly valuable and have attracted the attention of policymakers and institutions. To mention some examples, in 2018, the European Union (EU) General Data Protection Regulation (GDPR) to protect personal data was promulgated, and the State of California followed suit with the California Consumer Privacy Act. In 2020, the European Commission announced the EU Data Strategy (European Commission, 2020) to boost data sharing among firms and the recently proposed Digital Market Act includes mandatory data sharing as a crucial competition tool. The conventional view is that being non-rival, data can generate positive externalities, and the EU data strategy's vision is that data sharing has to be incentivized or even mandated.

If data are considered the fuel of the digital economy, 'data brokers' are its catalyst.[1] These often unknown actors are 'companies whose primary business is collecting personal information about consumers from a variety of sources and aggregating, analysing, and

1 The Economist (2017), 'Fuel of the future: data is giving rise to a new economy', 6 May 2017.

© Oxford University Press 2021.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

sharing that information' (Federal Trade Commission, 2014) and engage mostly in business-to-business relations. As they do not usually have any contact with final consumers, the latter are often unaware of their existence. A defining characteristic of this sector is that data brokers (DBs) transact and exchange data with each other, and more information is obtained this way than from direct sources. The Federal Trade Commission (2014) reports that seven out of nine DBs were buying and selling consumer data to each other. For example, Acxiom has partnerships with other DBs, including Corecom (specialized in entertainment data) and Nielsen (a global data company).

Yet, these sharing practices might not necessarily be consistent with the positive social role envisioned in the current regulatory debate and, more worryingly, may hide anti-competitive behaviours. As little is known about the behaviours of these DBs, investigations worldwide are taking place. For instance, the French National Commission on Informatics and Liberty (CNIL) carried out an in-depth investigation in the period 2017–9 auditing 50 DBs and ad-tech companies (Financial Times, 2019).

In this context, our main research question is to identify the incentives of DBs to share data in some markets and compete with others and how these relate to the nature of the data a DB has. This is relevant as, on the one hand, these companies compete to provide customers with specialized data, analytics, and market research; on the other hand, they also cooperate through partnerships and data-sharing agreements. Moreover, DBs may be particularly strong in different areas and specialize in some services, rendering the nature and type of data crucial for their strategies. For example, Acxiom and Datalogix profile consumers for targeting purposes, collecting information such as demographics, sociographics, and purchasing behaviours. DBs like Corelogic and eBureau mostly sell in-depth financial and property data analytics.

To this end, we present a simple yet rather general model to analyse how the nature of data and merging costs shape DBs' decisions. Our economy consists of two DBs and one data buyer. Throughout the article, we use 'the (data) buyer' and 'the downstream firm' interchangeably. The consumer-level information held by DBs potentially allows the downstream firm to increase its profits in its own market. For instance, a firm can use data to facilitate targeted advertising, to engage in price discrimination, or to adopt data-driven management practices.[2] DBs, on the other hand, can either share data and produce a consolidated report or compete to independently supply the downstream firm. If the DBs share data, they incur an upstream merging cost. If the DBs compete and the buyer acquires both datasets, then the buyer needs to merge them incurring a downstream merging cost.

We find that the underlying incentives to engage in either data sharing or competition crucially depend on whether the value of the merged dataset, net of the merging costs, shows forms of complementarities or substitutabilities. Indeed, data may be *super-additive* when combining two data sources, net of the merging costs, results in a more valuable dataset than the sum of the individual components. Combining the browsing history with email addresses, for example, would provide a detailed picture of the preferences of a certain consumer and enable targeted offers. In this example, data create synergies and become more valuable when merged.

2 Note that our stylized setting could still accommodate competition in the product market. Essentially, we assume that consumer level data creates extra value for the downstream firm and enhances its profitability in a given market environment, as if multiple buyers have independent interactions with the DBs.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

Data are *sub-additive* when aggregating two datasets leads to a new value, net of the merging costs, that is lower than the sum of the two separate datasets. For example, datasets might present overlapping information, diminishing marginal returns of data, correlated data points, or high merging costs. Finally, when combining two different data sources is extremely costly, a sharp reduction in the merged dataset's net value may occur. This represents a case of *extreme sub-additivity* and the value of the merged dataset is lower than the stand-alone value of its components.

Data sharing arises for two main reasons. First and foremost, to soften competition between DBs; secondly, to enable DBs to internalize potential merging cost inefficiencies on the buyer's side. The balance of these two effects drives our results. The former contrasts with the pro-competitive vision of data sharing, whereas the latter is consistent with the socially valuable perspective permeating the regulatory debate.

Suppose DBs are more efficient than the buyer in handling data. Then, when the data structure is sub-additive or extreme sub-additive both effects favour sharing. By merging sub-additive datasets, DBs can avoid granting the buyer the discount that results from competition and reflects the overlapping information and the buyer's merging cost. In the presence of an extreme sub-additive data structure resulting from a high merging cost, the mechanism is similar: as the buyer is only interested in one dataset, sharing avoids an intense, Bertrand-like, competition. When data complementarities are present, there exists a multiplicity of equilibria under competition and these render sharing less likely to occur: one DB may prefer to veto a sharing agreement when it expects to grab a larger share of the surplus than the sharing rule prescribes.

However, not always are DBs more efficient than buyers in merging datasets. For example, as a former partnership between Facebook and Acxiom suggests, a tech company may acquire information from DBs, and the former can be more efficient in handling data, given its expertise and computational capabilities.[3] In this case, the cost internalization incentive is clearly not present. However, an incentive to share data does exist when the value of the combined dataset is limited. Specifically, sharing avoids fierce competition when the datasets are extreme sub-additive. When instead the datasets are sub-additive, the two forces driving the incentives to share are now in contrast. On the one hand, DBs may be willing to share to soften competition and avoid discounting the overlapping component of the datasets. On the other hand, independent selling avoids the high merging cost facing the DBs.

Overall, depending on the nature of the data and merging costs, DBs may compete to supply a client firm in one market and, at the same time, cooperate and share data in another market. In this sense, our model successfully explains '*co-opetition*' between DBs, a characterizing feature of the sector.

Our modelling of data intermediaries is consistent with some distinguishing characteristics of the data market. First, our model captures that the value of data is contextual. For example, the same two datasets can be substitutes or complements depending on their final use and downstream market circumstances (Sarvary and Parker, 1997). While our model abstracts away from the specifics of the downstream market and sheds light on both substitute and complementary data, it is compatible with a market where DBs repeatedly interact to supply downstream buyers in different sub-markets and with buyer-specific projects. Secondly, combining and sharing data sources can be substantially more costly than

3 This partnership was in place between 2015 and 2018 (Acxiom, 2015).

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

bundling other products. This highlights a crucial difference between data, which can be merged and disposed, and product bundling.[4] For instance, merging datasets requires resource-intensive preparation of the data, and this may result in a very low net value of the final dataset. We highlight the importance of merging costs in shaping the data market outcome and characterize conditions for sharing to emerge in the unique subgame perfect Nash equilibrium. Finally, we discuss the possibility of data partitioning as, unlike many other products, a DB may be able to partly control the potential complementarity and substitutability when selling data.

1.1 Contribution to the literature

This article focuses on the market for data and the role of data intermediaries. The main contribution of our article is to capture the co-existence of competition and co-opetition between DBs, and identify the determinants of the transition between these. The closest papers to ours are Sarvary and Parker (1997), Bergemann *et al.* (2019), and Ichihashi (2021). Sarvary and Parker (1997) focus on the incentives of information sellers (e.g. consultancy, experts) to sell reports about uncertain market conditions to downstream firms, interested in finding the real state of the world. A crucial role is played by the reliability of information, data complementarity, or substitutability. In our framework, complementarity and substitutability are mediated by the presence of downstream and upstream merging costs, and data refer to individual characteristics rather than their reliability about the correct state of the world.

Instead, Bergemann *et al.* (2019) and Ichihashi (2021) analyse competition between DBs in obtaining data from consumers which can then be sold downstream. Similar to ours, Ichihashi (2021) considers a setting in which data intermediaries compete to serve a downstream firm with consumer data. However, he focuses on the welfare implications of data collection, whereas we explicitly study the incentives of data sharing and its implications for market actors.

Other studies have concentrated on related issues as privacy violations and anti-competitive practices stemming from access to data (Conitzer *et al.*, 2012; Casadesus-Masanell and Hervas-Drane, 2015; Clavorà Braulin and Valletti, 2016; Choi *et al.*, 2019; Gu *et al.*, 2019; Montes *et al.*, 2019; Belleflamme *et al.*, 2020; Ichihashi, 2020; Bounie *et al.*, 2021, and among others), strategic information sharing and signal jamming in oligopoly (Vives, 1984; Raith, 1996; Kim and Choi, 2010) and, more recently, the impact of data-driven mergers (Kim *et al.*, 2019; Chen *et al.*, 2020; De Cornière and Taylor, 2020; Prat and Valletti, 2021).

Our study also contributes to the recent law and economics literature on data sharing. In line with recent regulatory developments, this literature takes a mostly favourable view of the practice, based on the premise that, from a social perspective, there is not enough data sharing. For example, in Prüfer and Schottmüller (2021), data sharing might prevent tipping outcomes in data-driven markets. Graef *et al.* (2018) argue that the right to data portability, which enhances personal data sharing, should be seen as a new regulatory tool to stimulate competition and innovation in data-driven markets. Borgogno and Colangelo (2019) underline that data sharing via APIs requires a costly implementation process and to leverage their pro-competitive potential a regulatory intervention is necessary. Our results,

4 For the potential anti- and pro-competitive effects of bundling see, for example, Choi (2008).

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

instead, point to the possibility of excessive data sharing, through a harmful use of data to soften competition between data holding firms. This adds to other negative aspects of data sharing, as the overutilization of data pools or the reduced incentives for data gathering (Graef *et al.*, 2019; Martens *et al.*, 2020).

To a lesser extent, the issue we tackle shares similarities with patent pools (Lerner and Tirole, 2004, 2007) and how substitutability/complementarity might engender anti- or pro-competitive effects. In our framework, merging costs play an important role and interact with other forces in inducing data sharing. Moreover, a relevant difference between data and patent pools is that the latter can be considered as a structured combination of ideas whereas the former is a factor of production (Jones and Tonetti, 2020). Furthermore, unlike patents, data also have the characteristics of experience (Koutroumpis and Leiponen, 2013) and multipurpose goods (Duch-Brown *et al.*, 2017). While data and DBs have distinctive features that characterize them in general, our framework may be applicable in other settings featuring substitutability or complementarity. For example, the two upstream firms might be patent holders deciding to pool their technologies or license them independently to a downstream firm.

1.2 Outline

The rest of the article is organized as follows. Section 2 outlines the model. Our main results are presented in Section 3. Section 4 explores several extensions to our main model and Section 5 concludes with final remarks. A microfoundation of the data structure and all proofs can be found in the Appendix.

2. The model

2.1 The DBs

Consider an economy with two DBs, $k = 1, 2$, who are endowed with data on different individuals and attributes. Each DB may have independent access to a subset of the attributes.[5]

To fix ideas, let $f_k \geq 0$ be the extra surplus the buyer in question can generate by using the data owned by DB k, compared to a situation in which no data are available (i.e. $f(0) = 0$). The value function f can be interpreted as the monetary evaluation of the dataset from the perspective of the data buyer.

Data from different sources can be combined in a single dataset. This assembling process affects the value of the final dataset, depending on the underlying data structure, as defined below. In the *absence of merging costs*, a data structure is super-additive if $f_{12} \geq f_1 + f_2$ and sub-additive if $f_{12} < f_1 + f_2$, where $f_{12} \geq 0$ is the value of the merged dataset to the buyer in question.[6]

The data structure identifies a continuum of cases depending on the value of the merged dataset. It is super-additive when datasets are complements and their combination returns a final output whose value is at least as large as the sum of the individual components. There are indeed synergies in the data which lead to the creation of a more informationally powerful dataset. This may happen when the interaction between different types of data

5 For instance, this may result from a comparative advantage in different areas or from the different volumes of data they gathered. For more details, see, for example, Lambrecht and Tucker (2017).

6 More details about the microfoundation of the data structure can be found in the Appendix.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

plays a crucial role. For example, online purchasing history combined with credit card data collected offline can lead to data complementarity.

The data structure is sub-additive when the value of the merged dataset is lower than the sum of the values of individual datasets but is at least as large as either of the individual datasets. This happens when the two merging datasets have overlapping information.

The data structure is extreme sub-additive when the value of the merged dataset is lower than the value of an individual dataset. For instance, Dalessandro *et al.* (2014) suggest that, in some circumstances, adding additional data may be detrimental, and better predictions can be made with fewer data points. This is consistent with the seminal findings of Radner and Stiglitz (1984) who show theoretically that information can have a negative marginal net value. While a negative marginal value of information is caused by strictly positive information acquisition costs in Radner and Stiglitz (1984), in our framework the underlying force is the presence of non-negligible merging costs as we shall discuss below. Moreover, some customer attributes can be collinear or positively correlated (see, for example, Bergemann and Bonatti, 2019) and then lead to overlapping insights, whereas in other cases data can be difficult to integrate (see, e.g. health data in Miller and Tucker, 2014). Similar decreasing returns to scale are present in the recent literature on algorithms (Bajari *et al.*, 2019; Claussen *et al.*, 2019; Schäfer and Sapi, 2020).

DBs obtain revenues by selling their dataset. This can happen in two ways. First, DBs can sell their own dataset independently and simultaneously to the buyer. DB k's profit is then

$$\Pi_k = \begin{cases} 0 & \text{if the downstream firm does not buy } k\text{'s data} \\ p_k & \text{if the downstream firm buys } k\text{'s data} \end{cases}, \tag{1}$$

where p_k is DB k's price for its own data.

Alternatively, DBs can share their data and sell a single dataset. In this case, they jointly act as the unique data seller and make a take-it-or-leave-it offer to that specific buyer. In case of a sale, their joint profit is $P_{12} - c_{db}$, where P_{12} identifies the price jointly set by the two DBs, and $c_{db} > 0$ is the DBs' merging cost in the upstream. Let $s_k \in [0, 1]$ be k's share of the joint profit given by an exogenously fixed sharing rule. For our main analysis, we use a proportional sharing rule, to be specified in Section 3.2.1, which reflects the DBs' respective bargaining power. However, other desirable sharing rules, such as the Shapely value sharing rule can also be accommodated. We discuss this possibility in Section 4.1. DB k's individual profit when sharing is then

$$\Pi_k = \begin{cases} 0 & \text{if the downstream firm does not buy the merged data} \\ s_k \cdot (P_{12} - c_{db}) & \text{if the downstream firm buys the merged data} \end{cases}. \tag{2}$$

2.2 The data buyer

When DBs do not share data, the buyer's profits are as follows:

$$\Pi^b = \pi^0 + \begin{cases} 0 & \text{if the downstream firm does not buy data} \\ f_k - p_k & \text{if the downstream firm buys } k\text{'s data } \textit{only} \\ f_{12} - p_k - p_{-k} - c_b & \text{if the downstream firm buys data from both} \end{cases}, \tag{3}$$

where π^0 is the profit the buyer can make without data and c_b is the buyer's downstream merging cost.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

Alternatively, when DBs share their data and sell the merged dataset, the buyer obtains the following profit:

$$\Pi^b = \pi^0 + \begin{cases} 0 & \text{if the downstream firm does not buy the merged data} \\ f_{12} - P_{12} & \text{if the downstream firm buys the merged data} \end{cases}. \quad (4)$$

2.3 Timing

The timing of the game is as follows. In the first stage, the two DBs simultaneously and independently decide whether or not to share their data. Data sharing arises if, and only if, both DBs choose to share data. In the second stage, DBs jointly or independently set the price(s) for the dataset(s). Then, in the third stage, the buyer decides whether or not to buy the offered dataset(s). The equilibrium concept is Subgame Perfect Nash Equilibrium (SPNE).

3. Analysis

Before the analysis is presented, we first need to define the data structure taking into account the merging cost, occurring either at the upstream (DBs) or the downstream (the buyer) level. That is, our definition focuses on the net value of the final dataset when two different data sources are combined.

Assume, without loss of generality, that $f_2 \geq f_1$. We categorize the data structure as follows:

Definition 1. Under a given downstream merging cost c_b facing the buyer, the data structure is

- downstream super-additive, if $f_{12} - c_b \geq f_1 + f_2$,
- downstream sub-additive, if $f_2 \leq f_{12} - c_b < f_1 + f_2$, and finally
- downstream extreme sub-additive, if $f_{12} - c_b < f_2$.

The corresponding upstream data structure can be analogously defined by replacing c_b by c_{db}.

We note that the net benefit entailed by the combination of two datasets does not necessarily mirror the data structure in the absence of merging costs. For instance, a super-additive data structure without a merging cost may result in an extreme sub-additive data structure if the sharing activity takes place and its related cost is extremely high.

3.1 Independent data selling

We solve the game by backward induction. First, consider a second stage subgame where at least one DB has decided not to share data in the first stage and hence they simultaneously and independently set a price for their own data.

After observing the prices (p_1, p_2), the downstream firm decides whether to buy, and from whom, the dataset(s) so as to maximize its profit (3). This gives rise to the demand and revenue facing each DB for any given strategy profile (p_1, p_2).

Proposition 1

i. If the data structure is downstream super-additive, any pair of (p_1^*, p_2^*), such that $p_1^* + p_2^* = f_{12} - c_b$ and $p_k^* \geq f_k$, for $k = 1, 2$, constitutes a Nash equilibrium in this subgame. The downstream firm buys both datasets and merges them.

ii. If the data structure is downstream sub-additive, there exists a unique Nash equilibrium in this subgame in which $p_k^* = f_{12} - c_b - f_{-k}$, for $k = 1, 2$. The downstream firm buys both datasets and merges them.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

iii. If the data structure is downstream extreme sub-additive, there exists a unique Nash equilibrium in this subgame in which $p_1^* = 0$ and $p_2^* = f_2 - f_1$. The downstream firm does not merge the two datasets even when it buys both.

Proof: See Appendix.

The rationale of the above results is as follows. First, consider the data structure is downstream super-additive. In this case, the two datasets are characterized by strong synergies and complementarities persist even when considering merging costs c_b. This implies that rather than trying to price the rival out, each DB prefers the rival to sell its dataset too. This way, each DB hopes to appropriate some of the (positive) externalities the datasets produce downstream. As a result, in equilibrium, the buyer acquires data from both DBs and merge them on its own.

We note that in this case of downstream super-additivity, there is a continuum of competitive equilibria in which the DBs always extract the entire surplus from the buyer, that is, $\Pi_k^* + \Pi_{-k}^* = f_{12} - c_b$. This leaves the buyer 0 net benefit. Note also that the merging cost that the downstream firm faces is passed upstream because, in any equilibrium, the downstream firm will pay no more than $f_{12} - c_b$ in total.

Consider now the case where merging two datasets leads to downstream sub-additivity. In contrast to the super-additivity case, the DBs prefer undercutting the rival than accepting its own marginal value to the rival's dataset, an observation common in Bertrand-type price competition models. As a result, the unique equilibrium in (ii) emerges. Note that even if the downstream merging cost was negligible, the prices set by the DBs are limited by the substitutability of the datasets when the structure is sub-additive (e.g. overlapping information or high correlation between datasets).

In equilibrium, the buyer purchases from both DBs and pays a composite price of $p_1^* + p_2^* = 2f_{12} - f_1 - f_2 - 2c_b$, with a net benefit of $f_1 + f_2 - f_{12} + c_b > 0$. As a result, the buyer is better off: in competition, DBs have to discount the merging costs, which are incurred by the buyer only once, and also the overlapping component.

Finally, merging costs can be large for the buyer such that the data structure gets extreme sub-additive. This implies that combining different data sources becomes less appealing and the buyer would only need the most valuable dataset. Under the assumption of $f_2 \geq f_1$, only DB2 sells its data in equilibrium for sure. Its equilibrium price in this case equals the difference in the datasets' intrinsic values, whereas the rival is forced to set a zero price, as a result of competition. The buyer obtains a net benefit of f_1.

The following corollary summarizes the downstream firm's surplus and, for comparison, the industry profit of the DBs.

Corollary 1

i. If the data structure is downstream super-additive, $\Pi^b = \pi^0$ and $\Pi_1^c + \Pi_2^c = f_{12} - c_b$.
ii. If the data structure is downstream sub-additive, $\Pi^b = \pi^0 + f_1 + f_2 - f_{12} + c_b$ and $\Pi_1^c + \Pi_2^c = 2f_{12} - f_1 - f_2 - 2c_b$.
iii. If the data structure is downstream extreme sub-additive, $\Pi^b = \pi^0 + f_1$ and $\Pi_1^c + \Pi_2^c = f_2 - f_1$,

where Π_k^c denotes DB k's profit under competition.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

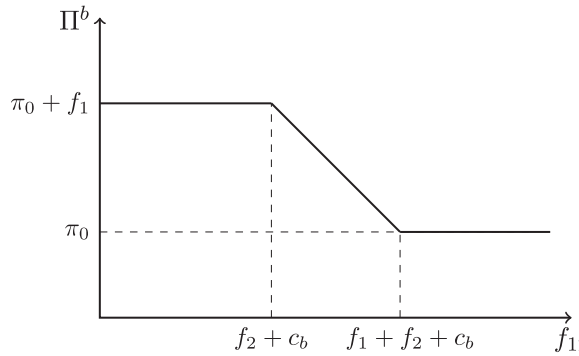

Fig. 1. The data buyer's surplus and the value of the merged dataset in the absence of a merging cost, f_{12}.

Figure 1 illustrates the buyer's surplus in relation to the gross value of the merged dataset, f_{12}. It is clear from the figure that the buyer is weakly worse off as the value of the merged dataset increases. It starts off with a positive net benefit of f_1 when the datasets are downstream extreme sub-additive and ends up with zero net surplus in the case of downstream super-additivity. The more synergy between the individual datasets, the worse it is for the downstream firm.

3.2 Data sharing

Consider the subgame when both DBs agreed to share their data. In this case, they act as an exclusive supplier to the downstream firm for its specific project.[7] As they jointly make a take-it-or-leave-it offer to the buyer, if the data structure is upstream super- or sub- additive, the total profit the DBs can obtain is $f_{12} - c_{db}$. If the data structure is upstream extreme sub-additive, DBs would not proceed to merging the datasets and simply sell the most valuable one to the buyer, jointly obtaining f_2. To sum up, under a given sharing rule s_k individual DB's profits are, for $k = 1, 2$,

$$\Pi_k^s = s_k \cdot \max\{f_{12} - c_{db}, f_2\}, \tag{5}$$

where Π_k^s denotes DB k's profit under data sharing.

We now analyse DBs' decision on data sharing. Figure 2 presents the normal form representation at the first stage of the game. To simplify the presentation, we assume $|c_b - c_{db}| \leq f_1$. That is, we exclude the less relevant cases where the cost difference is larger than the value of DB1's dataset.[8]

For data sharing to occur as an SPNE, the joint profit of the DBs when sharing their data has to be no less than those under competition, that is, $\Pi_1^s + \Pi_2^s \geq \Pi_1^c + \Pi_2^c$. Otherwise, sharing cannot be a mutual best response at the first stage.

7 Being an exclusive supplier of data for a specific project implies that the merged dataset cannot be sold by any of the two parties independently. For instance, data can be protected by non-disclosure agreements, binding contracts, or DBs can share data through an encrypted cloud or a sandbox (OECD, 2019, p.33).

8 If $|c_b - c_{db}| > f_1$, DB1 is very much disadvantaged and cooperation becomes a moot point.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

		DB 2: Share	DB 2: Compete
DB 1	Share	Π_1^s, Π_2^s	Π_1^c, Π_2^c
DB 1	Compete	Π_1^c, Π_2^c	Π_1^c, Π_2^c

Fig. 2. The normal form game at the first stage. (a) DBs are more efficient ($c_b > c_{db}$). (b) The buyer is more efficient ($c_{db} - f_1/2 \leq c_b < c_{db}$). (c) The buyer is much more efficient ($c_b < c_{db} - f_1/2$).

Proposition 2 (Joint Profits)

i. Suppose $c_b \geq c_{db}$. The joint profits of the DBs under data sharing are no less than those under independent selling, irrespective of the nature of the data structure.

ii. Suppose instead $c_b < c_{db}$. The joint profits of the DBs under data sharing are no less than those under independent selling if $f_{12} \leq \hat{f}_{12}$, where

$$\hat{f}_{12} = \begin{cases} f_1 + f_2 + 2c_b - c_{db} & \text{if } c_{db} - f_1/2 \leq c_b < c_{db} \\ f_1/2 + f_2 + c_b & \text{if } c_b < c_{db} - f_1/2 \end{cases}. \tag{6}$$

Proof: See Appendix.

Figure 3 provides a graphical representation of the findings presented in Proposition 2. Figure 3a focuses on the more natural case in which the buyer is less efficient than the DBs in merging the datasets, $c_b > c_{db}$. One illustration could be an insurance company that wants to access several potential clients' characteristics for credit scoring and profiling. For example, browsing history can be used to know an individual's habits and can be obtained through a DB specializing in marketing. Differently, data related to income and wealth can be accessed through a financial DB. These DBs routinely handle the latter data, whereas merging and cleaning separate databases may considerably be a harder task for the insurance company. The solid line (joint profits under sharing) is always above the dashed line (joint profits under competition). As a result, DBs are collectively better off when sharing data as it helps internalize downstream inefficiencies and avoid competition when their datasets overlap.

Figure 3(b, c) considers the cases where the buyer is more efficient than the DBs, $c_b < c_{db}$. For example, a dot com company, particularly effective in handling data, acquires new information from the DBs. Sharing in such cases is only an option if $f_{12} < \hat{f}_{12}$, that is, when the value of the merged datasets is sufficiently small. Intuitively, without the benefit of internalizing downstream merging inefficiencies, sharing only helps to increase joint profit when information overlapping is sufficiently severe.[9] The graphs also illustrate how the

9 While in the current setting there is no incentive to share data when the data structure is super-additive and the DBs are inefficient, such an incentive to share may be restored if the demand for data were downward sloping, for example as a result of buyer's private information about willingness to pay. In that case, individual sales of datasets would give rise to the well-known pricing externality (Cournot complementarity externalities) for which prices are too high as DBs do not internalize the externality caused by the rival. Sharing data in such a case would remove such inefficiency. However, it is important to note that this does not necessarily imply that DBs would share data. The reason is that the benefits from the internalization of the Cournot effect need to be weighed against the merging costs (which are higher upstream than downstream if $c_b < c_{db}$).

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

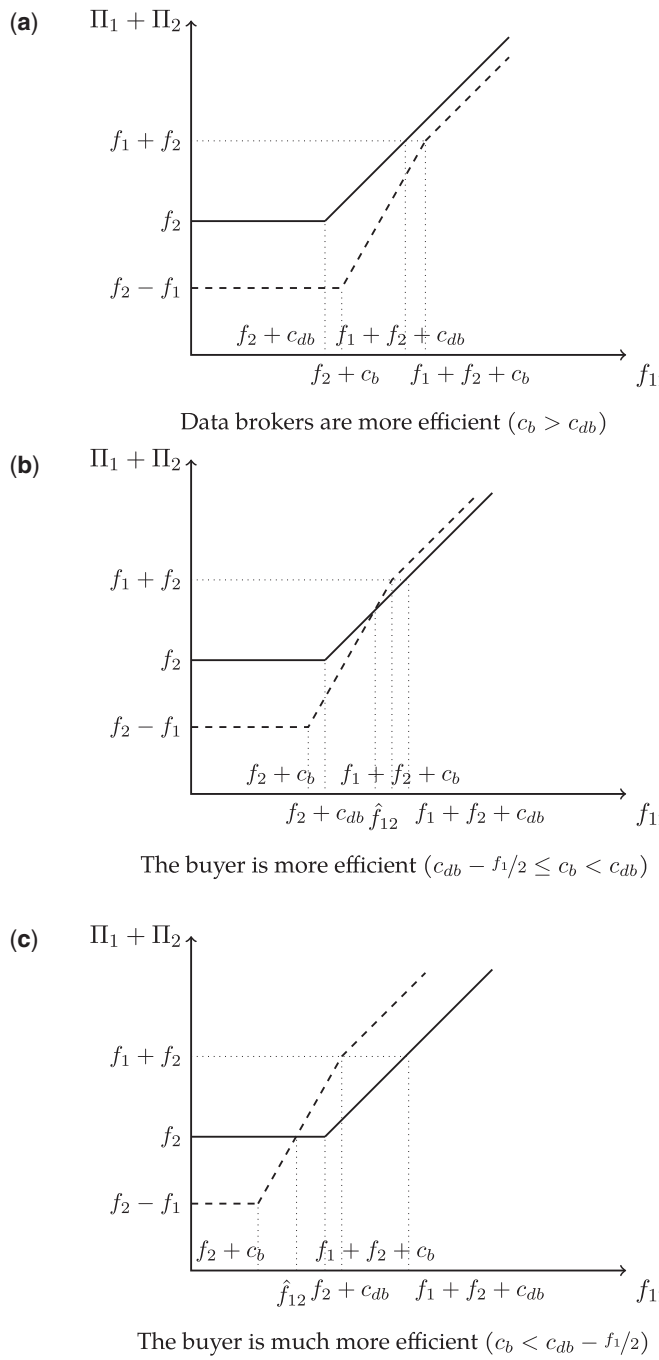

Fig. 3. DBs' joint profits from sharing (solid line) and from individual sales (dashed line), and the joint value of the datasets.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

cut-off value $\hat{f}_{12}$ is derived in these two scenarios, that is, when the downstream merging cost is relatively high or low compared to $c_{db} - f_1/2$.

3.2.1 Proportional sharing rule

Data sharing may not necessarily emerge even if joint profits are larger when sharing than under competition. For sharing to be a mutual best response, individual sharing profits must be no less than individual competition profits for both DBs. To compare these, we assume the following sharing rule that assigns a share of the joint profits to a DB that is proportional to the stand-alone value of its dataset. Namely, for $k = 1, 2$,

$$s_k = \frac{f_k}{f_k + f_{-k}}.$$

On the other hand, when data are downstream super-additive, competition between DBs leads to a multiplicity of equilibria and, similarly, only joint profits are identified. To enable the comparison, we introduce a parameter $\alpha \in [0, 1]$ to index the Nash equilibria in the competitive subgame when data are downstream super-additive. α captures the DBs' (common) belief about the share of the extra surplus assigned to DB2. Formally, we select the equilibrium where

$$p_1^* = \alpha f_1 + (1-\alpha)(f_{12} - c_b - f_2) \text{ and } p_2^* = (1-\alpha)f_2 + \alpha(f_{12} - c_b - f_1).$$

In this way, we capture all possible equilibria, ranging from the one in which the extra-surplus is allocated equally across DBs ($\alpha = 0.5$) to the ones characterized by a very asymmetric surplus reallocation ($\alpha = 1$ or $\alpha = 0$).

We are now ready to present the main result of our analysis.

Proposition 3 (Equilibrium Sharing)

i. Suppose $c_b \geq c_{db}$. Data sharing emerges in the unique Subgame Perfect Nash Equilibrium of the game, if and only if, $f_{12} < \tilde{f}_{12}$ where

$$\tilde{f}_{12} = f_1 + f_2 + c_b + \begin{cases} \dfrac{(c_b - c_{db})f_1}{(1-\alpha)f_2 - \alpha f_1} & \text{if } \alpha < \dfrac{f_2}{f_1 + f_2} \\ \infty & \text{if } \alpha = \dfrac{f_2}{f_1 + f_2} \text{ and } c_b > c_{db} \\ 0 & \text{if } \alpha = \dfrac{f_2}{f_1 + f_2} \text{ and } c_b = c_{db} \\ \dfrac{(c_b - c_{db})f_2}{\alpha f_1 - (1-\alpha)f_2} & \text{if } \alpha > \dfrac{f_2}{f_1 + f_2} \end{cases}. \tag{7}$$

ii. Suppose instead $c_b < c_{db}$. Data sharing emerges in the unique Subgame Perfect Nash Equilibrium of the game, if and only if, $f_{12} < \tilde{\tilde{f}}_{12}$ where

$$\tilde{\tilde{f}}_{12} = f_1 + c_b + \begin{cases} \dfrac{f_2^2}{f_1 + f_2} & \text{if } c_{db} - c_b < \dfrac{f_1^2}{f_1 + f_2} \\ f_2 - \dfrac{f_2}{f_1}(c_{db} - c_b) & \text{if } c_{db} - c_b > \dfrac{f_1^2}{f_1 + f_2} \end{cases}. \tag{8}$$

Proof: See Appendix.

Consider the case where DBs are more efficient than the buyer in handling data, that is, $c_b \geq c_{db}$. Suppose first that the data structure features some complementarities. The

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

previous proposition established that sharing could be industry-efficient, but this does not necessarily arise. As under competition, DBs may make very asymmetric profits (given the multiplicity of equilibria), and sharing would make one of them better off but penalize the other. In other words, for either a large or a small α, one DB vetoes a sharing agreement provided that the joint profits are sufficiently large. Only in the special case where the expected competitive profit shares are exactly in line with the sharing rule, do both brokers agree to share their data for any value of the joint dataset. To obtain the unique result, we differentiate whether $c_b > c_{db}$ or $c_b = c_{db}$ as in the latter case for any $f_{12} \geq f_1 + f_2 + c_b$, competition can also be an equilibrium outcome. The above discussion is reflected in the critical value of $\tilde{f}_{12}$ and in the conclusion that data sharing arises for $f_{12} < \tilde{f}_{12}$ as defined by (7).

Turning to a sub-additive data structure, data sharing allows for a surplus extraction that they would otherwise fail to implement fully with independent selling. Because competition leads DBs to provide a discount to the buyer (equal to downstream merging cost and the overlapping component of the datasets), sharing data can restore full surplus extraction. This way, DBs can soften competition and internalize downstream inefficiencies. A similar argument applies to an extreme sub-additive data structure. In this case, data sharing is optimal for DBs as it always allows them to coordinate on 'throwing away' DB1's dataset and extract all surplus generated by the most valuable dataset. Importantly, both DBs are better off with sharing under the assumed sharing rule than under competition.

Suppose now that the buyer is more efficient than the DBs. Note that in this case, the benefit of internalizing inefficient merging costs through sharing is absent and hence, at least one DB objects sharing when the data structure is super-additive.

When the data structure is sub-additive or extreme sub-additive, sharing can help DBs to appropriate some surplus otherwise left because of the overlapping component between their datasets. However, this appealing strategy constitutes an equilibrium only when the loss from the higher merging cost outweighs each DB's loss under competition. When the value of the merged dataset is sufficiently low, meaning substantial overlapping information, then sharing would be optimal for both DBs. As a result, there exists a critical value such that only for lower values of the joint dataset both DBs agree to share and to take on the higher upstream merging cost. This critical value is denoted by $\tilde{\tilde{f}}_{12}$.

An interesting result emerges from the above discussion. At first, one may expect that an incentive to share would emerge when complementarities between data are strong. For instance, combining email addresses (or postal codes) with the browsing history would provide the two DBs with powerful information to be sold in the market for data. Similarly, when data partially overlap or their joint use leads to quality deterioration, the incentive to share would decrease as the incremental benefit of the rival's database decreases too. On the other hand, joint selling may soften competition when data are substitutes, rendering sharing more appealing. Our model indicates that data sharing is most likely to arise when datasets present forms of substitutability and DBs are more efficient than buyers in handling data. On the contrary, competition arises more often when datasets are complements and there are upstream inefficiencies in merging data.

As noted previously, the value of data is often contextual. The same datasets held by the brokers can have different data structures, depending, for example, on the data already

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

possessed by the downstream buyer. Suppose there are three units of data, A, B, and C.[10] DB DB1 has data A and B while DB2 has B and C. Suppose further that the downstream buyer possesses A and C. With slight abuse of notation, it is easy to verify that $f_1 = f_2 = f_{12} = f(A,B,C) - f(A,C)$. Consequently, the upstream data structure is almost always sub-additive, and hence DBs face fierce competition in independent selling. As a result, in this example, data sharing between the DBs is very likely to emerge. Note also that the buyer's data make those of the DBs completely overlapping with each other although, on their own, they are complements. In this sense, the buyer's data substantially enhance the DBs' incentive to share data upstream.

Suppose the downstream buyer possesses data B instead. In this case, $f_1 = f(A,B) - f(B)$, $f_2 = f(B,C) - f(B)$ and $f_{12} = f(A,B,C) - f(B)$. In the absence of merging costs, it is easy to check that the upstream data structure can be either super- or sub-additive depending on the sign of $f(A,B,C) \lessgtr f(A,B) + f(B,C) - f(B)$. Thus, the same data held by the DBs can have different data structures depending on the context. Moreover, in this example, data B is available to all parties from the outset. By letting $f(A) = f(A,B) - f(B)$, $f(C) = f(B,C) - f(B)$, and $f(A,C) = f(A,B,C) - f(B)$, our baseline model can capture the same strategic situation without explicitly referring to B. In this sense, the buyer's data do not substantially alter the sharing incentives of the DBs.

Finally, consider the welfare implications of our analysis. Note that we abstract from explicitly modelling consumers, and this greatly simplifies the analysis. In fact, as prices are just transfers between DBs and the buyer, if datasets are merged under both regimes (competition and data sharing), welfare corresponds to the value of the data, f_{12}, net of the merging costs (c_b *and* c_{db}, respectively). As a result, the welfare gain of sharing vis-à-vis competition is simply the cost differential, $\Delta W = W^s - W^c = c_b - c_{db}$. For example, if $c_b \geq c_{db}$ then data sharing is welfare enhancing. Hence, according to Proposition 3, Part (i), for the values of f_{12} above the threshold $\tilde{f}_{12}$ the equilibrium featuring competition is inefficient.

Inefficiency in the opposite direction can take place when DBs are less efficient than the buyer ($c_b < c_{db}$). In fact, competition is welfare enhancing when $f_{12} \geq f_2 + c_b$.[11] However, the market outcome features socially inefficient sharing if f_{12} is between $f_2 + c_b$ and the threshold $\tilde{\tilde{f}}_{12}$ of Proposition 3, Part (ii). In case neither the buyer nor the DBs merge the two datasets, that is, if f_{12} is very low, then only a reallocation of surplus across parties occurs regardless of the scenario. In turn, $\Delta W = 0$ and no choice is strictly socially efficient. The previous discussion can be summarized as follows:

Proposition 4 (Welfare)

i. Suppose $c_b \geq c_{db}$. In equilibrium, welfare decreasing competition takes place for $f_{12} \geq \tilde{f}_{12}$. Otherwise, the equilibrium outcome is (weakly) socially efficient.

10 We thank one anonymous referee for suggesting this insightful example.

11 Note that the welfare gain of sharing vis-à-vis competition is simply the cost differential, $\Delta W = c_b - c_{db}$ for $f_{12} \geq f_2 + c_{db}$, whereas if $f_{12} \in [f_2 + c_b, f_2 + c_{db})$ the welfare gain of sharing is $\Delta W = f_2 - f_{12} + c_b < 0$, as under data sharing DBs do not merge datasets and jointly sell DB2's units at a price equal to f_2, whereas under competition the buyer continues to buy from both and incurs merging costs c_b.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

ii. Suppose instead $c_b < c_{db}$. In equilibrium, welfare decreasing data sharing takes place for values of $f_{12} \in [f_2 + c_b, \tilde{\tilde{f}}_{12})$. Otherwise, the equilibrium outcome is (weakly) socially efficient.

Proof: See Appendix.

4. Extensions

4.1 Alternative sharing rules

The sharing rule adopted in the previous section is just one among several possible alternatives. For example, s_k can follow the Shapley value implementation. Unlike the proportional rule, the Shapley value captures the average marginal contribution of a DB to a given coalition, that is, in our context, a data-sharing proposition. Indeed, the literature on the Nash implementation of the Shapley value demonstrates that it can be the equilibrium outcome of a properly constructed non-cooperative bargaining game (Gul, 1989).

The results obtained prove robust. Also in this context, data sharing arises for relatively low values of the combined dataset, whereas competition prevails if combining datasets generates high values. Moreover, sharing is more likely if DBs are relatively more efficient in handling the data and if the competitive equilibrium share of profits is expected to be balanced, that is, when α is close to the Shapley sharing rule. Overall, compared to a proportional rule, a Shapley value sharing rule contributes to realigning the choices of DBs with industry efficiency.

Still, both the proportional and the Shapley sharing rules may lead to a loss of surplus and inefficiency from the perspective of the DBs. Indeed, if c_b is larger than c_{db}, the joint profits always increase through sharing, but it is often the case that a proposed agreement is vetoed by one of the parties. These sharing rules have been considered so far as exogenously given. The sharing rule could be endogenized in several ways, and a take-it-or-leave-it offer by one of the DBs is a natural one. In such a setup, if the industry surplus is higher under sharing than competition, the proposer will make sure that the receiver will not veto industry efficient data sharing. Similarly, if the DBs engage in Nash bargaining with their profits in independent selling as their respective outside option, the outcome will also be efficient for them.

There are reasons to believe that both types of sharing rules (exogenous or endogenous as an outcome of a bargaining process) may characterize what happens in reality. In fact, given DBs' repeated interactions over time, a fixed sharing rule may act as a sort of (flexible) commitment, to be adjusted on a case-by-case basis. Indeed, whereas an endogenous rule leads to industry efficiency, always negotiating an endogenous agreement may be in itself overly costly for the involved parties.

4.2 Data can be partitioned

A key feature of data is its divisibility. That is, a dataset containing information regarding N consumers and M attributes can be 'repackaged' to contain information on alternative sets $\hat{N}$ of consumers and $\hat{M}$ of attributes. One may wonder whether DBs have an incentive to operate strategically such partitions when competition occurs. A rationale for partitioning might be that DBs try to soften the very harsh competition that occurs when data are sub-additive. In other words, if the original datasets feature some overlaps or correlation, the data may be restructured prior to competition in a way that eliminates or minimizes such issues.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

We note, however, that this would not affect the conclusions of our previous analysis for two reasons. First, as Part (ii) of Proposition 1 demonstrates, the DB that considered removing some overlapping information from its own dataset still obtains a profit equal to its net marginal contribution, whereas the other DB would now obtain a higher profit. Secondly, selectively repackaging some information can be particularly costly. For example, identifying specific variables and observations to remove can be time-consuming for a DB. This suggests that absent anti-competitive side-transfers, a DB may not have incentives to unilaterally reduce overlaps.

4.3 Sequential pricing

We also investigate whether DBs' incentive to share data changes when they set their prices sequentially. The timing is changed as follows. DB k first sets p_k and then DB—k sets p_{-k} after observing p_k. Given the resulting prices, the downstream firm decides whether to buy the dataset(s) and from which DB. Regardless of the order of moves, our main findings and intuitions remain qualitatively similar: data sharing emerges as a tool to soften the competition between DBs. However, as compared to the case in which prices are set simultaneously, sharing arises less often.

The intuition is as follows. A first-mover advantage is identified with a downstream super-additive data structure, which leads to the possibility of naturally selecting one equilibrium from the multiplicity identified in the benchmark. Formally, this implies selecting the equilibrium with $\alpha = \{0, 1\}$ from the benchmark model with $c_{db} \leq c_b$, and, hence, the most asymmetric surplus divisions. As a result, the first-mover has an incentive to veto any sharing agreement, rendering competition the most likely scenario.

5. Conclusion and discussion

This article sheds light on the quite obscure and relatively underexplored market for data. We present a model of data intermediaries and study their role as suppliers of valuable information to downstream firms. A distinctive aspect of the sector, prominently transpiring from the Federal Trade Commission's (2014) report, is the exchange and trade of data *between* brokers and how this relates to the particular properties of data, as compared to other products (contextual value, merging costs, and complementarities).

Our framework is compatible with a market for data in which DBs repeatedly interact to supply buyers in different sub-markets, and in which projects are buyer-specific. We highlight how the incentives for data sharing are crucially related to the nature of the data held by the brokers. Specifically, we find that data sharing can arise for two reasons. First, DBs can soften competition when data present some form of substitutability. Secondly, it allows DBs to internalize downstream inefficiencies, as buyers may be less efficient than DBs in merging multiple datasets. In turn, we identify a possible trade-off between the positive effects of cost internalization, consistent with the spirit of the EU Data Strategy (European Commission, 2020), and the negative effects of data sharing linked to reduced competition in this opaque market.

In particular, our analysis highlights the importance of the sub- or super-additive data structures, the data merging costs, and the selection of the competitive equilibrium for their decisions to cooperate on a shared project. These insights are also partly consistent with the literature on co-opetition, which has long held that companies may be collaborators with respect to value creation but become competitors when it comes to value capture (e.g.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

Nalebuff and Brandenburger, 1997). In the context of our model, collaboration may go beyond situations of value creation (efficiency savings) and can soften competition between DBs at the expense of their clients.

Our theoretical analysis rationalizes the large heterogeneity in the contractual arrangements and collaborations in this market, as also illustrated by the Federal Trade Commission (2014). For a client, our results provide two rather counter-intuitive implications. First, a firm may prefer to buy 'lower quality' (e.g. sub-additive, with overlapping information) data. This happens because competition between brokers intensifies and the firm can retain some of the surplus produced through the data. Secondly, downstream cost inefficiencies may prove to be an advantage as competition leads DBs to grant a discount to a downstream firm. This suggests that downstream firms may have less incentive to develop their digital skills when there is a functioning data market.

The sector is not particularly transparent and reliable information to conduct a proper empirical analysis of DBs' strategies is not easy to access. If data were available, however, our model delivers testable predictions. For example, the probability that DBs may exchange a dataset required by a buyer should positively relate to their relative efficiency in handling data compared to the buyers. The probability should also increase in the data homogeneity, and decrease when composite information from a variety of sources is usually in demand. At the same time, it might be inferred from highly asymmetric revenues in competitive segments of the market that data sharing has failed due to the profitable firm anticipating its dominant role.

Moreover, we shall note that the EU and the USA have followed different regulatory approaches on how data should be managed by intermediaries, third parties, and retailers. The EU has tackled the issue of privacy more strictly. More specifically, the EU GDPR has strengthened the conditions for consent by consumers, who need to be explicitly informed about the final use of the data collected.

In other words, data sharing among different DBs without prior authorization of consumers is deemed illegal, to the point that such regulation is often emphatically evoked as the 'death of third-party data'.[12] In the light of our analysis, the EU GDPR may have some unintended pro-competitive effect in the upstream data market. Specifically, the need for the explicit consent of the consumers to data sharing should reduce the prevalence of this practice, with the further consequence of enabling downstream firms to partially retain some of the data-generated surplus.

Additionally, most of the attention of the policymakers has been devoted to the final use of data and on how data sharing might create positive externalities and pro-competitive effects. Nevertheless, little attention has been given to data as an input, produced, managed, and traded by DBs. Our analysis highlights that the co-opetitive practices of DBs might require additional scrutiny from a regulator.

Finally, we conclude with a few possible extensions for future work. First, it is important to note that in our model we have assumed perfect information about buyer's evaluation. Related to the case of patent pools (e.g. Lerner and Tirole, 2004), incorporating uncertainty about the buyer's evaluation could create an additional incentive for DBs to share data in the presence of super-additive data structures. Secondly, for tractability, we have also assumed that there are only two DBs. We conjecture that our results in the spirit

12 See, for example, Wired (2018), 'Forget Facebook, mysterious DBs are facing GDPR trouble', 8 November 2018.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

of Proposition 3, that is, data sharing emerges as the unique equilibrium outcome when the merged value of individual datasets falls below a (possibly more demanding) critical level, are likely to hold with more DBs. However, data structures become more involved in the presence of several individual datasets, as complementarity and substitutability have to be specified among all possible merging decisions. This is akin to the specification of a characteristic function in a cooperative game. We leave this extension for future work where one can further explore the conditions under which data sharing only takes place among a strict subset of the DBs. Last but not least, our model is also parsimonious in the downstream and does not directly model consumers. A welfare analysis encompassing consumers would be of particular interest in this context as, besides the effect of data on product prices, data transfers, and sharing could affect the risks of data leakages and, more generally, influence consumers' privacy.

Supplementary material

Supplementary material is available on the OUP website. This consists of the online appendix.

Funding

This work was supported by the 'MOVE-IN Louvain' Incoming Fellowship Programme of the Université Catholique de Louvain to LM.

Acknowledgements

We are grateful to Alan Beggs (the Associate Editor) and two anonymous referees for insightful comments that have improved the paper. We also thank Paul Belleflamme, Nina Bobkova, Federico Boffa, Marc Bourreau, Emilio Calvano, Bruno Carballa-Smichowski, Elias Carroni, Alexandre de Cornière, Flavio Delbono, Antoine Dubus, Néstor Duch-Brown, Luca Ferrari, Juan-José Ganuza, Axel Gautier, Andreas Hervas-Drane, Alberto Iozzi, Johannes Johnen, Yassine Lefouili, Jan Krämer, Fabio Manenti, Andrea Mantovani, Bertin Martens, David Ronayne, Lorien Sabatino, Daniel Schnurr, Tommaso Valletti, John Van Reenen, Wouter Vergote, and Patrick Waelbroeck, alongside participants at the IX IBEO Workshop on Media Economics (Alghero), 45th EARIE Conference (Athens), 11th TSE Digital Economics Conference, X NERI Workshop (Rome), 9th Florence School of Regulation Seminar on Media and the Digital Economy, 2019 EEA-ESEM Conference (Manchester), 2019 IIOC (Boston), 2020 MACCI/EPoS Virtual IO Seminars, and seminar participants at different universities, for useful comments and suggestions. The usual disclaimer applies.

References

Acxiom. (2015) *Acxiom becomes Audience Data Provider Facebook Marketing Partner Program*. https://www.acxiom.co.uk/news/acxiom-becomes-audience-data-provider-facebook-marketing-partner-program (Last accessed July 24, 2021).

Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. (2019) The impact of big data on firm performance: an empirical investigation, *AEA Papers and Proceedings*, **109**, 33–7.

Belleflamme, P., Lam, W.M.W., and Vergote, W. (2020) Competitive imperfect price discrimination and market power, *Marketing Science*, **39**, 996–1015.

Bergemann, D. and Bonatti, A. (2019) Markets for information: an introduction, *Annual Review of Economics*, **11**, 85–23.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

Bergemann, D., Bonatti, A., and Gan, T. (2019) The economics of social data. Discussion Paper No. 2203R, Cowles Foundation, New Haven, CT.

Borgogno, O. and Colangelo, G. (2019) Data sharing and interoperability: fostering innovation and competition through APIs, *Computer Law & Security Review*, **35**, 105314.

Bounie, D., Dubus, A., and Waelbroeck, P. (2021) Selling strategic information in digital competitive markets, *RAND Journal of Economics*, **52**, 283–313.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015) Competing with privacy, *Management Science*, **61**, 229–46.

Chen, Z., Choe, C., Cong, J., and Matsushima, N. (2020) *Data-Driven Mergers and Personalization*. Discussion Paper No. 1108, Institute of Social and Economic Research, Osaka, Japan.

Choi, J.P. (2008) Mergers with bundling in complementary markets, *Journal of Industrial Economics*, **56**, 553–77.

Choi, J. P., Jeon, D. S., and Kim, B. C. (2019) Privacy and personal data collection with information externalities, *Journal of Public Economics*, **173**, 113–24.

Claussen, J., Peukert, C., and Sen, A. (2019) *The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News*. Working Paper No. 3479854, Social Science Research Network (SSRN), Rochester, NY.

Clavorà Braulin, F. and Valletti, T. (2016) Selling customer information to competing firms, *Economics Letters*, **149**, 10–4.

Conitzer, V., Taylor, C., and Wagman, L. (2012) Hide and seek: costly consumer privacy in a market with repeat purchases, *Marketing Science*, **31**, 277–92.

Dalessandro, B., Perlich, C., and Raeder, T. (2014) Bigger is better, but at what cost? Estimating the economic value of incremental data assets, *Big Data*, **2**, 87–96.

De Cornière, A. and Taylor, G. (2020) *Data and Competition: A General Framework with Applications to Mergers, Market Structure, and Privacy Policy*. Discussion Paper No. DP14446, Centre for Economic Policy Research (CEPR), London, England.

Duch-Brown, N., Martens, B., and Mueller-Langer, F. (2017) *The Economics of Ownership, Access and Trade in Digital Data*. Working Paper No. 2017–01, JRC Digital Economy, Seville, Spain.

European Commission. (2020) *A European Strategy for Data*. https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data (Last accessed July 24, 2021).

Federal Trade Commission. (2014) *Data Brokers: A Call for Transparency and Accountability*. Report, Federal Trade Commission, Washington, DC.

Financial Times. (2019) *Data Brokers: Regulators try to Rein in the 'Privacy Deathstars'*. https://www.ft.com/content/f1590694-fe68-11e8-aebf-99e208d3e521 (Last accessed July 24, 2021).

Graef, I., Husovec, M., and Purtova, N. (2018) Data portability and data control: lessons for an emerging concept in EU law, *German Law Journal*, **19**, 1359–98.

Graef, I., Tombal, T., and De Streel, A. (2019) *Limits and Enablers of Data Sharing. An Analytical Framework for EU Competition, Data Protection and Consumer Law*. Discussion Paper No. DP 2019–024, Tilburg Law and Economics Center (TILEC), Tilburg, The Netherlands.

Gu, Y., Madio, L., and Reggiani, C. (2019) *Exclusive Data, Price Manipulation and Market Leadership*. Working Paper No. 7853, CESifo, Munich, Germany.

Gul, F. (1989) Bargaining foundations of shapley value, *Econometrica*, **57**, 81–95.

Ichihashi, S. (2020) Online privacy and information disclosure by consumers, *American Economic Review*, **110**, 569–95.

Ichihashi, S. (2021) Competing data intermediaries, *RAND Journal of Economics, Forthcoming*.

Jones, C. I. and Tonetti, C. (2020) Nonrivalry and the economics of data, *American Economic Review*, **110**, 2819–58.

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022

Kim, B. C. and Choi, J.P. (2010) Customer information sharing: strategic incentives and new implications, *Journal of Economics & Management Strategy*, **19**, 403–33.

Kim, J. H., Wagman, L., and Wickelgren, A. L., (2019) The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing, *Journal of Economics & Management Strategy*, **28**, 373–91.

Koutroumpis, P. and Leiponen, A. (2013) Understanding the value of (big) data. *2013 IEEE International Conference on Big Data*, 38–42. IEEE, CA.

Lambrecht, A. and Tucker, C. (2017) Can big data protect a firm from competition? Antitrust Chronicle 1, Competition Policy International (CPI). https://www.competitionpolicyinternational.com/can-big-data-protect-a-firm-from-competition/.

Lerner, J. and Tirole, J. (2004) Efficient patent pools, *American Economic Review*, **94**, 691–711.

Lerner, J. and Tirole, J. (2007) Public policy toward patent pools, *Innovation Policy and the Economy*, **8**, 157–86.

Martens, B., De Streel, A., Graef, I., Tombal, T., and Duch-Brown, N. (2020) *Business-to-Business Data Sharing: An Economic and Legal Analysis*. Working Paper No. 2020–05, JRC Digital Economy, Seville, Spain.

Miller, A. R. and Tucker, C. (2014) Health information exchange, system size and information silos, *Journal of Health Economics*, **33**, 28–42.

Montes, R., Sand-Zantman, W., and Valletti, T. (2019) The value of personal information in online markets with endogenous privacy, *Management Science*, **65**, 1342–62.

Nalebuff, B. J. and Brandenburger, A. M. (1997) Co-opetition: competitive and cooperative business strategies for the digital economy, *Strategy & Leadership*, **25**, 28–33.

OECD. (2019) *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use Across Societies*. https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm (Last accessed July 24, 2021).

Prat, A. and Valletti, T. (2021) Attention oligopoly, *American Economic Journal: Microeconomics*, Forthcoming.

Prüfer, J. and Schottmüller, C. (2021) Competing with big data, *Journal of Industrial Economics*, Forthcoming.

Radner, R. and Stiglitz, J. (1984) A nonconcavity in the value of information, in M. Boyer, and R. Kihlstrom (eds) *Bayesian Models of Economic Theory*, Elsevier, Amsterdam, the Netherlands, 33–52.

Raith, M. (1996) A general model of information sharing in oligopoly, *Journal of Economic Theory*, **71**, 260–88.

Sarvary, M. and Parker, P. M. (1997) Marketing information: a competitive analysis, *Marketing Science*, **16**, 24–38.

Schäfer, M. and Sapi, G. (2020) *Learning from Data and Network Effects: The Example of Internet Search*. Discussion Paper No. 1894, German Institute for Economic Research (DIW), Berlin, Germany.

The Economist. (2017) Fuel of the future: data is giving rise to a new economy. https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy. (Last accessed July 24, 2021).

Vives, X. (1984) Duopoly information equilibrium: cournot and Bertrand, *Journal of Economic Theory*, **34**, 71–94.

Wired. (2018) *Forget Facebook, Mysterious Data Brokers are Facing GDPR Trouble*. https://www.wired.co.uk/article/gdpr-acxiom-experian-privacy-international-data-brokers. (Last accessed July 24, 2021).

Downloaded from https://academic.oup.com/oep/article/74/3/820/6369180 by guest on 31 October 2022