

The microeconomics of data – a survey

*Original*

The microeconomics of data – a survey / Pino, Flavio. - In: ECONOMIA E POLITICA INDUSTRIALE. - ISSN 0391-2078. - ELETTRONICO. - (2022). [10.1007/s40812-022-00220-6]

*Availability:*

This version is available at: 11583/2969584 since: 2022-07-06T09:35:03Z

*Publisher:*

Springer

*Published*

DOI:10.1007/s40812-022-00220-6

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)



# The microeconomics of data – a survey

Flavio Pino<sup>1</sup> 

Received: 5 November 2021 / Revised: 13 April 2022 / Accepted: 21 June 2022  
© The Author(s) 2022

## Abstract

In recent years, academia, institutions, and policymakers have been focusing their attention on the impact of data in digital markets. The economic literature that explicitly models data and their collection as strategic variables is growing, but most studies focus on distinct settings with specific data uses. This survey aims to organise this literature to extract general insights that hold across different models and assumptions. To do so, I identify three classes of models according to the way they model data collection. I find that each class is characterised by a specific impact of data on the market outcomes, regardless of the specific data use. First, when firms obtain data without strategic interactions, their use has a pro-competitive effect on the market. However, firms fail to fully internalise the data externalities, leading to data overuse and, in turn, privacy concerns. Second, when firms collect data from their interaction with consumers, data can facilitate market tipping, especially if firms are asymmetric in their starting positions. Third, when firms acquire data from data intermediaries, data are strategically sold to temper competition in the downstream market, allowing intermediaries to extract most of the surplus at the expense of firms and consumers. These general insights can facilitate future research and help policymakers to have a more general understanding of the competitive effects of data, depending on the situation at hand.

**Keywords** *Data · Data sale · Survey · Microeconomics · Data effects*

---

✉ Flavio Pino  
flavio.pino@polito.it

<sup>1</sup> Department of Management, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy

## 1 Introduction

The role of data in economics has seen a significant increase in importance in recent years, primarily due to the ever-increasing relevance of digital markets (Cr mer et al., 2019). The use of data is widespread across every sector, thanks to their versatility; typical uses include improving products or services' quality and efficiency, personalisation, matching, and discriminating between different consumers groups or individuals (Goldfarb & Tucker, 2019). Moreover, the rapid growth of online platforms has raised new challenges to competition and privacy authorities, who need to assess the potential outcomes of data-driven business models correctly.

Over the years, many authors have developed models to study the effects of data on various aspects, such as market structure, competition, welfare, and privacy. However, most of these works have only focused on a specific data effect in peculiar settings. This, in turn, has created a conundrum: while the impact of data is widely analysed, the specificity of most studies makes it difficult to abstract more general insights. Moreover, since little is known about the data collection and sale processes (Montes et al., 2019), the theoretical models currently outnumber the empiric papers where the analysed strategies can be observed in action. Thus, there can be a perceived detachment between the analysed models and real-world situations.

Data have been described as 'the oil of the digital era'<sup>1</sup>, and there is widespread consensus that their use significantly influences the economy. However, data have a considerable number of applications and reviewing how all of them impact market outcomes may result in ambiguous insights that would be of little help for policymakers or for suggesting future research developments. To set a boundary on the scope of this work, I focus my attention on models where data are explicitly modelled as an input to a decision problem. This survey is thus positioned in the strand of literature commonly referred to as digital economics. This choice excludes most of the literature regarding artificial intelligence and machine learning, as that strand of literature often focuses on data-enabled technologies rather than on how different quantities (or qualities) of data affect such technologies. The reader can refer to Agrawal et al., (2019) for a broad analysis of artificial intelligence and economics and Abrardi et al., (2021) for a comprehensive survey on artificial intelligence and machine learning.

Recent contributions have aimed to review the growing literature on digital economics by finding common characteristics that could help the authors abstract from the individual models. Goldfarb & Tucker (2019) organise the literature by identifying five types of cost reductions that stem from digital technologies and how they impact market outcomes. Bergemann & Bonatti (2019) instead focus on the characteristics of information products and their sale and the interaction between firms and data intermediaries. The contribution of the present review is twofold. First, the literature on digital economics has widely expanded in the last years, bringing new approaches and insights that could help direct future research and policy action. This survey aims to organise these recent additions and link them to previous research developments. Whenever available, I also present related evidence from empirical

<sup>1</sup> The Economist, May 6, 2017, "Regulating the internet giants. The world's most valuable resource is no longer oil, but data".

papers to better frame the insights described in the theoretical models. Second, I find that the assumptions regarding data collection are a strong driver for the models' market outcomes, regardless of the specific data use. Thus, I organise the literature depending on how data collection is modelled. This approach allows me to extract general insights that hold across different models and assumptions.

The survey is organised in three sections, each dedicated to a class of models. First, I analyse the studies where firms collect data without strategically interacting with other actors. Examples include models where firms exogenously have data from the start or when firms can acquire data by paying a marginal cost. Second, I focus on papers where firms acquire data through single or repeated interactions with consumers. In this class, firms consider the trade-off between data collection and its effect on consumer behaviour and the intertemporal effects of data acquisition (Chen, 1997; Fudenberg & Tirole, 1998). Third, I analyse models where firms can acquire data from strategic third parties, referred to as *data intermediaries*. These actors usually function as data collectors and aggregators, compounding different sources to better profile consumers (FTC, 2014). As data intermediaries serve multiple firms, their selling strategies consider how selling data to a given firm impacts its rivals. Thus, data intermediaries have a higher degree of internalisation of the overall data effects when compared to the first class of models.

The class subdivision based on the data collection process allows me to abstract general insights within the class itself, regardless of the specific data uses. When firms collect data without strategic interactions, data have a pro-competitive effect. The increase in competition is due to firms' overcollection and overuse of data, as they have a limited internalisation of data externalities. On the other hand, data overcollection raises privacy concerns that policymakers should consider when accounting for the effects of data. When firms obtain data from consumers, the effects of data strongly depend on firms' symmetry. Data acquisition and use in repeated periods can exacerbate a firm's starting advantage, potentially increasing concentration and even leading to market tipping. Moreover, firms can strategically trade or share data to limit their interaction with consumers and reduce the compensation they pay to consumers for data. This strategy is especially relevant when consumer data are correlated, as even small datasets can help firms infer information on consumers who did not disclose their data. Policymakers should thus pose particular attention to data sharing and data-driven mergers. Finally, data acquisition through intermediaries results in various outcomes that should be a concern for policymakers. Data intermediaries strategically sell their datasets to temper competition in the downstream markets to extract more profits at the expense of both firms and consumers. Moreover, the high concentration of the data intermediaries' industry (FTC, 2014) grants them substantial market power, and competing data intermediaries can strategically coordinate their actions to temper competition between them. These insights suggest that further research is needed to assess better if and which policy interventions should be implemented to limit consumer harm in these scenarios.

The survey is organised as follows: in Sect. 2, I focus on models where firms acquire data without strategic interactions with other actors. Section 3 describes works where firms acquire data from their interaction with consumers. Section 4 analyses papers where firms acquire data from data intermediaries. In each of these

sections, I briefly describe the development in the theoretical models, highlighting the main differences and findings. I then abstract from the individual models to gather more general insights and policy implications that hold true for the entire class and better assess the effects of data that arise from each data acquisition method. Finally, Sect. 5 concludes.

## 2 Data acquisition with no strategic interactions

I first analyse the strand of literature where firms acquire data without strategically interacting with other actors. Examples include models where data are exogenously available to firms or where firms incur a marginal cost when acquiring data. For ease of exposition and to better highlight connections between works, I separately analyse models where data have different effects. In particular, I distinguish three cases: when data allows price discrimination, targeting ads and more general effects like cost reduction or revenue increase.

### 2.1 Price discrimination

Price discrimination is a practice where firms can identify consumers, leading to targeted offers that better extract surplus. Data can thus be seen as a tool that allows customer identification, enabling price discrimination. Usually, price discrimination is a profitable strategy for firms: however, it can also lead to increased competition, dissipating profits.

Consider a duopoly in a spatial competition setting, where two symmetric firms exogenously have data on all consumers and can operate first-degree price discrimination. (Thisse & Vives, 1988; Shaffer & Zhang, 1995; Bester & Petrakis, 1996; Taylor, 2003). On the one hand, firms can calibrate their targeted offers depending on consumers' willingness to pay: this allows them to extract higher profits from consumers and is commonly referred to in the literature as the *surplus extraction effect* (Liu & Serfes, 2004). On the other hand, since both firms can send targeted offers to all consumers, they anticipate their rival strategy and engage in price wars. Firms thus lower their tailored prices until they match the difference in consumers' willingness to pay between the two firms. This effect dissipates profits and is commonly referred to in the literature as *intensified competition effect* (Liu & Serfes, 2004). When firms are symmetric, the *surplus extraction effect* is lower than the *intensified competition effect*: thus, firms realise lower profits under price discrimination than under mill pricing, while consumer surplus increases due to increased competition. However, firms' best response is always to commit to first-degree price discrimination: if the rival does not commit to it, a firm's profits increase with price discrimination, while if the rival commits to price discrimination, the firm limits its losses by also committing to it. In other words, firms face a prisoner's dilemma: while they would be better off by not engaging in price discrimination, it is their best response to the rival's strategy.

These results suggested that privacy could be detrimental to consumers since it would temper competition between firms. However, further analysis showed that the insight heavily depended on the assumption that firms could identify all consumers

from the start. Introducing a cost of information acquisition (Chen and Iyer, 2002; Shy & Stenbacka 2013) leads firms to identify fewer consumers in the market, tempering the *intensified competition effect* and increasing profits while making consumers worse off. Similar results are obtained under different variations of the basic model with costly data acquisition. These variations include scenarios where information sharing is allowed (Shy & Stenbacka, 2016), where consumers are loyal to one firm (Anderson et al., 2019) or where consumers can actively hide from firms (Belleflamme & Vergote, 2016; Chen et al., 2020). Moreover, Taylor & Wagman (2014) examine the effects of privacy by comparing the market outcomes of various fundamental models when firms can or cannot operate first-degree price discrimination. They find that first-degree price discrimination favours consumers in a multi-unit symmetric demand model while it harms them in a Hotelling setting (Hotelling, 1929), a Salop setting (Vickrey, 1964; Salop, 1979) and a vertical differentiation setting (Tirole, 1988).

Market asymmetries can also influence the *surplus extraction effect* and *intensified competition effect*, leading to diverse outcomes. When two firms exogenously have data on all consumers and are vertically differentiated in a spatial competition setting, the high-quality firm can expand its market share, increasing profits (Shaffer & Zhang, 2002). Instead, if only one firm has data, semi collusive behaviour can arise through a first-mover advantage (Gu et al., 2019): the informed firm sets a high price, enabling his rival to undercut him and thus avoid a price war. Switching to a homogeneous product market, asymmetries become crucial for profitability. If both firms have the same level of data accuracy, or only one firm is informed, firms end up in the Bertrand paradox (i.e. firms set prices equal to their marginal costs of production and achieve zero profits). However, when both firms have imperfect tracking with different accuracies, they can achieve positive profits at equilibrium, making consumers worse off (Belleflamme et al., 2020).

Other authors have instead focused their attention on third-degree price discrimination. In these models, data enable firms to observe a consumer's type, often identified with the consumer's loyalty to one of the firms. When compared to first-degree price discrimination, third-degree price discrimination leads to a lower increase in competition between firms, as they do not engage in price wars over individual consumers. Moreover, third-degree price discrimination could allow firms only to identify consumer types, tempering even further competition. As a guiding example, consider a case where firms only identify their loyal consumers. In this scenario, firms can extract higher profits from them but do not try to poach their rivals' consumers as they do not identify them. Only identifying loyal consumers results in lower competition and higher firms' profits (Shaffer & Zhang, 2000; Iyer et al., 2005). Firms can also escape the prisoner's dilemma identified by Thisse & Vives (1988) if information quality is low enough: in this situation, firms' equilibrium strategy involves committing not to price discriminate, even if the information is free. However, this strategy gets dominated once information quality rises (Liu & Serfes, 2004). Moreover, third-degree price discrimination can itself temper competition enough so that firms find it profitable to discriminate all consumer types, making them worse off (Armstrong & Zhou, 2010). An interesting analysis has been recently carried out by Bergemann et al. (2015), who focus on third-degree price discrimination in a

single-product monopoly setting. In particular, they analyse the welfare effects of market segmentation and demonstrate that market segmentation can achieve any combination of consumer surplus and producer surplus as long as (i) consumer surplus is nonnegative, (ii) producer surplus is greater or equal than his surplus under no segmentation and (iii) total surplus does not exceed the total value that consumers receive from the good.

Finally, introducing inaccuracies in the price discrimination process has ambiguous effects on industry profits and consumer surplus, depending on market specifics and the starting accuracy level (Chen et al., 2001; Esteves, 2014; Mauring, 2021). In particular, information inaccuracy is crucial when competing firms have asymmetric market shares (Colombo et al., 2021). If the inaccuracy is high, firms are incentivised to deviate from the equilibrium, as not offering tailored prices avoids price wars. If instead inaccuracy is low, different equilibria arise depending on firms' starting differences in market shares, and the firm with the highest starting market share ends up with lower profits than its rival. Moreover, the effect of information inaccuracy on welfare also depends on market shares' asymmetry: an increase in accuracy benefits consumers when asymmetry is high and harms them when asymmetry is low. Similar results are found when considering the allocation of data property rights to firms or consumers (Hermalin & Katz, 2006); however, enabling consumers to fully control data sharing makes all consumers better off under a monopoly setting (Ali et al., 2020).

## 2.2 Ad targeting

The targetisation of advertisement has seen a significant improvement in both scope and accuracy during the internet era, primarily due to the widespread availability of consumer data (Athey et al., 2013). Through consumer profiling, firms can reach precise targets, making specific consumers aware of the firm's presence in the market (FTC, 2014).

This strand of literature mainly focuses on a specific data effect: targeting can help firms reach consumers who otherwise would be unaware of their existence. Thus, unlike under price discrimination, data improves matching between firms and consumers and, in turn, the social value of advertising (Bergemann & Bonatti, 2011). However, some open questions remain regarding the benefits of data: do firms reach the social optimum when investing in ad targeting? Is there a threat of increasing market concentration?

A first result is given in a homogenous product market where advertising is costly (Roy, 2000): firms opt to create local monopolies when marginal ad costs are low, only targeting a share of the market and obtaining positive profits. This strategy allows them to avoid duplication costs (i.e. two firms sending an ad to the same consumer), which would increase competition and dissipate profits. While this strategy maximises welfare, as the market is fully covered, all surplus is appropriated by firms due to market segmentation. Another valuable setting is competition between advertisers for ad slots, especially when one advertiser possesses data on users that visualise that specific slot. This information asymmetry allows the advertiser to identify valuable opportunities better. While the identification of *peaches* (i.e. high valuation

consumers) does not have a substantial impact on the bidding process, the ability to recognise *lemons* (i.e. low-valuation consumers) allows the informed firm to gain a considerable advantage over its competitors (Abraham et al., 2020).

The introduction of targeting inaccuracy, as well as richer settings, allowed for further development of these insights. First, the link between an increase in targeting accuracy and a rise in firms' concentration shows to be robust even when considering firms competing over several markets (Bergemann & Bonatti, 2011) or when vertically separating firms between media firms, who show ads to consumers, and advertisers, who buy ad space (Rutt, 2012). Moreover, these additions allow to study the effects of firms' strategies regarding the social optimum: when targeting is poor (good), ad intensity is too high (low) with respect to the social optimum.

Other additions include limiting the effect of targeting through other means: the introduction of hiding technologies that allow consumers to block ads with a certain probability reduce consumer surplus due to consumers' inability to internalise the negative externalities they have on their peers (Johnson, 2013), while the introduction of a cap on available ad space can lead firms to prefer targeted advertising over general ads, increasing welfare (Athey & Gans, 2010). Another valuable addition is *malvertising*, i.e. malicious advertising such as spam or phishing. Suppose that a firm offers a free service to consumers and can sell targeted ad slots to advertisers. If malicious advertisers buy the ad slots, consumers refrain from participating again in the service, as they experience a disutility. Depending on the probability of advertisers being malicious, the firm could be deterred from selling ad slots or be incentivised to screen advertisers to retain more consumers in future periods (Jullien et al., 2020).

Finally, a strand of literature recently focuses on mergers between digital platforms: these studies are rather timely due to recent developments in this sector, like the Facebook and Instagram merger in 2012. Concerning ad targeting, merged platforms are incentivised to limit the ad supply to maximise its value, leading to a bottleneck (Prat & Valletti, 2021). Thus, this effect should be considered when analysing platform mergers.

### 2.3 Cost reduction, revenue increase

Another strand of literature has analysed the effects of data when they lead to cost reduction or revenue increase. In particular, the latter could be due to increased revenues extracted from consumers or data monetisation.

An example is provided by the search engine market, where accumulated search history improves search accuracy and thus decreases the cost of investing in quality. If one of the search engines has a more extensive search history log, this advantage can enable it to tip the market, reducing competition and welfare. These results also hold when considering a dynamic setting (Argenton & Prüfer, 2012; Prüfer & Schottmüller, 2021). Mandatory sharing of search history would level up the field between firms to maintain competition, allowing a competitive oligopoly.

Data can also increase a product's quality premium through customisation. Even when considering consumers who have a distaste for their data being used, the quality premium provided by data could be enough to attract new consumers in the market:

thus, data can have a positive welfare effect when they are quality-increasing (Campbell et al., 2015).

Finally, an even more general approach is provided by setting up a model where firms directly supply utility to consumers, as in Armstrong & Vickers (2001). Through the competition in utilities, data effects can be identified as pro-competitive or anti-competitive, with minimal or no information required on market demand and competition intensity (De Cornière & Taylor, 2020). These results give insights regarding data collection policies: restricting data collection is only desirable when data are anti-competitive.

## 2.4 Discussion and existing evidence

In this section, I have reviewed models where data are already available to firms, or where firms can source them at a marginal cost. This assumption often drives firms to overuse data, due to the threat of facing rivals who could use data more aggressively (Thisse & Vives, 1988; Shaffer & Zhang, 1995; Bester & Petrakis, 1996; Taylor, 2003). The effects of data use strongly depend on firms' starting positions. While symmetric firms usually end up in a prisoner's dilemma, with lower industry profits and an increased consumer surplus, asymmetric firms can increase their advantage thanks to data: this can lead to dominant positions or even market tipping (Shaffer & Zhang, 2002; Gu et al., 2019; Belleflamme et al., 2020).

Regarding policy implications, the seminal works in this class of models often highlighted how welfare increased under data use, due to more intense competition (Taylor & Wagman, 2014). However, this result changes when introducing features that narrow the gap between the models and real-world situations. The presence of imperfect targeting (Chen et al., 2001; Esteves, 2014; Mauring, 2021), loyal consumers (Anderson et al., 2019), limited ad capacity (Athey & Gans, 2010) or consumers' nuisance when their data are used (Jullien et al., 2020) lead to situations where consumers (or even the market as a whole) are worse off. Some papers have focused on data ownership to address these trade-offs: while results strongly depend on the models' assumptions, all papers where data have multi-faced effects conclude that introducing privacy policies would favour consumers (Hermalin & Katz, 2006; Ali et al., 2020). Other works have instead suggested acting on privacy policies to increase welfare. However, there is disagreement regarding the optimal level of enforcement: while some papers advocate for a total ban of specific data uses (Shy & Stenbacka, 2016; Anderson et al., 2019), like price discrimination or ad targeting, others assess how limiting their use, like through the GDPR, would give better results (Mauring, 2021; Belleflamme & Vergote, 2016). However, some recent empiric papers argue that soft privacy intervention could disproportionately harm smaller firms, increasing asymmetries in the market. Batikas et al., (2020) highlight this phenomenon in the market of web providers after implementing GDPR: the policy introduction lowered the market shares of all firms in favour of the market leader, Google. Garrett et al., (2022) also observed a similar result regarding web technology vendors (e.g. advertising, web hosting, audience measurement): after implementing GDPR, websites reduced their demand for vendors while concentration in the vendors' market

increased. However, most of these shocks were reabsorbed by 2018 due to the market's growth or lack of enforcement related to GDPR.

Finally, another critical policy aspect regards market concentration. Due to the increased competition, some authors find how newcomers' entrance could be hindered by data use or how the number of firms in the market would reduce after adopting data (Bergemann & Bonatti, 2011; Taylor & Wagman, 2014). This effect can facilitate mergers between firms, which in turn can create asymmetries that could result in dominant positions amplified by the data use. Binns & Bietti (2019) show how the market of online third-party trackers significantly increased its concentration over time (with Alphabet firms being present in more than 70% of the analysed sample). The same trend is observed by Batikas et al., (2020) in the web provider market.

### 3 Data from consumers

A second strand of literature has considered cases where data are endogenously created. For example, firms can gather consumer data on their first interaction with them, which can then be used in a later period; or consumers can actively decide how much data they want to share with firms. The endogenous creation or diffusion of data strengthens the link between firms and consumers: firms need to consider their externalities on consumers to predict their behaviour and act accordingly. Regarding the data effects, a substantial part of this strand of literature focuses on behaviour-based price discrimination: firms store data regarding consumers' past purchases and can use this information to distinguish recurring consumers from new ones. Other effects, like quality improvement or data monetisation, are presented in subsequent sections.

#### 3.1 Behaviour-based price discrimination

Behaviour-based price discrimination (referred to as BBPD in this section) is a relatively recent form of price discrimination, which does not fall under the standard degrees of price discrimination described in the seminal work of Pigou (1920). The novelty of this practice lies in its dynamic feature: firms need to observe consumers' past purchasing behaviour to discriminate them in a later period. The observation of consumer behaviour can be seen as a stage of data collection, where firms compete to obtain information that they can use later. Thus, BBPD falls within the boundaries of this survey. In the first works in this strand of literature, BBPD was defined as the ability to 'segment customers on the basis of their purchasing histories and to price discriminate accordingly' (Esteves, 2009a). This type of price discrimination differs from those presented in Sect. 2.1, where data allowed firms to pinpoint consumers' preferences (first-degree price discrimination) or learn their characteristics (third-degree price discrimination), as BBPD only allowed firms to distinguish between new and recurring consumers. However, in more recent works, the term BBPD has been used more broadly to describe models where firms acquire data from consumers and use it to price discriminate, regardless of the type of price discrimination (see, for example, Choe et al., 2018). To better follow the developments in this literature, in this section, I analyse models that follow the more recent and broader definition of

BBPD. Comprehensive surveys on the subject can be found in Fudenberg and Villas Boas (2006) and Esteves (2009b).

The first studies on BBPD focused on competitive two-period settings where consumers can switch firms between periods (Esteves 2009b). In these models, firms compete in a first stage and distinguish between previous and new consumers in the second. Firms thus offer better deals to their rival's previous customers without lowering the profits made on their previous consumer segment. This strategy has been described as *paying customers to switch* (Chen, 1997) or *customer poaching* (Fudenberg & Tirole, 2000). Offering discounts to the rival's customers increases competition, lowering firms' profits. Moreover, the increased customer switching leads to higher deadweight loss, lowering total welfare (Chen, 1997; Esteves, 2010; Choe et al., 2018) expand on this model by allowing firms to operate first-degree price discrimination on the consumers they identified in the first period. Their results highlight how this scenario leads to two asymmetric equilibria where one of the two firms prices more aggressively than the other to maximise its market share in the first period. However, both firms are still worse off with respect to the basic model where firms can only distinguish between new and recurring consumers, and total welfare is reduced with respect to the model without BBPD. The reduction in total welfare linked to the increased consumer switching is robust to various extensions, such as considering an infinitely-lived game with two firms (Villas Boas, 1999) or considering behaviour-based customisation of a product instead of BBPD (Zhang, 2011). On the other hand, consumers do not switch in equilibrium if the switching costs are split between transaction costs (that consumers pay whenever they switch) and learning costs (that consumers only pay during their first interaction with a firm). An increase in transaction costs results in consumer harm since firms increase their price as they realise that consumers face a higher cost when switching (Nilssen, 1992).

Other settings have instead highlighted how BBPD can be desirable for both firms and consumers. In a homogenous good market, BBPD allows firms to escape the Bertrand paradox (i.e. pricing at marginal cost) and obtain positive profits while increasing consumer surplus; on the other hand, total welfare still decreases due to excessive switching (Taylor, 2003; Esteves, 2009a). Another viable strategy is allowing firms to offer long-term contracts that lock consumers for both periods under a breach penalty threat: this strategy reduces switching, increasing consumer surplus and total welfare (Fudenberg & Tirole, 2000). Moreover, in an experience goods market, a monopolist can maximise his profits through BBPD depending on consumers' mean evaluation of the good (Jing, 2011). BBPD can also be a driver for mergers: when BBPD is allowed, a two-to-one merge in a triopoly setting allows the merged firm to extract more surplus. The merger harms consumers but does not cause deadweight loss (Esteves & Vasconcelos, 2015).

Other authors have extended the basic setting by introducing features for either or both actors: these include allowing data sharing between firms, allowing more nuanced consumer behaviour, and allowing consumers to have more control over their data, letting them decide over the amount of shared data or allowing them to use anonymising technologies to avoid being identified.

Data sharing can allow firms to identify consumers better, enabling them to increase their focus on the high willingness to pay ones. Data sharing has proven

beneficial when firms' products are positively correlated (Taylor, 2004) and when firms are uncertain about the substitutability or complementarity of their products (Kim & Choi, 2010). Data sharing can also help firms infer consumers' willingness to pay if they learn how much a consumer paid for a rival's product (Baye & Sappington, 2020). If firms collect data from consumers in the first period and operate first-degree price discrimination in the second, then data sharing only occurs if firms are (enough) vertically differentiated. In particular, the low-quality firm always sells its dataset to the high-quality one. The sale occurs because some consumers purchase from the low-quality firm in the first period even if they have a stronger preference for the high-quality one. Thus, in the second period, the high-quality firm benefits from identifying those consumers by purchasing its rival's dataset. While data sharing is Pareto-improving for firms, it is harmful to consumers, who would be better off if data sharing was banned (Liu & Serfes, 2006). Similarly, data sharing harms consumers when it occurs between identical banks that compete on credit contracts and consumers (entrepreneurs) incur positive switching costs. Under this scenario, banks can operate third-degree price discrimination on consumers they serve in the first period, identifying if their customers are talented or untalented entrepreneurs. By sharing their datasets, banks can target their offers exclusively to talented entrepreneurs, increasing their aggregate profits. Moreover, as banks anticipate data sharing in the second period, competition in the first period is also relaxed, and talented entrepreneurs pay higher prices (Gehrig & Stenbacka, 2007). Instead, the conditions under which data sharing is profitable change when a consumer sequentially buys correlated products from two sellers and the first one can sell information to the second. The first seller is better off by not sharing data if (i) its profits are independent of the second seller's profits, (ii) products' valuations are positively correlated, and (iii) the optimal contract between the second seller and the consumer is independent on the decisions of the first seller (i.e. preferences in the downstream relation are additively separable). In this case, the effect of data sharing is ambiguous: while it increases efficiency in the downstream contract, it reduces it in the upstream contract (Calzolari & Pavan, 2006; Argenziano & Bonatti, 2021) expand on this model by allowing for heterogeneous sources and uses for data and allowing the consumer to distort the information revealed to the first agent as a way to influence the second agent's behaviour. Their main result highlights how the consumer can benefit from data sharing when the quality offered by the two agents is similar, as this scenario minimises the consumer's incentive to distort his revealed information.

Considering more nuanced consumer behaviour also gives valuable results. A first strand of literature has introduced *forward-looking consumers* in a BBPD setting: these consumers anticipate firms' BBPD in the second period and update their first period behaviour accordingly. Suppose that consumers' preferences for two products are correlated and that the first firm can sell the purchasing history of its customer base to the second. Then, some high valuation consumers would prefer not to buy from the first firm to avoid being identified from the second one (Taylor, 2004). Similarly, *forward-looking consumers* can also use future data sharing to credibly signal their low willingness to pay, leading to price concessions (Baye and Sappington, 2020). A related aspect of consumer behaviour is analysing how it changes when a firm discloses that it adopts BBPD. If the firm does disclose this information, results

align with those of Fudenberg & Tirole (2000) and Fudenberg & Villas-Boas (2006): consumers' anticipation of the firm's BBPD in the second period drives the firm's profits downwards. If instead, the firm does not disclose this information, consumers form beliefs about the firm's use of BBPD based on the price observed in the first period. In this scenario, the firm would be better off by committing not to adopt BBPD. However, consumers form the belief that the firm will use BBPD in the second period, as the firm cannot credibly commit to not using it, and the firm's profits are still lower than the benchmark case where BBPD is not feasible or where he can credibly commit to not adopting it (Li et al., 2020). Similarly, when considering a product that gets improved between two periods, a firm would be better off without using BBPD: the segmentation would incentivise consumers to delay their purchase to only period 2, decreasing firms' profits (Fudenberg & Tirole, 1998). The same result holds when considering an infinite game without product improvement (Villas Boas, 2004) or a situation where goods are horizontally differentiated in period 1 and homogenous in period 2 (Jeong & Maruyama, 2009). Firms can also use consumers' anticipation of their strategies to induce self-selection: in a market with loyal consumers and lowest price buyers, setting high prices in the first period allows firms to drive the lowest price buyers out of the market, softening competition. This strategy allows firms to expand their market, with ambiguous effects on consumer welfare (Chen & Zhang, 2009). Other additions on consumer behaviour regard heterogeneity in purchase quantity and stochastic preferences. The former captures an empirically observed characteristic: a low share of consumers contributes to a large share of profits (Schmittlein et al., 1993). The latter better describes how consumer preferences change over time due to the specific purchase situation (Wernerfelt, 1994). When consumers show high heterogeneity or preferences exhibit enough stochasticity, BBPD can increase firms' profits even when consumers are forward-looking. Moreover, when consumers are heterogeneous, and their preferences change over time, then firms use BBPD as a tool to reward their own best consumers and prevent poaching; instead, if only one of the two persists, then firms are better off by rewarding their rivals' consumers (Shin & Sudhir, 2010; Shin et al., 2012) expand on this model by modelling consumer heterogeneity as heterogeneity on the cost to serve a specific consumer. The intuition is that some consumers overuse the services they pay for, resulting in a firm's profit loss over those specific consumers. Their result shows that if customer heterogeneity in their cost to serve is sufficiently high, then BBPD is profitable for the firm as it allows it to 'fire' the high-cost customers. On the other hand, low levels of heterogeneity drive the firm's profits downwards, as in Villas-Boas (2004).

A more recent strand of literature has instead focused on consumers' information disclosure when data improves matching and allows the firm to operate price discrimination. This literature adopts a setup similar to Bergemann et al. (2015) but allows consumers to disclose data instead of exogenously giving it to the firm. On the one hand, consumers are incentivised to disclose information to allow the firm to offer them a more valuable good. On the other hand, information disclosure also allows the firm to tailor its prices and extract more surplus from the trade. In equilibrium, consumers disclose the least informative segmentation that still guarantees a trade with the firm. However, total welfare would be higher with more information

disclosure, as the consumer-product matching would be improved (Hidir & Vellodi, 2021; Ichihashi, 2020) expands on this model by allowing the firm to commit not to use consumer information for pricing. Through this commitment, consumers are incentivised to share more data, and the additional revenues generated from better matching more than offset the losses caused by not being able to price discriminate. Perhaps surprisingly, consumers are worse off under this commitment, as they cannot strategically disclose data to influence the firm's prices downwards.

Adding an anonymising cost, together with the possibility of enhancing the product in period 2 and consumer heterogeneity, leads instead to ambiguous results: selling an improved product to high-valuation consumers can improve both firms' profits and welfare thanks to induced self-selection. However, BBPD is not always convenient: its profitability depends on the ratio between consumers' product valuations in both periods (Acquisti & Varian, 2005). Similar results are observed when past purchases create a consumer score, which is correlated to their willingness to pay: consumers can benefit from BBPD when the willingness to pay is high. However, this situation hurts consumers when their scores are concealed from them (Bonatti & Cisternas, 2020). Other studies also show how lowering the hiding cost can improve consumer welfare (Conitzer et al., 2012).

Finally, a crucial feature that has been recently studied is that of *data externality*, i.e. the correlation between data of different users. This topic has seen significant media coverage, primarily because of the Cambridge Analytica scandal, where the firm could infer data of 87 million users while only 320 thousand had given their consent (Schneble et al., 2018). This effect significantly impacts the market outcomes. Suppose that consumer data collected by a firm in the first period can also disclose information about consumers who have not shared their data with that firm and interact with it in the second period. Consumers do not anticipate the negative effect their disclosing has on other consumers, so their data disclosing strategy leads to consumer harm (Garrat and van Oordt, 2021). Other works regarding *data externality* that do not focus on BBPD are described in the next Section.

### 3.2 Data monetisation and as a revenue increasing factor

While price discrimination has been the most common focus when considering data collection from consumers, other relevant effects have been analysed in the literature. Here I offer a brief overview of some of the most notable ones: data as a valuable good (i.e. *data monetisation*), data as a quality increasing factor, and data as a more general revenue increasing factor.

Data monetisation has been primarily analysed in online platform settings, where it is the primary source of revenue for firms. The basic result is that consumers can't correctly evaluate the value of their data in the transactions, leading platforms to collect too much data compared to what is socially desirable (Fainmesser et al., 2020). One suggested solution to this inefficiency is shifting data ownership to consumers (Dosis & Sand-Zantman, 2019; Jones & Tonetti, 2020). However, recent literature highlighted that shifting data ownership is not always welfare enhancing once the public benefit of data is considered (i.e. data generated from one consumer benefits other consumers). Depending on the magnitude of this public benefit with respect to

the individual disutility created by data, either firm's ownership or consumers' ownership can be welfare enhancing (Markovich & Yehezkel, 2021). Another tool to temper data overcollection would be to enforce stricter privacy policies, such as default opt-out policies regarding data usage (Economides and Lianos, 2021). Another valuable scenario is where firms collect consumer data by providing a service at a price and then monetise data. This leads firms to have two possible revenue streams, as they can either focus on making profits through their service by charging positive prices or subsidise consumers to maximise data collection. Under vertical differentiation, a high-quality firm opts for the former, while a low-quality firm opts for the latter. The effects of competition on consumer privacy (i.e. the amount of disclosed data) are ambiguous: while fewer data are disclosed under competition than under monopoly, a high level of competition leads to more disclosed data, as low-quality firms increase subsidies to consumers (Casadesus-Masanell & Hervas-Drane, 2015).

Data can also improve services' quality: the most common example is search engines, where both individual and collective search histories improve search accuracy. When considering markets for homogenous products in an infinitely lived game, firms benefit more from a steeper learning curve than a larger starting data stock. In this situation, consumers would benefit from data sharing since both across-user and within-user learning would be improved. However, if firms can anticipate such a policy, they opt to soften competition and decrease consumer subsidies, leading to a decrease in welfare (Hagi & Wright, 2020). A similar application is when consumers' gross utility depends on the data they disclose and a firm's investment in quality: for example, a social media platform could create new and more effective sharing tools, leading consumers to increase their use of the platforms and share more data. Under this scenario, the platform under-provides quality and over collects data with respect to the social optimum. A regulator could impose a cap on consumer data disclosure levels (e.g. setting limits on which data types may be disclosed or posing conditions regarding the type of third parties the platform can sell data to). However, the effect of a disclosure cap on welfare is ambiguous, depending on whether the cap introduction reduces consumer participation on the platform and on the complementarity between disclosed data and quality investment in determining consumers' utility (Lefouili & Toh, 2017).

Finally, data has also been modelled as a more generic revenue-increasing factor, allowing for more general insights that abstract from specific effects. In particular, some of these works have focused on *data externality*, highlighting additional implications that stem from this effect. First, consumers can anticipate how data sharing can be harmful to them but do not consider the (positive or negative) spillover effect their sharing has on other consumers, leading to potentially inefficient outcomes. Moreover, *data externality* also affects non-users, as the firm infers non-users data from its users. This leads to excessive loss of privacy compared to the social optimum (Choi et al., 2019). Second, consumers with a low valuation of privacy share their data more easily; however, *data externality* undermines the value of privacy for all users, as firms can infer high-valuation consumers' data without their consent. Thus, consumers' data valuation is reduced, again leading to too much data collected in the market compared to the social optimum. An effective solution to overcollection would be to add a third-party mediator between the user and the firm. If a user wants

to share its data with the firm, the third-party mediator could transform these data so they are still informative regarding the user but do not reveal information on users who do not want to share their data. This intermediation would effectively avoid harmful *data externalities*, increasing welfare. (Acemoglu et al., 2022).

### 3.3 Discussion and existing evidence

In this section, I have reviewed models where firms obtain data from consumers: data can then be used in subsequent periods or directly be the firms' income source (i.e. data monetisation). One of the most analysed data effects is Behaviour-Based Price Discrimination (BBPD): firms collect data on consumers who buy from them and can thus distinguish recurrent consumers in subsequent periods. In recent years, the term BBPD has instead been used more broadly to identify models where prior interaction with consumers allows some form of price discrimination from firms, regardless of the type of discrimination. BBPD leads to increased competition in markets with horizontal differentiation, dissipating profits (Chen, 1997; Fudenberg & Tirole, 2000; Esteves, 2010; Choe et al., 2018). However, BBPD can be beneficial when considering homogenous goods (Taylor, 2003; Esteves, 2009a) or experience goods markets (Jing, 2011). Moreover, BBPD can also favour mergers in high-concentration markets (Esteves & Vasconcelos, 2015). Most of these models also describe a dead-weight loss caused by excessive consumers' switching when firms can distinguish previous consumers.

The increase in competition caused by BBPD benefits consumers, who obtain lower prices. However, firms can gain an advantage by sharing their data: this strategy allows for better targeting, hurting consumers (Taylor, 2004; Kim & Choi, 2010; Baye & Sappington, 2020; Liu & Serfes, 2006; Gehrig & Stenbacka, 2007). This outcome thus emphasises the role of consumer privacy in market outcomes; a survey on this subject can be found in Acquisti et al., (2016). Many of the models where data sharing is analysed called for stricter policies on consumer privacy to rebalance the market outcomes in favour of consumers. On the other hand, a study from Aridor, Nelson and Saltz (2020) shows how implementing more straightforward and more efficient ways to protect consumer privacy, such as the explicit opt-out introduced by the GDPR, can have unintended consequences: by abandoning more inefficient ways to protect their privacy, like browser-based privacy protection, privacy-conscious consumers reduce the noise created by short and spurious consumer histories, increasing the value and the traceability of the remaining consumers.<sup>2</sup>

However, in markets where data can improve goods, such as search engines, mandatory data sharing between firms can benefit consumers, as a lack of starting data can act as an entry barrier, limiting competition in the market (Hagiu & Wright, 2020; Schäfer & Sapi, 2020). However, empirical evidence shows how the search engines learning curve has significantly more impact than the initial data stock (Chiou & Tucker, 2017): thus, mandatory sharing could result in ambiguous effects.

Finally, a relevant aspect when firms obtain data from consumers is data externality: this can easily undermine consumers' valuation of privacy, leading to excessive

<sup>2</sup> For a survey regarding empiric studies on consumer attention towards privacy, see Acquisti et al., (2015).

data in the market (Choi et al., 2019; Acemoglu et al., 2022). This issue raises concerns about privacy policies: if firms can infer consumer data from users who are correlated with them, local policies such as GDPR would be heavily hindered. The topic of data externality has been central in discussing the Google/Fitbit merger case, which the European Commission cleared in 2020. Indeed, one of the main points of the merger's opposers is that Google's acquisition of Fitbit's health data could result in Google inferring information regarding non-users, leading to consumer harm (Bourreau et al., 2020). Moreover, the efficiency gains from the merge could result in a drop in prices in the market where data is collected (i.e. Fitbit) and in a raise in prices in the market where data is applied (i.e. Google advertising or the digital health sector), where personalised pricing increases the per-consumer revenue. If the efficiency gains are significant enough, the merger could lead to monopolies in both markets (Chen et al., 2022). On the other hand, the DGCOMP Chief Economist agreed with the Commission's decision, stressing how the Commission posed remedies to possible threats, such as Google limiting interoperability of Fitbit's wearables with non-Android devices by asking Google for a commitment to maintaining interoperability for at least ten years. Moreover, he highlights how the theory of harm regarding data externality lacks evidence, as there are no studies on the marginal value of data on advertising revenues, let alone on digital health data (Regibeau, 2021).

#### 4 Data from intermediaries

The most recent and steadily growing literature strand focuses on data intermediaries: these firms collect massive amounts of data that they then sell in downstream markets. Their upstream position allows them to consider data externalities to their full extent: while an individual firm aims at maximising its profits, a data intermediary's goal is to maximise the value of data in the downstream market that he can then extract from purchasing firms. In particular, the literature has been splitting into two major strands: one regarding data brokers and the other regarding attention platforms. While these data intermediaries gather massive amounts of data that they can then strategically sell in downstream markets, their main difference lies in the data collection process.

As highlighted by a recent FTC (2014) report, data brokers' data collection process rarely includes direct interaction with web users, as they collect consumer data through online and offline means (Bounie et al., 2021a). Consumers are often unaware that their personal data is present in their databases, and data brokers do not have to consider this negative externality when collecting data. Transferring this approach in theoretic models, data brokers are usually modelled as actors who already have consumer data or collect them by paying a marginal cost: the strategic interaction only happens between the data broker and downstream firms, not consumers.

On the other hand, attention platforms are at the forefront of data collection, as their business model focuses on gathering data by providing services, usually for free (Prat & Valletti, 2021). A distinctive trait of these platforms is that they usually need to strike a balance between the service quality they provide and their revenue stream: while increasing the latter would result in higher profits in the short term, it could

result in less traffic on the platform in the long run, leading to fewer data gathered and also fewer transactions (Evans, 2019). Attention platforms thus need to consider both consumer-side externalities in the data collection process and firm-side externalities in the data sale since a misuse could drive users away from the platform.

#### 4.1 Data brokers

The inclusion of data brokers in competition models introduces an additional level of strategic decisions, regardless of the effect of data at hand. Data brokers can serve more firms that belong to the same market: as such, these actors strategically choose to whom and which data to sell. Compared to settings where firms obtain data without strategic interactions (as in Sect. 2), data brokers consider additional externalities when choosing their strategy. The main difference is that data brokers consider how selling data to one firm impacts the others, as this externality impacts firms' data valuation and, in turn, the data broker's profits. A first strand of literature has focused on monopolistic data brokers: this assumption is not far-fetched, as the data broker industry is highly concentrated, and their combined market value is estimated at USD 156 billion per year (Pasquale, 2015). In these scenarios, data brokers choose their strategy by only focusing on the resulting outcomes in the downstream market, as they do not have to worry about competitors.

As a guiding example, consider a setting as in Thisse & Vives (1988)<sup>3</sup>, but where firms can buy data regarding all consumers from a data broker instead of having them exogenously like in the basic model. The data broker can predict the market outcome depending on his actions: for instance, serving both firms would result in an excessive increase in competition in the downstream market as in Thisse & Vives (1988) model, leading firms to deplete their profits and reducing their willingness to pay for data. Instead, the data broker excludes one of the two firms from his services to ensure higher profits that he can then extract through the data price (Montes et al., 2019). When instead data allows improving firms' products through differentiation, a data broker opts only to sell data that increase the product value for consumers who are loyal to one of the two firms. This strategy allows him to temper the rise in competition given by introducing data in the downstream market. On the other hand, if forced to sell data to both firms, the data broker would also sell competition-increasing data that would allow a firm to conquer its rival's customers. However, this strategy does not raise competition in the downstream market, as firms are better off buying all the data and then refraining from using competition-increasing ones. The fact that their rival has all the data is used by the data broker as a tool to threaten a potential non-buyer, increasing firms' willingness to pay for data (Yier and Soberman, 2000). Another available strategy is separating the access to consumers from the sharing of their information. For example, a data broker can do this by auctioning ad slots and deciding whether to grant access to consumers or share consumer information with the winning firms. Disclosing consumer data implicitly reveals information regarding their valuation of the winning firm's product to its rivals, increasing prices in the market. As such, information disclosure can be beneficial for the data

<sup>3</sup> For a detailed description of the model, see Sect. 2.1.

broker, especially when the number of downstream firms is large (De Cornière and De Nijs, 2016).

When firms are heterogeneous, or the data broker has incomplete information regarding them, the data broker can also guide the market outcome by offering a menu of contracts to achieve firms' self-selection (Arora & Fosfuri, 2005; Bergemann et al., 2018). Another way to handle firms' heterogeneity by the data broker is to sell access to individual consumers through tailored queries (i.e. cookies). Through this selling mechanism, firms can opt for positive targeting (buying information of consumers they want to reach) or negative targeting (buying information of consumers they want to avoid), and the data broker achieves their self-selection (Bergemann & Bonatti, 2015).

On the other hand, when considering long-lived data that can be used over multiple periods, a data broker opts to sell low-accuracy data in the first period to temper the competition with his future self (Cespa, 2008). Toggling the data accuracy can also be used to reduce competition in the downstream market to extract more profits: a data broker would opt to sell low-accuracy data to all firms or high-accuracy data only to a subset of them. (Garcia & Sangiorgi, 2011; Kastl et al., 2018). Similarly, a search engine is incentivised to provide suboptimal matching to competing firms: even if the marginal cost of advertising is passed to consumers by firms, the increase in competition given by optimal matching would still result in a decrease in firms' profits and, in turn, of the search engine's ones (De Cornière, 2016). Choosing the correct level of data accuracy also depends on the type of competition. When firms' actions are strategic complements (i.e. competition is *à la Bertrand*), the data broker opts for high accuracy, as data increases firms' profits that he can then extract through the price of data. On the other hand, when firms' actions are strategic substitutes (i.e. competition is *à la Cournot*), the data broker opts for low accuracy, as data would negatively affect firms' profits due to a considerable increase in downstream competition (Bimpikis et al., 2019).

Other strategies emerge when considering peculiar settings for data usage. A first example is when a data broker offers a search service that allows consumers to buy from their preferred stores. As the data broker is paid depending on the total visits in both stores, he has an incentive to divert consumers towards their least favourite store: this increases the number of consumers who visit both stores and influences firms pricing strategies downward as they try to retain some of the misplaced consumers (Hagiu & Jullien, 2011). A similar setting is also studied when downstream firms can pay the data broker with their own consumer data instead of money to gain more prominence and attract more consumers: consumer data is then sold in an external market by both the firms and the data broker, providing an additional level of competition. Moreover, the data broker already has some consumer data: thus, data obtained from firms are deemed exclusive if the data broker does not possess them and non-exclusive otherwise. While both exclusive and non-exclusive data are valuable in the external market, exclusive data grant a higher marginal revenue. When the value of non-exclusive data is high, the data broker can achieve higher profits by making firms pay for prominence with consumer data rather than money, while the opposite occurs if the value of non-exclusive data is low. In both cases, introducing

paid prominence weakens the competition between firms, leading to lower consumer surplus (Bourreau, Hofmann and Kramer, 2021).

Another strand of literature has focused on the effect of data brokers in spatial competition settings where data allow firms to price discriminate. When the data broker is forced to sell all consumer data and has all bargaining power, he opts only to serve a subset of firms, even if they are vertically differentiated: which of the firms is served depends on their quality-adjusted cost differential (Braulín and Valletti, 2016; Montes et al., 2019). Maintaining the same setting and allowing consumers to hide from the data broker and thus not appearing in his dataset does not change the data broker's incentive to underserve the downstream market. Moreover, consumer hiding would decrease consumer surplus, as their data increases competition in the market. Instead, forcing the data broker to sell data to all firms would benefit consumers (Montes et al., 2019). The data broker's strategy might change when he is provided more freedom in his data sale: when removing the assumption that a DB has to sell his entire dataset, he only sells a part of the dataset to a subset of firms. This strategy allows him to further temper the competition in the downstream market, increasing firms' profits that he can then extract through the data price (Bounie et al., 2021a). Moreover, if the data broker's bargaining power is reduced, his optimal strategy switches to providing all firms with data: as the data broker can extract fewer profits from individual firms, he opts to maximise the number of buyers at the expense of individual firms' willingness to pay (Bounie, 2020).

The previous studies focused on duopoly settings and highlighted how exclusive deals in data sales emerge when DBs have enough bargaining power. However, expanding this setting to a circular city with three firms highlights how a DB always benefits from selling to more than one of them (Delbono et al., 2021). Moreover, allowing endogenous firm entry might lead to different results. Under this scenario, the high data price results in a barrier to entry, reducing competition in the downstream market and harming consumers regardless of the data broker's bargaining power (Abrardi et al., 2022).

Consumer data availability can also ambiguously influence mergers between firms: in a two-to-one merger, the presence of a data broker exacerbates the anti-competitive effect, leading to consumer harm. Instead, in a three-to-two merger, consumers benefit from the increased efficiency caused by the merger and data. However, the data broker still benefits by only serving the merged firm, while consumers would be better off if he also serves the non-merged firm (Kim et al., 2019).

Finally, a strand of literature considers competition in the data brokers' market. Bounie et al., (2021b) obtained a first set of results, expanding the setting in Bounie et al., (2021a) by introducing competition between DBs. Their main findings highlight how competition between DBs increases competition between firms, benefiting consumers and lowering the amount of data collected. Other works have instead focused on the non-rivalrous nature of data, considering the possibility of overlaps and synergies between two data brokers' datasets. First, the degree of data accuracy and the correlation between datasets influences the buyers' behaviour: a high correlation, and likewise a high degree of accuracy, would result in firms seeing the datasets as substitutes, increasing the competition between data brokers. Moreover, a high level of competition between downstream firms further increases the degree of data-

sets' substitutability: as such, data brokers would prefer opting for exclusive deals with downstream firms to obtain higher profits. This strategy is reinforced by the fact that a first-mover advantage is observed: as such, a firm would pay a higher price for data if it can guarantee that its rival remains uninformed (Sarvary & Parker, 1997; Xiang & Sarvary, 2013). On the other hand, when rival data brokers can merge their datasets before interacting with firms, a form of co-opetition emerges. If data are sub-additive (i.e. the value of the merged dataset is lower than the sum of the individual values of the pre-merge datasets due to overlaps), data brokers merge their datasets. As the downstream firms see the datasets as substitutes, merging them allows the data brokers to avoid fierce competition. Instead, when data are supra-additive due to synergies between the datasets, data brokers opt not to merge them as firms see them as complements (Gu et al., 2021).

Other models have instead focused on data effects resulting from types of competition typical of attention platforms. A first notable effect is the control data brokers have over the firms' ability to reach consumers. The typical real-world examples are social networks, where platforms allow firms to show targeted ads to specific consumer segments. While platforms are paid per ad, firms' valuation of the ads depends on the competition they face for consumers: platforms face a trade-off between the number of ads sold and their individual value. If competition between platforms is mild, platforms opt to limit the ad supply, artificially creating a bottleneck that maximises the ads' value at the expense of entrant firms (Prat & Valletti, 2021).

Another peculiar practice recently analysed in the literature is that of social logins. This practice offers consumers a fast registration channel to many online services and allows information sharing between these services and the corresponding platform that offers the social login. The information sharing has ambiguous effects on the online services, depending on the targeting improvement. Moreover, competition between platforms that offer social logins benefits consumers and increases the number of situations where social logins are offered to online services (Kramer et al., 2019).

## 4.2 Attention platforms

The business model of data-centred attention platforms may seem relatively simple: the platform offers content, usually for free, to consumers who agree to disclose some of their data to the platform. The platform then monetises this information by selling datasets or selling access to the identified consumers to advertisers. This second strategy usually involves targeted advertising: the platform auctions a series of ad slots shown to specific consumer segments, and advertisers compete to obtain them. However, this simple mechanism hides a series of nuances that must be considered to model its functioning correctly. First, the content offered by the platform is valuable for consumers, and they choose which platform to home based on this evaluation. Second, advertisements are a nuisance to consumers, as they get in the way of the content's consumption. As such, the platform must consider the trade-off between the increasing revenues given by additional ads and the decreasing consumers' utility. Third, the advertisement demand by advertisers and the content supply by the platform are positively correlated, creating a feedback loop: a higher demand for ads

increases the value of individual ads and thus incentivises the platform to increase the amount of content provided to attract more consumers (Evans, 2019, 2020).

Starting from this basic mechanism, the existing literature has expanded upon it by introducing additional factors observed in the real world. First, consumers can have negative externalities on their peers when disclosing their data: this happens when an individual's information is predictive of the behaviour of others. This negative externality can allow the platform to obtain consumer information at a lower cost, as it reduces consumers' data evaluation. On the other hand, the platform opts to sell data to advertisers at an aggregate level to partially preserve consumers' privacy. This strategy allows the platform to capture the total value of information as the number of consumers becomes large (Bergemann et al., 2020). Second, allowing advertisers to also reach consumers through secondary channels instead of only through the platform allows to better assess the platform's influence on the market. In this scenario, advertisers can better target consumers by serving them through the platform but must pay a fee in return. On the other hand, consumers can decide how much information to share with the platform and whether they prefer buying through it or the outside market. As information becomes more precise, the value of data increases: consumers prefer to buy on the platform more, and the platform's market power over firms increases. This leads to two peculiar effects. First, as advertisers' profits in the offline market decrease, showing up on the platform becomes necessary. As such, the platform has almost complete control over their access to consumers, leading to a *gatekeeper effect*. Second, if the platform can compete with the advertisers it hosts, a *copycat effect* is observed: the platform can use the competitive advantage given by data to outclass them on their respective products, further decreasing their profits. Combining these two effects leads to a decrease in advertisers' participation in the market, as only a subset of them can sustain the cost of being active on the platform (Kirpalani and Philippon, 2020).

Another strand of literature has instead focused on competition between platforms. A key characteristic is whether consumers are single-homing (i.e. they only use one of the platforms) or multi-homing (i.e. they can use more than one platform). If consumers single home, they self-select into the platforms depending on the content they provide: this, in turn, creates a competitive bottleneck, as both platforms are the only channel through which an advertiser can reach a specific consumer. Perhaps surprisingly, if the two platforms were allowed to merge, this would increase the number of ads, while the effect on total welfare is ambiguous. Moreover, the entry of a new platform would result in higher prices for advertisement, as each platform still holds an exclusive channel to a subset of consumers (Anderson & Coate, 2005). Two key assumptions mainly drive these results: consumers can absorb any number of ads without them losing their value, and they single-home.

First, assuming that consumers can absorb any number of ads does not consider consumers' limited attention span. Introducing a limit on the number of ads that consumers register creates an additional externality between platforms: increasing the number of ads on a given platform negatively affects all the other platforms, as consumers reach their limit faster. As such, the entry of a new platform reduces the value of advertisement, as the higher the number of firms, the less they internalise the negative congestion externality, leading to a lower price per ad. The opposite is

observed if two firms merge (Anderson et al., 2012). By also introducing a limit on the time available to consumers, increased platform entry has an ambiguous effect on consumer welfare. On the one hand, consumers benefit from the increased differentiation; on the other, more entry increases the number of ads, thus reducing the time that consumers can spend consuming content. Which of the two effects dominate depends on the level of platforms' asymmetry (Anderson & Peitz, 2020).

Second, while the single-homing assumption can be a good approximation for some markets, like traditional TV broadcasting, the rapid growth of online advertising and streaming allows a sizeable share of consumers to multi-home. Allowing for consumer multi-homing solves the aforementioned competitive bottleneck, as advertisers now have multiple channels to reach specific consumers. For example, consumers could single-home in a given period but switch platforms between periods. This creates an overlap, as advertisers could reach the same consumers multiple times, wasting some advertisements. In turn, also advertisers must choose between single-homing and multi-homing: specifically, high-value advertisers opt for multi-homing, while low-value single-home (Athey et al., 2013). Allowing for multi-homing in the same period also brings similar results. In particular, introducing a correlation between consumer tastes shows how a positive correlation between platforms' contents increases the share of multi-homing consumers, leading to lower advertising levels (Ambrus et al., 2016). Similarly, allowing platforms to choose the genre of their content shows how platforms would opt for a high level of horizontal differentiation to maximise their share of single-homing consumers (Anderson et al., 2018). However, this last result depends on the assumption that the amount of multi-homing consumers is exogenous. If instead each consumer can choose whether to single or multi-home, platforms opt for head-on competition by reducing differentiation. While this strategy lowers platforms' demand, only a handful of consumers will multi-home, as the platforms' services are close substitutes. As such, platforms obtain larger shares of single-homing consumers, granting them market power when bargaining with advertisers (Haan et al., 2021). Other remarkable insights emerge when considering the addictiveness of platforms. This feature is different from the content quality provided by platforms: while content quality influences consumers' utility of participation on a platform, the addictiveness influences their marginal utility in allocating attention. Platforms can control the level of addictiveness through many UI choices, such as intrusive notifications systems or infinite scrolling. When platforms can endogenously choose the level of addictiveness, they opt to sacrifice quality for attention when consumers have a tight constraint on their attention level: in this situation, addictiveness does not change total attention. However, it influences how consumers split their attention between platforms: as such, increased competition could incentivise platforms to raise addictiveness and steal consumers from their rivals (Ichihashi & Kim, 2021).

Finally, consumer multi-homing can also be detrimental when platforms anticipate it and know whether advertisers' use of consumer data benefits or harms them. If data are non-rivalrous, consumers can share their data with multiple platforms and earn compensation from all of them. Anticipating this, platforms compete less aggressively for data to reduce the amount of overlap between their datasets and increase

their market power with intermediaries. This strategy allows platforms to earn positive profits at the expense of consumers (Ichihashi, 2021).

### 4.3 Discussion and existing evidence

In this section, I have analysed models where data intermediaries can strategically collect and sell data in a downstream market and how their presence can influence market outcomes. Due to their position in the market, data intermediaries internalise a larger share of data externalities than other actors such as consumers and downstream firms (Bergemann & Bonatti, 2019). As observed in most of the literature, data often have a pro-competitive effect on firms since they allow them to identify consumers better (Thisse & Vives, 1988). This increase in competition can effectively deplete firms' profits, reducing their willingness to pay for data. Thus, data intermediaries adopt strategies that allow them to temper competition in the downstream market to increase firms' profits that the intermediaries can then extract. Data intermediaries can temper competition in multiple ways: an intermediary could, for example, sell data of different consumers to different firms (Yier and Soberman, 2000), only serve a subset of firms (Braulín and Valletti, 2016; Montes et al., 2019; Bounie et al., 2021a; Bounie et al., 2021b; Abrardi et al., 2022), or achieve firms' self-selection through various levers when firms are heterogeneous (Arora & Fosfuri, 2005; Bergemann & Bonatti, 2015; Bergemann et al., 2018). In all these cases, firms' profits are reduced as they either spend most of their profits on buying data or compete against rivals who obtained data and thus have a competitive advantage over them. This effect can effectively limit firms' entry into the market, further decreasing competition and harming consumers (Abrardi et al., 2022).

On the other hand, data intermediaries can also have an incentive to increase competition, depending on how firms pay for their services. If a data intermediary is paid depending on the number of consumers he brings to the firms, he opts to divert consumers towards their less preferred firm so that a single consumer visits both firms, allowing double marginalization (Hagiu & Jullien, 2011). In this setting, the increase in competition between downstream firms is beneficial for the data intermediary, as it attracts more consumers that use his service to reach firms. This practice of consumer steering has also been observed in empirical works when a platform can compete with the firms it hosts. For example, Amazon products sold over Amazon are more recommended than its competitors' products; moreover, third-party products get fewer recommendations whenever Amazon runs out of the corresponding product (Chen & Tsai, 2019). A similar practice of steering has also been observed as a tool to maximise the value of ads: Facebook advertisements are skewed towards certain demographic groups despite neutral targeting settings, leading to discriminatory ad delivery (Ali et al., 2019).

Strategic interactions between data intermediaries also allow them to temper competition: the literature highlights scenarios in which data intermediaries can increase their profits by merging their datasets (Gu et al., 2021). If firms view data intermediaries' datasets as substitutes, a merge becomes a strategic tool to temper substitutability and, thus, competition. These results align with the observed practice of data trading between intermediaries (FTC, 2014).

Although the literature agrees on the inherent risks stemming from the figure of data intermediaries, empirical evidence is mostly lacking. This is due to the secretive nature of data transactions (Montes et al., 2019) and the black-box mechanisms through which platforms assign ads to consumers (Pasquale, 2015). Some recent studies have first tried to isolate the effect of data on both firms' performance and ads accuracy, to better understand the amount of market power data intermediaries have. First, they highlighted how product data could improve demand forecast, especially if the dataset contains many time periods (Bajari et al., 2019). Second, they showed how targeting consumers through data intermediaries is more inaccurate than commonly assumed in the literature, with platforms outperforming data brokers regarding accuracy levels (Neumann et al., 2019).

The dominant positions of data intermediaries in the digital markets have raised policymakers' concerns on whether and how to intervene, bringing data intermediaries to the forefront of competition policy. Antitrust authorities have thus opened many investigations and requested various reports on data brokers and attention platforms to better assess the underlying market dynamics and their perils (for a survey, see Lancieri & Sakowski 2021). One of the main problems in these analyses is the zero-price nature of attention platforms toward consumers: this feature disables many instruments of antitrust analysis, such as the SSNIP test, that rely on pricing dynamics (Newman, 2020). Recent literature has thus advanced some proposals to overcome this issue. First, an Attention-SSNIP test has been proposed as a tool to better assess market relevance when considering attention platforms. Increasing the consumers' nuisance when using a free product, for example, introducing a mandatory time-out before launching a search query on Google, would allow mapping how consumers react to this figurative increase in cost and which services they opt for instead (Wu, 2017). A second proposal has been to use the amount of data shared by consumers as a proxy for the service's price. An exploratory study in this direction showed that Facebook overcharges consumers with data, hinting at a possible dominant position in the market (Summers, 2020). Finally, a very recent development aims at subverting the way these platforms are analysed. By focusing on consumer attention instead of the derived data, attention can be seen as a scarce, rivalrous and tradeable product, more in line with standard economic definitions. Moreover, situating humans as attention producers and data intermediaries as distributors allows advertisers to be seen as the final consumers, making attention markets closely resemble familiar top-down distribution systems (Newman, 2020). These changes in how we think about platforms' business models could help policymakers better frame their market dynamics and better understand instances of platforms' abuse.

## 5 Conclusions

In this paper, I have reviewed the effects of data on market outcomes. This theme is especially relevant due to the fast expansion of the digital economy, which heavily relies on data collection and consumption and the growing concerns regarding the market leaders controlling much of the data market. By organising the literature based on how data collection is modelled, I extracted more general insights that can

help inform policymakers on if and how to intervene in case of controversies. Moreover, this organisation can help authors better identify gaps in the literature to guide future research.

When data are widely available, they often have a pro-competitive effect in the market. This can, however, hinder newcomers' entrance into the market, leading to higher concentration. Firms also have a limited internalisation of data externalities, often leading to data overuse: this raises privacy concerns that policymakers should consider when accounting for the effects of data.

I also observe similar outcomes when firms need to strategically interact with consumers to obtain data. However, these models' repeated nature also highlights how data can facilitate market tipping, especially if firms are asymmetric in their starting positions. In these situations, data use exacerbates the advantage of the high-value (or better informed) firm, increasing concentration. Firms can also strategically trade data to limit their interaction with consumers, allowing them to extract more profits: this is especially relevant when consumer data are correlated since a small dataset allows firms to infer information regarding large shares of the consumer base.

Finally, the introduction of data intermediaries further underlines the perils of unregulated data collection and usage. Data intermediaries can strategically sell their datasets to temper competition in the downstream market, allowing them to extract more profits at the expense of both firms and consumers. Their high market power stems from their quasi-monopolistic positions and the ability to strategically coordinate their actions in the case of competition between intermediaries. These concerns emerge regardless of the specific data effect, suggesting that further research is needed to better regulate these actors.

Overall, the theoretical literature is mostly in line with the scarce empirical evidence on the subject: however, the high specificity of most models limits their applicability, as it is difficult to untangle the individual effects of the various assumptions. From this point of view, models that make milder assumptions on both the type of competition and the effects of data allow for broader insights that can be extremely helpful when trying to understand the overall effect of data (for example, see De Cornière & Taylor 2020). Further research is also needed on the empirical side, as real-world evidence can help uncover novel mechanisms that theoretical models can then investigate.

**Funding** Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abraham, I., Athey, S., Babaioff, M., & Grubb, M. D. (2020). Peaches, lemons, and cookies: Designing auction markets with dispersed information. *Games and Economic Behavior*, 124, 454–477
- Abrardi, L., Cambini, C., & Rondi, L. (2021). Artificial intelligence, firms and consumer behavior: A survey. *Journal of Economic Surveys* forthcoming. <https://doi.org/10.1111/joes.12455>
- Abrardi, L., Cambini, C., Congiu, R., & Pino, F. (2022). User data and endogenous entry in online markets. Mimeo
- Acemoglu, D., Makhdoui, A., Malekian, A., & Ozdaglar, A. (2022). Too Much Data: Prices and Inefficiencies in Data Markets. *American Economic Journal: Microeconomics*, forthcoming
- Acquisti, A., & Varian, H. R. (2005). Conditioning Prices on Purchase History. *Marketing Science*, 24(3), 305–523. <https://doi.org/10.1287/mksc.1040.0103>
- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and Human Behavior in the Age of Information. *Science*, 347(6221), 509–514. <https://doi.org/10.1126/science.aaa1465>
- Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442–492. <https://doi.org/10.1257/jel.54.2.442>
- Agrawal, A., Gans, J., & Goldfarb, A. (2019). *The economics of artificial intelligence: An agenda*. Chicago and London: University of Chicago Press
- Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How Facebook’s Ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–30
- Ali, S. N., Lewis, G., & Vasserman, S. (2020). Voluntary disclosure and personalized pricing. In *Proceedings of the 21st ACM Conference on Economics and Computation* (pp. 537–538)
- Ambrus, A., Calvano, E., & Reisinger, M. (2016). Either or Both Competition: A “Two-Sided” Theory of Advertising with Overlapping Viewerships. *American Economic Journal: Microeconomics*, 8(3), 189–222. <https://www.aeaweb.org/articles?id=10.1257/mic.20150019>
- Anderson, S. P., & Coate, S. (2005). Market Provision of Broadcasting: A Welfare Analysis. *The Review of Economic Studies*, 72(4), 947–972. <http://www.jstor.org/stable/3700696>
- Anderson, S. P., & Peitz, M. (2020). Ad Clutter, Time Use, and Media Diversity. CEPR Discussion Paper No. DP15130, Available at SSRN: <https://ssrn.com/abstract=3674907>
- Anderson, S. P., Baik, A., & Larson, N. (2019). Price discrimination in the information age: Prices, poaching, and privacy with personalised targeted discounts. CEPR Discussion Paper No. DP13793. Available at SSRN: <https://ssrn.com/abstract=3428313>
- Anderson, S. P., Foros, Ø., & Kind, H. J. (2018). Competition for Advertisers and for Viewers in Media Markets. *The Economic Journal*, 128, 34–54. <https://doi.org/10.1111/ecoj.12428>
- Anderson, S. P., Foros, Ø., Kind, H. J., & Peitz, M. (2012). Media market concentration, advertising levels, and ad prices. *International Journal of Industrial Organization*, 30(3), 321–325. <https://doi.org/10.1016/j.ijindorg.2011.11.003>
- Argenton, C., & Prüfer, J. (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics*, 8(1), 73–105. <https://doi.org/10.1093/joclec/nhr018>
- Argenziano, R., & Bonatti, A. (2021). *Data Linkages and Privacy Regulation*. Working Paper
- Aridor, G., Che, Y. K., Nelson, W., & Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. Discussion Paper 26900, National Bureau of Economic Research
- Armstrong, M., & Vickers, J. (2001). Competitive Price Discrimination. *The RAND Journal of Economics*, 32(4), 579–605. <https://doi.org/10.2307/2696383>
- Armstrong, M., & Zhou, J. (2010). *Conditioning Prices on Search Behaviour: MPRA Paper 19985*. Germany: University Library of Munich
- Arora, A., & Fosfuri, A. (2005). Pricing diagnostic information. *Management Science*, 51(7), 1015–1064. <https://doi.org/10.1287/mnsc.1050.0362>
- Athey, S., & Gans, J. S. (2010). The Impact of Targeting Technology on Advertising Markets and Media Competition. *The American Economic Review*, 100(2), 608–613. <http://www.jstor.org/stable/27805067>
- Athey, S., Calvano, E., & Gans, J. (2013). The Impact of the Internet on Advertising Markets for News Media. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2180851>

- Bajari, P., Chernozhukov, V., Hortaçsu, A., & Suzuki, J. (2019). The Impact of Big Data on Firm Performance: An Empirical Investigation. *AEA Papers and Proceedings*, 109, 33–37. <https://www.aeaweb.org/articles?id=10.1257/pandp.20191000>
- Batikas, M., Bechtold, S., Kretschmer, T., & Peukert, C. (2020). European Privacy Law and Global Markets for Data. CEPR Discussion Paper No. DP14475. Available at SSRN: <https://ssrn.com/abstract=3560282>
- Baye, M., & Sappington, D. (2020). Revealing transactions data to third parties: Implications of privacy regimes for welfare in online markets. *Journal of Economics & Management Strategy*, 29(2), 260–275. <https://doi.org/10.1111/jems.12337>
- Belleflamme, P., & Vergote, W. (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters*, 149, <https://doi.org/10.1016/j.econlet.2016.10.027>
- Belleflamme, P., Lam, W. M. W., & Vergote, W. (2020). Competitive imperfect price discrimination and market power. *Marketing Science*, 39(5), 996–1015
- Bergemann, D., & Bonatti, A. (2011). Targeting in advertising markets: implications for offline versus online media. *The RAND Journal of Economics*, 42(3), 417–443. <http://www.jstor.org/stable/23046807>
- Bergemann, D., & Bonatti, A. (2015). Selling cookies. *American Economic Journal: Microeconomics*, 7(3), 259–294. <https://doi.org/10.1257/mic.20140155>
- Bergemann, D., & Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, 11, 85–107
- Bergemann, D., Bonatti, A., & Gan, T. (2020). The Economics of Social Data. Cowles Foundation Discussion Paper No. 2203R, Available at SSRN: <https://ssrn.com/abstract=3548336>
- Bergemann, D., Bonatti, A., & Smolin, A. (2018). The design and price of information. *American economic review*, 108(1), 1–48. <https://www.aeaweb.org/articles?id=10.1257/aer.20161079>
- Bester, H., & Petrakis, E. (1996). Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization*, 14(2), 227–242. [https://doi.org/10.1016/0167-7187\(94\)00469-2](https://doi.org/10.1016/0167-7187(94)00469-2)
- Bimpikis, K., Crapis, D., & Tahbaz-Salehi, A. (2019). Information Sale and Competition. *Management Science*, 65(6), 2646–2664. <https://doi.org/10.1287/mnsc.2018.3068>
- Binns, R., & Bietti, E. (2019). Acquisitions in the third-party tracking industry: competition and data protection aspects. Preprint. <https://doi.org/10.31228/osf.io/fe8u7>
- Bonatti, A., & Cisternas, G. (2020). Consumer Scores and Price Discrimination. *The Review of Economic Studies*, 87(2), 750–791. <https://doi.org/10.1093/restud/rdz046>
- Bounie, D., Dubus, A., & Waelbroeck, P. (2020). Market for Information and Selling Mechanisms. CESifo Working Paper No. 8307, Available at SSRN: <https://ssrn.com/abstract=3618830>
- Bounie, D., Dubus, A., & Waelbroeck, P. (2021a). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, 52, 283–313. <https://doi.org/10.1111/1756-2171.12369>
- Bounie, D., Dubus, A., & Waelbroeck, P. (2021b). Competition and Mergers with Strategic Data Intermediaries. Available at SSRN: <https://ssrn.com/abstract=3918829>
- Bourreau, M., Caffarra, C., Chen, Z., Choe, C., Crawford, G. S., Duso, T. ... Vergé, T. (2020). *Google/Fitbit will monetise health data and harm consumers*. Centre for Economic Policy Research
- Bourreau, M., Hofmann, J., & Krämer, J. (2021). Prominence-for-Data Schemes in Digital Platform Ecosystems: Economic Implications for Platform Bias and Consumer Data Collection. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems. WI 2021* (48 vol.). Cham: Springer. Lecture Notes in Information Systems and Organisation [https://doi.org/10.1007/978-3-030-86800-0\\_36](https://doi.org/10.1007/978-3-030-86800-0_36)
- Calzolari, G., & Pavan, A. (2006). On the Optimality of Privacy in Sequential Contracting. *Journal of Economic Theory*, 130(1), 168–204. <https://doi.org/10.1016/j.jet.2005.04.007>
- Campbell, J., Goldfarb, A., & Tucker, C. (2015). Privacy Regulation and Market Structure. *Journal of Economics & Management Strategy*, 24, 47–73. <https://doi.org/10.1111/jems.12079>
- Casadesus-Masanell, R., & Hervas-Drane, A. (2015). Competing with privacy. *Management Science*, 61(1), 229–246
- Cespa, G. (2008). Information sales and insider trading with long-lived information. *The Journal of Finance*, 63(2), 639–672. <https://doi.org/10.1111/j.1540-6261.2008.01327.x>
- Chen, N., & Tsai, H. (2019). Steering via Algorithmic Recommendations Available at SSRN: <https://ssrn.com/abstract=3500407>
- Chen, Y. (1997). Paying Customers to Switch. *Journal of Economics and Management Strategy*, 6(4), 877–897. <https://doi.org/10.1111/j.1430-9134.1997.00877.x>
- Chen, Y., & Iyer, G. (2002). Consumer Addressability and Customized Pricing. *Marketing Science*, 21(2), 197–208. <http://www.jstor.org/stable/1558067>

- Chen, Y., Narasimhan, C., & Zhang, Z. J. (2001). Individual Marketing with Imperfect Targetability. *Marketing Science*, 20(1), 23–41. <https://doi.org/10.1287/mksc.20.1.23.10201>
- Chen, Y., & Zhang, Z. J. (2009). Dynamic Targeted Pricing with Strategic Consumers. *International Journal of Industrial Organization*, 27(1), 43–50. <https://doi.org/10.1016/j.ijindorg.2008.03.002>
- Chen, Z., Choe, C., & Matsushima, N. (2020). Competitive Personalised Pricing. *Management Science*, 66(9), 4003–4023. <https://doi.org/10.1287/mnsc.2019.3392>
- Chen, Z., Choe, C., Cong, J., & Matsushima, N. (2022). Data-driven mergers and personalization. *The RAND Journal of Economics*, forthcoming. <https://doi.org/10.1111/1756-2171.12398>
- Chiou, L., & Tucker, C. (2017). Search Engines and Data Retention: Implications for Privacy and Antitrust. NBER Working Papers 23815, National Bureau of Economic Research, Inc. <https://doi.org/10.3386/w23815>
- Choe, C., King, S., & Matsushima, N. (2018). Pricing with cookies: Behavior-based price discrimination and spatial competition. *Management Science*, 64(12), 5669–5687
- Choi, J. P., Jeon, D. S., & Kim, B. C. (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*, 173, 113–124
- Clavara Braulin, F., & Valletti, T. (2016). Selling customer information to competing firms. *Economics Letters*, 149, 10–14. <https://doi.org/10.1016/j.econlet.2016.10.005>
- Colombo, S., Graziano, C., & Pignataro, A. (2021). History-Based Price Discrimination with Imperfect Information Accuracy and Asymmetric Market Shares. CESifo Working Paper No. 9049, Available at SSRN: <https://ssrn.com/abstract=3837777>
- Conitzer, V., Taylor, C. R., & Wagman, L. (2012). Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases. *Marketing Science*, 31(2), 277–292. <http://www.jstor.org/stable/41488023>
- Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). Competition policy for the digital era. Report for the European Commission Retrieved from <https://op.europa.eu/it/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1>
- De Cornière, A., & de Nijs, R. (2016). “Online Advertising and Privacy”. *The RAND Journal of Economics*, 47(1), 48–72. <https://doi.org/10.1111/1756-2171.12118>
- De Cornière, A., & Taylor, G. (2020). Data and Competition: A General Framework with Applications to Mergers, Market Structure, and Privacy Policy. CEPR Discussion Paper No. DP14446. Available at SSRN: <https://ssrn.com/abstract=3547379>
- De Cornière, A. (2016). “Search Advertising.”. *American Economic Journal: Microeconomics*, 8(3), 156–188. <https://doi.org/10.1257/mic.20130138>
- Delbono, F., Reggiani, C., & Sandrini, L. (2021). *Strategic data sales to competing firms* (pp. 2021–2005). Joint Research Centre (Seville site)
- Dosis, A., & Sand-Zantman, W. (2019). The Ownership of Data. Available at SSRN: <https://ssrn.com/abstract=3420680>
- Economides, N., & Lianos, I. (2021). Restrictions on Privacy and Exploitation in the Digital Economy: A Market Failure Perspective. *Journal of Competition Law and Economics*, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=3686785>
- Esteves, R. (2014). Price Discrimination with Private and Imperfect Information. *Scandinavian Journal of Economics*, 116, 766–796. <https://doi.org/10.1111/sjoe.12061>
- Esteves, R. B. (2009a). Customer Poaching and Advertising. *The Journal of Industrial Economics*, 57(1), 112–146. <http://www.jstor.org/stable/25483452>
- Esteves, R. B. (2009b). A Survey on the Economics of Behaviour-Based Price Discrimination. NIPE Working Papers 5/2009, NIPE - Universidade do Minho
- Esteves, R. B. (2010). Pricing with customer recognition. *International Journal of Industrial Organization*, 28(6), 669–681
- Esteves, R. B., & Vasconcelos, H. (2015). Price discrimination under customer recognition and mergers. *Journal of Economics & Management Strategy*, 24(3), 523–549. <https://doi.org/10.1111/jems.12107>
- Evans, D. S. (2019). Attention Platforms, the Value of Content, and Public Policy. *Review of Industrial Organization*, 54, 775–792. <https://doi.org/10.1007/s11151-019-09681-x>
- Evans, D. S. (2020). The Economics of Attention Markets. Available at SSRN: <https://ssrn.com/abstract=3044858>
- Fainmesser, I., Galeotti, A., & Momot, R. (2020). Digital Privacy. Discussion paper, Johns Hopkins University
- Fudenberg, D., & Tirole, J. (1998). Upgrades, Tradeins, and Buybacks. *The RAND Journal of Economics*, 29(2), 235–258. <https://doi.org/10.2307/2555887>

- Fudenberg, D., & Tirole, J. (2000). Customer Poaching and Brand Switching. *The RAND Journal of Economics*, 31(4), 634–657. <https://doi.org/10.2307/2696352>
- Fudenberg, D., & Villas-Boas, M. (2006). Behavior-Based Price Discrimination and Customer Recognition. *Handbook on Economics and Information Systems*, 1, 377–436
- Garcia, D., & Sangiorgi, F. (2011). Information sales and strategic trading. *The Review of Financial Studies*, 24(9), 3069–3104. <https://doi.org/10.1093/rfs/hhr041>
- Garratt, R. J., & Van Oordt, M. R. (2021). Privacy as a public good: a case for electronic cash. *Journal of Political Economy*, 129(7), 2157–2180
- Garrett, J., Shriver, S., & Goldberg, S. (2022). Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR. Available at SSRN: <https://ssrn.com/abstract=3477686>
- Gehrig, T., & Stenbacka, R. (2007). Information Sharing and Lending Market Competition with Switching Costs and Poaching. *European Economic Review*, 51(1), 77–99. <https://doi.org/10.1016/j.eurocorev.2006.01.009>
- Goldfarb, A., & Tucker, C. (2019). Digital Economics. *Journal of Economic Literature*, 57(1), 3–43. <https://doi.org/10.1257/jel.20171452>
- Gu, Y., Madio, L., & Reggiani, C. (2019). Exclusive Data, Price Manipulation and Market Leadership. CESifo Working Paper No. 7853. Available at SSRN: <https://ssrn.com/abstract=3467988>
- Gu, Y., Madio, L., & Reggiani, C. (2021). Data Brokers Co-Opetition. *Oxford Economic Papers*, 2021;, gpab042, <https://doi.org/10.1093/oepp/gpab042>
- Haan, M. A., Zwart, G., & Stoffers, N. (2021). Choosing Your Battles: Endogenous Multihoming and Platform Competition. TILEC Discussion Paper No. DP2021-011, University of Groningen Faculty of Law Research Paper Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3847216>
- Hagiu, A., & Jullien, B. (2011). Why Do Intermediaries Divert Search? *The RAND Journal of Economics*, 42(2), 337–362. <https://doi.org/10.1111/j.1756-2171.2011.00136.x>
- Hagiu, A., & Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. Working Paper
- Hermalin, B. E., & Katz, M. L. (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics*, 4, 209–239. <https://doi.org/10.1007/s11129-005-9004-7>
- Hidir, S., & Vellodi, N. (2021). Privacy, personalization, and price discrimination. *Journal of the European Economic Association*, 19(2), 1342–1363
- Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153), 41–57
- Ichihashi, S. (2020). Online Privacy and Information Disclosure by Consumers. *American Economic Review*, 110(2), 569–595. <https://doi.org/10.1257/aer.20181052>
- Ichihashi, S. (2021). Competing data intermediaries. *The RAND Journal of Economics*, 52, 515–537. <https://doi.org/10.1111/1756-2171.12382>
- Ichihashi, S., & Kim, B. C. (2021). Addictive platforms. Available at SSRN: <https://ssrn.com/abstract=3853961>
- Iyer, G., Soberman, D., & Villas-Boas, J. M. (2005). The Targeting of Advertising. *Marketing Science*, 24(3), 461–476. <https://doi.org/10.1287/mksc.1050.0117>
- Jeong, Y., & Maruyama, M. (2009). Commitment to a strategy of uniform pricing in a two-period duopoly with switching costs. *Journal of Economics*, 98(1), 45–66. <http://www.jstor.org/stable/41795638>
- Jing, B. (2011). Pricing Experience Goods: The Effects of Customer Recognition and Commitment. *Journal of Economics and Management Strategy*, 20(2), 451–473. <https://doi.org/10.1111/j.1530-9134.2011.00294.x>
- Johnson, J. P. (2013). Targeted advertising and advertising avoidance. *The RAND Journal of Economics*, 44(1), 128–144. <http://www.jstor.org/stable/43186411>
- Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9), 2819–2858
- Jullien, B., Lefouili, Y., & Riordan, M. H. (2020). Privacy protection, security, and consumer retention. *Security, and Consumer Retention* (June 1, 2020). Available at SSRN: <https://ssrn.com/abstract=3655040>
- Kastl, J., Pagnozzi, M., & Piccolo, S. (2018). Selling information to competitive firms. *The RAND Journal of Economics*, 49(1), 254–282. <https://www.jstor.org/stable/45147433>
- Kim, B. C., & Choi, J. P. (2010). Customer Information Sharing: Strategic Incentives and New Implications. *Journal of Economics & Management Strategy*, 19, 403–433. <https://doi.org/10.1111/j.1530-9134.2010.00256.x>

- Kim, J. H., Wagman, L., & 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(3), 373–391
- Kirpalani, R., & Philippon, T. (2020). Data Sharing and Market Power with Two-Sided Platforms. NBER working paper 28023, available at <https://www.nber.org/papers/w28023>
- Kramer, J., Schnurr, D., & Wohlfarth, M. (2019). Winners, losers, and facebook: The role of social logins in the online advertising ecosystem. *Management Science*, 65(4), 1678–1699. <https://doi.org/10.1287/mnsc.2017.3012>
- Lancieri, F., & Sakowski, P. M. (2021). Competition in digital markets: A review of expert reports. *Stanford Journal of Law Business & Finance*, 26(1), 65–170. <https://doi.org/10.2139/ssrn.3681322>
- Lefouili, Y., & Toh, Y. L. (2017). Privacy and quality. *TSE Working Paper*, 17 (795)
- Li, X., Li, K. J., & Wang, X. (2020). Transparency of behavior-based pricing. *Journal of Marketing Research*, 57(1), 78–99
- Liu, Q., & Serfes, K. (2004). Quality of Information and Oligopolistic Price Discrimination. *Journal of Economics & Management Strategy*, 13, 671–702. <https://doi.org/10.1111/j.1430-9134.2004.00028.x>
- Liu, Q., & Serfes, K. (2006). Customer information sharing among rival firms. *European Economic Review*, 50(6), 1571–1600. <https://doi.org/10.1016/j.eurocorev.2005.03.004>
- Markovich, S., & Yehezkel, Y. (2021). Data Regulation: Who Should Control Our Data? Available at SSRN: <https://ssrn.com/abstract=3801314>
- Mauring, E. (2021). Search and Price Discrimination Online. CEPR Discussion Paper No. DP15729. Available at SSRN: <https://ssrn.com/abstract=3783955>
- Montes, R., Sand-Zantman, W., & Valletti, T. (2019). The Value of Personal Information in Online Markets with Endogenous Privacy. *Management Science*, 65(3), 955–1453. <https://doi.org/10.1287/mnsc.2017.2989>
- Neumann, N., Tucker, C., & Whitfield, T. (2019). How Effective Is Third-Party Consumer Profiling and Audience Delivery? Evidence from Field Studies. *Marketing Science*, 38(6), 918–926. <https://doi.org/10.1287/mksc.2019.1188>
- Newman, J. (2020). Antitrust in Attention Markets: Definition, Power, Harm. University of Miami Legal Studies Research Paper No. 3745839, Available at <https://doi.org/10.2139/ssrn.3745839>
- Nilssen, T. (1992). Two Kinds of Consumer Switching Costs. *RAND Journal of Economics*, 23, 579–589
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press
- Prat, A., & Valletti, T. M. (2021). Attention Oligopoly. *American Economic Journal: Microeconomics*, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3197930>
- Prüfer, J., & Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, 69(4), 967–1008
- Regibeau, P. (2021). Why I agree with the Google-Fitbit decision. Available at <https://voxeu.org/article/why-i-agree-google-fitbitdecision>
- Roy, S. (2000). Strategic segmentation of a market. *International Journal of Industrial Organization*, 18(8), 1279–1290. [https://doi.org/10.1016/S0167-7187\(98\)00052-6](https://doi.org/10.1016/S0167-7187(98)00052-6)
- Rutt, J. (2012). Targeted Advertising and Media Market Competition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2103061>
- Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics*, 10(1), 141–156
- Sarvary, M., & Parker, P. M. (1997). Marketing information: A competitive analysis. *Marketing science*, 16(1), 24–38. <https://doi.org/10.1287/mksc.16.1.24>
- Schäfer, M., & Sapi, G. (2020). Learning from Data and Network Effects: The Example of Internet Search. DIW Berlin Discussion Paper No. 1894. Available at SSRN: <https://ssrn.com/abstract=3688819>
- Schmittlein, D. C., Cooper, L. G., & Morrison, D. G. (1993). Truth in Concentration in the Land of (80/20) Laws. *Marketing Science*, 12(2), 167–183
- Schneble, C. O., Elger, B. S., & Shaw, D. (2018). The Cambridge Analytica affair and Internet-mediated research. EMBO reports, 19(8)
- Shaffer, G., & Zhang, Z. (2000). Pay to Switch or Pay to Stay: Preference-Based Price Discrimination in Markets with Switching Costs. *Journal of Economics and Management Strategy*, 9(3), 397–424. <https://doi.org/10.1111/j.1430-9134.2000.00397.x>
- Shaffer, G., & Zhang, Z. J. (1995). Competitive Coupon Targeting. *Marketing Science*, 14(4), 395–416. <http://www.jstor.org/stable/184137>

- Shaffer, G., & Zhang, Z. J. (2002). Competitive One-to-One Promotions. *Management Science*, 48(9), 1143–1160. <http://www.jstor.org/stable/822606>
- Shin, J., & Sudhir, K. (2010). A customer management dilemma: When is it profitable to reward one's own customers? *Marketing Science*, 29(4), 671–689
- Shin, J., Sudhir, K., & Yoon, D. H. (2012). When to “fire” customers: Customer cost-based pricing. *Management Science*, 58(5), 932–947
- Shy, O., & Stenbacka, R. (2013). Investment in customer recognition and information exchange. *Information Economics and Policy*, 25(2), 92–106. <https://doi.org/10.1016/j.infoecopol.2013.03.002>
- Shy, O., & Stenbacka, R. (2016). Customer privacy and competition. *Journal of Economics & Management Strategy*, 25(3), 539–562. <https://doi.org/10.1111/jems.12157>
- Summers, M. (2020). Facebook isn't free: Zero-price companies overcharge consumers with data. *Behavioural Public Policy*, 1–25. doi:<https://doi.org/10.1017/bpp.2020.47>
- Taylor, C. R. (2003). Supplier Surfing: Competition and Consumer Behavior in Subscription Markets. *The RAND Journal of Economics*, 34(2), 223–246. <https://doi.org/10.2307/1593715>
- Taylor, C. R. (2004). Consumer Privacy and the Market for Customer Information. *The RAND Journal of Economics*, 35(4), 631–650. <https://doi.org/10.2307/1593765>
- Taylor, C. R., & Wagman, L. (2014). Consumer Privacy in Oligopolistic Markets: Winners, Losers, and Welfare. *International Journal of Industrial Organization*, 34, 80–84. <https://doi.org/10.1016/j.ijindorg.2014.02.010>
- Thisse, J. F., & Vives, X. (1988). On The Strategic Choice of Spatial Price Policy. *The American Economic Review*, 78(1), 122–137. <http://www.jstor.org/stable/1814702>
- Tirole, J. (1988). *The theory of industrial organization*. MIT press
- Vickrey, W. S. (1964). *Microstatics*. Harcourt. New York - Burlingame: Brace & World, Inc. USA
- Villas-Boas, J. M. (1999). Dynamic Competition with Customer Recognition. *The RAND Journal of Economics*, 30(4), 604–631. <https://doi.org/10.2307/2556067>
- Villas-Boas, J. M. (2004). Price Cycles in Markets with Customer Recognition. *The RAND Journal of Economics*, 35(3), 486–501. <https://doi.org/10.2307/1593704>
- Wernerfelt, B. (1994). On the function of sales assistance. *Marketing Science*, 13(1), 68–82
- Wu, T. (2017). Blind Spot: The Attention Economy and the Law. *Law & Society: Public Law - Antitrust eJournal*. Available at [https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3030&context=faculty\\_scholarship](https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3030&context=faculty_scholarship)
- Xiang, Y., & Sarvary, M. (2013). Buying and selling information under competition. *Quantitative Marketing and Economics*, 11(3), 321–351. <https://doi.org/10.1007/s11129-013-9135-1>
- Zhang, J. (2011). The Perils of Behavior-Based Personalisation. *Marketing Science*, 30(1), 170–186. <https://doi.org/10.1287/mksc.1100.0607>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.