Data-driven Envelopment with Privacy-Policy Tying*

Daniele Condorelli[†], Jorge Padilla[‡]

June 19, 2023

Abstract

We present a theory of monopoly protection by means of entry in adjacent markets that have a common customer base (i.e., envelopment). A firm dominant in its market enters a data-rich secondary market and engages in predatory pricing and privacy-policy tying. We define the latter as conditioning service provision to the subscription of a privacy-policy that allows bundling of user data across all sources. Acquiring data from the secondary market confers an advantage in the data-intensive primary market that shields the dominant firm from entry, thus harming consumers. We discuss potential remedies, including data unbundling and portability.

Keywords: Entry-deterrence, Predatory pricing, Platform Envelopment, Data, Privacy-policy Tying

*We thank Heski Bar-Isaac, Florian Ederer, Chiara Fumagalli, Alkis Georgiadis-Harris, Francesco Nava, Salvatore Piccolo, several anonymous referees and seminar audiences at the Toulouse Digital Economics Conference, the CCP Annual Conference and at the UK Competition and Market Authority for many valuable comments. We are grateful to Tanay Venkata and Cindy Kim for research assistance.

[†]Department of Economics, University of Warwick, UK. Email: d.condorelli@gmail.com

[‡]Compass Lexecon, Spain. Email: JPadilla@compasslexecon.com

1 Introduction

We present a theory of entry deterrence in digital markets. A *dominant* firm preemptively enters a *data-rich secondary market* and acquires customers by offering products at below-cost prices. This strategy is intended to build a data advantage that deters entry in its monopolized *data-intensive primary market*.[1] The bundling of user data enhances profits in the primary market and, consequently, converts firms operating in the secondary market into potential entrants. Hence, gaining control of data from the secondary market entrenches the dominant firm's position in the primary market and protects it from entry.

This foreclosure motive represents an incentive to ecosystem building by big tech companies, especially those that monetise user attention through advertising. As opposed to bundling of complementary products (e.g., Whinston (1990) and Carlton and Waldman (2002)), our theory relies on tying of data collected in unrelated markets to motivate entry by dominant platforms into adjacent markets, a strategy coined as *platform envelopment* by Eisenmann et al. (2011).[2]

The objective of *combining data across markets* may have been a leading force behind a number of acquisitions conducted in recent years, primarily by Google and Facebook (now Meta). These two firms dominate the USD 500 billion a year digital advertising market, and depend heavily on data to enhance their services and advertising delivery. We highlight three landmark cases.

- In 2005, Google acquired Android, a company developing an operating system (OS) for mobile phones, for $50 million. Google launched its OS in 2007, licensing it for free to mobile phone manufacturers. The acquisition of a start-up was a natural path for Google to enter the (licensable) mobile OS market, during a time when Microsoft was a competitor in this sector. Over time, Google secured control of the mobile OS market, becoming able to track users' locations, which is crucial information for delivering search results and advertising. Arguably, the acquisition of Android curtailed the growth of Bing, an entrant search engine by Microsoft that also had the prospect of gaining access to location data.
- In 2008 Google acquired DoubleClick, a company providing ad-serving services to websites, for $3.2 billion. The transaction was cleared by the European Commission (EC) and the Federal Trade Commission (FTC), on the grounds that Google and DoubleClick were not competitors. However, as many commentators noted, acquiring the leading ad-serving company allowed Google to track user browsing activities outside of its walled garden. This, in turn, improved Google's search-advertising capabilities and consolidated its position.[3]

[1] A data-rich business is one that enables the harvesting of extensive datasets on user behavior, while a data-intensive business centers on the exploitation of such data. These definitions are, of course, not mutually exclusive.

[2] This motive complements those highlighted in the literature on conglomeration, such as complementarity and economies of scope. See Bourreau and de Streel (2019) for a recent survey that focuses on digital conglomerates and Condorelli and Padilla (2020) for a discussion of envelopment strategies by digital platforms.

[3] For instance, in a 2007 US Senate hearing about the merger (J-110-25), the following statement was made: "Google collects an enormous amount of information on computer users' search history and Internet preferences. DoubleClick also collects a vast amount of information regarding consumers' Internet preferences. While DoubleClick assures us today that this information is shared with no one other than the advertiser or the website carrying the advertising, what will happen to this treasure trove of consumer data once Google gains control of DoubleClick?".

- In 2021 Google acquired Fitbit for \$2.1 billion. Fitbit produces wearable devices that track a number of health indicators of its users. Complaints were raised about the possibility of Google using Fitbit's data to improve its targeted search advertising. The EC and FTC ultimately cleared the deal, but Google committed not to use Fitbit's data for advertising.

Empirical evidence confirming our hypothesis that many big-tech acquisitions were data-driven has been published by Affeldt and Kesler (2021). By web-scraping the Google Play Store, they examined 400 acquisitions carried out in the last ten years by big tech companies. They show that about half of the acquired applications have been discontinued. Regarding those that are continued, which tend to be those most privacy-intrusive, they observe that "the monetization strategy seems to change as apps become free of charge but request more privacy-sensitive permissions."

In practice, bundling of user data is often achieved by means of *privacy-policy tying*. Privacy policies are legal documents describing how user data is handled and shared by companies that collect them while providing their services. In most cases, they are embedded in terms of service, so that when users accept the latter, they have also accepted the former. Typically, these policies allow firms to collect and combine user data from their various, often unrelated, services as well as multiple third-party sources. To give an idea, Facebook collects information from "Facebook (including the Facebook mobile app and in-app browser), Messenger, Instagram (including apps like Direct and Boomerang), Portal-branded devices, Bonfire, Facebook Mentions, Spark AR Studio, Audience Network, NPE Team apps and any other features, apps, technologies, software, products, or services offered by Facebook Inc. or Facebook Ireland Limited under our Data Policy. The Facebook Products also include Facebook Business Tools, which are tools used by website owners and publishers, app developers, business partners (including advertisers) and their customers to support business services and exchange information with Facebook, such as social plug-ins (like the "Like" or "Share" button) and our SDKs and APIs."[4]

Our theory hinges on two key drivers. The first is the hypothesis that controlling data harvested in a secondary market increases the profitability of serving a primary market. There are several ways in which improved decision making as a result of more data may increase profits: by enabling the creation of better products, thus increasing users' willingness to pay; by reducing the cost of service provision; or by lessening information asymmetries in a way that increases profit (e.g., via price discrimination). In our model, this *data advantage*, proportional to the customers served in the secondary market, turns incumbents in this market into threats for the dominant firm in the primary one. Crucially for our conclusion, entry in the primary market must be feasible *if and only if* a data advantage in the secondary market is accrued by the entrant and not by the monopolist.

[4]See Facebook's Data Policy, dated April 19, 2018, available at `https://facebook/policy.php` (Downloaded on 27 March 2020). As another example, Google collect information about (a) the apps, browsers, and devices used to access Google services; (b) users' activity in Google services, including search terms, watched videos, views and interactions with content and ads, voice and audio information, purchase activity, activity on third-party sites and apps that use our services, Chrome browsing history, phone number, calling-party number, receiving-party number, forwarding numbers, time and date of calls and messages, duration of calls, routing information, and types of calls; (c) location, etc. See Google's Privacy and Terms, dated October 15, 2019, available at `https://policies.google.com/privacy?hl=en-US#infocollect` (Downloaded on 27 March 2020).

Our leading example is that of an advertising platform which monetises user attention. For instance, when suggesting purchases to users, a search engine benefits greatly from having data on the brick-and-mortar shops they have just recently visited, which may be collected by a different application that requires granular location data to function (e.g., as hinted, a mobile phone operating system). Likewise, a social network may expand faster if it knows the identities of its users' friends, which may be obtained via a messaging application where users store their contacts. Such data, of course, would also benefit greatly new entrants in search or social networking, to the point that their availability might be essential for successful entry. The diagram below illustrates this environment.

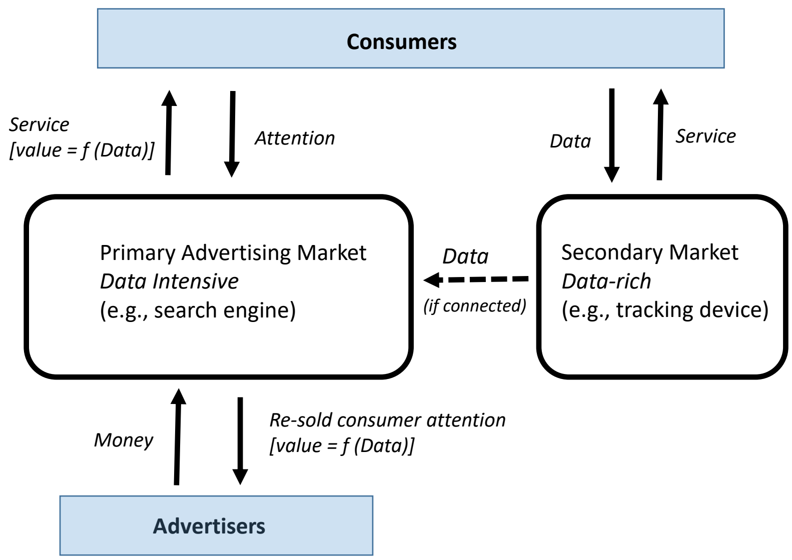


Figure 1: Diagram of the Environment

Beyond casual observation, a revealed preference argument suggests that accumulating data about consumers increases advertising returns, given the substantial effort such advertising platforms go through in order to acquire, compile, and store such data. In fact, it is reported that immediately after Apple made it easier for users to opt out of data sharing across mobile apps, major advertising platforms lost more than \$140 billion in collective valuation.[5] The same conclusion is reached by more in-depth industry studies. For instance, in its 2020 "Online platforms and digital advertising market study" (Appendix F), the CMA argues that "Data gives platforms a competitive advantage in the provision of digital advertising. [...] For these purposes, detailed data on consumers' demographics, interests, preferences and behaviours is most valuable in terms of profiling consumers, predicting consumers' potential response to advertising and tailoring advertising messages."[6]

[5] "Apple's ad changes wiped \$142 billion off snap, facebook, and other online ad giants" (Forbes, http://bitly.ws/EGgs).

[6] See also Wernerfelt et al. (2022), who estimate the cost for an advertiser of a acquiring a customer on Meta raises by almost 40% when the use of data collected by Meta outside its platform is restricted. For a more cautious view on the value of user data for advertising see Lambrecht and Tucker (2017).

The second driver is an initial exogenous asymmetry, as the dominant firm in the primary market is endowed with a *first-mover advantage* over potential entrants in its own primary market. This more traditional advantage, which is a common hypothesis in theories of strategic entry deterrence, may be conferred by dominance in the primary market, sheer size, or availability of resources. For instance, leaked emails indicate that Mark Zuckerberg, CEO of Facebook, was extremely concerned about the growth of Tinder, a dating app which exploited Facebook data in addition to collecting its own data. Facebook subsequently entered the dating space with its own service, at a point in time when Tinder did not have the know-how to expand into a full-fledged social network, an attempt it made later with Tinder Social.[7]

Then, our story goes, by preemptively entering the secondary market and conquering it through below-cost prices (e.g., by offering a free product in line with evidence from Affeldt and Kesler (2021)), the dominant firm may be able to deter entry in the primary market, even from *more efficient* firms already operating in the secondary market. Although the dominant firm sustains a loss to undercut other firms in the secondary market, the strategy is likely to be profitable if it deters entry in the primary. In fact, the dominant firm is willing to outbid the secondary market's incumbent because it stands to lose monopolistic profits if entry in the primary market takes place. On the other hand, a firm's incentive to retain control of the secondary market and then enter the primary market is diluted by the expectation of competitive prices following entry in the latter. The positive wedge between the value of retaining a monopoly and the value of entering, also known in the literature as *efficiency effect*, makes the exclusionary strategy profitable for the monopolist.[8]

Our theory echoes growing concerns that data accumulation and combination by big tech companies may erect barriers to entry into digital markets that harm consumers.[9] Indeed, our results also illustrate that if *additional* competition in the primary market is substantially more valuable to consumers than it is in the secondary market, then the exclusionary conduct we describe harms not only competitors but also consumers, compared to the counter-factual scenario in which entry in the secondary market is blocked. As an example, consider the DoubleClick and Fitbit deals mentioned above. Conceivably, consumers are worse-off from a monopoly in web-search than they are from a monopoly in the intermediation of display advertising or in the market for tracking bracelets. (However, note that this observation could be reversed if the primary market exhibited, say, strong network effects that made it invariably prone to tipping to monopoly even in the case of entry.)

[7]Zuckerberg wrote "Tinder's growth is especially alarming to me because their product is built completely on Facebook data, and it's much better than anything we've built for recommendations using the same corpus." (Forbes reporting on leaked emails, see `https://bit.ly/2TVr5X8`)

[8]Fudenberg and Tirole (2013) first used the term "efficiency effect" referring to a monopolist as being more *efficient in extracting profits*, assuming all firms have the same technology. This effect, typically assumed positive in light of the argument just offered, is key in models of innovation attempting to explain the persistence of monopoly. In contrast to the "Arrow effect", it suggests that a monopolist's incentive to innovate may exceed that of potential competitors, if successful innovation triggers entry. See footnote 19 for cases when the efficiency effect may be negative.

[9]This has been emphasized in a number of recent reports, including the UK Furman Report "Unlocking Digital Competition" (see 1.71), Crémer et al. (2019) (see 5.IV) prepared for the European Commission and the Final report of the Stigler Committee on Digital Platforms (see II.a.ii.1).

In light of the above, if pricing in the secondary market is below-cost (i.e., entry can't be justified by commercial reasons), our results can be used to construct an antitrust *theory of harm* against privacy-policy tying (i.e., a potential abuse of dominance) or against a data-driven merger of firms operating in apparently unrelated markets. Steps in this direction have been taken by the CMA and the EC. In 2022, a deal between Facebook and Kustomer, a customer relationship management company, received the green light but the CMA investigated the possibility that "the Merger would raise barriers to entry and expansion by increasing Facebook's data advantage in online display advertising, leading to reduced competition in that market".[10] In its Google/Fitbit decision, the European Commission stated that "although pre-Transaction Fitbit is not competing in the same markets as Google, "the transaction would give Google control over an important asset, the Fitbit data, that would further strengthen Google's dominance in the markets for the supply of online search advertising".[11]

Within the scope of our model, we also perform some simple policy analyses, by comparing the exclusionary equilibrium outcome with the outcomes that would arise from various policies under the same assumptions. Primarily, we consider imposing a *prohibition to tying of the privacy policy* and *enforcing portability of user data*. Our results highlight that, in order to evaluate the effect of such policies, a key element to assess is whether a data advantage accrued by serving some secondary market is essential for entry in the primary market or not. We say that a *data advantage is essential for competition* in the primary market if entry does not take place *unless* the new entrant holds the secondary-market data in an exclusive fashion. As we elaborate later on, we believe that having access to user location data might be essential for competition in the search market. That is, having had exclusive access to tracking the location of consumers might have allowed profitable entry into the search market, but only at a time when Google did not have such access as well, which it gained through Android.

When data is *not* essential for competition, both a prohibition to tying of the privacy policy and enforcing portability of user data are effective in preventing the dominant platform in the primary market from implementing its exclusionary strategy. However, we show that when data *is* essential for competition these two policies may backfire if imposed also on potential entrants, because they make it impossible for such entrants to gain exclusive access to data. In particular, a generalized prohibition to privacy-policy tying makes the entrant unable to bundle data, while portability implies that users can migrate their data to the dominant operator if they wish so after the new entrant has made the investment to enter. In both cases the entrant in the primary market is left with no advantage over the incumbent and loses incentives to enter. The end result is that consumers will be worse off than under the exclusionary equilibrium, as they will be left with a monopoly in both the primary and secondary markets. This negative effect might be amplified when combining data actually benefits consumers by reducing production costs, for instance because it improves product quality.

[10] See "Theory of Harm 1", CMA's decision "Anticipated acquisition by Facebook, Inc. of Kustomer, Inc." ME/6920/20.

[11] See Commission Decision of 17.12.2020 on Case M.9660 – GOOGLE/FITBIT, page 98.

The discussion above indicates that imposing such restrictions *only* on the primary market incumbent is preferable. Even more so when data combination is likely to have a negative effects on consumers, for example as a result of enabling price-discrimination. While traditionally regulation has focused on creating a level playing field, the emergence of powerful big tech has oriented regulators worldwide toward imposing more asymmetric provisions. Indeed, data unbundling and portability requirements are now part of the EU Digital Markets Act (DMA), which imposes a set of obligations only on larger firms operating core platform services, who are identified as gatekeepers.

Finally, we also look at the possibility of allowing firms to trade data they acquired. As one would expect, free trading of data between firms eliminates some of the strategic externalities that arise from competition. Ultimately, allowing data trading will either have no effect, or otherwise it will allow the dominant firm in the primary market to buy data from the potential entrant, thus preventing entry into the primary market without the need to compete in the secondary. When the latter happens, consumers are worse off than in the exclusionary equilibrium due to the persistence of a double monopoly. Similar effects on consumers arise when the purchase of data is accomplished via a merger or an acquisition, a topic we touch upon in the conclusions.

The rest of paper is structured as follows. Next, we discuss the literature. In sections 3 and 4 we formalize our theory. Section 5 is dedicated to policy analysis. Section 6 contains some conclusive remarks.

2 Related literature

Our work contributes primarily to the literature on strategic entry deterrence, which we informally describe as strategic behaviour that is expected to be profitable only if it prevents entry.

The literature is quite vast but can be broadly subdivided into two strands. One strand emphasizes asymmetric information and, in particular, how an incumbent can persuade potential entrants that competition will be tough (e.g., see Kreps and Wilson (1982) and Milgrom and Roberts (1982)). Another, to which our paper is closely connected, focuses on the ability of a dominant firm to neutralize incentives to enter by gaining an advantage (or denying it to others) that reduces the profit potential entrants may expect. Following the Fudenberg and Tirole (1984)'s taxonomy, in our paper the monopolist attempts to turn itself into a "top-dog" to deter entry. The early literature has focused on investment in capacity (e.g., Spence (1977)), learning by doing (e.g., Williamson (1979), Spence (1981), and Cabral and Riordan (1994)) and R&D (e.g., Gilbert and Newbery (1982)).[12] Some more recent work emphasised exclusive dealing as a way to deny scale to competitors (e.g., Rasmusen et al. (1991), Bernheim and Whinston (1998), Segal and Whinston (2000) and Fumagalli and Motta (2006)). Efficiency-reducing exclusion is possible in our model because, as in these papers, buyers in the secondary market do not internalize the losses suffered by those in the primary.

[12] In Gilbert and Newbery (1982) we see the efficiency effect appearing for the first time in the literature. That is, preemptively investing in *R&D* is potentially able to deter entry in a market that requires an innovation, because the monopolist stands to lose a monopoly while the new entrant only expects competitive profits.

A seminal contribution to this second strand of the entry deterrence literature is Carlton and Waldman (2002) (CW). They consider a model where some firm is initially dominant in two markets, but a competitor can enter the secondary market in period one and both markets in period two only. For the case in which the products in the secondary and primary market are perfect complements, they show how tying can reduce incentives to enter. In particular, by committing to tying two products that have no value unless consumed together, the dominant firm forecloses the secondary market in the first period and forces the potential entrant to either enter both markets in period two or none at all. Then, in a version of their theory, entry is deterred because operating in one period only is not sufficient to cover entry costs. In another, in which the bundled products enjoy network externalities, foreclosure operates by preventing competitors to acquire critical mass in the secondary market in period one, which makes it harder to compete in the primary market, regardless of entry cost. Either way, in CW the focus is on strategic tying of complementary products, while we analyse data pooling in unrelated markets, which does not require tying of products.[13]

Within the research spurred by CW, the most closely related paper is Fumagalli and Motta (2013) (FM). They formalize the idea that an incumbency advantage, accrued in serving incremental customers, may allow a firm with market power to deter entry by a firm who would be more efficient at scale. Repackaged to fit our context, their main assumption is that a dominant firm has an efficiency advantage in serving either the primary or the secondary market in isolation, but the potential entrant would be more efficient in serving both. Then, if the dominant firm engages in aggressive pricing in the secondary market before competing in the primary, it could deny scope to other firms. In contrast to FM, our two firms might be equally efficient in serving both markets, but it's the first-mover advantage that creates an asymmetry.

There are important analogies between CW, FM and our paper. In particular, the general strategic force that shapes behaviour is fundamentally equivalent. That is, the dominant firm is willing to temporarily sustain losses, either by reducing its profit in the primary market (as in CW) or by pricing below cost to consumers in the secondary market (as in FM and our paper), in order to later on prevent effective competition in the primary market.[14] However, as we explained, the specific incentives are different and our model is not a special case of either. Furthermore, policy analysis hinges on the properties of data, as opposed to alternate explanations grounded on network externalities, learning-by-doing, or generic shared inputs that produce economies of scale and scope. Notably, data can be easily transferred from one firm to another, can be subject to portability requirements and it is technically straightforward to enforce a mandate preventing two divisions of a firm from sharing data. Finally, in contrast to classic economies of scale or scope (see Panzar and Willig (1981)), consumers need not always benefit from firms combining data. Because data are an input that can improve decision making, data can be extractive. That is, by combining data, firms may not expand their technological frontier but, instead, may change their relationship with consumers in a way that disadvantages the latter, such as by price discrimination.

[13] A related argument is developed by Choi and Stefanadis (2001), where firms need to engage in successful R&D to enter a market. By tying products, the incumbent makes it riskier (and costlier) for a competitor to attempt entry.

[14] See also Fumagalli and Motta (2020), where potentially unprofitable vertical foreclosure serves to a barrier to entry in the upstream market.

The consequences of data generating advantages in unrelated markets is also explored in Prufer and Schottmuller (2022), Farboodi et al. (2019) Hagiu and Wright (2020), de Cornière and Taylor (2020) and Chen et al. (2022). The first three papers focus on long run competitive dynamics taking place in a single market and generated by data-enabled learning (i.e., firms that improve their products by dynamically learning about customers). In contrast, we focus on pooling of information from unrelated markets and its potential foreclosing effects. Moreover, the decision to acquire data is endogenous in our paper. Like us, de Cornière and Taylor (2020) model data as producing a shift in the per-consumer revenue obtained by firms. They define them as "unilaterally pro-competitive" (anti-competitive) if more data induce firms to offer a higher (lower) level of utility to consumers, *ceteris paribus*. In contrast to us, they show how equilibrium utility of consumers is affected in mergers, depending on the nature of data. Finally, Chen et al. (2022) share with us the presence of two markets with data-generated spillovers, but their focus is on welfare effects of mergers between firms that are already competing (à la Hotelling) in both markets they operate in.

3 Model: A two-period data-driven preemption game

There are two profit-maximizing firms A and B. At the outset, both are monopolists in their respective markets, the primary market, P, and the secondary, S. Firms A and B play the following two-period entry-game. In the first period, A decides whether to enter S. After the entry decision, competition (or lack of it) determines outcomes in both markets. In the second period, B decides whether to enter P. Then, as in the previous period, outcomes in both markets are determined. At the end of second period, the game ends. We assume that there are non-negative fixed costs of entry in market P and S, namely f_P and f_S.[15]

To facilitate application of the analysis to a more realistic multiple-period environment, we rescale *fixed costs* and what we *interpret* as one-period-only profits, that is *first-period profits in P* and *second-period profits in S*, by a factor of $1-\beta$, with $\beta \in [0,1]$ representing a common discount factor. As β grows, costs and profits we deem sustained for one period only become less relevant. The reader interested in the long-term horizon should assume $\beta = 1$, while the reader interested in the two-period game should assume $\beta = 0$.

Both the *first-mover advantage* of A and its *market power* are essential assumptions. However, they are natural hypotheses in the set of situations we aim to model. Online advertising markets are currently led by two mature platforms, Facebook whose primary market is social-networking and Google whose primary market is search. These big conglomerates continuously venture in adjacent markets, which are populated by firms often incapable, in their startup phase, of challenging them in their search and social networking primary markets. However, our analysis can be extended to the scenarios involving competition in S and the restriction to two periods is for simplicity.

To close the model, we describe next how competition operates in the two markets in both periods. An important feature of our setup is that profits in P in the second period depend *both* on

[15]In contrast to FM, as long as $f_P \leq f_S$ (i.e., entry into the primary market is less costly than entry into the secondary one), the monopolist need not enjoy any specific asymmetric cost advantage.

the entry decision of B *and* on the outcome of competition in S in the first period. In particular, the profit of a firm in market P in the second period is larger if such a firm holds a *data advantage*, which is conferred by serving consumers, the more the better, in S in the first period.

Secondary Market. We assume it is populated by a unit-mass of consumers with a constant value, v_S, for a product which both firms produce identically. When two firms are in the market, they engage in Bertrand price-competition. Remarkably, in our environment, a Bertrand price-offer could also be negative, thus representing a positive transfer to consumers, which may be also interpreted as taking the form of additional product quality or features. We focus on the case where firms are *not* capacity-constrained and have heterogeneous but constant marginal cost, c_S^i for $i \in \{A, B\}$, with $0 < c_S^i < v_S$. These assumptions impose useful structure to our analysis but are mainly made for simplicity of exposition. In particular, due to Bertrand competition, the market will be wholly served by the single firm capable of providing the best deal to consumers. Hence, barring ties and policy interventions that will be discussed later, we can model the advantage provided by data in the primary market as being either held by A or by B.

Crucially, in the secondary market consumers buy from the firm who offers the lowest price (or highest transfer). This assumption requires elaboration. In fact, while it is clear that atomistic consumers will not internalize their collective impact on market structure, it may be argued that they may sometimes internalize the effect that buying from one firm or the other will have on the deal they will receive in the primary market and on their privacy. We suggest several ways to rationalize our assumption in our setup. First and of easiest interpretation, the primary and secondary markets might serve two different consumer pools, but learning about consumers in the secondary market is informative about consumers in the primary. For example, consider two geographically separated markets with similar populations of consumers. Second, the same atomistic consumers might patronize both markets but any of the following applies. One, even if consumers refuse to buy in the secondary market, with the aim of preventing the firm in the primary market to combine their data, the relevant information they would like to protect could still be inferred using the data collected from close-enough consumers.[16] Two, consumers might be myopic and not able to recognize the cross-market externality implied by the tying of the privacy policies.[17]

If B remains a *monopolist* in the secondary market, it extracts all value from consumers and we can denote its profit, $v_S - c_S^B$, with π_S^m. However, since the outcome in S determines who holds the data advantage in the future, if entry takes place, profits from competition in the secondary market are endogenous and do not *only* depend on production cost. This is a central aspect of the equilibrium analysis in the next section. Nonetheless, to simplify the interpretation of our main

[16] A basic externality problem arises because data acquired from a set of consumers can be used to gain information about other consumers with similar characteristics. For instance, the location of a person might be identified from the location of its closest contacts. This has been discussed in the legal literature (see Fairfield and Engel (2015)) and in economics (see Acemoglu et al. (2019), Bergemann et al. (2019) and Choi et al. (2019)).

[17] For example, consumers might not be aware that Facebook owns Instagram or that Google owns Android. There is ample evidence that users pay little or no attention to privacy policies and are unlikely to internalize externalities arising from the sharing of personal data. See Economides and Lianos (2019) for a perspective at the intersection of law and economics.

result, let's denote with Π_S^d and π_S^d the profit that firm A and B, respectively, would obtain if one-shot *duopolistic* Bertrand competition took place with firms having marginal costs c_S^A and c_S^B. That is, $\Pi_S^d = \max\{0, c_S^B - c_S^A\}$ and $\pi_S^d = \max\{0, c_S^A - c_S^B\}$.

Primary Market. Since behaviour in the primary market does not spill-over to other markets, we can treat outcomes in reduced form. This simplifies modelling and, arguably, makes the analysis more applicable. In particular, we denote with Π_P^D and π_P^d the competitive profits of A and B in the primary market when A is the only firm with a data advantage. Conversely, we denote with Π_P^d and π_P^D the competitive profits of A and B in the primary market when B is the only firm with a data advantage. We indicate with Π_P^M and Π_P^m the monopoly profit of A with and without the advantage. Naturally, we maintain that $\Pi_P^M > \Pi_P^m$, $\Pi_P^D > \Pi_P^d$ and $\pi_P^D > \pi_P^d$. That is, obtaining data improves the profitability of a monopolist and provides an advantage in case of competition.

Bertrand-like scenario While, at the purpose of the next section, there is no loss in treating outcomes in the primary market in reduced form, it is helpful to derive sharper conclusions to specialize the model to the case where competition also takes place a' la Bertrand in the primary market. To this end, we can specify total value generated in the primary market as v_P, the marginal cost of the two firms as c_P^A and c_P^B, and let the data advantage take the form of a shifter of costs. In particular, we can assume a firm that has data has its cost reduced by α, with $\alpha > |c_P^A - c_P^B|$. In this setup, we can easily compute $\Pi_P^D = \alpha + c_P^B - c_P^A > 0 = \Pi_P^d$, $\pi_P^D = \alpha + c_P^A - c_P^B > 0 = \pi_P^d$ and $\Pi_P^M = v_P + \alpha - c_P^A > v_P - c_P^A = \Pi_P^m$, consistently with the reduced-form version of the model.

4 Exclusionary Entry

The characterization of the sub-game perfect equilibria of the entry-game, which encapsulates the central insight of the paper, forms the core of this section. The building of a data advantage, implemented through entry into a data-rich secondary market followed by predatory pricing *and* privacy-policy tying, may increase the strength of the incumbent and give rise to exclusion of potentially more efficient competitors in the primary data-intensive market. Two conditions are required for this result. First, that a data advantage is pivotal for the entry of B in the primary market, see (PDA) below. Second, that the efficiency-effect is large enough to make the exclusionary strategy ex-ante rational (EE-IR) and incentive compatible upon entry (EE-IC). The proposition is followed by a short heuristic explanation. A formal proof is consigned to the appendix.

Proposition 1 *In any subgame-perfect equilibrium the following holds.*

[a.] *If* $\pi_P^D < (1-\beta)f_P$, *then* B *stays out of* P*;*

If $\pi_P^d > (1-\beta)f_P$, *then* B *enters* P.

[b] *Assume instead that the data advantage is pivotal for entry, that is,*

$$\underbrace{\pi_P^D > (1-\beta)f_P > \pi_P^d.}_{\textit{Pivotality of data advantage}} \tag{PDA}$$

[b.1.] *If, in addition, the efficiency effect is large enough so that the exclusionary strategy is ex-ante rational, that is,*

$$\underbrace{\Pi_P^M - \Pi_P^d - \pi_P^D}_{\text{efficiency effect}} > \left(\pi_S^d - (2-\beta)\Pi_S^d\right) - (1-\beta)(f_P - f_S) \tag{$EE-IR$}$$

and incentive compatible upon entry, that is

$$\underbrace{\Pi_P^M - \Pi_P^d - \pi_P^D}_{\text{efficiency effect}} > \left(\pi_S^d - \Pi_S^d\right) - (1-\beta)f_P, \tag{$EE-IC$}$$

then A enters S and obtains the data advantage, while B stays out of P;

[b.2.] *If either ex-ante rationality $(EE-IR)$ or incentive compatibility $(EE-IC)$ of the exclusionary strategy does not hold, then B enters P.*

Part [a.] of the proposition identifies scenarios in which a data advantage is *not* pivotal for entry, in the sense that B has a dominant strategy of either entering or staying out. As the behavior of A has no bearing on the entry decision of B, we can say that, unless (PDA) holds, A lacks the *ability* to exclude B. Instead, exclusionary behavior can arise when data is *pivotal for entry*, that is when the (PDA) condition holds. In this case, A has the ability to avoid entry of B by denying it the data advantage. Then, conditions (EE-IR) and (EE-IC) further identify when the *efficiency effect* is large enough that such a strategy of entering the secondary market *and* securing the data advantage, thus preventing entry of B in the primary market, is *individually rational* for A and, indeed, *incentive compatible upon entry.* On the other hand, if, under (PDA), conditions (EE-IR) or (EE-IC) do not hold, then B will maintain the data advantage and enter the primary market.

To parse case [b.1.] let's start by considering the outcome of Bertrand competition in the secondary market upon entry of A. Due to pivotality of the data advantage, B will enter the primary market if A stays out of the secondary *or* if A enters but does not win competition. Then, as standard in the analysis of Bertrand competition, we begin by identifying the price that makes both firms *indifferent* between serving the entire market *and* letting the competitor serve it. That is, the cost of production *minus* the difference in future payoff from owning the data versus leaving the data in the hands of the competitor. This is equal to $c_S^A - (\Pi_P^M - \Pi_P^d)$ for A because A stands to lose the difference between monopoly and competitive profits in the primary market; and it is equal to $c_S^B - (\pi_P^D - (1-\beta)f_P)$ for B, because B stands to lose the competitive profits in the primary market with the data advantage net of the cost of entry. Now, we can determine the winner of Bertrand competition by looking at the firm for which the minimum acceptable price is lowest. Crucially, if $(EE-IC)$ holds, then A wins Bertrand competition in the secondary market.[18] Our backward-induction argument is concluded by showing that (EE-IR) implies A will want to enter the secondary market in the first period. We refer the reader to the proof for details.

[18] A's minimum acceptable price in S is lower than that of B if $c_S^A - (\Pi_P^M - \Pi_P^d) < c_S^B - (\pi_P^D - (1-\beta)f_P)$. After observing that $c_S^A - c_S^B = \max\{c_S^A - c_S^B, 0\} + \min\{c_S^A - c_S^B, 0\} = \max\{c_S^A - c_S^B, 0\} - \max\{c_S^B - c_S^A, 0\} = \pi_S^d - \Pi_S^d$, we conclude that, upon entry, A wins competition in S and forecloses B in P if $(EE-IC)$ holds.

From a practical standpoint, verifying pivotality of the data advantage (PDA) may be complex. The chain of inequalities can be interpreted as requiring the existence of a wedge between the profit a firm can attain when it competes with the data advantage, versus the profit it attains when it competes with an opponent that has the advantage. Such a wedge must be large enough to bound fixed costs in the long-run. However, we expect the individual rationality (EE-IR) and incentive compatibility (EE-IC) assumptions to be widely satisfied in the set of markets we are studying. In particular, while the right-hand side of both inequalities is likely to be small if not negative, because competitive profits of the two firms in the secondary market are likely to be similar and cost of entry in the primary market larger than cost of entry in the second, the *efficiency effect* on the left-hand side is likely to be positive and substantial in many cases of interest. Intuitively, if all firms have access to a similar technology, a monopolist platform is able to generate substantially more profit than competing firms sharing the market.[19] For instance, if the time horizon is very long ($\beta = 1$), firms compete intensely in the secondary market ($\Pi_S^d \geq \pi_S^d$) and entry cost in the primary market is larger than that in the secondary ($f_P \geq f_S$), then it is sufficient that the efficiency effect is positive for the exclusionary equilibrium to take place under pivotal data.

We stress that the exclusionary equilibrium we exhibited would survive, with only slight modification to (EE-IC) and (EE-IR), in the presence of privacy-concerned consumers that patronise both markets. That is, suppose consumers incur small but relevant privacy loss, expressed in monetary units, from being served by the same firm in both markets. Then, the indifference condition would imply that there is an equilibrium where A wins over B if its minimum acceptable price, minus the privacy loss, is above the minimum acceptable price of B. Note that, because A wins over B, the expectation of consumers of incurring the privacy loss in the primary market would be justified. In light of this extra transfer to consumers, also the cost of implementing the exclusionary strategy would be raised by an amount equal to the privacy loss. In sum, exclusion becomes more difficult with privacy-concerned consumers but it is nonetheless possible.

Now suppose all the conditions for exclusion from Proposition 1 holds. As we established, in equilibrium B will not enter the primary market even though it would have done so if A stayed away from the secondary market. However, even if this is the outcome, we can't conclude that the actions of A were led uniquely by exclusionary objectives. In particular, A might have wished to enter *regardless* of whether its entry produces exclusion or not. We will therefore say that behavior is *explainable by non-exclusionary concerns* if *both* entry *and* acquisition of the data advantage would take place whether or not B entered the primary market.

Remark 1 *Suppose the conditions for exclusion are present, that is, data are pivotal for entry* (PDA) *and the exclusionary strategy is both incentive compatible* ($EE-IC$) *and ex-ante rational* ($EE-IR$)*. Behavior of A is explainable by non-exclusionary concerns if, and only if,*

$$\Pi_S^d = \max\{c_S^B - c_S^A, 0\} \geq \frac{1-\beta}{2-\beta} f_S. \qquad (ENC)$$

[19] The efficiency effect is negative if the industry structure that maximizes total profit is not a monopoly. For instance, this might be the case if consumer adoption is driven by expectations (e.g., due to network effects under interconnectedness) which in turn can be influenced by industry structure (e.g., see Condorelli and Padilla (2021)).

In words, as long as, A is more efficient than B in serving the secondary market and such cost advantage, which determines A's profit, is large enough to justify entry, then A would enter regardless of the impact of its actions on B. Remarkably, behavior explainable by non-exclusionary concerns corresponds to the case where incentive to undercut B in the secondary market following entry implies that entry will ultimately yield a positive return. That is, (EE-IC) implies (EE-IR).

In a logic analogous to that we use to separate predatory behavior from competitive pricing, in markets that are connected by data economies of scope, entry of A in a secondary market cannot be attributed to the desire of protecting monopoly power, if A expects to make sufficient profits in the secondary market. In the parlance of competition law, foreclosure would not be anti-competitive when (ENC) holds. Empirically, in order to corroborate a theory of harm based on data-driven envelopment one would have to conduct a test based on production costs in the secondary market.

The cases that are most likely to represent intentional exclusionary behavior are those where B is more efficient than A in serving the secondary market. This raises two questions: Can the exclusionary equilibrium arise if B is more efficient than A also in the primary market? If that is the case, can the exclusionary equilibrium also generate too little entry in the primary market from a total welfare perspective? Both questions can be answered in the affirmative.

Remark 2 *The conditions for an exclusionary equilibrium,* (PDA)*,* $(EE-IC)$ *and* $(EE-IR)$*, can arise even in cases where B is more cost-efficient than A in both markets and the efficiency gain is such that it would be socially optimal for B to enter the primary market.*

Not only a more efficient competitor may be deterred by the conduct of A, but also the resulting outcome may be welfare reducing. This is explainable by a cross-market externality problem. By providing excessive rents to consumers in the secondary market, A is able to deter entry of B at the expenses of consumers in the primary market, who suffer from continued monopoly. Such externality is made possible by privacy-policy tying, which links the two markets, and by the atomistic nature of consumers. We demonstrate Remark 2 in Appendix, focusing on the Bertrand-like scenario described at the end of Section 3.

5 Policy Analysis

A central theme of this paper is that the building of a data advantage by a firm with market power, achieved through privacy-policy tying, may have exclusionary effects that harm consumers. In order to evaluate such harm, we shall compare consumer surplus in the exclusionary equilibrium with that arising in alternative outcomes that could be implemented through appropriate policies.

In this section we evaluate the effect of several policies, mainly under the three assumptions (PDA), (EE-IC) and (EE-IR) that produce the exclusionary equilibrium. First, we consider an outright restriction on entry by A, which serves as our main counter-factual scenario to evaluate harm to consumers from the exclusionary equilibrium. Then, we consider two policies that have received ample attention by policy-makers, that is *prohibition to data tying* and *mandating portability of data.* Finally, we consider an environment where firms are allowed *free trade of data.*

Before proceeding, we must introduce notation for consumer surplus and impose those restrictions that are consistent with the model, economic theory and the effects of data on outcomes. Our modeling of the secondary market leaves us little flexibility. The welfare of consumers in the secondary market is zero if B remains monopolist, while it is $(v_S - p_1^*) + (1-\beta)W_S^d$ if A enters, where p_1^* is the winning price-offer made to consumers in the first period and $W_S^d = v_S - \max\{c_S^A, c_S^B\} > 0$ is the welfare from final-period competition. Instead, consumer surplus in the primary market is not directly implied by other the variables in our model and must be exogenously specified. To begin with, we normalize to zero consumer welfare under monopoly also in the primary market, when A has no data. Then, we let W_P^M indicate consumer welfare in the primary market when A is a monopolist and holds the data advantage. Finally, we denote with $W_P^{\mathscr{D}}$ consumer surplus in the primary market under competition when both firms have the data, with W_P^D the surplus when only one of the two firms has the data advantage, and with W_P^d the welfare when no firm has it.

While the welfare variables for the primary market are arbitrary, some restrictions can be motivated by economic theory. First, we assume $W_P^d > 0$ and that $W_P^D > W_P^M$, that is, consumer surplus under competition is greater than welfare under monopoly, keeping endowments of data constant. Second, following the principle that more symmetric firms will compete more fiercely, we assume $W_P^{\mathscr{D}} > W_P^D$.

Finally, we consider two alternative scenarios regarding the effect of data. Borrowing terminology from de Cornière and Taylor (2020), we say that the data advantage is *pro-competitive* if $W_P^M > 0$ and $(W_P^{\mathscr{D}} >)W_P^D > W_P^d$ while it is *anti-competitive* if, instead, $W_P^M < 0$ and $(W_P^D <)W_P^{\mathscr{D}} < W_P^d$. For example, data can be pro-competitive when it improves the efficacy of targeted advertising or allows to customise customer experience, allowing firms to produce higher quality products at a lower cost. But it can also be anticompetitive, for instance, if it enables extraction of additional consumer surplus by means of price discrimination. While we do not model it explicitly, anti-competitiveness may also arise out of privacy considerations.

5.1 Entry restrictions

Suppose A is forced to stay out of the target market. This represents our main counter-factual scenario to evaluate harm to consumers and social welfare from the exclusionary equilibrium in the case where (ENC) does not hold and, therefore, entry of A would not be justified on purely commercial grounds.

In this modified model, B will enter the primary market holding a data advantage. Compared to the exclusionary equilibrium, there will be competition in primary market, while competition in the secondary market will be absent. Consumers benefit from this policy if, and only if

$$W_P^D \geq W_P^M + (1-\beta)W_S^d + v_S - c_S^B + \pi_P^D - (1-\beta)f_P \equiv W^E. \tag{1}$$

The exclusionary entry-strategy harms consumers when the (stationary) increase in surplus due to competition in the primary market, $W_P^D - W_P^M$, is larger than that in the secondary, measured as $v_S - c_S^B + (1-\beta)W_S^d$, plus the transfer accrued as a result of the price-war, π_P^D. In turn, this is more likely to be true when the value generated in the primary market is high and profits from competition are small. For instance, in the case of intense Bertrand-like competition over a long time horizon, that is with $c_P^B = c_P^A$ and $\beta = 1$, we have $W_P^D = v_P - c_P^A$, $W_P^M = 0$, and the condition boils down to $v_P - c_P^A > v_S - c_S^B$. In summary, we make the following observation.

Remark 3 *Restricting entry into the secondary market by the primary market monopolist is more likely to benefit consumers when the value produced in the primary market is larger than the value produced in the secondary and competition is intense.*

5.2 Prohibition to data-tying

Suppose regulation prevents *all* firms from pooling consumer data across markets. This prohibition, which could be implemented by requiring the un-bundling of privacy policies, would, in effect, neutralize any data advantage.

To analyse this scenario it is necessary to introduce another piece of notation and a definition. We denote with $\pi_P^{\mathscr{D}}$ the profit of B when competing in the primary market if it does not have a data advantage over A, either because both firms have the advantage or neither of them has it. Then, we say that the *data advantage is essential for competing* in the primary market if

$$(1-\beta)f_P > \pi_P^{\mathscr{D}}. \tag{DEC}$$

That is, B would not enter the primary market unless it has an edge over A via the data advantage.

Whether a data advantage is essential for competition or not will depend on the specific application. However, the analysis of some business cases may shed some light on which data might be essential for which industry. For example, it is conceivable that the development of the Android operating system by Google was also intended to obtain access to user-location data, which were already available to Microsoft through its Windows operating system. Arguably, by linking location data with its wealth of information, Google was able to limit the growth of Bing, its main rival search-engine, owned by Microsoft. Therefore, we posit that a data-advantage in the form of location data was, and may still be now, essential for competition in the web-search market.

When data is essential for competition (i.e., (DEC) holds), a generalized prohibition of data pooling will backfire. There will be no entry in either market or, possibly, there will be entry only in the secondary market if entry of A is not guided by exclusionary concerns (i.e., (ENC) holds). This policy would therefore defy its purpose and induce consumer welfare of at most

$$(2-\beta)W_S^d,$$

which is bound to be *lower* than the consumer welfare generated by the exclusionary equilibrium (i.e., W^E) as it does not offer the additional value from the data-driven price-war.

Instead, if the data advantage is *not* essential for competition, then a prohibition to data tying will not impede entry of competitors in the primary market. However, it will prevent A from implementing its exclusionary strategy. Consumers will harness the benefit from increased competition in the primary market, although entry of A in the secondary market may not take place. On the one hand, if entry of A does not take place, consumers benefit from a prohibition to bundle data across markets if

$$W_P^d \geq W^E$$

which is akin to condition (1) from the previous sub-section, except in the left-hand side welfare from competition with the data advantage in the primary market is replaced with the welfare from competition with no data advantage. On the other hand, if entry in the secondary market takes place anyway because (ENC) holds, then consumers are more likely to benefit from a prohibition to data tying. In particular, they benefit from this policy if

$$W_P^d + (2-\beta)W_S^d \geq W^E = W_P^M + (1-\beta)W_S^d + v_S - c_S^B + \pi_P^D - (1-\beta)f_P.$$

We argue that this condition is likely to hold, especially if the two firms have similar production technologies and data is anti-competitive. First, we might conjecture $W_P^d - W_P^M > \pi_P^D$, because we expect the consumers' loss from monopolization of the primary market to be greater than the profits of any of the competitors under competition. Second, it will be often the case that W_S^d will be close to $(v_S - c_S^B)$, since the welfare from competition in the secondary market is going to be equal to v_S minus the price charged, which, in turn, will be equal to the cost of one of the two firms, given Bertrand competition.

Because it may be difficult verifying when data is essential for competition, we might conclude that a blanket prohibition of data tying, especially so when data is pro-competitive, is a risky policy that may, in some circumstances, end up harming consumers. However, an *asymmetric prohibition to data-tying*, which only applies to the dominant company, will carry a lower risk of harming consumers. As long as data is *not* essential for competition, it implements the same outcome as a generalized ban on data-pooling, thus neutralizing the ability of A to exclude B, but allowing B to harness data and use such data in the primary market. This is unambiguously welfare enhancing over a generalized ban if data is pro-competitive.

When a data advantage is essential for competition, instead, an asymmetric ban will reduce, although not completely eliminate, the incentives of A to carry out an exclusionary strategy: A can acquire the data with the only purpose of denying them to B. Now A stands to gain $\Pi_P^m - \Pi_P^d$ in the primary market rather than $\Pi_P^M - \Pi_P^d$ and so the conditions for an exclusionary equilibrium might remain in place. If exclusion takes place, then the outcome for consumers is still better than in the case of a generalized ban, because a price-war now takes place in the secondary market. If exclusion does not take place, then the outcome is preferable to the one arising from a generalized ban and is the same as with a ban on entry of A, except A might now enter the secondary market thus generating higher welfare for consumers.

Remark 4 *A policy that prevents the dominant operator in the primary market to pool data is preferable to policy that imposes a generalized ban on privacy-policy tying, more so if data is pro-competitive. It reduces the incentive to foreclosure when a data advantage is essential for competition and eliminates them when data is not essential.*

5.3 Data Portability

Now suppose data collected in the secondary market is portable across firms and non-rival, but consumers remain in control of the ability to transfer it from one firm to another. We envisage that, after entry decisions have been made, consumers can *transfer their data to any firm operating in the same market*, possibly in exchange for a fee or at a cost, which we treat as negligible. For example, social-networking profiles built within the incumbent platform, might be made portable to other social-networking platforms. To make progress, we focus on a model where, whenever A and B are both in the primary market, consumers will either all always want to give away their data to the firm that does not have it or all refuse to do that.

In particular, consumers will share their data if data is pro-competitive, that is if and only if $W_P^{\mathscr{D}} > W_P^d$. Hence, when data is anti-competitive, the possibility of data-sharing will have no effect on equilibrium behavior. The exclusionary equilibrium arises, which likely leaves consumers worse off than if they shared the data. But, in our model as in practice, consumers are unable to collectively commit to share data with B to encourage its entry.

Let's therefore focus on pro-competitive data. In this case, giving data to all firms induces firms to compete the data advantage away, at least partially. Hence, all consumers port their data if there is competition in the primary market. This implies that, as long as a data advantage is *not* essential for competition, firm B will always enter, making it impossible for A to exclude it. In turn, A enters the secondary market only if it is profitable to do so (i.e., ENC holds). In this scenario, where the data advantage is not essential, consumers benefit from data portability, as it neutralizes the risk of foreclosure.

On the other hand, consider the case where the data advantage is essential for competition. In this case, B will never enter the primary market, as it is unable to gain an exclusive data advantage given consumers' optimal ex-post behavior. In light of the above A will stay out of the secondary market, unless, again, entry would anyway take place. In any event, no price war will take place in the secondary market. As a result, consumers are left with a double monopoly, unless A enters. Remarkably, due to strategic considerations, assigning ownership of data to consumers may not always alleviate the exclusionary concern we present and, in fact, it may even aggravate it.

Remark 5 *Data portability has no effect when data is anti-competitive. Instead, when data is pro-competitive, data portability will prevent exclusion if data is not essential for competition but it will harm consumers when a data advantage is essential for competition.*

We have seen that enforcing unrestricted portability may end up damaging consumers when data is essential for competition. Instead, a requirement to offer portability of data only levied on the dominant operator would both restore incentives to entry and eliminate those to foreclose — analogously to the case of a unilateral prohibition of data tying. In contrast to what happens with a unilateral ban on data-tying, if there is no threat of entry in the primary market, then data portability requirements still allow the dominant firm to enter the secondary market and vie to acquire data, which can happen if data raises monopoly profits enough in the primary market.

5.4 Data Trading

Finally, assume data is the property of the firm that collects it and that it is transferable and non-rival (i.e., the data advantage can be enjoyed by all firms in the same way). We focus on the case where the data-owner can sell the data collected in the secondary market at the beginning of the second period, but after all entry decisions have been made. That is, we maintain that B cannot credibly commit not to enter the primary market as part of the data selling deal.[20] To simplify the exposition, we also make two, largely technical, assumptions: (i) the efficiency effect is non-negative and (ii) the sum of profits when no firm has an advantage over the other in the primary market is below the sum of profits when one firm only has the advantage (i.e., competition is fiercest in the primary market when the data endowments of the two firms are symmetric).

Data trading is common among tech firms. For example, it is well known that Facebook used to share its data with other tech companies such as Amazon, in exchange for partnerships of various sorts.[21] However, the legal status of such practices is to a large extent uncertain and subject to debate in many jurisdictions, where the nature of property rights over data remains uncertain. It is therefore important to evaluate some of the consequences of assigning such rights to firms.

Again, whether a data advantage is essential for competition or not affects the outcome. Suppose, first, that a data advantage is *not* essential for competition in the primary market. Then, B enters the primary market, unless A happens to be the *only* firm with the data advantage. In this case, as we shall argue next, exclusion remains an equilibrium. On the one hand, if A stays out of the secondary market *or* does not acquire the data advantage, B will enter the primary market without selling the data to A. In particular, a trade of data from B to A would only be profitable to both parties if $\Pi_P^{\mathscr{D}} - \Pi_P^d > \pi_P^D - \pi_P^{\mathscr{D}}$, which is never true if (ii) above holds. On the other hand, if A enters, it has the same incentives as in the baseline model to win competition in the secondary market and obtain the data advantage, as it will then avoid entry of B. Note that A would have no incentive to sell the data once it acquires them, because $\Pi_P^M - \Pi_P^{\mathscr{D}} > \pi_P^{\mathscr{D}} - (1-\beta)f_P$ when (i) and (ii) above hold. Hence, when data is not essential for competition, allowing data trade will have, under the stated assumptions (i) and (ii), no impact on the outcome.

Instead, suppose a data advantage is essential for competition. Entry by B in the primary market is profitable if and only if it enjoys the data advantage exclusively. In this case, the possibility of trading data changes the strategic outlook of the game, because now B can credibly commit to remain out of the primary market if it sells its data to A. To analyse this case, suppose first that A remains out of the secondary market. On the one hand, it would be willing to pay up to $\Pi_P^M - \Pi_P^d$ for data in order to avoid entry of B. On the other hand, B would be willing to accept $\pi_P^D - (1-\beta)f_P$ (i.e., the profit it makes by entering the primary market) to relinquish the data advantage. Because the efficiency effect is positive, we have $\Pi_P^M - \Pi_P^d > \pi_P^D - (1-\beta)f_P$. Hence, a trade of data will be feasible, i.e., acceptable to both parties, at any price in the interval $[\pi_P^D - (1-\beta)f_P, \Pi_P^M - \Pi_P^d]$.

[20] Such an agreement would clearly be illegal, as it would allow both firms to maintain monopoly in their respective markets. Consumers would not benefit.

[21] See Solon and Farivar (2019), "Mark Zuckerberg leveraged Facebook user data to fight rivals and help friends, leaked documents show", Nbcnews.

Let's denote with p^D the price at which both firms expect the trade of data to take place. Then A's payoff from remaining out of the market is $(1-\beta)\Pi_P^m + \Pi_P^M - p^D$.

Let's now consider what happens if A enters. In this case, either A or B will own the data advantage at the end of the first period. If A owns it, it will not sell it to B as data trading is now not feasible, due to the assumption that the efficiency effect is positive. If B owns the data advantage, instead, a trade of data will become feasible at any price in the interval $[\pi_P^D - (1-\beta)f_P, \Pi_P^M - \Pi_P^d]$, as in the case where A remains out of the secondary market. Let's denote again with p^D the price at which both firms expect the trade of data to take place. With this in mind, we can now evaluate the pricing behaviour of A and B in the secondary market. It takes just a bit of algebra to demonstrate the intuitive observation that, with data trading, the most efficient firm in serving the secondary market will win competition in the secondary market. In particular, the lowest acceptable price of A will be x that solves

$$\Pi_P^M - c_S^A + x = \Pi_P^M - p^D,$$

that is $x = c_S^A - p^D$, while, analogously, the lowest acceptable price for B will be $y = c_S^B - p^D$.

It follows that if A is less efficient than B in serving the secondary market, that is $\Pi_S^d < \pi_S^d$, then it will remain out of such market and buy the data from B instead. Assuming the price paid for the data is the same whether it stays out of the secondary market or enters and loses competition there, there is no reason for A to sustain fixed costs only to end up making zero profit in the secondary market and paying for data. Instead, suppose A is more efficient than B. By entering it will make payoff $(1-\beta)\Pi_P^m + \Pi_P^M + (2-\beta)\Pi_S^d - (1-\beta)f_P - p^D$, where we used the fact that $\pi_S^d = 0$ as A is more efficient. Hence, we conclude that A will enter if and only if the behaviour of A is explainable by non-exclusionary concerns, and in that case will win competition in the secondary market.

The main conclusion we draw is that allowing firms to sell data they acquire does not eliminate the incentives of A to exclude B from the primary market, when such incentives are present to begin with. In particular, A now either out-prices B in the secondary market and acquires a data advantage this way or, whenever it's more convenient, it avoids entry and buys data from B instead. A data purchase is possible when data is essential for competition, as in this case selling the data credibly commits B not to enter in the primary market. Moreover, a data trade will not happen when A would want to enter the secondary market anyway. Remarkably, when such a data sale happens, consumers are left worse-off than in the exclusionary equilibrium, because there is no entry in any of the two markets, while both B and A retain monopoly profits. Summing up, we have the following.

Remark 6 *Data trading reduces strategic externalities between firms and facilitates monopoly protection. It has no impact if data is not essential for competition while, otherwise, it may result in monopoly in both markets and lower consumer surplus than in the exclusionary equilibrium.*

6 Conclusions

In this paper, we have explored conditions that could prompt a firm dominant in a primary market to enter an unrelated secondary market and set below-cost prices with the aim to acquire data that makes entry of potential competitors in the primary market less desirable. We have highlighted that consumers can be worse off from this strategy, compared to the scenario in which entry in the secondary market does not take place. Furthermore, we have provided guidance as to the right remedies in such cases. We found that data unbundling and portability obligations imposed on the dominant player are likely to benefit consumers. We end the paper with two remarks about two issues that we have so far neglected, the effects on total welfare (the sum of consumer surplus and profits) and the role of mergers.

We focused on consumer welfare, since it is the standard adopted by competition authorities. However, it is worthwhile asking whether, despite reducing consumer surplus, exclusion can raise total welfare. Moreover, one may wonder whether the corrective measure we proposed may end up burning overall more surplus than the exclusionary equilibrium does. We argue that the answer is negative in both cases. To see this, consider the Bertrand scenario we described at the end of Section 3. This is an easy case to deal with, because there is no deadweight loss from monopoly and, as a consequence, total welfare is determined by entry decisions alone. Assuming the conditions that generate the exclusionary equilibrium hold, consider entry in the secondary market. Because the entrant appropriates all the efficiency gains, if any, there will be always entry whenever it is efficient, that is when it is explainable by non-exclusionary concerns (ENC). However, due to the additional value of foreclosing, the exclusionary equilibrium might result in too much entry. Then, by reducing incentives to foreclosure, unbundling and portability requirements mitigate the risk of inefficient entry in the secondary market. The converse occurs in the primary market. The exclusionary equilibrium can only induce too little entry. Again, the proposed remedies move incentives in the right direction.

Data can also be acquired by merging with an already successful firm operating in the secondary market.[22] As we elaborated in the introduction, we believe this has been a common phenomenon in recent years. In our framework, an acquisition has similar welfare effects similar to data-selling agreement discussed in the previous paragraph. In fact, it is always more profitable to merge for both the target firm in the secondary market and the dominant firm in the primary market, as long as the purchase price is kept low enough by the target's shareholders realizing that refusing to sell will trigger entry. This observation suggests that some acquisitions may take place under the threat of envelopment and predation we describe. When a merger happens for this reason, consumer surplus is even lower than with preemptive entry, as there will be no competition for the data advantage and consumers will end up with a monopoly in both markets. This indicates that competition authorities should be especially vigilant about data-driven acquisitions by big tech companies, even when firms do not *prima facie* compete in the same market.

[22]This motive needs to be distinguished from acquisitions that have the purpose of eliminating future competition, also called "killer acquisitions". Some recent contributions include Motta and Peitz (2020), Kamepalli et al. (2019) and Cunningham et al. (2019)). Here, the acquiring firm has no incentive to shut down the target firm.

Appendix

Proof of Proposition 1 Consider [a.] Net of fixed costs, π_P^D is the most B expects from entry and π_P^d is the least. Hence, if $\pi_P^D < (1-\beta)f_P$ then B stays out and if $\pi_P^d > (1-\beta)f_P$ it enters.

Therefore, if neither of the two conditions above holds, since $\pi_P^D > \pi_P^d$, we must have $\pi_P^D > (1-\beta)f_P > \pi_P^d$. Case [b.] imposes two additional assumptions. To analyse equilibrium, we proceed by backward induction. First, we compute equilibrium payoffs in the subgame that starts after A decides to stay out of S. In period one A and B make discounted monopoly profits $(1-\beta)\Pi_P^m$ and π_S^m, respectively. Then, if B does not enter, in the second period they both achieve again monopoly profits. Instead, suppose B enters. In this case B still obtains $(1-\beta)\pi_S^m$ in the second period, as it remains monopolist in the target market. In addition, by competing in the primary market, it makes π_P^D. We conclude that, in the subgame that starts after A stays out, B enters, as (PDA) implies $\pi_P^D > (1-\beta)f_P$. In the equilibrium of this subgame A obtains

$$(1-\beta)\Pi_P^m + \Pi_P^d. \qquad (1)$$

The result of this argument is illustrated in figure 3(a).

To determine the initial action of A, let's now analyze the equilibrium entry decision of B in the subgame where A enters. There are two cases to consider. First, B wins Bertrand competition for the target market in period one. Then it will make net second-period (incremental) profit equal to $\pi_P^D - (1-\beta)f_P + (1-\beta)\pi_S^d$ if it enters the primary market, while it will make $(1-\beta)\pi_S^d$ if it stays out. As it was the case for the other branch of the tree, (PDA) implies B will enter the market. This point is illustrated Figure 3(b).

Now suppose A wins the data advantage in S. Consider again the decision of B. By staying out it makes $(1-\beta)\pi_S^d$ incremental profit. If it enters, it competes with A in P. Given the data advantage acquired by A, B will make second-period (incremental) profit equal to $(1-\beta)\pi_S^d + \pi_P^d - (1-\beta)f_P$. It follows from (PDA) that B will prefer to stay out if A enters the target market and wins the data advantage. This is illustrated in Figure 3(c).

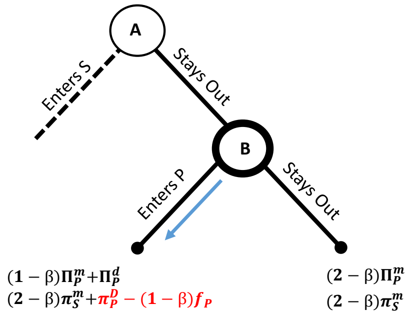
We are now ready to wrap-up. We first determine the outcome of Bertrand competition if A enters and then we compare it with the outcome from staying out. As illustrated in the main text, under $(EE-IC)$ the outcome of Betrand competition is such that A serves the market at price $c_S^B - \pi_P^D + (1-\beta)f_P$. As illustrated above, B stays out. The profit of A is

$$(1-\beta)\Pi_P^m + \Pi_P^M + (1-\beta)(\Pi_S^d - f_S) - c_S^A + c_S^B - \pi_P^D + (1-\beta)f_P. \qquad (2)$$

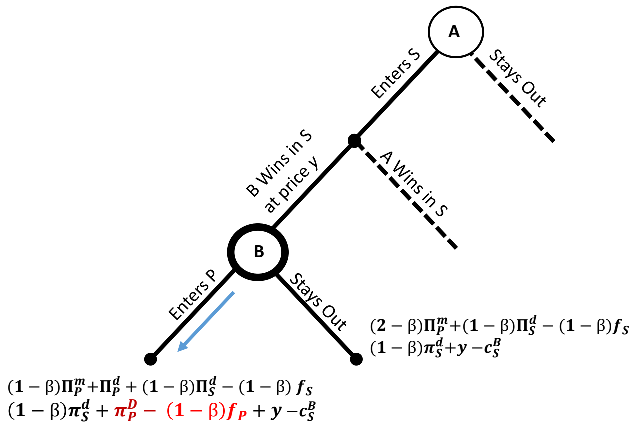
To conclude the proof of [b] we need to verify that entering the secondary market is indeed optimal for A. That is, we need to show that (2) > (1). Assumption $(EE-IR)$ guarantees that this is the case, after we substitute $c_S^A - c_S^B$ with $\pi_S^d - \Pi_S^d$ (see footnote 18).

Finally consider [c]. We argued that if $(EE-IC)$ does not hold, than B enters. If $(EE-IC)$ holds and $(EE-IR)$ fails, it follows from the observation in the paragraph above that A is better off remaining out of the market. If A stays out of the secondary market, then B enters. $\square$

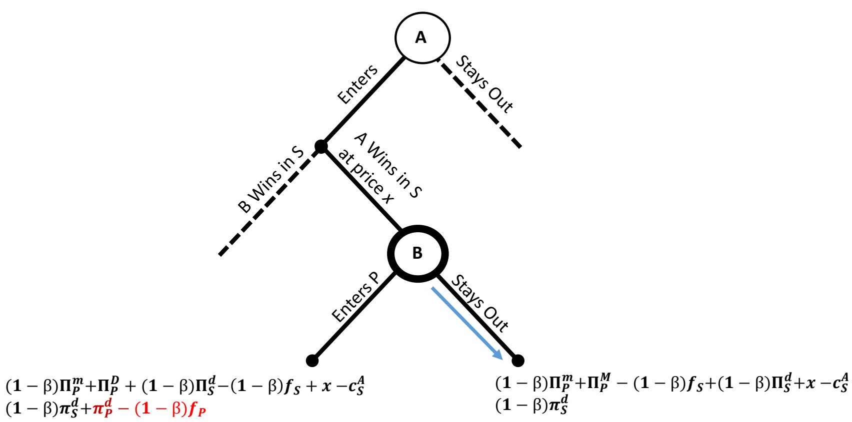
Proof of Remark 2 We demonstrate the statement in Remark 2 above by means of examples, focusing on the Bertrand-like scenario described at the end of Section 3. Because in the Bertrand-like scenario there is no dead-weight loss from monopoly, social optimality boils down to whether entry decisions are optimal in equilibrium or not. Let's begin with the first question. Firm B is more efficient than firm A in P if $c_P^B < c_P^A$. Then, the (PDA) condition becomes $\alpha + c_P^A - c_P^B > (1-\beta)f_P > 0$, which is satisfied as long the cost-advantage of holding data, α, is sufficiently large. The $(EE-IC)$ and $(EE-IR)$ conditions become $(v_P - c_P^A) - (c_P^A - c_P^B) > (c_S^A - c_S^B) - (1-\beta)f_P$ and $(v_P - c_P^A) - (c_P^A - c_P^B) > (c_S^A - c_S^B) + (1-\beta)(f_S - f_P - \max\{0, c_S^B - c_S^A\})$, respectively. For v_P sufficiently large, both can hold even when $c_P^A > c_P^B$ and $c_S^A > c_S^B$. Let's now discuss the second question. Whether entry by A followed by exclusionary behaviour is a welfare-optimal outcome or not depends on whether entry decisions are the same as those that a welfare-maximizing central authority would mandate. Clearly, a central authority would mandate entry of B into the primary market if the efficiency gain would be enough to cover fixed costs, that is if $(c_P^A - c_P^B) > (1-\beta)f_P$. It is then not difficult to see that (PDA),(EE-IC) and (EE-IR) can hold when that condition holds. This is obvious for the case of (PDA). By assuming $(c_P^A - c_P^B) = (1-\beta)f_P$ we can see that both (EE-IC) and (EE-IR) above can hold strictly when v_P is large enough. $\square$



(a) Equilibrium decision of B in the subgame after A stays out



(b) Equilibrium decision of B in the subgame where A enters but B wins the target market



(c) Equilibrium decision of B in the subgame where A enters and wins the target market

Figure 2: Equilibrium Construction: The decision node being analyzed is highlighted in bold; in red, the difference in payoffs between the two outcomes of the decision; the arrow indicates the optimal decision; x and y are arbitrary payments.

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar, "Too Much Data: Prices and Inefficiencies in Data Markets," Working Paper 26296, NBER September 2019.

Affeldt, Pauline and Reinhold Kesler, "Big Tech Acquisitions — Towards Empirical Evidence," *Journal of European Competition Law & Practice*, 04 2021, *12* (6), 471–478.

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan, "The Economics of Social Data," Cowles Discussion Papers 2203, Cowles Foundation, Yale University September 2019.

Bernheim, B. Douglas and Michael D. Whinston, "Exclusive Dealing," *Journal of Political Economy*, 1998, *106* (1), 64–103.

Bourreau, Marc and Alexandre de Streel, "Digital Conglomerates and EU Competition Policy," Working Paper 2019.

Cabral, Luis M. B. and Michael H. Riordan, "The Learning Curve, Market Dominance, and Predatory Pricing," *Econometrica*, 1994, *62* (5), 1115–1140.

Carlton, Dennis W. and Michael Waldman, "The Strategic Use of Tying to Preserve and Create Market Power in Evolving Industries," *The RAND Journal of Economics*, 2002, *33* (2), 194–220.

Chen, Zhijun, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima, "Data-driven mergers and personalization," *The RAND Journal of Economics*, 2022, *53* (1), 3–31.

Choi, Jay Pil and Christodoulos Stefanadis, "Tying, Investment, and the Dynamic Leverage Theory," *The RAND Journal of Economics*, 2001, *32* (1), 52–71.

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

Condorelli, Daniele and Jorge Padilla, "Harnessing Platform Envelopment Through Privacy Policy Tying," *forthcoming at Journal of Competition Law & Economics*, 2020.

Crémer, Jacques, Yves-Alexandre de Montjoye, and Heike Schweitzer, *Competition Policy for the digital era : Final report*, Luxembourg: Publications Office of the European Union, 2019.

Cunningham, Colleen, Florian Ederer, and Song Ma, "Killer acquisitions," *Mimeo*, 2019.

de Cornière, Alexandre and Greg Taylor, "Data and Competition: a General Framework with Applications to Mergers, Market Structure, and Privacy Policy," *Mimeo*, 2020.

Economides, Nicholas and Ioannis Lianos, "Restrictions on Privacy and Exploitation in the Digital Economy: A Competition Law Perspective," WP 19-15, NET Institute October 2019.

Eisenmann, Thomas, Geoffrey Parker, and Marshall Van Alstyne, "Platform Envelopment," *Strategic Management Journal*, 2011, *32* (12), 1270–1285.

Fairfield, Joshua A. T. and Christoph Engel, "Privacy as a Public Good," *Duke Law Journal*, December 2015, *65* (3), 385.

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

Fudenberg, Drew and Jean Tirole, "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look," *The American Economic Review*, 1984, *74* (2), 361–366.

_ and _ , *Dynamic Models of Oligopoly* Dynamic Models of Oligopoly, Taylor & Francis, 2013.

Fumagalli, Chiara and Massimo Motta, "Exclusive Dealing and Entry, when Buyers Compete," *The American Economic Review*, 2006, *96* (3), 785–795.

_ and _ , "A Simple Theory of Predation," *The Journal of Law & Economics*, 2013, *56* (3), 595–631.

_ and _ , "Dynamic Vertical Foreclosure," *The Journal of Law and Economics*, 2020, *63* (4), 763–812.

Gilbert, Richard J. and David M. G. Newbery, "Preemptive Patenting and the Persistence of Monopoly," *The American Economic Review*, 1982, *72* (3), 514–526.

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

Kamepalli, Sai Krishna, Raghuram Rajan, and Luigi Zingales, "Kill Zone," 2019.

Kreps, David M and Robert Wilson, "Reputation and Imperfect Information," *Journal of Economic Theory*, August 1982, *27* (2), 253–279.

Lambrecht, Anja and Catherine Tucker, "Can Big Data Protect A Firm From Competition?," *CPI Chronicle*, 2017.

Milgrom, Paul and John Roberts, "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, 1982, *50* (2), 443–459.

Motta, Massimo and Martin Peitz, "Big Tech Mergers," *Information Economics and Policy*, August 2020, *27*, 100868.

Panzar, John C. and Robert D. Willig, "Economies of Scope," *The American Economic Review*, 1981, *71* (2), 268–272.

Prufer, Jens and Cristoph Schottmuller, "Competing with Big Data," *Journal of Industrial Economics*, 2022.

Rasmusen, Eric B., J. Mark Ramseyer, and John S. Wiley, "Naked Exclusion," *The American Economic Review*, 1991, *81* (5), 1137–1145.

Segal, Ilya R. and Michael D. Whinston, "Naked Exclusion: Comment," *The American Economic Review*, 2000, *90* (1), 296–309.

Spence, A. Michael, "Entry, Capacity, Investment and Oligopolistic Pricing," *The Bell Journal of Economics*, 1977, *8* (2), 534–544.

_ , "The Learning Curve and Competition," *The Bell Journal of Economics*, 1981, *12* (1), 49–70.

Wernerfelt, Nils, Anna Tuchman, Bradley Shapiro, and Robert Moakler, "Estimating the Value of Offsite Data to Advertisers on Meta," 2022.

Whinston, Michael D., "Tying, Foreclosure, and Exclusion," *The American Economic Review*, 1990, *80* (4), 837–859.

Williamson, Oliver E., "Assessing Vertical Market Restrictions: Antitrust Ramifications of the Transaction Cost Approach," *University of Pennsylvania Law Review*, 1979, *127*, 953–9933.