

The Incumbency Advantage In The Digital Era: Why Data Barriers Matter

Author : Matteo Bulgarelli

Thesis Director : Éric Toulemonde

Thesis Reader : Paolo Garella

Academic Year 2019-2020

In order to obtain the joint degree:

Master 120 en Sciences économiques, Orientation générale, Finalité approfondie (UCL/UNamur)

and

Dottore magistrale in Economics and Political Science (UNIMI)

The Incumbency Advantage In The Digital Era: Why Data Barriers Matter

Matteo Bulgarelli

August 17, 2020

Abstract

I adopt a three stage game to describe the interaction between data driven effects and cross group network externalities in the competition among two homogeneous platforms. I show that, if users singlehome and hold pessimistic beliefs on new entrants, the existence of data cost synergies provides to incumbents an advantage that translates in positive profits. When beliefs are symmetric, as data renders competition fiercer, firms might subsidize their users excessively, incurring both in negative profits, with aggregate innovation levels reaching the maximum. Antitrust authorities should consider how data reshape the competition and the incumbency advantage in digital industries.

1 Introduction

The competitive dynamics at the heart of platform industries have been one of the most debated topic among antitrust legislators and academics over the last 20 years; as the spread of internet in the everyday life of consumers kick-started many business solutions that have the intermediation among users as fundamental pillar, the need to sketch clear regulatory boundaries to avoid inefficiencies and market failures has become urgent.

In economic terms, platforms are defined as middlemen between two different type of agents; what actually distinguishes them from common retailers is the interdependence between the utility an agent derives and the participation of another user to the mediator's network. These effects are defined as *network externalities*, and can be classified in two distinct and non-mutually exclusive groups; first, we talk about *within group* effects if the presence of users of the same type augments the willingness to participate of a single agent (as typical of social networks such as Facebook, Twitter or Instagram), whereas, when the participation of users of a different type (such as typical buyer-seller relationship) increases the utility of an agent of the other type, we talk about *cross group* effects (e.g., e-commerce

marketplaces, such as eBay, video or music streaming, like Twitch or Spotify, or financial service intermediaries, like VISA). Besides, network effects can be *direct* or *indirect*: in the former case, the presence of an additional user directly affects the utility of the others. The latter arises typically for gaming consoles, where a greater gamer base on a given platform ensures an increased amount of sales to firms that develop videogames for that platform. These revenues are expended in the development of better products, boosting the utility from participation of the buying side. Lastly, network externalities can be positive, as in the cases described above, or negative, as in the relationship between advertisers and users on social media.

The array of activities carried out by two-sided intermediaries, as set forth in Belleflamme and Toulemonde (2018), is considerable: *hardware & software systems* facilitate the interaction between users and developers (Mac OS, Playstation, Android, Windows); *transaction systems* allow monetary exchanges between participants (Bitcoin, Monero, PayPal, Mastercard); *matchmakers* or *exchange platforms* help agents of a side to find the desired counterpart on the other side (Meetic, Tinder, Alibaba, Amazon, eBay); *peer-to-peer marketplaces* ease the sharing of specific items among "peers" (Uber, WeWork, Airbnb, Couchsurfing); *crowdfunding platforms* allow start-ups to raise funds for their projects (Kickstarter); *social media* and *search engines* sell advertising space and provide content to users.

Economic theory, when users *singlehome* (i.e. they visit only one intermediary), predicts a high tendency to tip for markets characterized by cross-group, positive network effects and homogenous consumers; as the users derive utility from mutual participation, followers are driven out of the market, in so far producing migratory effects requires to "get everyone on board". However, as migration from one side produces a flee also from the other, the leader cannot rest on her laurels, and must compete fiercely to maintain her privileged position. This consideration triggers a Bertrand argument: as agents of a type are fundamental for conquering the demand of the others, the incumbent will need leave all of the surplus to her users. The straightforward implication of this line of reasoning is that agglomeration equilibria, which feature a monopolistic platform that serves the both sides, should be efficient, and result in zero economic profits. However, when considering the individual performance within many two-sided industries, such as the e-commerce marketplaces, this perspective is hardly challenged; Figure 1 displays the Net Income of the 5 greatest listed marketplaces in the United States, using data from financial statements retrieved from Orbis; Amazon, conglomerate that combines her marketplace with direct retailing activities and is active also in cloud computing services; eBay, e-commerce marketplace founded 1995; Etsy, marketplace for vintage items; Overstock.com, intermediary specialized in closeout merchandise; and Wayfair, which facilitates sales of home decor, bedding and similar goods. It can be immediately noticed how the two leaders of the market are also very profitable, with rivals

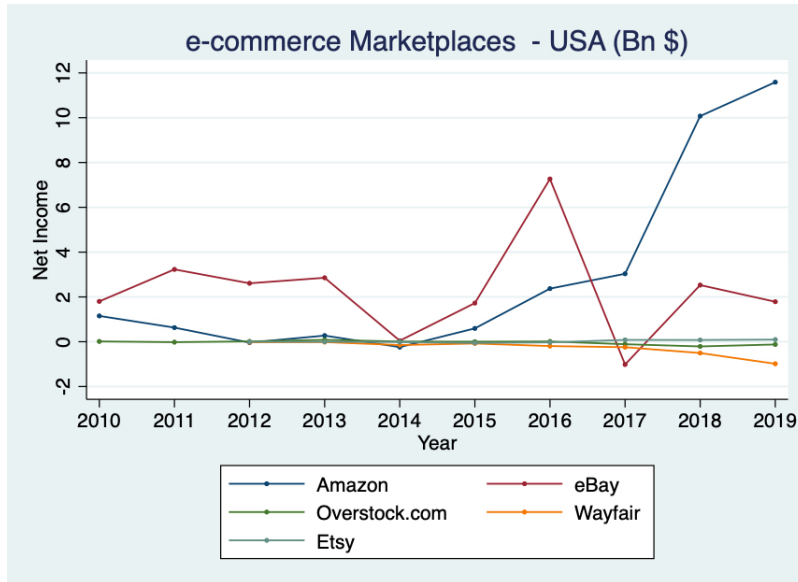


Figure 1: Net Income of the five biggest listed two-sided Marketplaces in the United States, excluding brick-and-mortar businesses. Source: Orbis.

lagging far behind, fluctuating around the zero bound; hence, data seem to rebut the zero profit conclusion.

It's worth spending a few words on this graph, which shall not be interpreted as a market analysis, but rather a simple comparison of similar business models that operate within the same industry: indeed, these firms face harsh competition also from traditional brick and mortar stores such as Walmart, Target or Best Buy. Many of these retailers, in recent times, have started their own marketplace for third party sellers, as part of their e-commerce activities. However, the biggest third marketplace managed from brick and mortar store, i.e. Walmart Marketplace, accounts today for just above 50000 sellers; given that Etsy alone, hosts more than 2.5 million sellers on her network, and given that it is not possible to obtain an income statement that discloses the profits from the marketplace alone, I decided not to include them. Two additional remarks: first, the 5 companies reported also compete with the Canadian multinational Shopify, which provides sellers an Application Programming Interface (API) that allows them to create and manage easily their own online store. However, Shopify cannot be considered a platform in the economic sense of the term, in so far her practices do not feature network externalities; in fact, sellers do not derive utility from participation of other sellers, and end-customers are not targeted directly by Shopify. Besides, the company, which went public in 2014, has always operated at loss. Another important player is Rakuten, Japanese e-commerce marketplace active in the United States. In this

case, however, no income statement for the American (non-listed) subsidiary is available, and Rakuten presumptively earns the most of her revenue in Asia. Lastly, three more privately held competitors are Wish, Bonanza and Newegg, which do not disclose their income statements. Second, Amazon's data should be taken with a pinch of salt: indeed, as the *The Economist* reports, the company derives the most of her net income (slightly above 50% in 2019) from her cloud computing service, Amazon Web Services (AWS). Amazon earns half of its revenue from pure e-commerce retailing; however, estimates by Bernstein reveal negative operating profit margins for the segment, suggesting a loss-leading conduct. In a nutshell, Amazon's profits arise from three sources: third party sellers, advertising, and AWS.

While it seems clear that these markets favor the entrenchment of quasi-monopolistic positions, it's less clear, on a theoretical perspective, what are the drivers of such rent creation. One possible explanation could lie in the role that personal data play in moulding the competitive playground within these industries.

I refer to personal data as "digital records created by and about people" (Schwab, Marcus, Oyola and Hoffman, 2011) : they can be represented as *user information* that firms can collect from the utilization of their services or devices. As explained exhaustively by Crémer, de Montjoye and Schweitzer (2019) , individual level data can be acquired by firms through different channels; first, they can be *volunteered* if they are intentionally contributed by users by mean of their activities on the platform, as emails, reviews on a product, or posts on a social media; second, they can be *observed* if they are obtained automatically from users' behavior, as in the case of movements of individuals traced by mobile phones, or data records on road trips for drivers; third, data can be *inferred* when they are derived from observed or volunteered records, as it happens frequently to profile music users on streaming platforms to propose them playlists that are coherent with their preferences. Individual data can also be distinguished with respect to how firms use them; first, their use can be *non-anonymous* if a firm exploits individual level data to provide a service, such a playlists for music streamers; second, use can be *anonymous* when data are encrypted in such a way that controllers cannot retrieve any personal identifier from the information they dispose of; for example, when developers use volunteered data from likes on social medias for training algorithms for demand targeting purposes, they do not need to "feed" them with true identifiers; it's sufficient to record that the user 12a4kz8 liked a given product or a given topic. Lastly, personal data can be *aggregated*, as it happens to draft national statistics. As non-anonymous use of data introduces privacy concerns that are beyond the scope of this paper, I will refer uniquely to anonymous use of either observed, volunteered or inferred data.

Industries characterized by huge flows of user information exhibit a unique combination of

increasing return to scale and economies of scope; the former arises in so far the costs of providing services to the user base are much less than proportional to the number of customers served; in particular, R&D expenses related to the development of machine learning algorithms are *decreasing* in the number of individual customers. The latter is consequence of the ability of firms to leverage their own data driven business model into other, related segments; as firms' costs to provide innovational, data intensive solutions drop the bigger the amount of user's information available to them, being active in a diversified array of services allows platforms to innovate at lower costs. These effects have been traditionally overlooked by economic theory, and could create insurmountable barriers to new entrants. Two questions follow naturally:

- *Do data-driven cost synergies favor rent creation by rooted incumbents in two-sided settings?*
- *Are segmentation equilibria, characterized by two (or more) active firms with similar market shares, sustainable, if we consider Big Data?*

In this paper, I study the interplay between data driven business models and cross group network effects to determine the magnitude of the incumbency advantage faced by dominant two-sided platforms. The work is intended to investigate what are the implications in terms of efficiency and innovation of dominant positions in platform markets. These problems are particularly relevant for competition policy in at least two fields:

- **Merger Analysis.** Two sided markets have always presented many challenges to antitrust authorities with respect to market definition: as one side of the market can be subsidized completely, i.e. can have access to the platform for free, common measures such the SSNIP test need to be corrected for the specific framework under analysis: scholars refer to the dualism between the common, *single market* approach, where the effects of a price increase are considered separately on two distinct markets, and a *multi-market* approach, where the two sides of the market are pooled together and anti-competitive effects are evaluated holistically (see, for example, Filistrucchi (2018) or Franck and Peitz (2019)). The main problem of quantitative methods stems from the fact that they might overlook strategic decisions of incumbents aimed at eliminating possible future, innovative, entrants due to inherent difficulties in determining which revenues should be inspected to define the boundaries of the industry. One key example is *Facebook-Instagram*¹; as the relevant market was considered the one for photo sharing applications, Camera+ and Hipstamatic were identified as credible

¹see <https://www.ftc.gov/news-events/press-releases/2012/08/ftc-closes-its-investigation-facebooks-proposed-acquisition>

competitors, and the FTC ruled in favour of the merger, without taking into account the prospective future of Instagram as an innovative social network (see Prat, Valletti, 2019) . Academics refer to fusions aimed at hindering future rivals as *killer mergers*. Intensive debate is unfolding on whether policymakers should weigh more the harms generated by False Negatives, as a consequence of the loss in terms of potential competition. ² Yet, relatively little attention has been paid to the strengthening of the competitive advantage that would result due to increasing returns to scale after mergers with other horizontal competitors: as data driven business models display cost synergies that are proportional to the users' information of the firms, merger analysis should also consider if such unions could generate behemoths that would discourage potential rivals due to inability to reach the minimum efficient scale to credibly compete, and should also assess whether there could be efficiency gains or not by making the market more contestable, both in terms of innovation and consumer surplus. In this context, mergers might not be killer in their strategic sense: they might be driven by an intrinsic feature of the industry.

- **Determination of Exclusionary Conducts.** In digital economies, data are regarded as the new oil: Li, Nirei and Yamana (2019) underline the non-depreciating nature of users' information; as the process of data fusion allows statisticians to gain further insights on customers' behavior, they can be thought as an R&D investment that does not incur in obsolescence. Therefore, they represent a consistent value adding asset.

Policymakers have generally given little consideration to the possibility of data overaccumulation for exclusionary purposes; Nonetheless, as Khan (2016) reports, Amazon has been running in the red for almost 20 years before incurring in profits, while her market capitalization kept on skyrocketing. Legal scholars have been often debating on the mildness of policy prescriptions on predatory pricing and on the lack of instruments in the hand of the authorities. As it has been widely documented that the adoption of algorithmic pricing allows for a massive correction of prices on a daily basis³, understanding whether the market delivers *naturally* incentives to predatory behavior that differ from simple pricing tactics might be fundamental to prohibit certain practices that today could be deemed as pure loss leading. For example: if overinvestment in R&D today could be aimed to exclude competitors tomorrow, and

²see Argentesi, Buccirosi, Calvano, Duso, Marazo, Nava <https://voxeu.org/article/mergers-and-merger-policy-digital-markets> and Caffarra, Crawford, Valletti <https://voxeu.org/content/how-tech-rolls-potential-competition-and-reverse-killer-acquisitions>

³see Quartz, 2013

<https://qz.com/157828/amazon-changes-its-prices-more-than-2-5-million-times-a-day/>

if such practice would result inextricably in lower innovation levels or lower consumer surplus, antitrust authorities could consider the market structure to impose specific remedies on transgressors, without considering the pricing behavior.

The rest of the paper is organized as follows: in section 2, I describe the economic literature on network externalities, platform markets and the incumbency within it, and on data driven effects. In section 3, I develop a model based on singlehoming (exclusive intermediation) to describe competition under asymmetric beliefs (i.e. with an incumbency advantage) and with symmetric beliefs. Section 4 will be reserved to conclusions.

2 Literature

The interplay between network externalities and competition has been deeply analyzed in the influential article of Katz and Shapiro (1985) . Academics have focused extensively on describing how network effects reshape the competitive dynamics in platform industries. Three seminal works serve as road maps: Rochet and Tirole (2003) develop a model for platform competition in a two-sided setting where the intermediary charges a per transaction fee. They find out that multihoming on one side of the market results in a pricing structure more favorable to the other side of the market. Armstrong (2006) uses a Hotelling specification where platforms cannot observe transactions and levy a membership fee. In his model, multihoming reduces the market tipping effect produced by cross group network externalities. Caillaud and Jullien (2003) adopt a framework characterized by transaction and membership fees, where both sides of the market differ only in their probability of being matched by the intermediary with desired counterpart. In recent extensions, Belleflamme and Peitz (2018) defined the pricing structure in platform markets in a general setting, while Cabral (2011) developed a dynamic model for identifying prices under the provision of network goods. Belleflamme and Toulemonde (2009) unravel the effect of negative intra-group externalities, i.e. of rivalry among one group of users: they argue that, for an intermediate range of negative network effects, a new platform has no incentive to enter the industry, in so far it would not be able to profitably attract each of the two sides.

As competitive dynamics have been investigated deeply over the past two decades, researchers are shifting their attention toward the determinants and the magnitude of the incumbency advantage, which can be represented as "the fact that an incumbent, that is, a firm with already an installed base, will be able to generate higher profits than a new firm (an entrant) even if the entrant offers identical terms to consumers, or even better" (Biglaiser, Calvano, Crémer (2019)).

It's worth noting that competition, in industries characterized by network externalities, en-

tails unavoidably a tradeoff between lower prices and the so called segmentation equilibria: if we assume that users would benefit from the participation of other users, agglomeration on the same platform represents an optimum, absent membership fees. To capture the relation between socially optimal allocations of users and entrance, and in order to inspect the impact on the incumbent's profits, Biglaiser and Crémer (2016) adopt an infinite horizon game featuring within and cross group effects; in their model, each user of the platform assesses whether it would gain positive utility from migrating toward a competitor given the belief that no other consumer would migrate, and they define as equilibrium outcomes where no one would switch. They conclude that entrance, counterintuitively, fosters equilibria where every user selects the incumbent intermediary. They motivate the finding asserting that the threat of entrance and the presence of network externalities induces the incumbent to fight for its user base more fiercely, frightened by potential mass migration. They also note that the incumbency advantage, in terms of profits, should not be substantial, in so far payoffs of the incumbent do not vary with or without entrance, regardless from the fact that in the former case she is forced to compete more intensely.

Halaburda, Jullien and Yehekel (2020) introduce the notion of focality, which can be described as the ability of a lower quality incumbent to maintain its dominant position independently from the existence of an higher quality potential rival. Using a dynamic specification with one-sided network effects, they note that advantages arising from a focal position are a specific feature of industries sharing network externalities, as equilibria where a focal platform monopolizes the market in both stages of their game do not arise in an revised version of their model that features switching cost; this finding is motivated by the exogenous nature of the latter, which are costs that, upon migration to a competitor, the individual user inflicts only to himself, whereas the former entails a social component that result in costs for every other user in case of departure of a single one. Halaburda and Yehekel (2016) explore the interplay between partial coordination bias, i.e. partial beliefs held by both sides of the market that migration to a focal platform will take place, and the decision of each intermediary to subsidize a given group of users. They assume negative within group externalities for the sellers' group, and that buyers suffer saturation from purchase: their marginal utility becomes negative after acquiring a given number of sellers' products. Their findings suggest that the equilibrium choice of the business model of a platform, intended as its decision to "milk" one of the two groups and subsidize the other, is highly dependent on the strength of the coordination bias toward the focal intermediary and on the development costs of the seller.

The theoretical literature on data competition in digital economies is still scant. One of the few analytical works that addresses the problem of how data driven industries reshape the competitive dynamics is Prüfer and Schottmüller (2017) . Using a Hotelling refine-

ment, they define demand as a function of quality difference between two firms. In their framework, data are modelled as *users' information* that can be stored for free, and hence increase in the demand of the intermediary. They define costs of these platforms as a function of investment in quality innovation: the interesting feature of their specification is that costs are decreasing in the demand of each firm, and hence, in the amount of data they dispose. In a dynamic repetition of the game, they show that, if the incumbent doesn't discount excessively future profits, the market will tip and the winner will "rest on its laurels", refraining from investing any further in quality improvement. In their model, network effects are present only indirectly, and they are not result of cross group interaction: indeed, participation is enhanced conditionally on quality increase, not on increased participation. Besides, they consider a setting with no fees, where the platform cares only about the participation of one side, leaving the sellers' group unmodeled; the relevant context is the one of either social networks or search engines, in their own words.

Argenton and Prüfer (2012) develop a model for search engine competition, which abstracts from direct network effects to inspect uniquely the interdependence between market concentration and cost synergies arising from users' information. They prove that markets characterized by innovation costs that are inversely proportional to the installed user base of a firm have a natural tendency to tip, and that both the aggregate welfare and the consumer surplus of a monopolistic structure are lower than a competitive oligopoly scenario, where search engines are obliged to "share their (anonymized) data on clicking behavior of users following previous search queries among each other and among new entrants". De Cornière and Taylor (2019) analyze a setting where the intermediary competes actively against the same sellers it serves, by mean of vertical integration. The main goal of their article is to ascertain whether the process of biased intermediation (for example, a better positioning of the vertically integrated sellers' product in the queries) can harm consumers' surplus. They show that if end-users and platforms have congruent interests, i.e. if an higher mark-up might be associated with higher utility to consumers due to higher product quality, then intermediaries enjoying the bias might propose better deals to consumers, possibly improving their welfare, whereas if the parties display conflicting interests, i.e. if consumer's utility can be augmented only by mean of price reductions, than the intermediation bias can only be detrimental. The problem they analyze is orthogonal to the extreme return to scale effect that datafication produces. To the best of my knowledge, no theoretical work that exhibits direct cross-group network effects in a data driven setting exists.

3 Baseline Setting with Singlehoming

Framework The following model is a revised version of Caillaud and Jullien (2003) and Prüfer and Schottmüller (2017). I assume a mass one of homogenous type-1 users and a mass one of homogenous type-2 users that decide which platform $k = \{I, E\}$ to visit. Each i -user is interested in being matched only with one j -user, and derives utility from intermediation only throughout a successful match. Absent the intermediation, users- i cannot interact with users- j . For the rest of the section, assume that both users singlehome, i.e. intermediation is exclusive: platform's customers visit only one matchmaker. Users of side i split across platforms so that $n_i^k = 1 - n_i^{-k}$. The total gross gain from trade between matched partners is normalized to 1, and trade benefits of type- i agents, u_i are shared linearly across users, so that $u_1 + u_2 = 1$. I further assume that type-2 users value more gains from trade than type 1 users: $u_1 \leq 1/2 \leq u_2$. Platform's k can invest $x^k \in [0, 1]$ in the development of a matching technology; this innovation can be thought as an algorithm that, by exploiting users' information available to the matchmaker, improves her ability to couple one side with the coherent member on the other bank of the river. It follows that x^k can be interpreted as an R&D investment. In a second stage, the intermediary sets membership fees p_i^k to allow i -agents to access to her network. She also levies a total transaction fee t^k , conditionally on the occurrence of the trade.

When visiting platform k , each i -user faces a random probability $\lambda^k n_j$ of being matched with the desired counterpart on the other side of the market. In particular, λ^k represents the quality of the matching technology adopted by the platform; it can be viewed as an algorithm that helps each side find the proper counterpart of the other type. We assume that λ^k is uniformly distributed, and that the effectiveness of the matching algorithm depends on the number of users participating to the network:

$$\lambda^k \sim \mathcal{U} \left(\frac{n_i^k + n_j^k + x^k}{3}, 1 \right) \quad (1)$$

The distribution of (1) implies that the expected probability of match is increasing in the demand of the firm and on the innovative investment. The former feature is a key component of, e.g., algorithms such as PageRank, which require great amount of data points to deliver satisfactory results. Another example of algorithms that facilitate the trade among sides are the pricing algorithms that are adopted by sellers on different marketplaces, such as Amazon: while these algorithms are developed independently either by the sellers themselves or by third parties that provide turnkey solutions, the backbone infrastructure for their development can be supplied by the marketplace (an example being Amazon Marketplace Web Service), and it requires a huge amount of data itself to facilitate the process of price

automation. For an exhaustive discussion on pricing algorithms and algorithmic collusion, see Calvano, Calzolari, De Nicolò and Pastorello (2019).

As the matching technology described here has the scope to maximize the probability of a successful encounter, defining λ^k as increasing in the demand of each matchmaker implies that, for improvements to be possible, apart from technical upgrades, which are embedded in x^k , the algorithm needs to be "fed" with users' information.

This modelling choice is consistent with the representation of data as *users' information* rather than a direct *quality augmenting* investment; while the latter scenario is more apt to describe the competitive framework in industries such as the social media, where data are indeed used as a tool to offer better demand targeting services to advertisers and hence are fundamental to improve objectively the product sold to the non-subsidized group, the former allows to abstract from product differentiation. Indeed, in this setting each side is intrinsically homogenous in their preferences: the only relevant feature for ascertaining the quality of the service of the intermediary is its ability to pair users. While this refinement is generally adopted to competition between *matchmaking* platforms such as Meetic or Tinder *exchange* platforms, such as eBay, Amazon or Alibaba, should be analyzed through the same lens, as the mounting degree of standardization in the service they provide is not accordant with the formal definitions of vertical or horizontal differentiation. The classification of the search engine industry, within the frameworks represented by the aforementioned models, is more blurred; while the idea of matching-improving investment seems coherent with the nature of the service offered to users, as an algorithm such as PageRank displays results of keyboard searches based on relevance scores, and hence increases the probability of finding the desired link, it could be argued that users hold the same preferences with respect to providers, in so far they understand objectively which search engine grants the better service. This consideration is conceptually aligned with the definition vertical differentiation, and leaves room for the identification of hybrid models, which are out of the scope of this work.

I focus on price vectors $P^k = (p_1^k, p_2^k, t^k)$ such that agents enjoy non-negative surplus from trading, i.e.

$$\lambda^k u_i (1 - t^k) - p_i^k \geq 0 \tag{2}$$

Let $P = \{P^I, P^E\}$ be the price system resulting from the optimal decision of each firm. Let $\mathcal{N} = \{n_i^I, n_i^E\}_{i=1,2}$ denote a distribution of agents across platforms, with n_i^k being the proportion of type- i users participating to network k . Finally, let $\mathcal{X} = \{x^I, x^E\}$ be a vector of profit maximizing investments. Then, the net indirect utility of an i -agent registering with matchmaker k will be given by

$$U_i(P, k, \mathcal{N}, \mathcal{X}) = \lambda^k u_i n_j^k (1 - t^k) - p_i^k \quad (3)$$

for $j \neq i$. Equation (3) points out that i -users derive utility from the participation of j -users, i.e. that we are describing a setting characterized by cross-group externalities. Notice that the transaction fee t^k doesn't enter linearly the disutility the agent faces when visiting a platform. This feature arises because agents derive positive utility from transacting on the platform, and hence the transaction fee, being an intermediation fee, must be weighted for the surplus users derive from a successful match. Platforms can store information on users for free through their servers, and their expenditure depends only on the investment in the development of a new matching technology; to characterize competition among *data-driven* business models, I define the cost function of each firm as linearly decreasing in her demand, i.e. in the total number of participating users:

$$C(x^k, n_1^k + n_2^k) = \frac{\gamma}{2}(x^k)^2 + \alpha x^k [2 - (n_1^k + n_2^k)],$$

where $\gamma > 0$ and $\alpha \in [0, 1]$ are two parameters that capture, respectively, the difficulty to innovate, and hence to produce valuable insights on customers' behavior (lower values γ imply lower difficulty) and the intensity of data-driven effects, or the degree to which information on new users smooth costs of producing an innovative matching technology (lower values of α imply strong data-driven effects).

Firm's k profits are given by:

$$\pi^k(P^k, \mathcal{N}, \mathcal{X}) = \sum_i^2 n_i^k p_i^k + E(\lambda^k) n_1^k n_2^k t^k - C(x^k, n_1^k + n_2^k)$$

Timing and Equilibrium Definition I model the interaction between the incumbent and the follower as 3 stage game, which resembles a sequential entry game:

1. In stage 1, platforms set simultaneously and non cooperatively the R&D investments $\mathcal{X} = (x^I, x^E)$.
2. In stage 2, platforms set simultaneously and non cooperatively $P^k = (p_1^k, p_2^k, t^k)$.
3. In stage 3, users decide which platform they will visit based on their expected indirect utility, V_i^k .

It is important to stress that the decision of users is determined by the expected value of (3), i.e. $V_i(P, k, \mathcal{N}, \mathcal{X}) = E[U_i(P, k, \mathcal{N}, \mathcal{X})]$. I assume that every user visits one of the two intermediaries, which implies that $\sum_{k=I,E} \sum_i^2 n_i^k = 2$. The timing of the game is justified

by the lower flexibility that R&D investments typically display when compared to pricing decisions.

To characterize the Subgame Perfect Nash Equilibria of the game, we need first to formalize a definition of optimal market allocation.

Definition 3.1. An allocation of users \mathcal{N} is an equilibrium distribution for a price system P and an investment vector \mathcal{X} if, for all $k \in \{I, E\}$

$$n_i^k \implies V_i(P, k, \mathcal{N}, \mathcal{X}) = \max_{h \in \{E, I\}} V_i(P, h, \mathcal{N}, \mathcal{X})$$

A market allocation is a mapping $\mathcal{N}(\bullet)$ that associates, for a given price system P , a distribution $\mathcal{N}(P)$ of users across platforms. To rule out demand functions that are increasing in P^k , I formally state that a candidate equilibrium distribution \mathcal{N} must be non-increasing in P^k .

Definition 3.2. A market allocation $\mathcal{N}(\bullet)$ is monotone if, $\forall k$, $n_i^k(P^k, P^{-k})$ is non-increasing in P^k .

Definition 3.3. An equilibrium is a triple $(P_*, \mathcal{N}, \mathcal{X})$ such that:

- i $\mathcal{N}(\bullet)$ is a monotone market allocation.
- ii P_*^I and P_*^E are Nash equilibria of the respective pricing subgame.
- iii \mathcal{X} are Nash equilibria of the respective investment subgame.

Competition under pessimistic beliefs To identify the Subgame Perfect Nash Equilibria of the game, I start by focusing on settings where users hold ex-ante bad expectations regarding the entrant platform; in a nutshell, they expect that every user will visit platform I , implying $n_i^E = 0$ and $n_i^I = 1$. Pessimistic beliefs are representation of the incumbency advantage that an entrenched dominant faces due to network externalities in many two-sided, digital industries; as users understand that the value they can get from another matchmaker exists only in so far a substantial number of agents of the other side participates to the new network, in principle they are skeptical with respect to other, non-focal intermediaries. The following proposition characterizes the optimal behavior of firm E .

Proposition 3.1. *Under pessimistic beliefs, there exists a dominant firm Subgame Perfect Nash Equilibrium where I captures the whole market by setting any combination of prices that satisfies (i) $p_1^I + p_2^I = -\frac{5}{6}$ with $p_1^I < p_2^I$, $t^I = 1$ and (ii) investment $x^I = \frac{1}{6\gamma}$, while E exits the market. In this setting, (iii) I earns strictly positive profits, $\pi^I = \frac{1}{72\gamma}$.*

Proof. The characterization of the pricing strategy is adapted from Caillaud and Jullien (2001a and 2001b).

For a bad expectation distribution of users to exist, it must be that the expected indirect utility granted by platform I always exceeds the one provided by visiting platform E , for every user. This will hold if:

$$E(\lambda^I) u_i (1 - t^I) - p_i^I \geq -p_i^E$$

It follows that the entrant, in order to get a positive market share, must adopt a Divide and Conquer (hereafter, DC) strategy: it must subsidize a group of users to get them on board, and then profitably attract the other, exploiting the existence of network externalities. As $u_1 \leq 1/2 \leq u_2$, it is rational to subsidize type-1 agents, which have lower willingness to pay, to "milk" type-2 users. E subsidizes type-1 users by setting:

$$p_1^E < p_1^I - E(\lambda^I) u_i (1 - t^I) \leq 0 \quad (4)$$

For this pricing strategy, E will overcome pessimism and obtain $n_1^E = 1$, regardless from the ex-ante expectations. Hence, to attract type-2 agents, E will adopt a pricing policy such that $V_2^E > V_2^I$. This condition yields:

$$p_2^E + E(\lambda^E) u_2 t^E < E(\lambda^E) u_2 + \inf\{p_2^I, 0\} \quad (5)$$

Letting $r_i^k = p_i^k + E(\lambda^k) n_j^k u_i t^k$, (5) can be re-written as:

$$r_2^E < E(\lambda^E) u_2 + \inf\{p_2^I, 0\} \quad (6)$$

Notice that t^E doesn't appear in (4): besides, as p_i^k enters linearly in the (dis)utility function of i -users, while they pay only a fraction u_i of disutility due to trade gains, attracting customers by lowering membership fees is much less costly than relying on transaction fees to achieve the same purpose. It follows that E will set maximal transaction fee, $t^E = 1$, as stated in part (i). Therefore, E will attract 1-users by setting $p_1^E = p_1^I - E(\lambda^I) u_1 (1 - t^I)$ (slightly below) and $r_2^E = E(\lambda^E) u_2 + \inf\{p_2^I, 0\}$ (slightly below); indeed, it will be sufficient for E to give to type-1 users a subsidy greater than the amount of utility that platform I

could grant them if everyone agglomerates on I to get everyone on board, whereas type-2 users need to be priced so that the amount of utility they receive from participating in platform E , assuming now agglomeration on the entrant's network, is strictly greater than the utility they would receive from the incumbent.

To conquer the industry, I must drive E 's gains to 0, earn non-negative profits and grant users non-negative utility. These considerations yield the following system of conditions:

$$\begin{aligned}\pi^E &= p_1^E + E(\lambda^E) u_1 t^E + \underbrace{p_2^E + E(\lambda^E) u_2 t^E}_{r_2^E} - C(x^E, 2) \leq 0 \\ \pi^I &= p_1^I + p_2^I + E(\lambda^I)(u_1 + u_2)t^I - C(x^I, 2) \geq 0 \\ V_i^I &= E(\lambda^I) u_i(1 - t^I) - p_i^I \geq 0\end{aligned}$$

Plugging p_1^E and r_2^E in the system, and recalling that it must be that $t^E = 1$, we can re-arrange it as:

$$\pi^E = p_1^I + \inf\{p_2^I, 0\} + E(\lambda^E)(u_1 + u_2) - E(\lambda^I) u_1(1 - t^I) - C(x^E, 2) \leq 0 \quad (7)$$

$$\pi^I = p_1^I + p_2^I + E(\lambda^I)(u_1 + u_2)t^I - C(x^I, 2) \geq 0 \quad (8)$$

$$V_i^I = E(\lambda^I) u_i(1 - t^I) - p_i^I \geq 0 \quad (9)$$

Applying an analogous reasoning as above, and noting that t^k is non-negative, I will react to the DC strategy by setting maximal transaction fee, $t^I = 1$. This implies, from (9), that p_i^I will be non-positive. Given these considerations, and noting that it must be that $p_2^I \leq 0$, the system can be rewritten as:

$$p_1^I + p_2^I \leq C(x^E, 2) - E(\lambda^E) \quad (10)$$

$$p_1^I + p_2^I \geq C(x^I, 2) - E(\lambda^I) \quad (11)$$

$$E(\lambda^I) u_i(1 - t^I) - p_i^I \geq 0 \quad (12)$$

Conditions (10) and (11) imply:

$$C(x^I, 2) - E(\lambda^I) \leq p_1^I + p_2^I \leq C(x^E, 2) - E(\lambda^E) \quad (13)$$

As $p_1^I + p_2^I \leq C(x^E, 2) - E(\lambda^E)$, and as I 's profits must be non negative, it must be that:

$$0 \leq \pi^I \leq C(x^E, 2) - E(\lambda^E) - C(x^I, 2) + E(\lambda^I) \quad (14)$$

(14) implies that I 's profits are non negative if:

$$E(\lambda^I - \lambda^E) \geq C(x^I, 2) - C(x^E, 2) \quad (15)$$

Taking expectation in (15) at $n_i^k = 1$, we have:

$$\begin{aligned} \frac{1}{2} \left[1 + \frac{1}{3}(2 + x^I) - 1 - \frac{1}{3}(2 + x^E) \right] &\geq \frac{\gamma}{2}(x^I)^2 - \frac{\gamma}{2}(x^E)^2 \\ \frac{1}{3}(x^I - x^E) &\geq \gamma(x^I - x^E)(x^I + x^E) \end{aligned} \quad (16)$$

Solving (16) for x^E , it must be that I 's profit are non-negative if:

$$\begin{cases} x^E \leq \frac{1}{3\gamma} - x^I, & \text{if } x^I > x^E \\ x^E \geq \frac{1}{3\gamma} - x^I, & \text{if } x^I < x^E. \end{cases}$$

It follows that E 's best response is defined as:

$$x^E(x^I) = \begin{cases} x^E > \frac{1}{3\gamma} - x^I, & \text{if } x^I > x^E \\ x^E < \frac{1}{3\gamma} - x^I, & \text{if } x^I < x^E. \end{cases}$$

Maximizing $E(\lambda^I) - C(x^I, 2)$, the profit maximizing investment for I is identified at $x^I = \frac{1}{6\gamma}$ (it can be easily shown that this is a maximum and that the second order conditions hold). Thus, consider the candidate equilibrium $x^I = \frac{1}{6\gamma}$:

- If $x^E > \frac{1}{6\gamma}$, by definition of maximizer, $C(x^I, 2) - E(\lambda^I) < C(x^E, 2) - E(\lambda^E)$. Therefore, in the second stage of the game, I can always set $p_1^I + p_2^I = C(x^I, 2) - E(\lambda^I)$ and win the market. Due to the investment in period 1, E 's profits will be strictly negative.
- If $x^E < \frac{1}{6\gamma}$, an analogous reasoning applies.
- If $x^E = \frac{1}{6\gamma}$, the LHS of (13) implies that I sets $p_1^I + p_2^I = -\left(\frac{1}{72\gamma} + \frac{5}{6}\right)$ and wins the market, while earning zero profits. Firm E , however, would still be strictly in the red, because of the initial investment.

Therefore, it must be that $\mathcal{X} = \{\frac{1}{6\gamma}, 0\}$. Recalling that (13) implies that the upper bound on $p_1^I + p_2^I$ is $C(x^E, 2) - E(\lambda^E)$, so that E cannot profitably subsidize i -agents without investing, and as it must be that $x^E = 0$, prices will need to satisfy:

$$\begin{aligned} p_1^I + p_2^I &= C(x^E, 2) - E(\lambda^E) \\ &= -\frac{1}{2} \left[1 + \frac{2}{3} \right] \\ &= -\frac{5}{6} \end{aligned} \tag{17}$$

Any combination that attains (17), with $p_1^I < p_2^I$, will be optimal. Hence, I will set $p_1^I + p_2^I$ slightly below $-\frac{5}{6}$.

Finally, I profits are identified as:

$$\begin{aligned} \pi^I &= p_1^I + p_2^I + E(\lambda^I) - C(x^I, 2) \\ &= -\frac{5}{6} + \frac{1}{2} \left[1 + \frac{1}{3} \left(2 + \frac{1}{6\gamma} \right) \right] - \frac{\gamma}{2} \left(\frac{1}{6\gamma} \right)^2 \\ &= \frac{1}{72\gamma}. \end{aligned} \tag{18}$$

■

Proposition 3.1 outlines the relevance of the incumbency advantage in two-sided markets when data-driven business models are considered; indeed, differently from Caillaud and Jullien (2003), the incumbent's profits are strictly positive; the results is explained in so far the competitive pressure generated by the cross-group network externalities is partially buffered by the increasing surplus that the platform can deliver to their user due to her ability to derive insights on their preferences. An implication of this outcome is that the incumbent can subsidize *less* her users, compared to an analogous setting with exogenous matching technology. It should be noted that the incumbent's profitability doesn't arise from a strategically deterrent investment decision; indeed, the exclusion of firm E from the market arises from a profit maximizing conduct. This finding suggests that industries characterized by both data cost synergies and network externalities should be scrutinized carefully by antitrust authorities, in so far they display a natural tendency to generate failures. In particular, horizontal mergers in these settings should be inspected reversing the famous error-cost paradigm. According to this view, absent a strong presumption of dominance outlined by an excessively high HHI or CR index, competition policy in highly innovative scenarios should be always aimed at minimizing the losses, in terms of decreased

R&D expenditure, from False Positives, i.e. authorities should maintain a laissez-faire profile and intervene only when inefficiencies can be clearly showed, imposing the burden of the proof on the plaintiff. In the framework discussed previously, however, a potential misspecification of the relevant market could create an established incumbent that would very likely enjoy the incumbency advantage encapsulated by the pessimistic beliefs. Advocates of the mainstream error-cost framework claim that the competitive push *for* the market favored by the cross-group network externalities would determine an efficient outcome, and hence that these industries should be considered as highly contestable. However, this result shows that data barriers can relax substantially the competition *in* the market, and the argument of constestability might not apply. Hence, the cost of potential type I errors is two-fold: on one hand, the loss in consumer surplus that arises from this monopolistic structure: on the other, the potential disincentive to produce innovative solutions that both the leader and new start ups would face over time, which has been discussed by Bryan and Hovenkamp (2020) . I will deal with the relation between innovation levels and competition *for* the industry in the next paragraph.

Therefore, the task of showing the pro-competitiveness of acquisitions should lie on the merging entity's shoulders, and, given the fallacy of the constestability defense in this context, players should show that the remaining competitors in the market can represent a credible threat.

Negative membership fees can be thought as free access to a platform, or access granted at cost for the platform itself. This is a common feature of many two-sided intermediaries: for instance, Airbnb provides residential sharing services without imposing any membership fee neither on hosts nor on guests. However, to ensure the functioning of the customer service for 24 hours per day, the company incurs in costs that are recouped from a 3% fee on hosts on every booking, while a 14 to 20% on the amount of the reservation is charged to both sides ⁴. Even if the former fee is levied on each transaction, it is *de facto* a masked membership fee, in so far it covers indirect participation costs. The meaning $p_1 < p_2$ is enshrined in this differential treatment: indeed, guests are receiving participation services at cost for the firm, implying they are receiving a surplus transfer in economic terms.

Lastly, it should be pointed out that Proposition 3.1 holds for $\gamma > 1/3$; indeed, to characterize a feasible best response function for E , it must be that the difficulty to innovate is sufficiently high, otherwise the model would predict investments above 1.

After characterizing the nature of the incumbency advantage, the next step is to analyze whether segmentation equilibria exist under singlehoming.

⁴see https://www.airbnb.co.uk/help/article/1857/what-is-the-airbnb-service-fee?_set_bev_on_new_domain=1591216545_NTE1NGNkNDYzZjkz

Segmentation Equilibria To prove the existence of equilibria with two matchmakers active, I will proceed by excluding unsustainable scenarios; I will restrict the focus of the analysis on strategies that are stable against deviation, and, at the end of the paragraph, I will identify the Subgame Perfect Nash Equilibrium of the game under symmetric beliefs. The next proposition rules out asymmetric equilibria, where one firm k holds a fraction of type- i agents greater than the fraction of type- j agents that participate to her network. To facilitate the discussion, let Δn_i be the differential between the leader on a side of the market and the follower: for example, if platform I has the lead on side 1, whereas platform E has the lead on side 2, it will be $\Delta n_1 = n_1^I - n_1^E$ and $\Delta n_2 = n_2^E - n_2^I$. Besides, let $E(\lambda_D^k)$ denote the expected effectiveness of the matching technology under a duopolistic outcome, and $E(\lambda_M^k)$ be the monopolistic equivalent. It shall also be reminded that every user, in the last stage of the game, will select an intermediary she will visit, and hence that there is no user unserved. It's worth anticipating how the logical path of the analysis will unfold: Proposition 3.2 will rule out the existence of equilibria where one firm holds a fraction of the market, on either side, different from 1/2. Proposition 3.3 will restrict even further the set of possible solutions, eliminating scenarios where both firms set a transaction fee different from the maximal one, and split the market. At this point, we will be left with one set of solutions, where both firms engage in DC strategies: Proposition 3.4 will serve to characterize the Subgame Perfect Nash Equilibrium under symmetric beliefs.

Proposition 3.2. *If no firm faces an incumbency advantage, i.e. if beliefs are symmetric, no segmentation equilibrium with $n_i^k \neq n_j^k$ exists.*

Proof. Firstly, consider the case where $n_1^I > n_2^I$, $n_1^E < n_2^E$, $n_1^I > n_2^E$. Noting that these assumptions imply that $\Delta n_1 > \Delta n_2$, the profits of firm k can be written as:

$$\pi^k = n_1^k p_1^k + n_2^k p_2^k + E(\lambda_D^k) n_1^k n_2^k t^k - \frac{\gamma}{2} (x^k)^2 - \alpha x^k [2 - (n_1^k + n_2^k)] \quad (19)$$

For any segmentation equilibrium to be sustainable in this setting, as agents of the same type are homogenous, it must be that users are indifferent between the two matchmakers, i.e. $V_1^I = V_1^E$ and $V_2^I = V_2^E$. It follows that the following system of conditions must be satisfied:

$$E(\lambda_D^I) n_2^I u_1(1 - t^I) - p_1^I = E(\lambda_D^E) n_2^E u_1(1 - t^E) - p_1^E \quad (20)$$

$$E(\lambda_D^I) n_1^I u_2(1 - t^I) - p_2^I = E(\lambda_D^E) n_1^E u_2(1 - t^E) - p_2^E \quad (21)$$

Let $\rho_i^k = p_1^k + E(\lambda_D^k) n_j^k u_i t^k$ be the revenue per user that firm k derives from type- i agents. Then, (20) and (21) can be written as:

$$\begin{aligned}\rho_1^I - E(\lambda_D^I) n_2^I u_1 &= \rho_1^E - E(\lambda_D^E) n_2^E u_1 \\ \rho_2^I - E(\lambda_D^I) n_1^I u_2 &= \rho_2^E - E(\lambda_D^E) n_1^E u_2\end{aligned}$$

Re-arranging:

$$\rho_1^E - \rho_1^I = u_1 [E(\lambda_D^E) n_2^E - E(\lambda_D^I) n_2^I] \quad (22)$$

$$\rho_2^I - \rho_2^E = u_2 [E(\lambda_D^I) n_1^I - E(\lambda_D^E) n_1^E] \quad (23)$$

Taking the first order conditions of (19) with respect to x^k , we find that the profit maximizing investment is increasing in the demand of firm k :

$$x^k = \frac{n_1^k n_2^k t^k - 6\alpha [2 - (n_1^k + n_2^k)]}{6\gamma} \quad (24)$$

As $\Delta n_1 > \Delta n_2$ and $n_1^I > n_1^E$, it must be that the demand of firm I is greater than the demand of firm E . It follows that $x^I > x^E$ and $E(\lambda_D^I) > E(\lambda_D^E)$.

Comparing the RHS of (22) and (23), as $u_1 < u_2$, $\Delta n_1 > \Delta n_2$ and $E(\lambda_D^I) > E(\lambda_D^E)$, it must necessarily be that:

$$\rho_2^I - \rho_2^E > \rho_1^E - \rho_1^I$$

Then,

$$\rho_1^I + \rho_2^I > \rho_1^E + \rho_2^E \quad (25)$$

(19) can be written as:

$$\pi^k = n_1^k \rho_1^k + n_2^k \rho_2^k - \frac{\gamma}{2} (x^k)^2 - \alpha x^k [2 - (n_1^k + n_2^k)] \quad (26)$$

(25) and (26) imply that asymmetric equilibria cannot be sustained: indeed, for any investment level x^E , in stage 2 E would always have an incentive to copy the incumbent's pricing structure, slightly undercut it on side 1 and get the whole market. As costs are decreasing in the demand of the firm, if E conquers the whole market, for the same investment level, her total expenditure will be lower; besides, (25) implies that the total revenues that E would get if she would monopolize the market copying I 's pricing structure will be strictly greater than the actual revenue: her profits, as a consequence, will be greater, if she deviates from her actual strategy. Formally: if E sticks to the segmentation strategy, she would get profits π_S^E equal to:

$$\pi_S^E = n_1^E \rho_1^E + n_2^E \rho_2^E - \frac{\gamma}{2} (x^E)^2 - \alpha x^E [2 - (n_1^E + n_2^E)]$$

If she subsidizes slightly more one of the two groups to make the market tip, keeping the same pricing structure, she would get the deviation profits $\pi_{D^E}^E$:

$$\pi_{D^E}^E = \rho_1^E + \rho_2^E - \frac{\gamma}{2} (x^E)^2$$

However, if she deviates from the segmentation strategy copying the incumbents pricing structure, her payoff $\pi_{D^I}^E$ is described as:

$$\pi_{D^I}^E = \rho_1^I + \rho_2^I - \frac{\gamma}{2} (x^E)^2$$

As (25) implies that $\pi_{D^I}^E > \pi_{D^E}^E$, it follows that any equilibrium with $\rho_1^I + \rho_2^I > \rho_1^E + \rho_2^E$ cannot be sustained, as E would always prefer to copy the incumbents pricing structure.

Comparing (22) and (23), as $u_1 = u_2$, we can see that even in the symmetric case, $\Delta n_1 = \Delta n_2$ and $E(\lambda_D^E) = E(\lambda_D^I)$, inequality (25) would hold, implying that the same conclusion follows.

Secondly, consider the case where $n_2^I > n_1^I$, $n_2^E < n_1^E$, $n_2^I > n_1^E$, which imply that $\Delta n_2 > \Delta n_1$. Noting that these inequalities imply also that the demand of firm I will exceed the demand of firm E , it must be that $x^I > x^E$, so that $E(\lambda_D^I) > E(\lambda_D^E)$. Then, using the indifferent type- i agent condition, we can rearrange (22) and (23) as follows:

$$\rho_1^I - \rho_1^E = u_1 [E(\lambda_D^I) n_2^I - E(\lambda_D^E) n_2^E] \quad (27)$$

$$\rho_2^E - \rho_2^I = u_2 [E(\lambda_D^E) n_1^E - E(\lambda_D^I) n_1^I] \quad (28)$$

At this point, recalling that $u_1 + u_2 = 1$ the main task is to compare the RHS of (27) and (28):, we want to establish what kind of relation exists between $\rho_1^I - \rho_1^E$ and $\rho_2^E - \rho_2^I$. Recalling that $u_1 < \frac{1}{2} < u_2$ and $u_1 + u_2 = 1$, we have:

$$\begin{aligned} u_1 [E(\lambda_D^I) n_2^I - E(\lambda_D^E) n_2^E] &\stackrel{?}{=} u_2 [E(\lambda_D^E) n_1^E - E(\lambda_D^I) n_1^I] \\ E(\lambda_D^I) n_2^I - E(\lambda_D^E) n_2^E &\stackrel{?}{=} u_2 [E(\lambda_D^I) (n_2^I - n_1^I) - E(\lambda_D^E) (n_1^E - n_2^E)] \\ E(\lambda_D^I) [n_2^I - u_2 (n_2^I - n_1^I)] &\stackrel{?}{=} E(\lambda_D^E) [n_2^E - u_2 (n_1^E - n_2^E)] \end{aligned}$$

As $E(\lambda_D^I) > E(\lambda_D^E)$, to prove that (27) > (28) it is sufficient to show that $n_2^I - u_2 (n_2^I - n_1^I) > n_2^E - u_2 (n_1^E - n_2^E)$:

$$\begin{aligned}
n_2^I - u_2 (n_2^I - n_1^I) &\stackrel{?}{=} n_2^E - u_2 (n_1^E - n_2^E) \\
n_2^I &\stackrel{?}{=} n_2^E + u_2 (1 - n_2^E - 1 + n_1^E - n_1^E + n_2^E) \\
n_2^I &> n_2^E
\end{aligned} \tag{29}$$

(29) implies that $\rho_1^I + \rho_2^I > \rho_1^E + \rho_2^E$. It follows that an analogous argument as the one exposed in the first case applies. As we have ruled out every possible scenario, the statement is proved. ■

Proposition 3.2 implies that if segmentation equilibria exist, they will only be consistent with $n_i^k = n_j^k = \frac{1}{2}$. The aim of the next proposition will be to further restrict the focus on DC strategies symmetric equilibria, in order to finally derive the Subgame Perfect Nash Equilibria of the game.

Proposition 3.3. *Under symmetric beliefs, any equilibrium where firms set $t^k \neq 1$ and split the market cannot be sustained.*

Proof. Consider the market allocation $\mathcal{N} = \{\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$. Due to symmetry, for this allocation to be an equilibrium, it must be that users are indifferent between the two intermediaries. Additionally, profits of firm k must be non-negative. These considerations leave us with the following system of conditions:

$$\begin{aligned}
\pi^k \geq 0 &\iff \frac{1}{2} (p_1^k + p_2^k) + \frac{1}{4} E (\lambda_D^k) t^k - \frac{\gamma}{2} (x^k)^2 - \alpha x^k \geq 0 \\
V_i^k = V_i^{-k} &\iff E (\lambda_D^k) \frac{1}{2} u_i (1 - t^k) - p_i^k = E (\lambda_D^{-k}) \frac{1}{2} u_i (1 - t^{-k}) - p_i^{-k}
\end{aligned} \tag{30}$$

Our goal is to rule out equilibria where both firms do not adopt DC strategies and split the market: hence, we need to compare the outcome π_C^k when firms decide to split the market, in a cooperative fashion, to the payoff each player would receive if they would deviate and monopolize the industry, π_D^k . Recalling that $V_i^k \geq 0$, it must be that, in the cooperative setting, firms set $t^k = 0$ and $p_i^k = E (\lambda_D^k) \frac{1}{2} u_i$; as the marginal revenue from the transaction fee is always lower than the one from the membership fee, if firms do not opt for exclusionary strategies in stage 2, then their optimal choice will be to set t^k at the minimum possible value. Hence, by plugging the expression for p_i^k into (30), the profit maximizing investment is identified by taking the first order conditions, which yield:

$$\begin{aligned}\frac{1}{24} - \gamma x^k - \alpha &= 0 \\ x^k &= \frac{1}{24\gamma} - \frac{\alpha}{\gamma}\end{aligned}\tag{31}$$

Plugging (31) and p_i^k into (30), profits are:

$$\begin{aligned}\pi_C^k &= \frac{1}{6} + \frac{1}{24} \left(\frac{1-24\alpha}{24\gamma} \right) - \frac{\gamma}{2} \left(\frac{1-24\alpha}{24\gamma} \right)^2 - \alpha \left(\frac{1-24\alpha}{24\gamma} \right) \\ &= \frac{1}{6} + \left(\frac{1-24\alpha}{24\gamma} \right) \left(\frac{1}{24} - \frac{1-24\alpha}{48} - \frac{48\alpha}{48} \right) \\ &= \frac{1}{6} \left\{ 1 + \frac{1}{8} \left[\frac{(1-24\alpha)^2}{24\gamma} \right] \right\}\end{aligned}\tag{32}$$

If firm k sets $p_1^k = E(\lambda_D^k) \frac{1}{2} u_1$ (slightly below), it gets:

$$\begin{aligned}\pi_D^k &= \frac{5}{24} + \frac{1}{24} \left(\frac{1-24\alpha}{24\gamma} \right) - \frac{\gamma}{2} \left(\frac{1-24\alpha}{24\gamma} \right)^2 \\ &= \frac{5}{24} + \left(\frac{1-24\alpha}{24\gamma} \right) \left(\frac{1+24\alpha}{48} \right) \\ &= \frac{1}{6} \left\{ \frac{5}{4} + \frac{1}{8} \left[\frac{(1-24\alpha)(1+24\alpha)}{24\gamma} \right] \right\}\end{aligned}\tag{33}$$

Comparing (32) and (33), by straightforward calculation it holds that the deviation profits are greater than the market sharing ones. ■

An important remark must accompan Proposition 3.3: as equation (31) suggests, for a positive R&D investment to exist in a setting with more than one player active, it must be that $\alpha < \frac{1}{12}$: as the marginal benefit of investing is $\frac{1}{6}$, for firms to invest, it must be that the strength of the data driven cost synergies is sufficiently strong.

Proposition 3.3 also suggests that collusive equilibria, in settings where both network externalities and data driven effects are present, can hardly be a threat; indeed, as the cost benefit from conquering the market, for any given R&D expenditure, is so enticing, the prospect of market sharing are negligible.

After excluding alternatives that are not robust against deviations, it shall be remarked that the remaining equilibria imply DC strategie; as firms cannot credibly use membership fees to extract surplus from their users, they are forced by the existence of network externalities to compete aggressively. The next proposition will finally characterize the Subgame Perfect Nash Equilibria of the game under symmetric beliefs.

Proposition 3.4. *Under symmetric beliefs, there exists no Subgame Perfect Nash Equilibrium in pure strategies. There exists mixed strategies equilibria were both matchmakers (i) set $t^k = 1$, (ii) invest $\mathcal{X} = \{\frac{1}{6\gamma}, \frac{1}{6\gamma}\}$, and (iii) randomize membership fees such that $(p_1^k + p_2^k) \in \left[-\frac{5}{6} - \frac{1}{72\gamma}, -\frac{2}{3} - \frac{1-12\alpha}{72\gamma}\right]$.*

Proof. As proved in Proposition 3.1, DC strategies equilibria are only consistent with $t^k = 1$. To win the market, from (10) and (11), firm k has to satisfy the following system of conditions:

$$p_1^k + p_2^k \leq C(x^{-k}, 2) - E(\lambda_M^{-k}) \quad (34)$$

$$p_1^k + p_2^k \geq C(x^k, 2) - E(\lambda_M^k) \quad (35)$$

$$E(\lambda_M^k) u_i(1 - t^k) - p_i^k \geq 0 \quad (36)$$

Thus, following the same line of reasoning that brought to (13) the profits of firm k will be slightly below the following expression:

$$\pi^k = C(x^{-k}, 2) - E(\lambda_M^{-k}) + E(\lambda_M^k) - C(x^k, 2) \quad (37)$$

Consider a candidate equilibrium were firm $-k$ invests an amount different from the monopolistic profit maximizing investment, $x^k = \frac{1}{6\gamma}$. As $\frac{1}{6\gamma}$ maximizes $E(\lambda_M^k) - C(x^k, 2)$, it must be that it minimizes $C(x^k, 2) - E(\lambda_M^k)$. It follows that for any investment decision $x^{-k} \neq \frac{1}{6\gamma}$, firm k will always be able to set $p_1^k + p_2^k \leq C(x^{-k}, 2) - E(\lambda_M^{-k})$, so that firm $-k$ is driven out of the market and (37) is strictly positive: indeed, as costs are decreasing in the demand of the firm, a profit maximizing investment under monopolistic agglomeration will grant a profit margin greater than any possible alternative, and firms will use it to subsidize a side of the market sufficiently to drive out any competitor that didn't invest as much. Since network externalities will trigger, the winner will successfully attract the other bank of the river, provided that she doesn't set membership fees on the second side so high that it would be profitable for the entrant to apply a DC strategy. Hence, it must be that both firms invest $\frac{1}{6\gamma}$. From Proposition 3.1, we now that the lowest possible total subsidy that a monopolist can offer to her user base is $p_1^k + p_2^k = -\frac{5}{6} - \frac{1}{72\gamma}$. However, if both firms offer this combination of membership fees, agents would be indifferent between the two platform, and hence would split. If such separation verifies, both firms would incur in losses:

$$\begin{aligned}
\pi^k &= -\frac{5}{6} - \frac{1}{72\gamma} + \frac{4}{6} + \frac{1}{36\gamma} - \frac{\gamma}{2} \left(\frac{1}{6\gamma}\right)^2 - \alpha \left(\frac{1}{6\gamma}\right) \\
&= -\frac{1}{6} + \frac{1}{6\gamma} \left(\frac{1}{6} - \frac{1}{12} - \frac{1}{12} - \alpha\right) \\
&= -\frac{1}{6} \left(1 + \frac{\alpha}{\gamma}\right)
\end{aligned} \tag{38}$$

Therefore, both firms have an incentive to deviate from this strategy. The upper bound of the combined subsidy must be given by the combination of negative membership fees that would drive profits to zero if both firm are active and invest $\frac{1}{6\gamma}$:

$$\begin{aligned}
p_1^k + p_2^k &= \frac{\gamma}{2} \left(\frac{1}{6\gamma}\right)^2 + \alpha \left(\frac{1}{6\gamma}\right) - \frac{4}{6} - \frac{1}{6} \left(\frac{1}{6\gamma}\right) \\
&= -\frac{2}{3} + \left(\frac{1}{6\gamma}\right) \left(\frac{1}{12} + \alpha - \frac{1}{6}\right) \\
&= \frac{2}{3} - \frac{1 - 12\alpha}{72\gamma}
\end{aligned} \tag{39}$$

However, setting (39) can't be optimal as well: indeed, firm k would earn strictly positive profits by subsidizing slightly more group one and get the whole market. Thus, it must be that firms randomize fees in the interval $\left[-\frac{5}{6} - \frac{1}{72\gamma}, -\frac{2}{3} - \frac{1-12\alpha}{72\gamma}\right]$. ■

Lemma 3.1. *Under symmetric beliefs, every segmentation equilibria, i.e. where two match-makers are active, is characterized by both firms earning weakly negative profits, $\pi^k \leq 0$.*

Proof. Given the interval for $p_1^k + p_2^k$ specified in Proposition 3.4, and given that it must be $x^k = x^{-k} = \frac{1}{6\gamma}$ it follows that for every price system P where $p_1^k + p_2^k = p_1^{-k} + p_2^{-k}$ is below the upper bound, firms will earn negative profits. As discussed in Proposition 2.4, profits are zero when both the firms select prices consistent with the upper bound. ■

Lemma 3.2. *Under exclusive intermediation services (singlehoming), there exist no segmentation equilibria that can be socially optimal.*

Proof. Consider equilibria where firms set $p_1^k + p_2^k = -\frac{2}{3} - \frac{1-12\alpha}{72\gamma}$. As firms earn 0 profits, it is sufficient to prove that the consumer surplus is below the level identified in the pessimistic belief scenario to rule out this case:

$$2 \left(\frac{2}{3} + \frac{1 - 12\alpha}{72\gamma} \right) \stackrel{?}{=} \frac{10}{6}$$

$$\frac{1}{6} \left(\frac{1}{6\gamma} - 2 \right) - \frac{\alpha}{3\gamma} < 0$$

The last inequality is ensured by the condition $\gamma \geq \frac{1}{6}$, which is necessary to hold x^k below 1. As the social welfare under bad expectations is $\frac{5}{6} + \frac{1}{72\gamma}$, consider every other segmentation equilibria: in every case, firms would earn strictly negative profits, and the maximum consumer surplus can be attained when both set $p_1^k + p_2^k = -\frac{5}{6} - \frac{1}{72\gamma}$. Therefore, the statement must be true. ■

Proposition 3.4 and Lemmas 3.1 and 3.2 propose novel insights that deserve to be discussed thoroughly. First, the aggregate innovation levels are lower if an incumbent firm faces a belief advantage; the result is in line with Prüfer and Schottmüller (2017), which showed that, in a dynamic extension of their model, the entrant would stop fighting as the lead of the incumbent became impossible to overcome.

This finding indicates that applying Schumpeterian arguments to industries where data-driven effects are significant might not be appropriate; even if welfare maximization arises naturally from agglomeration, antitrust authorities should be wary of the entrenchment of excessively strong positions in this setting, in so far the spillover effects that high R&D intensity generates, as the empirical work of Levin (1988) suggests, can be conducive of technical progress.

Second, Lemma 2.1 propounds that competition with big data might feature *rational* strategies undertaken by *both* actors that lead to outcomes where the whole market is incurring in losses. The results arises because of data synergies; indeed, as the demand of the firms favors cost abatement, the pricing game in stage 2 will be characterized by a increased subsidies to users, in comparison with a setting where only cross group network effects are present. The investment, in turn, will increase user's willingness to pay for the service, thus making the prize of market tipping more appealing; actors are overinvesting in R&D (they are spending as monopolists, yet they face competition), in order to subsidy as much as possible each side, even at the cost of incurring in losses. These dynamics might provide new insights on a wide set of extravagant facts that are characterising some emerging firms within the boundaries of the sharing economy: Figure 2, retrieved from Orbis, exhibits the net income of the two existing firms in the United States in the ride hailing industry. As reported by *The Economist*, Uber, leader with a market share around 70%⁵, has been accumulating almost

⁵see Second Measure <https://secondmeasure.com/datapoints/rideshare-industry-overview/>

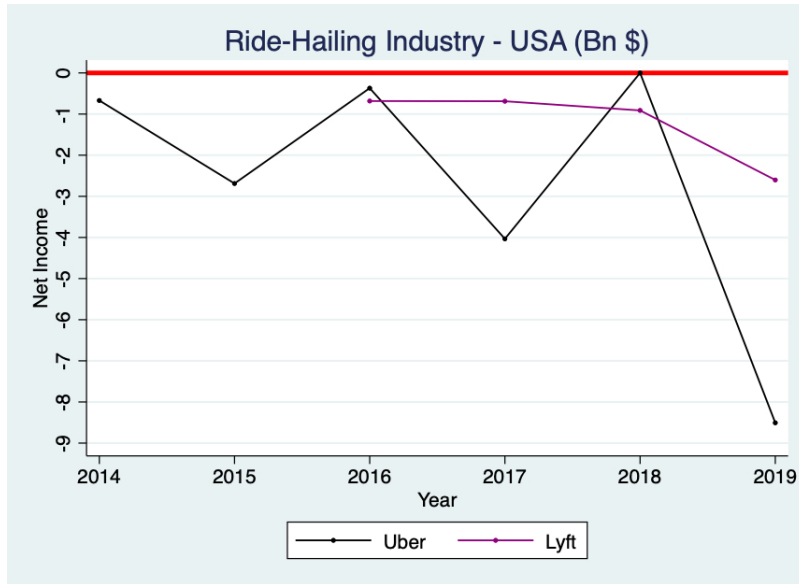


Figure 2: Net Income of the five biggest listed two-sided Marketplaces in the United States, excluding brick-and-mortar businesses. Source: Orbis.

30 bn \$ dollar of losses from her inception, yet the companies' IPO in 2019, despite being a complete fiasco compared to the expectations, was still the richest of the year⁶. Uber's main competitor, Lyft, fares no better: as of 2018, the San Francisco based company incurred in cumulative losses above 40 bn \$ in just 7 years of activity , while ending her first day on stock markets with a valuation above 20 bn \$. In the jargon of traders and financial analysts, these companies, which include many two-sided intermediaries such Sea, WeWork or Spotify, have been labeled as "tech unicorns", due to their lack of profitability, and have generally devoted most of their efforts to ensure user's base growth. These results suggests that their business models have a glimpse of rationality, and that network externalities alone can tell just a part of the story.

Third, the pessimistic belief outcome is socially optimal, yet there exists symmetric belief equilibria that could grant consumers all of the surplus that is extracted by the producer, i.e. if one firm sets $p_1^k + p_2^k = -\frac{5}{6} - \frac{1}{72\gamma}$ and the other sets a membership fee combination that exceeds this sum. This finding hints that while agglomeration is efficient *per se*, the first best allocation can be reached only if markets are contestable; as competition *for* the market, in my setting, arises simply from a behavioral assumption, there's room for

⁶see <https://markets.businessinsider.com/news/stocks/19-firms-who-have-bought-most-uber-shares-post-ipo-2019-8-1028450956>

future research to inspect the linkages between migratory behavior and data driven effects. In general, however, competition authorities, as already mentioned earlier in this section, should look carefully at mergers that could trigger data driven cost synergies within an industry, as there is an important trade-off: on one hand, they foster efficiency and can make R&D expenditure valuable; on the other, they could eliminate competition forever and grant monopolistic rents to incumbents.

4 Conclusion

The competitive dynamics in markets that feature inter-group network externalities and homogenous users have been widely discussed; indeed, the theory suggests that agents, when singlehoming, should agglomerate uniquely on one network. This monopolistic outcome has been deemed as socially efficient by many commentators; as the leader cannot afford to lose a side of the market, she must leave all the surplus to her users. In this paper, I showed how the extreme return to scale that datafication entails can reshape the playground; as data-driven effects create an important barrier to entry for new challengers, incumbents can extract a monopolistic rent, subsidizing less their users. While the performance when the incumbent disposes of a belief advantage is still welfare maximizing, aggregate innovation levels are higher if players are symmetric. This peculiarity leads to a paradoxical results; indeed, when considering Big Data, firms might fight so hard that they may *all* end up incurring in losses due to a combination excessive R&D expenditure and subsidies to users. Therefore, data barriers matter substantially for two reasons; first, they configure an advantage in terms of profitability for leaders; second, they outline the competitive strain that characterizes the digital economies; antitrust authorities should always bear this in mind, especially when assessing horizontal mergers. There are a number of question marks that these findings leave behind: first, as many two-sided markets exhibit more than one player earning a great amount of money, how does multihoming interact with data-driven business models? Caillaud and Jullien (2003) highlighted how, in a setting with no data cost synergies, the usage of multiple platforms by a set of users induces positive profits for both competitors: it would be fundamental to assess how their result are affected by considering datafication.

Second, what are the long run implications of a competitive framework in which might rationally commit to negative profits? The main risk is that these aggressive strategies might dynamically end-up establishing a leader that will enjoy the incumbency advantage described above, rendering the market impossible to dispute. In turn, it is not clear how dynamic segmentation equilibria could lead to efficient results; indeed, I've showed that, in the static case, the only segmentation equilibrium where firm earned profits that were not

strictly negative entailed a lower consumer welfare than the pessimistic belief scenario. Third, how could conglomerate or vertical takeovers from other dominant, data drive players affect competition in these settings? In this work, I outlined the synergies that data brings in a purely horizontal scenario. However, data-driven business models generate economies of scope: can actors from an unrelated sector leverage their own business model to produce an incumbency advantage in the sector in which they perform the takeover? I leave these questions for future research.

References

- [1] ARGENTON, C., AND PRÜFER, J. Search engine competition with network externalities. *Journal of Competition Law and Economics* 8, 1 (2012), 73–105.
- [2] ARMSTRONG, M. Competition in two-sided markets. *The RAND Journal of Economics* 37, 3 (2006), 668–691.
- [3] BELLEFLAMME, P., AND PEITZ, M. Platforms and network effects. In *Handbook of Game Theory and Industrial Organization, Volume II*. Edward Elgar Publishing, 2018.
- [4] BELLEFLAMME, P., AND TOULEMONDE, E. Negative intra-group externalities in two-sided markets. *International Economic Review* 50, 1 (2009), 245–272.
- [5] BELLEFLAMME, P., AND TOULEMONDE, E. Tax incidence on competing two-sided platforms. *Journal of Public Economic Theory* 20, 1 (2018), 9–21.
- [6] BIGLAISER, G., CALVANO, E., AND CRÉMER, J. Incumbency advantage and its value. *Journal of Economics & Management Strategy* 28, 1 (2019), 41–48.
- [7] BIGLAISER, G., AND CRÉMER, J. The value of incumbency in heterogeneous platforms. *CEPR Discussion Paper No. DP11207* (2016).
- [8] BRYAN, K. A., AND HOVENKAMP, E. Antitrust limits on startup acquisitions. *Review of Industrial Organization* (2020), 1–22.
- [9] CABRAL, L. Dynamic price competition with network effects. *The Review of Economic Studies* 78, 1 (2011), 83–111.
- [10] CAILLAUD, B., AND JULLIEN, B. Chicken and egg: Competing matchmakers. *CEPR Discussion Paper* (2001).
- [11] CAILLAUD, B., AND JULLIEN, B. Competing cybermediaries. *European Economic Review* 45, 4-6 (2001), 797–808.

- [12] CAILLAUD, B., AND JULLIEN, B. Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics* (2003), 309–328.
- [13] CALVANO, E., CALZOLARI, G., DENICOLO, V., AND PASTORELLO, S. Artificial intelligence, algorithmic pricing and collusion. *Algorithmic Pricing and Collusion (April 1, 2019)* (2019).
- [14] CRÉMER, J., DE MONTJOYE, Y.-A., AND SCHWEITZER, H. Competition policy for the digital era. *Report for the European Commission* (2019).
- [15] DE CORNIÈRE, A., AND TAYLOR, G. A model of biased intermediation. *The RAND Journal of Economics* 50, 4 (2019), 854–882.
- [16] FILISTRUCCHI, L. Market definition in multi-sided markets. *Rethinking antitrust tools for multi-sided platforms* (2018), 37.
- [17] FRANCK, J.-U., AND PEITZ, M. Market definition and market power in the platform economy. *CERRE Report* (2019).
- [18] HALABURDA, H., AND YEHEZKEL, Y. The role of coordination bias in platform competition. *Journal of Economics & Management Strategy* 25, 2 (2016), 274–312.
- [19] HALABURDA, H., JULLIEN, B., AND YEHEZKEL, Y. Dynamic competition with network externalities: how history matters. *The RAND Journal of Economics* 51, 1 (2020), 3–31.
- [20] KATZ, M. L., AND SHAPIRO, C. Network externalities, competition, and compatibility. *The American economic review* 75, 3 (1985), 424–440.
- [21] KHAN, L. M. Amazon’s antitrust paradox. *Yale LJ* 126 (2016), 710.
- [22] LEVIN, R. C. Appropriability, r&d spending, and technological performance. *The American Economic Review* 78, 2 (1988), 424–428.
- [23] LI, W. C., NIREI, M., AND YAMANA, K. Value of data: there’s no such thing as a free lunch in the digital economy. *US Bureau of Economic Analysis Working Paper, Washington, DC* (2019).
- [24] PRAT, A., AND VALLETTI, T. M. Attention oligopoly. *Available at SSRN 3197930* (2019).
- [25] PRUFER, J., AND SCHOTTMÜLLER, C. Competing with big data. *Tilburg Law School Research Paper*, 06 (2017).

- [26] ROCHET, J.-C., AND TIROLE, J. Platform competition in two-sided markets. *Journal of the european economic association* 1, 4 (2003), 990–1029.
- [27] SCHWAB, K., MARCUS, A., OYOLA, J., HOFFMAN, W., AND LUZI, M. Personal data: The emergence of a new asset class. In *An Initiative of the World Economic Forum* (2011).
- [28] THE ECONOMIST. Tech’s new stars have it all—except a path to high profits, apr 17th. *The Economist* (2019).
- [29] THE ECONOMIST. The wave of unicorn IPOs reveals silicon valley’s groupthink, apr 17th. *The Economist* (2019).
- [30] THE ECONOMIST. And on the second day... , june 20th. *The Economist* (2020).