Does Data Portability Facilitate Entry? *

Wing Man Wynne Lam[†] Xingyi Liu[‡]

November, 2019

Abstract

Data portability rules are generally thought to encourage consumers to switch between different service providers and facilitate entry of new firms. Some of these rules, however, only apply to data "provided by" the consumer (data subject), e.g., purchasing patterns. Data "derived by" a firm (data controller) with the help of data analytics, e.g., recommendations derived from purchasing patterns, does not fall under data portability rules. We show that, under the current legislation along with extensive use of data analytics, data portability may hinder switching and entry due to the *demand-expansion* effect: the prospect of easier switching due to data portability may entice consumers to provide even more data to the incumbent, which strengthens the incumbency advantage. Hence, the effectiveness of data portability in fostering competition will depend on what types of data are portable. More generally, in analysing the effectiveness of polices aiming at reducing *ex post* switching costs, it is important to take into account their impacts on *ex ante* actions that build up endogenous entry barrier.

Keywords: Data portability, GDPR, Entry barrier

JEL Classification: K2, L5, L8

---

*We would like to thank Jacques Crémer, Morten Hviid, Gerard Llobet, Bruce Lyons, Frank Verboven, and seminar and conference participants at CRESSE 2018, EARIE 2018, Manchester, Norwich, and Toulouse. We are especially grateful to the co-editor and anonymous referees for their helpful comments that have greatly improved the paper.

[†]Norwich Business School and Centre for Competition Policy, University of East Anglia, Norwich Research Park, NR4 7TJ, Norwich, United Kingdom, e-mail: wing.m.lam@uea.ac.uk

[‡]Aston Business School, Aston University, B4 7ET Birmingham, United Kingdom, e-mail: x.liu29@aston.ac.uk

1 Introduction

Competition in non-price characteristics, such as functionality and data services, has become increasingly common in Internet markets. Many platforms offer consumers free services in exchange for consumer data being used for data analytics (e.g., Google search and Facebook). On one hand, this has prompted concerns over consumer privacy; on the other hand, data analytics may give platforms a competitive advantage and market power that affect future competition and innovation.[1]

In face of these challenges, the General Data Protection Regulation (GDPR) has come into force since May 2018, which grants consumers a set of rights with more control over the collection and use of their data. Notably, consumers are now given a new right to data portability under Article 20 of the GDPR, defined as follows:[2]

> "*The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the data have been provided.*"

A clear aim of data portability is to facilitate consumer switching between different service providers, prevent lock-in, and foster entry and competition. However, in contrary to many other existing markets where switching costs are largely exogenous (e.g., physical costs of opening a new bank account) or determined by firms (e.g., coupons offered to loyal consumers, high degree of incompatibility between firms' products), consumers play an important role when it comes to data related services. Specifically, when consumers provide more data to an incumbent, they may find themselves more locked-in with the incumbent, due to a range of services offered by the incumbent that analyse their data and encourage stickiness. Hence, it is important to understand the impact of data porta-

[1] "Big tech faces competition and privacy concerns in Brussels", The Economist, March 2019. Available at https://www.economist.com/briefing/2019/03/23/big-tech-faces-competition-and-privacy-concerns-in-brussels (Accessed: 23 October 2019).

[2] Article 20, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

bility on switching and entry, when consumers react to the new policy and adjust their behaviour. More generally, we attempt to look into the impact of policies that aim at reducing switching cost, which can be endogenously generated by consumers' consumption decisions.

Specifically, we consider a two-period model, where an incumbent acts as the monopolist in the first period and an entrant can potentially enter and compete in the second period. Both firms can provide a basic data service to a unit mass of homogeneous consumers. The incumbent can provide, if available, additional big data service in the second period, which reflects the advantage of the incumbent in analysing data over the entrant. The entrant can enter in the second period if it provides a better service than the incumbent, i.e., when the value of its service is above a certain threshold. We are interested in how data portability affects this threshold.

Certainly, entry becomes easier with data portability, should everything be portable. There are, however, boundaries to the right to data portability. In particular, it applies only to data "provided by" the data subject but not data "inferred or derived by" the data controller. For instance, whereas data on a consumer's search history fall within the scope of data portability, inferred consumer data for personalising products or making recommendations fall outside. Another example is that "the outcome of an assessment regarding the health of a user or the profile created in the context of risk management and financial regulations [...] are inferred or derived from the analysis of data provided by the data subject [...and hence] will not be within the scope of this new right".[3] Therefore, data analytics enable firms (also referred to as data controllers) to provide non-portable value added services to consumers (also referred to as data subjects), which can lock customers in a relationship with the data controller. Thus, data portability does not completely eliminate the incumbency advantage, and we show that it may even enhance such advantage under certain conditions.

To be more specific, we find that data portability affects entry in two ways. First, for a given level of data provision in the first period, it facilitates consumer switching and entry. This is the *switch-facilitating* effect, which is one of the most compelling reasons for promoting data portability. Second, allowing the level of data provision to vary, data portability encourages consumers to provide more data in the first period as the value

[3] See European Commission (2016).

of data becomes higher when they can be ported across service providers. This is the indirect *demand-expansion* effect, which raises the value of the incumbent's service and strengthens the incumbency advantage. More generally, data provision can be seen as an investment made by the consumers to increase the value of the relationship with a data controller, and this investment is relationship-specific when data cannot be ported. With data portability, it reduces investment specificity, which facilitates *ex post* switching. However, data portability also raises the value of data provided in the first period and increases consumers' *ex ante* incentive to invest and, hence, leads them to provide more data. Most notably, such a *demand-expansion* effect is recognised in the discussion on the introduction of data portability in Singapore:[4] "*The introduction of data portability [...] may in turn encourage consumers to share more data [...] due to increased ease in replicating existing data.*"[5]

The latter effect is largely ignored in the literature on exogenous switching cost, where switching cost creates consumer lock-in only on the extensive margin. However, in our model, data portability changes the intensive margin of consumer demand, which may create endogenous entry barrier. More specifically, we find that without data analytics, data portability facilitates switching and entry, as the *switch-facilitating* effect dominates the *demand-expansion* effect. With data analytics in addition to data portability, the *demand-expansion* effect dominates if the big data service is valuable enough, in which case data portability can make entry more difficult. Interestingly, this is more likely to be the case with network effects (i.e., when the value of big data service depends on large population data and hence increases with the size of the user base) compared to without. The reason is that with network effects, an individual consumer ignores the positive externality of his data provision on other customers and hence provides too little data, which weakens the *switch-facilitating* effect. On the other hand, less data provision and a higher degree of data portability means that a consumer is more likely to switch and port their data, which makes their data provision more responsive to enhanced data portability, i.e., the *demand-expansion* effect is stronger. Combining

[4] "Discussion Paper on Data Portability", Personal Data Protection Commission in collaboration with Competition and Consumer Commission of Singapore, February 25, 2019.

[5] A similar effect is also found in the US energy market. To facilitate consumers to access and port energy usage data, the "Green Button" project was initiated in 2011. Evidence suggests that people are more willing to provide data when there is a greater degree of data portability.

both effects, data portability is more likely to raise entry barrier when network effects are at work. Furthermore, entry deterrence is more likely when the entrant adopts a more innovative strategy, i.e., when the entrant is more likely to obtain a sufficiently innovative product and enter the market. The reason is that anticipating a better firm will enter with high probability, consumers are less likely to stay with the incumbent in the future, which reduces the value of providing data to the incumbent. This reduces first period data provision and weakens the *switch-facilitating* effect, compared to the situation with a less innovative entrant. However, even if the amount of data provided in the first period is smaller, consumers are more likely to port these data to the more innovative entrant, which strengthens the *demand-expansion* effect. Hence, entry becomes more difficult. In addition, the availability of data portability by itself is sufficient for the above effects to emerge. In GDPR, the right to data portability comes together with a set of other rights that grants consumers more control over their data and alleviates consumers' privacy concerns (see, for instance, Tucker 2014), which may further amplify the *demand-expansion* effect, due to higher willingness to provide data in the post-GDPR era, and make entry even more difficult.

Thus, although data portability may benefit consumers in the short run, it can have an adverse effect on entry and long-run efficiencies. Under certain circumstances, it can result in "excess inertia" which locks generations of consumers in with the incumbent (see also, for instance, Farrell and Saloner 1986). This becomes more prominent when consumers enter the market sequentially and a sufficient scale is necessary for successful entry (e.g., a large enough database for data analytics).[6] In such situations, early generations of consumers are incentivised to provide more data when data portability becomes viable, which allows the incumbent to accumulate even more data and makes future entry increasingly more difficult. This could have the further consequence of slowing down innovation.

The results point to the potential limit of data portability in fostering competition under the current framework of legislation, especially when established firms such as the GAFAM (Google, Amazon, Facebook, Apple, and Microsoft) rely more and more on derived data services. In fact, we also see incumbent firms actively and voluntarily working together on data portability, such as the launch of the Data Transfer Project by

[6]We provide such an example in Appendix C.4.

Microsoft, Facebook, Google, and Twitter in 2017.[7] Furthermore, our results shed light on why Google remains popular in spite of the introduction of its data portability service, Google Takeout, for 27 products in 2011, which was extended to other core services such as Google Search in 2016.[8]

In summary, in accordance with the recent Stigler Center Report (2019) (pp. 26 and 88) and the Vestager Report[9] (p. 58), the effectiveness of the right to data portability will depend on the way it will be implemented in practice, specifically, what types of data can be ported. Our results point out that the role of data portability in facilitating entry and competition may be limited, when inferred data are not covered under the current legislation. This is more likely so when we take consumer behaviour into account. Similar ideas may also apply to other markets where consumers can build up their own switching costs. For instance, in markets where reputation is important (e.g., online trading, peer-to-peer sharing), consumers may be incentivised to trade more on these platforms to build a better reputation, when their endorsements such as customer feedbacks, credit scores, trust scores become portable to potential entrants, and this can make future entry of new providers harder. This may also apply to markets for professional advice (e.g., medical, legal, financial services), where better services rely on information provided by clients. When these information become portable, clients may use their current service provider more intensively (e.g., stick with the same doctor, lawyer or mortgage advisor) and become more reluctant to switch in the future. Hence, it is important to understand how consumers react to policies that intend to lower switching cost, and our paper attempts to pave the way for further studies on the implementation of these policies.

---

[7] "Microsoft, Facebook, Google and Twitter Introduce the Data Transfer Project: An Open Source Initiative for Consumer Data Portability". Available at https://blogs.microsoft.com/eupolicy/2018/07/20/microsoft-facebook-google-and-twitter-introduce-the-data-transfer-project-an-open-source-initiative-for-consumer-data-portability (Accessed: 23 October 2019).

[8] "How I tried and failed to quit Google?", CBS News, December 18, 2018. Available at https://www.cbsnews.com/news/how-i-tried-and-failed-to-quit-google/ (Accessed: 23 October 2019).

[9] "Competition Policy for the Digital Era", a report by Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, European Commission, May 2019. Available at https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf (Accessed: 23 October 2019).

1.1 Related Literature

Our work contributes to the economic analysis of the impact of data portability on market competition. The right to data portability has been extensively discussed in the legal literature; see, e.g., Graef (2015), De Hert et al. (2018), and Van der Auwermeulen (2017). However, economic analyses are rare, except Christensen et al. (2013), who studies the impact of data protection regulation on small and medium sized enterprises, emphasising on the fixed costs of providing such protection. Our work also relates to the discussion on *number portability*; see, e.g., Bühler et al. (2006) for a survey on its implementation in Europe and Viard (2007) for the case of toll-free numbers in the US. Our paper, however, provides a new perspective that has not been discussed in number portability in that it analyses the effect of data portability on the intensive margin added on with the use of big data analytics.

Our focus on switching with data analytics also relates our work to the large literature on either switching costs or network effects; see Farrell and Klemperer (2007) for an excellent survey. Strikingly, however, there are few works that analyse both issues together. In addition to this paper, other recent contributions that attempt to fill this gap include Biglaiser et al. (2013) and Lam (2017). Moreover, in contrast to most of the literature taking switching cost as exogenous, data analytics in this paper generate endogenous barrier of switching.

This brings our paper in close relation to the literature on endogenous switching cost. For instance, Caminal and Matutes (1990) show in a two-period model that firms commit in equilibrium to lower prices for loyal customers, which creates endogenous switching costs in the second period. Similarly, Chen and Pearcy (2010) and Shin and Sudhir (2010) show that such a reward for loyal consumers can arise when consumer preference changes over time. When switching cost is positively related to product differentiation, Gehrig and Stenbacka (2004) show that firms have incentives to choose maximal differentiation in order to raise switching cost. Shi (2013) further demonstrates substitutability between exogenous switching cost and endogenous switching cost. When switching cost comes from incompatibility of system goods, Marinoso (2001) shows that producers of system goods have incentives to make their systems incompatible to create switching cost for the complementary parts in the second period, when the primary parts are sold in the first period. Most of this literature assumes that consumers have unit demands (e.g., in a Hotelling model) and focuses on the extensive margin of demand, whereas our analysis focuses more on the intensive margin of demand by assuming elastic demands on the consumer side.

Furthermore, the impact of data portability on entry relates our work to the large literature on exclusion and entry deterrence. This literature dates back to the seminal contribution of Aghion and Bolton (1987), which shows that the incumbent can deter entry with strategic contracting. Similarly, Tremblay (2019) shows that an incumbent platform can strategically subsidise content providers in earlier periods to limit the entry of competing platforms in later periods. Firms may also offer consumers loyalty rebates to create demand-side linkage and raise entry barrier. For instance, Cairns and Galbraith (1990) shows how this can be achieved by frequent flyer programmes in the airline industry.[10] More recent contribution from Calzolari and Denicolò (2018) also identifies a *demand-expansion* effect of loyalty rebates in the *Intel* case.[11] Different from this strand of literature, where entry barriers are generated through firms' pricing, technology, or contractual arrangements (e.g., contract breaching fees, exclusivity clauses, quantity discounts), in our model, the higher endogenous entry barrier originates from the consumer side, i.e., how consumers respond to market conditions by adjusting their demands.

The paper is organised as follows: Section 2 sets up the model; Section 3 presents a full analysis of the linear-quadratic case; Section 4 discusses the main mechanism underlining the model and several implications; Section 5 provides further extensions; Section 6 concludes with policy implications. All omitted proofs and additional materials are presented in the Appendix.

2 The Model

We consider a two-period model, where an incumbent I is present in the market for both periods and an entrant E can enter in the second period potentially.[12] Both firms can provide services to consumers at zero costs, and we assume that there is no discounting.

[10] Hartmann and Viard (2008), instead, shows that most consumers leave the programme offered by a golf course before they reach the critical threshold for reward.

[11] See Commission Decision COMP/C-3/37.990-Intel.

[12] The main results will not change if we consider more than one entrant, because in a model of homogeneous products, only the entrant providing the highest quality product can enter the market.

There is a continuum of consumers with a total mass of one.

The First Period

The incumbent I is the only firm present in the first period. It provides a basic service to consumers. In addition, if consumers provide data to the incumbent, it generates additional value from data services. As an example, we can think of Facebook's social networking service as a basic service, which allows users to connect to friends. When consumers provide data such as photos, messages, shopping preferences, trips and holiday plans, they also obtain values from sharing memories, experiences, etc., which depend on the amount of data they provide.[13] Specifically, we assume that a consumer obtains a utility of

$$u_1 = v_I + v(q_1) - C(q_1),$$

where v_I is the valuation of the basic service of I, q_1 is the total amount of data provided, $v(q_1)$ is the utility derived from data services, and $C(q_1)$ is the cost of data provision. The cost includes time and opportunity cost spent on providing data and not just the cost of one click. Moreover, we may interpret it as the perceived costs of privacy when consumers provide their personal data to firms.

The Second Period

If the entrant enters in the second period, and a consumer chooses to switch to the entrant and provide an amount of data of q_2, he obtains a utility of

$$u_2^E = v_E + \nu(\lambda q_1 + q_2) - C(q_2),$$

where $v_E = v_I + \delta$ is the valuation of the basic service of the entrant, and δ is a random draw from the distribution $F(\delta)$ on the support $[0, \Delta]$, which is known by the incumbent and the consumers.[14] That is, the entrant provides a random level of improvement over the incumbent's basic service. The utility from data service depends on the amount of data provided in the first period that is portable and new data provided in the second period. The degree of data portability is measured by λ: if $\lambda = 0$, there is no data portability; if $\lambda = 1$, there is full data portability; if $\lambda \in (0, 1)$, there is partial data

[13] Other examples include search engines, wearable devices, traffic and location data, etc. See "Guide to the GDPR", Information Commissioner's Office, UK.

[14] Extending the support of v_E to values below v_I does not affect our analysis, as the entrant drawing such a low value of v_E would not be able to enter the market.

portability.[15]

If the entrant does not enter in the second period, or if the entrant enters but a consumer decides not to switch and continues to use the service of the incumbent, this consumer obtains a utility of

$$u_2^I = v_I + \nu(q_1 + q_2) - C(q_2) + v_B(Q_1),$$

where $v_B(Q_1)$ is the additional utility derived from big data services. For instance, such big data services can be personalised recommendations and advertisements.[16] The value of big data services $v_B(Q_1)$ depends on the database Q_1 and we consider two interpretations of Q_1. First, the incumbent can infer a consumer's preference based on his/her own data, e.g., previous search and browsing histories. In this case, the value of big data service only depends on data collected from this consumer, that is, $Q_1 = q_1$. We treat this case in Section 3.2 as *individual switching cost.* Second, the inference can also depend on, for instance, other consumers' search and browsing histories, i.e., "people like me".[17] In this case, the value of big data service depends on the aggregate data provided to the firm by all consumers, that is, $Q_1 = \int_0^1 q_1^i di$, where q_1^i is the data provided by consumer $i \in [0, 1]$ in the first period. We treat this case in Section 3.3 as *collective switching cost.*[18]

The total utility from the data service $\nu(\lambda q_1 + q_2)$ in the second period depends on total data provided or ported in both periods ($\lambda = 1$ in the case of no switching), and

---

[15]In the main analysis, we focus on the case where consumers port data from the incumbent to the entrant, which is generally thought to be more conducive to entry. Our results still hold and may even be strengthened when data portability occurs in another direction, i.e., from the entrant to the incumbent, due to the fact that both effects work in favour of the incumbent.

[16]We could interpret $v_B(.)$ broadly to include not just valuable recommendations and advertisements but also nuisance costs such as receiving irrelevant and annoying ads. For the purpose of this paper, we do not need to make specific restrictions on the value of $v_B(.)$, as our main results identify conditions on $v_B(.)$ under which data portability hinders switching. However, in practice, we believe that its value is more likely to be positive, as under the current GDPR opt-in rules, a consumer will only opt-in for such big data services when they benefit from these services. Similarly, a consumer can exert the right to be forgotten if the value of such services falls below zero.

[17]For example, in collaborative filtering systems, recommendations of a product are made based on "people like me". See Bossenbroek and Gringhuis (2015).

[18]In the terminology of the Stigler Center Report (2019), the individual switching cost case is similar to *high dimensional data* and the collective switching cost case is similar to *large population data.*

we assume that the consumer only incurs the cost of providing the fresh data q_2. This is reasonable since the incumbent keeps all data from the first period, which need not be provided again but still generate value to the consumer. In the case of ported data, the consumer should be able to do this "*without hindrance*" at least according to the data protection rule. For instance, a record of past location data or a collection of past photos still generates utilities in the future, but there is no need for a consumer to incur the data provision cost again.[19]

When consumers do not switch, they obtain the additional service of big data from the incumbent, but when they switch, they no longer obtain this service. Our results would not change even if consumers could obtain a big data service from the entrant, as long as the value of it is smaller than what they would have obtained from the incumbent. This can capture the learning effects associated with big data services, i.e., the firm that has been in the market for a longer time accumulates more data and learns more from this acquired data, and thus can provide better services to consumers. This advantage in big data services can also come from the incumbent being active in multiple services, which we further discuss in Section 5.1.

We consider the following game: in the first period, each consumer decides how much data service to use (i.e., the amount of data, q_1, to provide). In the second period, the entrant draws its quality improvement δ and decides whether to enter the market. Finally, each consumer decides whether or not to switch to the entrant if entry occurs and how much data to provide (either to the incumbent or to the entrant). In the main analysis, we focus on the linear-quadratic case and we provide a generalisation in Appendix B.

Assumption 1. $\nu(q) = q$, $v_B(q) = v_B q$, *and* $C(q) = \frac{1}{2}cq^2$.

A few remarks: First, we assume that the basic service and the data service are independent, i.e., they enter the utility function of a consumer separately. This is mainly made for analytical convenience. This also allows us to separate the *ex post* effect of data portability from its *ex ante* effect on the entry barrier, by making consumers' second period data provision decision independent of their first period data provision. Never-

[19]In practice, a consumer may exercise the right to be forgotten to have his/her data deleted from a data controller. In such a case, data portability has no effect on the second period competition, as no historical data exists. However, as long as data generate positive value, a consumer would not exercise the right to be forgotten.

theless, our main insights remain valid when the basic service and the data service are complements, i.e., the value of these services enter the utility function in a multiplicative way, and we provide a detailed discussion and an example in Appendix C.2. As we show there, the *switch-facilitating* effect and the *demand-expansion* effect are still present, although the condition for data portability to monotonically raise entry barrier becomes stricter, as additional effects arise when consumers may provide different amount of data when they stay and when they switch in the second period.

Second, data are of different values in the real world, some of which depreciate more quickly than others. For instance, a dated profile of a consumer may have little relevance for prediction of his/her present behaviour. We can easily capture depreciation with a lower value of first period data in the second period or a lower value of v_B, implying a smaller incumbency advantage. At the extreme, if the value of the data fully depreciates in the second period, then the incumbency advantage disappears and the problem is reduced to the case without data analytics.

Third, we assume that data portability is enforced effectively, i.e., there is no technical obstacles to porting data, as motivated by the Guidelines on the right to data portability, which says: "these data should be received *in a structured, commonly used and machine-readable format*" and "*the GDPR prohibit[s] controllers from establishing barriers to the [data] transmission*".[20] This assumption is also reasonable to the extent that it allows us to address the problem that data portability may still hinder switching even in the ideal situation of perfectly enforced data portability. Furthermore, we assume that there are no fixed costs of data provision. If, however, there are some fixed costs, it is clear that data portability will allow a consumer to economise on these costs, which facilitates entry. This effect is well-known in the literature on exogenous switching cost. Hence, to focus on endogenous entry barrier, we assume away fixed costs of data provision.

Fourth, we assume that firms compete only in the quality dimension (value of the service), but not in the price dimension. This assumption can be justified in two ways. On one hand, a lot of basic data services are provided to consumers for free, e.g., Google search, maps and email, and Facebook. Similarly, some big data services, such as recommendations for relevant products and special offers, are free. On the other hand,

[20]Successful experiences from the *Open Banking Initiative* and the *Green Button Project* show that a well-enforced regulation can implement data interoperability.

price-related entry deterrence strategies have been the focus of much of the antitrust literature. The focus of this paper is on competition in non-price characteristics, such as data accumulation or analytics, and thus we choose to abstract from price competition.

Finally, we assume that consumers are homogeneous and single-home, i.e., when the entrant enters, a consumer either stays with the incumbent or switches to the entrant. This is a reasonable assumption under the current market situation, as discussed in the Stigler Center Report (2019) (p. 20), "*while users sometimes have the ability to employ multiple services, there is usually a convenience cost to doing so*". However, our general insights extend beyond. In digital markets, the decision about whether and how much to multi-home are both important. For instance, although consumers may subscribe to several social network services, these services are still competing for consumers' screen time and attention, which are crucially important for digital advertising.[21] We consider such an example in Appendix C.3 where a consumer allocates his time when multi-homing between several services. In such situations, our results can be more generally interpreted as the impact of data portability on the barrier to expansion. Consumer heterogeneity could be introduced in several dimensions (e.g., valuation for basic service, valuation for big data service, cost of providing data), but this would not affect our results as long as all consumers participate in the market. When the market is not fully covered, data portability has the additional effect of increasing the incumbent's market penetration, which may further raise entry barrier and strengthen our results.

3 The Analysis

We start the analysis with the case of no big data analytics, and then the case with individual switching cost and collective switching cost. All the proofs for results in this section are contained in Appendix A. To begin with, notice that with the linear-quadratic specification, in the second period, the consumer provides the same amount of data whether he stays with the incumbent or switches to the entrant, which is given by $q_2^* = 1/c$. This generates a second period utility from data service of $w_2^* = 1/(2c)$.

[21] "Facebook boasted of buying Instagram to kill the competition: sources", February 2019. Available at: https://nypost.com/2019/02/26/facebook-boasted-of-buying-instagram-to-kill-the-competition-sources/ (Accessed: 27 November 2019).

3.1 The Case without Big Data Analytics

When there is no big data analytics, i.e., $v_B = 0$, a consumer switches to the entrant in the second period if

$$v_I + q_1 + w_2^* < v_I + \delta + \lambda q_1 + w_2^*,$$

which simplifies to $\delta > \delta^o$ with

$$\delta^o = (1-\lambda)q_1. \tag{1}$$

In the first period, a consumer chooses q_1 to maximise his total utility U across two periods, given by

$$U(q_1) = v_I + q_1 - \frac{1}{2}cq_1^2 + \int_0^{\delta^o} [v_I + q_1 + w_2^*]dF(\delta) + \int_{\delta^o}^{\Delta} [v_I + \delta + \lambda q_1 + w_2^*]dF(\delta).$$

Assuming that c is large enough such that $U(q_1)$ is concave in q_1, the first order condition yields the optimal data provision in the first period:

$$cq_1 = 1 + \lambda + (1-\lambda)F(\delta^o). \tag{2}$$

The equilibrium δ^o satisfies

$$c\delta^o - (1-\lambda)(1+\lambda) - (1-\lambda)^2 F(\delta^o) = 0. \tag{3}$$

Using Implicit Function Theorem, we obtain

$$\frac{d\delta^o}{d\lambda} = \frac{-2\lambda - 2(1-\lambda)F(\delta^o)}{c - (1-\lambda)^2 f(\delta^o)},$$

which is always negative, that is,

Proposition 1. *Without big data analytics (i.e., $v_B = 0$), data portability facilitates entry, i.e., $d\delta^o/d\lambda < 0$.*

3.2 The Case of Individual Switching Cost

With big data service, let us consider the case of individual switching cost, i.e., when a consumer's value of the big data service only depends on his/her own data provision in the first period. In this case, a consumer switches in the second period if

$$v_I + q_1 + v_B q_1 + w_2^* < v_I + \delta + \lambda q_1 + w_2^*,$$

which defines a switching threshold δ^s, given by

$$\delta^s = (1 - \lambda + v_B)q_1. \tag{4}$$

The total utility in the first period is

$$U(q_1) = v_I + q_1 - \frac{1}{2}cq_1^2 + \int_0^{\delta^s} [v_I + q_1 + v_B q_1 + w_2^*]dF(\delta) + \int_{\delta^s}^{\Delta} [v_I + \delta + \lambda q_1 + w_2^*]dF(\delta).$$

The optimal data provision satisfies

$$cq_1 = 1 + \lambda + (1 - \lambda + v_B)F(\delta^s). \tag{5}$$

The equilibrium switching threshold, δ^s, solves

$$c\delta^s - (1+\lambda)(1 - \lambda + v_B) - (1 - \lambda + v_B)^2 F(\delta^s) = 0. \tag{6}$$

Applying Implicit Function Theorem, we obtain

$$\frac{d\delta^s}{d\lambda} = \frac{v_B - 2\lambda - 2(1 - \lambda + v_B)F(\delta^s)}{c - (1 - \lambda + v_B)^2 f(\delta^s)}. \tag{7}$$

Proposition 2. *With big data analytics and individual switching cost, data portability increases the switching threshold, i.e., $d\delta^s/d\lambda > 0$ for all $\lambda \in [0, 1]$, if and only if*

$$\frac{v_B - 2}{2v_B} \geq F(\frac{v_B^2 + 2v_B}{2c}). \tag{IS-L}$$

3.3 The Case of Collective Switching Cost

We now turn to the case where a consumer's value for the big data service depends on the aggregate data provided by all consumers in the first period. However, since each consumer is infinitesimal, when a consumer makes his data provision decision, he takes the aggregate data level and thus the value of big data service as given. This is closely related to the concept of network effect, where a consumer makes his decision taking the network size (the big data service) as given. In this situation, big data analytics create an endogenous barrier of switching, which depends on aggregate data provision (hence, a collective switching cost). To be more specific, given q_1, a consumer switches in the second period if

$$v_I + q_1 + v_B Q_1 + w_2^* < v_I + \delta + \lambda q_1 + w_2^*,$$

which simplifies to $\delta > \delta^c$ with

$$\delta^c = (1 - \lambda)q_1 + v_B Q_1, \tag{8}$$

where Q_1 is the aggregate data provided to the incumbent in the first period. The total utility in the first period becomes

$$U(q_1) = v_I + q_1 - \frac{1}{2}cq_1^2 + \int_0^{\delta^c} [v_I + q_1 + v_B Q_1 + w_2^*] dF(\delta) + \int_{\delta^c}^{\Delta} [v_I + \delta + \lambda q_1 + w_2^*] dF(\delta),$$

which yields the following first-order condition on q_1:

$$cq_1 = 1 + \lambda + (1-\lambda)F(\delta^c). \tag{9}$$

Lastly, the equilibrium requires rational expectations, i.e.,[22]

$$Q_1 = q_1. \tag{10}$$

The equilibrium switching threshold, δ^c, then satisfies

$$c\delta^c - (1+\lambda)(1-\lambda+v_B) - (1-\lambda)(1-\lambda+v_B)F(\delta^c) = 0,$$

which leads to

$$\frac{d\delta^c}{d\lambda} = \frac{v_B - 2\lambda - (2(1-\lambda)+v_B)F(\delta^c)}{c - (1-\lambda)(1-\lambda+v_B)f(\delta^c)}. \tag{11}$$

We have the following result:

Proposition 3. *With big data analytics and collective switching cost, data portability increases the switching threshold, i.e., $d\delta^c/d\lambda > 0$ for all $\lambda \in [0,1]$, if and only if*

$$\frac{v_B - 2}{v_B} \geq F(\frac{2v_B}{c}). \tag{CS-L}$$

Remark: Proposition 2 and 3 show the conditions under which improving data portability *monotonically* increases entry barrier. When the conditions are not satisfied, we may have a non-monotone or even a monotonically decreasing relationship between data portability and entry barrier. Indeed, when $v_B = 0$ (so neither Conditions (IS-L) nor (CS-L) is satisfied), Proposition 1 shows that improving data portability monotonically reduces entry barrier.

4 Discussions and Implications

4.1 The Main Effects of Data Portability

There are two main forces at work. On one hand, for a given amount of data provided in the first period, a high level of data portability always facilitates switching and entry

[22]This means that a consumer rationally anticipates the aggregate level of data provision in the market.

in the second period. To be more specific, in all three cases, we have

$$\frac{\partial \delta^o}{\partial \lambda} = \frac{\partial \delta^s}{\partial \lambda} = \frac{\partial \delta^c}{\partial \lambda} = -q_1.$$

We name this effect as the *switch-facilitating* effect. This is the intended effect of data portability that makes data more easily transferable between data controllers. In other words, data portability renders the data provided in the first period *less specific* to the incumbent and makes *ex post* switching easier.

On the other hand, as data become more portable and can be used across data controllers, the value of providing data in the first period becomes greater. That is, as data become *less specific* to the incumbent, the *ex ante* incentive to provide data becomes larger. More specifically, for a given switching threshold δ', we have

$$\frac{\partial q_1}{\partial \lambda} = \frac{1 - F(\delta')}{c},$$

in all three cases. This, however, has the effect of raising the switching threshold. In the case without big data analytics, we have

$$\frac{\partial \delta^o}{\partial q_1}\frac{\partial q_1}{\partial \lambda} = \frac{(1-\lambda)(1-F(\delta^o))}{c}.$$

In the case with big data analytics, we have

$$\frac{\partial \delta^s}{\partial q_1}\frac{\partial q_1}{\partial \lambda} = \frac{(1-\lambda+v_B)(1-F(\delta^s))}{c} \text{ and } \frac{\partial \delta^c}{\partial q_1}\frac{\partial q_1}{\partial \lambda} = \frac{(1-\lambda+v_B)(1-F(\delta^c))}{c}.$$

We name this effect of first period data provision on the switching threshold as the *demand-expansion* effect.

4.2 The Role of Big Data Analytics

It is clear that in the case without big data analytics, the *demand-expansion* effect vanishes in the limit as data become fully portable ($\lambda \to 1$) and hence only the *switch-facilitating* effect remains. Indeed, Proposition 1 shows that in this case, the *switch-facilitating* effect always dominates and the intended effect of data portability is obtained. However, with big data analytics, the *demand-expansion* effect becomes stronger and may even reverse the intended effect when big data analytics are valuable enough (a necessary condition is $v_B > 2$, which means that the marginal value of big data service is as high as the total marginal value of basic data service across the two periods under full data portability). This casts doubt on the role of data portability in fostering competition. Moreover, as we show in Section 5.2, if we endogenise the level of v_B by allowing the incumbent to invest in big data services, the incumbent will invest more in big data service (by choosing a higher v_B) when the degree of data portability increases, which can make entry even less attractive for the entrant.

Although the entry barrier is always higher when the big data service is more valuable (that is, $d\delta^s/dv_B > 0$ and $d\delta^c/dv_B > 0$) , the effect of the value of big data service on how data portability affects the entry barrier may not be monotone.[23] Take the collective switching cost case for example: on one hand, a higher v_B enhances the *demand-expansion* effect when holding q_1 and δ^c constant; On the other hand, a higher v_B also leads to a higher q_1 and hence the *switch-facilitating* effect is also stronger. In the linear-quadratic case, the trade-off between these two effects depend on the shape of $F(\delta)$. Nevertheless, for a class of distribution functions such as

$$F(\delta, \alpha) = \frac{\delta}{\delta + \alpha}, \text{ with } \alpha > 0 \text{ and } \delta \in [0, \infty),$$

we can show that Condition (CS-L) is satisfied if

$$v_B > \frac{2\alpha c}{\alpha c - 4}.$$

Thus, it is more likely that data portability raises the entry barrier when the value of big data service is higher.[24]

4.3 The Role of Network Effect and Cost of Data Provision

The case of collective switching cost resembles markets with network effects, where each consumer's data provision generates a positive externality on other consumers through the big data service. The following result immediately follows from the comparison between Conditions (IS-L) and (CS-L):

Corollary 1. *Data portability is more likely to raise entry barrier with collective switching cost compared to the case with individual switching cost, i.e., if $d\delta^s/d\lambda > 0$, then $d\delta^c/d\lambda > 0$.*

[23] That is, taking the collective switching cost case for example, if $\partial\delta^c/\partial\lambda > 0$ for $v_B = v_1$, it may not be true that $\partial\delta^c/\partial\lambda > 0$ for $v_B = v_2 > v_1$. One can construct such an example with, for instance, the uniform distribution.

[24] In general, this is the case as long as the entrant's quality is sufficiently dispersed (that is, $F(.)$ does not increase very rapidly).

To understand this result, notice that for a given q_1 and $\delta^s = \delta^c$, the *switch-facilitating* effect and the *demand-expansion* effect are the same in the cases of individual and collective switching costs. However, in the latter case, consumers choose too little data provision in the first period, as they do not take into consideration how their data provision benefits other consumers through 'correlated' big data analytics. This results in a lower q_1, which weakens the *switch-facilitating* effect, and $\delta^c < \delta^s$, which strengthens the *demand-expansion* effect. Similar reasoning applies to the cost of data provision. It is clear that both Conditions (IS-L) and (CS-L) are more easily satisfied when the cost of data provision, c, is larger. This is due to the fact that consumers provide less data when data provision is more costly, which weakens the *switch-facilitating* effect and strengthens the *demand-expansion* effect.

Furthermore, the right to data portability comes together with a set of other rights that give consumers more control over the collection and use of their data. This is likely to reduce the cost of providing data by, for instance, alleviating privacy concerns. We can capture this effect by making the cost of providing data negatively related to the extent of data portability, i.e., assuming that the cost of providing data is $c(\lambda)$ with $c'(\lambda) < 0$. This has the effect of further strengthening the *demand-expansion* effect. Specifically, for a given switching threshold δ', we have

$$\frac{\partial q_1}{\partial \lambda} = \frac{1 - F(\delta') - c'(\lambda)}{c(\lambda)},$$

which is increasing in λ when, for instance, $c(\lambda)$ is not too convex.

4.4 The Innovation Environment

We can also consider firms with different innovative technologies. For instance, some firms adopt more conservative strategies that produce moderate improvements over the incumbent, while other firms adopt more ambitious strategies that produce drastic improvements. We can capture the difference in innovation strategies by the following assumption on $F(\delta)$: assume that the entrant's technology is characterised by ϕ with $\partial F(\delta; \phi)/\partial \phi < 0$. That is, a higher ϕ represents a more ambitious strategy. It produces a large improvement with higher success rates in the sense of First-Order Stochastic Dominance. We can show that

Corollary 2. *Data portability is more likely to raise entry barrier in a more innovative environment. That is, for all $\phi_1 < \phi_2$, if Conditions* (IS-L) *and* (CS-L) *are satisfied for ϕ_1, they are satisfied for ϕ_2.*

The reason is that when the entrant is an ambitious innovator, a consumer anticipates a higher probability of switching in the future and hence a lower probability of staying with the incumbent. This reduces the value of providing data to the incumbent and therefore results in less data provision in the first period, compared to the situation with a less innovative entrant. As a consequence, the *switch-facilitating* impact of data portability on entry, which depends on the level of q_1, becomes weaker. On the other hand, the *demand-expansion* effect of data portability becomes stronger because the consumer is more likely to switch and port data to the more innovative entrant, so that the marginal effect of data portability on q_1 becomes stronger. Combining both effects, data portability is more likely to deter the entry of a more innovative newcomer to the industry. The potential long-term consequence is that data portability may reduce an entrant's incentives to invest in drastic innovations but instead divert its R&D investments to incremental innovations.

4.5 Proportion of Portable Data

In both cases of individual and collective switching costs, we have identified the conditions under which the degree of data portability (from no portability to full portability) monotonically raises the entry barrier. If the proportion of portable data is small, the condition for data portability to raise entry barrier is easier to be satisfied. To see this, let $\lambda' = \beta\lambda$, which means that only a proportion $\beta < 1$ of the data provided in the first period is portable. Then our previous discussion continues to hold as long as δ is monotonically increasing in λ' for $\lambda' \in [0, \beta]$. The only difference is that λ is replaced with $\lambda' \in [0, \beta]$. Thus, if data portability increases entry barrier when all first period data are portable, it also increases entry barrier when only part of the data is portable. The reverse, however, may not be true.

This has interesting implications for the design of data protection legislation. For instance, with the GDPR not covering inferred data, our result implies that data portability is more likely to raise entry barrier when the proportion of portable data is smaller. Furthermore, although we focus on personal data (i.e., the GDPR) in this paper, our results equally apply to non-personal data, which is another actively debated area.[25]

4.6 Consumer Welfare

We have shown that data portability may deter entry, which reduces the intensity of competition. However, this does not mean that consumers are worse-off. Indeed, we have $\partial U/\partial\lambda > 0$, that is, more data portability actually benefits consumers. This is because consumers are fully aware of and take into consideration the impact of their first period data provision on second period switching decisions, and thus they benefit from enhanced value of data brought by improved data portability.

Although data portability increases consumer welfare in this relatively static model, we should be cautious that in a more dynamic context, it may decrease welfare due to the possibility of deterring entry of a more efficient firm. More specifically, in this model, the consumer fully takes into account the externalities of first period data provision on his second period utility. In a more general dynamic model, where existing consumers do not fully take into account the impact of their data provision on future consumers, data portability may enable the incumbent to continuously accumulate data and make future entry increasingly more difficult, which potentially reduces total welfare. This is in the spirit of Farrell and Saloner (1986), where the incumbent can exploit such a coordination problem between different generations of consumers. This may not only lead to "excess inertia" in adoption of a new technology or a new service, but also reduce the pace of innovation and long-run efficiency. We present such an example in Appendix C.4.

5 Extensions

We now consider several extensions of the main model, including a multi-service provider, investments of the incumbent, and behavioural considerations.

5.1 Multi-Service Incumbent

In practice, multi-service providers are commonly observed, as big data technologies enable these providers to derive and analyse consumer data across different services. For instance, Google uses the information it learns from email communications and Facebook

[25] See "Building a European Data Economy", European Commission, COM(2017) 9 final.

uses the information about likes and dislikes to improve their recommendation algorithms. Other platforms that provide multiple services include Alibaba, which is active in retailing, banking and finance, etc. Hence, platforms are able to use the information they collect from one service to improve the quality of another service. Previously, we interpret v_B as a single big data service. A different interpretation is that v_B measures the benefit from multiple services provided by the incumbent.

Let us examine how the possibility of providing multiple services affects the impact of data portability on entry. Suppose the incumbent provides N different services and there is a potential entrant for each service, i.e., each entrant is a single-service provider. To emphasise cross-service network effects, we follow the analysis of collective switching cost. We consider two situations: the first is that when a consumer switches provider for service i, he only loses the big data service related to service i. In this case, a consumer switches in the second period if

$$v_I + v(q_1^i) + w_2^* + v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j) < v_I + \delta + v(\lambda q_1^i) + w_2^*.$$

As we can see from this condition, the big data service from service i analyses not only the data provided to service i but also the data provided to service j that is related to service i to a degree of $\gamma \in [0, 1]$. In this case, a consumer, who stops using service i, loses $v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j)$. For instance, when a consumer stops using Youtube, he loses the value of big data service from Youtube that may analyse the data provided to both Youtube and Gmail. The switching threshold is then given by

$$\delta^m = v(q_1^i) - v(\lambda q_1^i) + v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j).$$

In the second situation, when a consumer switches provider for service i, he loses both the value of big data service from i, $v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j)$, and the value from all other $N - 1$ big data services that analyse the data provided to service i, $(N - 1)v_B(\gamma Q_1^i)$. For example, when a consumer stops using Youtube, he loses the value of big data service from Youtube and also the value of big data service from Gmail that analyses the data provided to Youtube. In this case, a consumer switches in the second period if

$$v_I + v(q_1^i) + w_2^* + v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j) + (N - 1)v_B(\gamma Q_1^i) < v_I + \delta + v(\lambda q_1^i) + w_2^*,$$

which simplifies to

$$\delta^m = v(q_1^i) - v(\lambda q_1^i) + v_B(Q_1^i + \gamma \sum_{j \neq i} Q_1^j) + (N-1) v_B(\gamma Q_1^i).$$

In a symmetric equilibrium, it requires rational expectations, $Q_1^i = q_1^i = q$, where $i = 1, ..., N$. Together with the first order condition with respect to q_1, the equilibrium analysis is similar to the one in our basic model. The general lesson is that with multiple service provision, the marginal value of big data service v_B can be much higher than the marginal value of the data service, which is 1, and hence data portability is more likely to restrict entry.

Since the role of data portability in facilitating entry is undermined within a multi-service incumbent context, competition authorities should be careful about mergers of different services when there are spillovers to data collection and analytics between these services, which are particularly common in data-driven markets. See, for instance the *Facebook/WhatsApp* case, where the European Commission examined whether access to WhatsApp's data would give Facebook a competitive advantage over other competitors.[26]

5.2 Investment Incentives

We have seen in the previous section that data portability may discourage the potential entry of firms that adopt a more innovative strategy. It may also change the direction of the entrant's investment from drastic to incremental innovations, which may generate long-run inefficiencies. Let us now consider the investment incentives of the incumbent. Suppose the incumbent can choose to invest in the basic service value v_I and in the big data service value v_B to maximise δ, that is, to raise the threshold of entry. To illustrate the main idea, let us consider the linear-quadratic case with collective switching cost.

Suppose the incumbent can invest in the basic service at a cost of $D(v)$ to increase the basic service value by v, and it can invest in the big data service at a cost of $D(v_B)$ to achieve the level of big data service value v_B. Given v and v_B, the equilibrium is determined by

$$\begin{aligned} \hat{\delta} &= v + (1 - \lambda + v_B) q_1, \\ c q_1 &= 1 + \lambda + (1 - \lambda) F(\hat{\delta}). \end{aligned}$$

[26] Case COMP/M.7217, *Facebook/WhatsApp*, Commission Decision, 2014.

Thus, the switching threshold satisfies

$$c\hat{\delta} = cv + (1+\lambda)(1-\lambda+v_B) + (1-\lambda)(1-\lambda+v_B)F(\hat{\delta}).$$

The marginal return of investment in basic service is

$$\frac{\partial\hat{\delta}}{\partial v} = \frac{c}{c-(1-\lambda)(1-\lambda+v_B)f(\hat{\delta})}.$$

The marginal return of investment in big data service is

$$\frac{\partial\hat{\delta}}{\partial v_B} = \frac{1+\lambda+(1-\lambda)F(\hat{\delta})}{c-(1-\lambda)(1-\lambda+v_B)f(\hat{\delta})}.$$

Moreover, we have

$$\frac{\partial\hat{\delta}}{\partial \lambda} = \frac{v_B - 2\lambda - (v_B+2-2\lambda)F(\hat{\delta})}{c-(1-\lambda)(1-\lambda+v_B)f(\hat{\delta})}.$$

Thus, $\hat{\delta}$ is increasing in λ if

$$\frac{v_B-2}{v_B} > F(v+\frac{2v_B}{c}).$$

This condition is satisfied when v_B is large, c is large, and/or investment cost in basic service $D(v)$ is large (so the equilibrium level of v is low). Moreover, $\partial q_1/\partial\lambda$ is positive, which means that the equilibrium data provision in the first period is increasing in the degree of data portability λ. This implies that the ratio

$$\frac{\partial\hat{\delta}/\partial v_B}{\partial\hat{\delta}/\partial v} = q_1,$$

is increasing in λ, which, in turn, implies that data portability increases the relative marginal return of investing in big data service. Thus, data portability may reduce the incumbent's incentive to invest in basic service and raise its incentive to invest in big data service, which may further raise entry barrier.

5.3 Naive Consumers and Two-Sided Externality

We have assumed that consumers are sophisticated in the sense that they rationally anticipate how their first period data provision affects their second period switching decisions. In practice, however, consumers can be boundedly rational in different ways, which may prevent them from taking into consideration all the consequences of their data provision. We consider two such cases: shortsightedness and biased beliefs.

First, if consumers are shortsighted in the sense that they make first period data provision decisions to maximise their first period utility instead of total lifetime utility, then data portability will not have any effect on the first period choice of q_1. Since only the *switch-facilitating* effect remains, data portability always facilitates entry.

Second, if consumers are naive in the sense that they have biased beliefs about their future preferences, i.e., either they only consider how the basic data service affects their future behaviour ignoring the presence of the big data service or they underestimate the difficulty of switching, which is likely to be the case with a multi-service incumbent, then our result is strengthened. We illustrate this with the linear-quadratic case with collective switching cost, where consumers ignore the impact of big data service on the entry threshold.[27] The first period data provision satisfies:

$$cq_1 = 1 + \lambda + (1 - \lambda)F(\mathring{\delta}),$$

where $\mathring{\delta} = (1-\lambda)q_1$ is the belief of consumers on the switching threshold when they ignore the presence of big data service. However, in the second period, the actual switching threshold is

$$\delta^n = (1 - \lambda)q_1 + v_B q_1$$

in the presence of big data service. Combining these two equations, the actual equilibrium switching threshold satisfies

$$\frac{c\delta^n}{1 - \lambda + v_B} = 1 + \lambda + (1 - \lambda)F[(1 - \lambda)\frac{\delta^n}{1 - \lambda + v_B}].$$

Applying Implicit Function Theorem, we obtain

$$\frac{\partial \delta^n}{\partial \lambda} = \frac{v_B - 2\lambda - (v_B - 2\lambda + 2)F[(1 - \lambda)\frac{\delta^n}{1-\lambda+v_B}] - \frac{(1-\lambda)v_B\delta^n}{1-\lambda+v_B}f[(1 - \lambda)\frac{\delta^n}{1-\lambda+v_B}]}{c - (1 - \lambda)^2 f((1 - \lambda)\frac{\delta^n}{1-\lambda+v_B})}.$$

Thus, we have

$$\frac{\partial \delta^n}{\partial \lambda}|_{\lambda=1} = \frac{v_B - 2}{c},$$

which is positive whenever $v_B > 2$. Thus, compared to the case with sophisticated consumers (see Condition (CS-L)), full data portability is more likely to raise entry barrier when consumers are naive. A general lesson is that when consumers are boundedly rational, it is central to understand the type of naivety (whether consumers are shortsighted or naive) in order to assess the impact of data portability rules.

[27]Notice that with naive consumers, individual switching cost is the same as collective switching cost, as consumers' data provision does not depend on the big data service value.

Similar reasoning applies when there is two-sided externality and one group of agents do not take into account the impact of their activities on the other group. For instance, data provided by one group (consumers) may enable the platform to provide better services to the other group (advertisers). In this situation, successful entry requires the entrant to provide high enough quality to attract both groups. If consumers ignore their impact on the advertise side, they would expect to switch if $\delta > \mathring{\delta} = (1-\lambda)q_1$. However, since the data provided by the consumers enable the incumbent to offer better services to the advertisers (such as more precise targeting), denoted by a service premium $v_B Q_1$, then successful entry also requires the entrant to overcome this service premium on the advertiser side. The actual entry probability then becomes $(1 - F(\mathring{\delta}))(1 - G(v_B Q_1))$, where $G(\cdot)$ is the distribution of the entrant's service quality to the advertisers. Thus, if consumers do not take into consideration the impact of their data provision on the advertiser side, they will overestimate the probability of entry (expecting entry to occur with probability $1 - F(\mathring{\delta})$, which strengthens the *demand-expansion* effect and further increases the entry barrier.

6 Policy Implications and Conclusion

We have examined the impact of data portability on entry and competition. In addition to the direct *switch-facilitating* effect, we identify the indirect *demand-expansion* effect, where consumers can endogenously adjust their data provision in response to the new regulation on data portability. We show that when big data services are valuable enough, the latter effect (which focuses on the intensive margin) dominates the former effect (which focuses on the extensive margin), and an increased level of data portability can restrict future entry.

This paper is a helpful step in understanding the interactions between data portability, data analytics and data provision. These insights are relevant to the policy discussions surrounding the GDPR in the EU and the data protection legislation in the UK (see, for instance, Department for Business, Energy and Industrial Strategy (2018)). The results point to the potentially limited effectiveness of data portability in fostering market competition under the current legislation, and provide insights into the implementation of such policies. For instance, broadening the range of portable data might be useful.

Initiatives such as Open Data Institute and forced data sharing[28] may be helpful in allowing potential entrants to have sufficient access to data to train their algorithm, which eventually enables them to compete with the incumbents, instead of relying on consumers' motive to switch and port data. The results may also apply beyond digital markets where product or service quality depends on the information generated through consumer participation, for instance, online trading, sharing platforms, and markets for professional services. In such markets, the anticipation of easier switching in the future may induce consumers to use their current service provider more intensively, which eventually hinders future entry and switching.

On the theoretical side, although an extensive literature exists separately for switching costs and network effects, there is much less work that examines both issues together. Yet, in both Lam (2017) and this paper, we show that much of the literature that looks at these two issues separately provides an incomplete picture in the presence of both of them. Therefore, we believe that much more work is needed in this area to enable a deeper understanding of competition in markets where both switching costs and network effects are common. Moreover, whereas the impact of exogenous switching costs is well-established in the literature, the impact of endogenous barriers of switching is much less clear, and even more so when they are generated by both consumers and firms. Our paper makes a first attempt in this direction and could be extended in different directions to consider, for instance, the impact of alternative policies that aim at fostering *ex post* competition on the *ex ante* incentives to invest in data collection and data mining, when consumers play a more active part in the process.

References

Aghion, P. and P. Bolton (1987). Contracts as a barrier to entry. *American Economic Review*, 388–401.

Biglaiser, G., J. Crémer, and G. Dobos (2013). The value of switching costs. *Journal of Economic Theory 148*(3), 935–952.

Bossenbroek, H. and H. Gringhuis (2015). Recommendation in e-commerce.

[28]See, for instance, "Force Google, Apple and Uber to share mapping data, UK advised", Financial Times, November 2018. Available at https://on.ft.com/2zi63IT (Accessed: 23 October 2019).

Available at https://www.luminis.eu/wp-content/uploads/2015/08/White-Paper-Recommendation-in-e-commerce.pdf (Accessed: 28 March 2018).

Bühler, S., R. Dewenter, and J. Haucap (2006). Mobile number portability in Europe. *Telecommunications Policy 30*(7), 385–399.

Cairns, R. D. and J. W. Galbraith (1990). Artificial compatibility, barriers to entry, and frequent-flyer programs. *Canadian Journal of Economics*, 807–816.

Calzolari, G. and V. Denicolò (2018). Price-cost tests and loyalty discounts. *CEPR Discussion Paper DP12924*.

Caminal, R. and C. Matutes (1990). Endogenous switching costs in a duopoly model. *International Journal of Industrial Organization 8*, 353–373.

Chen, Y. and J. Pearcy (2010). Dynamic pricing: when to entice brand switching and when to reward consumer loyalty. *RAND Journal of Economics 41*(4), 674–685.

Christensen, L., A. Colciago, F. Etro, and G. Rafert (2013). The impact of the data protection regulation in the E.U. *Working Paper*.

De Hert, P., V. Papakonstantinou, G. Malgieri, L. Beslay, and I. Sanchez (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review 34*, 193–203.

Department for Business, Energy and Industrial Strategy (2018). Modernising consumer markets. Available at https://www.gov.uk/government/consultations/consumer-green-paper-modernising-consumer-markets (Accessed: 20 June 2018).

European Commission (2016). Guidelines on the right to data portability. Available at `http://ec.europa.eu/justice/data-protection/index_en.htm` (Accessed: 28 March 2018).

Farrell, J. and P. Klemperer (2007). Coordination and lock-in: Competition with switching costs and network effects. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*. North-Holland, Amsterdam, The Netherlands.

Farrell, J. and G. Saloner (1986). Installed base and compatibility: Innovation, product preannouncements, and predation. *American Economic Review*, 940–955.

Gehrig, T. and R. Stenbacka (2004). Differentiation-induced switching costs and poaching. *Journal of Economics & Management Strategy 13*(4), 635–655.

Graef, I. (2015). Mandating portability and interoperability in online social networks: regulatory and competition law issues in the European Union. *Telecommunication Policy 39*, 502–514.

Hartmann, W. R. and V. B. Viard (2008). Do frequency reward programs create switching costs? a dynamic structural analysis of demand in a reward program. *Quantitative Marketing and Economics 6*(2), 109–137.

Lam, W. (2017). Switching costs in two-sided markets. *Journal of Industrial Economics 65*(1), 136–182.

Marinoso, B. G. (2001). Technological incompatibility, endogenous switching costs and lock-in. *Journal of Industrial Economics 49*(3), 281–298.

Shi, M. (2013). A theoretical analysis of endogenous and exogenous switching costs. *Quantitative Marketing and Economics 11*(2), 205–230.

Shin, J. and K. Sudhir (2010). A customer management dilemma: When is it profitable to reward one's own customers? *Marketing Science 29*(4), 671–689.

Stigler Center Report (2019). Report. George J. Stigler Center for the Study of the Economy and the State, Committee for the Study of Digital Platforms, Market Structure and Antitrust Subcommittee. Available at https://www.judiciary.senate.gov/imo/media/doc/market-structure-report%20-15-may-2019.pdf (Accessed: 23 October 2019).

Tremblay, M. (2019). Platform competition and endogenous switching costs. *Journal of Industry, Competition and Trade*, 1–23.

Tucker, C. E. (2014). Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research 51*(5), 546–562.

Van der Auwermeulen, B. (2017). How to attribute the right to data portability in Europe: A comparative analysis of legislations. *Computer Law & Security Review 33*(1), 57–72.

Viard, V. B. (2007). Do switching costs make markets more or less competitive? The case of 800-number portability. *RAND Journal of Economics 38*(1), 146–163.

A Omitted Proofs

A.1 Proof of Proposition 2

From Equation (7), $\partial\delta^s/\partial\lambda > 0$ for all $\lambda \in [0,1]$ if and only if

$$\frac{v_B - 2\lambda}{v_B - 2\lambda + 2 + v_B} \geq F(\delta^s).$$

A necessary condition for the inequality to hold for any $\lambda \in [0,1]$ is $v_B > 2$, and thus we assume that this is the case in the following. Then, the left-hand side of the inequality is decreasing in λ. Moreover, $\partial\delta^s/\partial\lambda > 0$ implies that the right-hand side of the inequality is increasing in λ. Thus, for the inequality to hold for $\lambda \in [0,1]$, it is equivalent to

$$\frac{v_B - 2}{2v_B} \geq F(\delta^s|_{\lambda=1}).$$

At $\lambda = 1$, let $\delta(1) = \delta^s|_{\lambda=1}$, we have

$$c\delta(1) - 2v_B - v_B^2 F(\delta(1)) = 0. \tag{A.1}$$

Thus, the above inequality is equivalent to

$$c\delta(1) - 2v_B \leq \frac{v_B^2 - 2v_B}{2},$$

or

$$\delta(1) \leq \frac{v_B^2 + 2v_B}{2c}.$$

Since the left-hand side of Equation (A.1) is increasing in $\delta(1)$ (due to concavity of the utility function), the inequality is equivalent to

$$\frac{v_B^2 + 2v_B}{2} - 2v_B - v_B^2 F(\frac{v_B^2 + 2v_B}{2c}) \geq 0,$$

which simplifies to

$$\frac{v_B - 2}{2v_B} \geq F(\frac{v_B^2 + 2v_B}{2c}).$$

A.2 Proof of Proposition 3

From Equation (11), $\partial\delta^c/\partial\lambda > 0$ for all $\lambda \in [0,1]$ if and only if

$$\frac{v_B - 2\lambda}{v_B - 2\lambda + 2} \geq F(\delta^c).$$

As in the proof of Proposition 2, the left-hand side is decreasing in λ and thus the above inequality holds if and only if

$$\frac{v_B - 2}{v_B} \geq F(\delta^c|_{\lambda=1}).$$

Let $\delta^c(1) = \delta^c|_{\lambda=1}$, which must satisfy

$$c\delta^c(1) = 2v_B.$$

Therefore, $\partial\delta^c/\partial\lambda > 0$ if

$$\frac{v_B - 2}{v_B} \geq F(\frac{2v_B}{c}).$$

A.3 Proof of Corollary 1

Comparing Condition (IS-L) and Condition (CS-L), the left-hand side clearly shows

$$\frac{v_B - 2}{2v_B} < \frac{v_B - 2}{v_B}.$$

On the right-hand side, we have

$$\frac{v_B^2 + 2v_B}{2c} - \frac{2v_B}{c} = \frac{v_B(v_B - 2)}{2c},$$

which is positive as $v_B > 2$. Thus, when Condition (IS-L) is satisfied, Condition (CS-L) must also be satisfied.

B The General Setup

In this section, we present a general setup and show how our insights from the linear-quadratic case can be easily generalised. To deliver our main insights, we focus on the case where the second period utility is separable in data provided in the first period and second period, that is,

Assumption B.1. *$\nu(\lambda q_1 + q_2) = v(\lambda q_1) + v(q_2)$, for $\lambda \in [0, 1]$.*

This allows us to separate the data provision decisions in the first and second period, and thus separate the *ex post* effect of data portability on entry barrier from the *ex ante* effect of data portability on data provision. Specifically, a consumer provides the same amount of data, $q_2^* = argmax_{q_2} v(q_2) - C(q_2)$, in the second period no matter whether he switches or not. We denote this utility from second period data provision by $w_2^* = v(q_2^*) - C(q_2^*)$. More generally, data provision in the second period may either increase (i.e., data provision across the two periods are complements) or decrease (i.e., substitutes) with that provided in the first period. In the former case, our results are strengthened as this creates additional incentives for a consumer to stay with the incumbent. In the latter case, the incumbency advantage is weakened, but we show in Appendix C.1 that our main insights carry through.

In addition, we assume that the functions in our model satisfy the following conditions:

Assumption B.2. *(i) $v(q)$ and $v_B(q)$ are increasing and concave; (ii) $C(q)$ is increasing and convex.*

Assumption B.2 are standard assumptions of concave utility and convex cost functions, and we assume that the cost function is sufficiently convex such that all utility maximisation problems in the following analysis are well-defined.

B.1 Benchmark without Big Data Analytics

We start with the benchmark without big data analytics. When $v_B(\cdot) = 0$, in the second period, a consumer will switch if

$$u_2^I = v_I + v(q_1) + w_2^* < u_2^E = v_I + \delta + v(\lambda q_1) + w_2^*,$$

which simplifies to $\delta > \delta^o$ with

$$\delta^o = v(q_1) - v(\lambda q_1). \tag{B.2}$$

In the first period, a consumer chooses q_1 to maximise his total utility U across two periods, given by

$$U(q_1) = v_I + v(q_1) - C(q_1) + \int_0^{\delta^o} [v_I + v(q_1) + w_2^*]dF(\delta) + \int_{\delta^o}^{\Delta} [v_I + \delta + v(\lambda q_1) + w_2^*]dF(\delta).$$

The first order condition yields the optimal first period data provision:

$$C'(q_1) - v'(q_1) - v'(q_1)F(\delta^o) - \lambda v'(\lambda q_1)(1 - F(\delta^o)) = 0. \tag{B.3}$$

Thus, Equations (B.2) and (B.3) define an equilibrium with data provision q_1^o in the first period and a switching threshold δ^o in the second period.

The *switch-facilitating* effect is given by $\partial\delta^o/\partial\lambda = -q_1 v'(\lambda q_1) < 0$. The *demand-expansion* effect consists of two parts: the impact of data portability on data provision, $\partial q_1/\partial\lambda = \frac{[v'(\lambda q_1)+\lambda q_1 v''(\lambda q_1)](1-F(\delta^o))}{-U''(q_1)}$, and the impact of data provision on the entry threshold, $\partial\delta^o/\partial q_1 = v'(q_1) - \lambda v'(\lambda q_1)$. Thus, we have the following result,

Proposition B.1. *Without big data analytics, data portability raises the entry threshold δ^o, that is, $d\delta^o/d\lambda > 0$, if and only if*

$$\underbrace{q_1^o v'(\lambda q_1^o)}_{Switch-Facilitating} < \underbrace{\frac{[v'(\lambda q_1^o) + \lambda q_1^o v''(\lambda q_1^o)](1 - F(\delta^o))}{-U''(q_1^o)}(v'(q_1^o) - \lambda v'(\lambda q_1^o))}_{Demand-Expansion}, \tag{NB}$$

where $-U''(q_1^o) = C''(q_1^o) - v''(q_1^o)(1 + F(\delta^o)) - \lambda^2 v''(\lambda q_1^o)(1 - F(\delta^o))$. Moreover, data portability always lowers the entry threshold when λ is close to 1, i.e., $\frac{d\delta^o}{d\lambda}|_{\lambda\to 1} < 0$.

Proof. The equilibrium δ^o and q_1^o satisfy

$$\begin{aligned} v(q_1^o) - v(\lambda q_1^o) - \delta^o &= 0, \\ C'(q_1^o) - v'(q_1^o) - v'(q_1^o)F(\delta^o) - \lambda v'(\lambda q_1^o)(1 - F(\delta^o)) &= 0. \end{aligned}$$

Taking total differentiation with respect to λ, we obtain

$$-\frac{d\delta^o}{d\lambda} + \underbrace{[v'(q_1^o) - \lambda v'(\lambda q_1^o)]}_{A}\frac{dq_1^o}{d\lambda} = \underbrace{q_1^o v'(\lambda q_1^o)}_{B},$$

$$- \underbrace{[v'(q_1^o) - \lambda v'(\lambda q_1^o)]f(\delta^o)}_{\Gamma}\frac{d\delta^o}{d\lambda} + \underbrace{[C''(q_1^o) - v''(q_1^o)(1 + F(\delta^o)) - \lambda^2 v''(\lambda q_1^o)(1 - F(\delta^o))]}_{\Delta}\frac{dq_1^o}{d\lambda}$$
$$= \underbrace{[v'(\lambda q_1^o) + \lambda q_1^o v''(\lambda q_1^o)](1 - F(\delta^o))}_{E}.$$

By assumption, all A, B, Γ, Δ, E are positive. Using Cramer's Rule, we obtain

$$\frac{d\delta^o}{d\lambda} = \frac{det \begin{bmatrix} B & A \\ E & \Delta \end{bmatrix}}{det \begin{bmatrix} -1 & A \\ -\Gamma & \Delta \end{bmatrix}}.$$

The denominator is equal to $-\Delta + A\Gamma$, which is negative and follows from equilibrium stability. Specifically, from Equation (B.2), we have $\partial\delta^o/\partial q_1 = A$. From Equation (B.3), we have $\partial\delta^o/\partial q_1 = \Delta/\Gamma$. A stable equilibrium (δ^o, q_1^o) requires $\Delta/\Gamma > A$, hence, $A\Gamma - \Delta < 0$.

Thus, for $d\delta^o/d\lambda > 0$, we need $\Delta B - AE < 0$, that is,

$$\begin{aligned} & q_1^o v'(\lambda q_1^o)[C''(q_1^o) - v''(q_1^o)(1 + F(\delta^o)) - \lambda^2 v''(\lambda q_1^o)(1 - F(\delta^o))] \\ < \quad & [v'(q_1^o) - \lambda v'(\lambda q_1^o)][v'(\lambda q_1^o) + \lambda q_1^o v''(\lambda q_1^o)](1 - F(\delta^o)). \end{aligned}$$

Notice that $\Delta = -U''(q_1^o)$, thus the above condition simplifies to

$$q_1^o v'(\lambda q_1^o) < \frac{[v'(\lambda q_1^o) + \lambda q_1^o v''(\lambda q_1^o)](1 - F(\delta^o))}{-U''(q_1^o)}(v'(q_1^o) - \lambda v'(\lambda q_1^o)).$$

Furthermore, as $\lambda \to 1$, the right-hand side of Equation (NB) approaches zero as A approaches zero, and thus the condition fails to satisfy, i.e., data portability always lowers the entry threshold.

□

Proposition B.1 generalises Proposition 1, although with a small difference due to the curvature of $v(.)$: with general value and cost functions, the *switch-facilitating* effect dominates the *demand-expansion* effect when the degree of data portability is sufficiently strong. That is, the relationship between data portability and entry threshold can be either monotonically decreasing or inverted-U shape in the general case, whereas data portability always monotonically reduces entry barrier without big data analytics in the linear-quadratic case.

B.2 Big Data Analytics and Individual Switching Cost

Now we proceed to the situation with big data analytics and we start with the analysis of individual switching cost. Given q_1, the consumer switches in the second period if

$$v_I + v(q_1) + v_B(q_1) + w_2^* < v_I + \delta + v(\lambda q_1) + w_2^*,$$

which defines a switching threshold δ^s, given by

$$\delta^s = v(q_1) - v(\lambda q_1) + v_B(q_1). \tag{B.4}$$

The total utility in the first period is

$$\begin{aligned} U(q_1) =& v_I + v(q_1) - C(q_1) \\ &+ \int_0^{\delta^s} [v_I + v(q_1) + v_B(q_1) + w_2^*] dF(\delta) + \int_{\delta^s}^{\Delta} [v_I + \delta + v(\lambda q_1) + w_2^*] dF(\delta). \end{aligned}$$

The optimal data provision satisfies

$$C'(q_1) - v'(q_1) - [v'(q_1) + v_B'(q_1)]F(\delta^s) - \lambda v'(\lambda q_1)[1 - F(\delta^s)] = 0. \tag{B.5}$$

Equation (B.4) and (B.5) define a new equilibrium (δ^s, q_1^s) under individual switching cost. The *switch-facilitating* effect is now given by

$$\frac{\partial \delta^s}{\partial \lambda} = -q_1 v'(\lambda q_1),$$

whereas the *demand-expansion* effect becomes

$$\frac{\partial q_1}{\partial \lambda} = \frac{[v'(\lambda q_1) + \lambda q_1 v''(\lambda q_1)](1 - F(\delta^s))}{-U''(q_1)},$$

and

$$\frac{\partial \delta^s}{\partial q_1} = v_B'(q_1) + v'(q_1) - \lambda v'(\lambda q_1).$$

Together, we show that

Proposition B.2. *With big data analytics and individual switching cost, data portability increases the switching threshold δ^s, that is, $d\delta^s/d\lambda > 0$ for all $\lambda \in [0,1]$, if and only if*

$$\underbrace{q_1^s v'(\lambda q_1^s)}_{Switch-Facilitating} < \underbrace{\frac{[v'(\lambda q_1^s) + \lambda q_1^s v''(\lambda q_1^s)](1 - F(\delta^s))}{-U''(q_1^s)} [v_B'(q_1^s) + v'(q_1^s) - \lambda v'(\lambda q_1^s)]}_{Demand-Expansion}, \tag{IS}$$

where $-U''(q_1^s) = C''(q_1^s) - v''(q_1^s) - [v''(q_1^s) + v_B''(q_1^s)]F(\delta^s) - \lambda^2 v''(\lambda q_1^s)(1 - F(\delta^s))$.

Proof. Similar to the the proof of Proposition B.1, we take total differentiation with respect to λ for Equation (B.4) and (B.5) and obtain

$$-\frac{d\delta^s}{d\lambda} + \underbrace{[v_B'(q_1^s) + v'(q_1^s) - \lambda v'(\lambda q_1^s)]}_{A} \frac{dq_1^s}{d\lambda} = \underbrace{q_1^s v'(\lambda q_1^s)}_{B},$$

$$- \underbrace{[v_B'(q_1^s) + v'(q_1^s) - \lambda v'(\lambda q_1^s)] f(\delta^s)}_{\Gamma} \frac{d\delta^s}{d\lambda}$$

$$+ \underbrace{[C''(q_1^s) - v''(q_1^s) - [v''(q_1^s) + v_B''(q_1^s)] F(\delta^s) - \lambda^2 v''(\lambda q_1^s)(1 - F(\delta^s))]}_{\Delta} \frac{dq_1^s}{d\lambda}$$

$$= \underbrace{[v'(\lambda q_1^s) + \lambda q_1^s v''(\lambda q_1^s)](1 - F(\delta^s))}_{E}.$$

Hence, $d\delta^s/d\lambda > 0$ if $\Delta B < AE$. Notice also that $\Delta = -U''(q_1^s)$ and the condition can be rewritten as

$$q_1^s v'(\lambda q_1^s) < \frac{[v'(\lambda q_1^s) + \lambda q_1^s v''(\lambda q_1^s)](1 - F(\delta^s))}{-U''(q_1^s)} [v_B'(q_1^s) + v'(q_1^s) - \lambda v'(\lambda q_1^s)].$$

□

Proposition B.2 generalises Proposition 2. Specifically, with big data analytics, data portability may monotonically raise entry barrier, due to the fact that big data analytics enhance the *demand-expansion* effect, that is, $v_B'(q_1^s) > 0$. Thus, even if λ approaches 1, the *demand-expansion* effect does not vanish.

B.3 Big Data Analytics and Collective Switching Cost

In the case of network effects, a consumer's value for the big data service depends on the aggregate data provided by all consumers in the first period. That is, given q_1, a consumer switches in the second period if

$$v_I + v(q_1) + v_B(Q_1) + w_2^* < v_I + \delta + v(\lambda q_1) + w_2^*,$$

which simplifies to $\delta > \delta^c$ with

$$\delta^c = v(q_1) - v(\lambda q_1) + v_B(Q_1). \tag{B.6}$$

The total utility in the first period becomes

$$\begin{aligned} U(q_1) =& v_I + v(q_1) - C(q_1) \\ &+ \int_0^{\delta^c} [v_I + v(q_1) + v_B(Q_1) + w_2^*] dF(\delta) + \int_{\delta^c}^{\Delta} [v_I + \delta + v(\lambda q_1) + w_2^*] dF(\delta), \end{aligned}$$

which yields the following first-order condition:

$$C'(q_1) - v'(q_1) - v'(q_1) F(\delta^c) - \lambda v'(\lambda q_1)(1 - F(\delta^c)) = 0. \tag{B.7}$$

The equilibrium also requires rational expectations, that is,

$$Q_1 = q_1. \tag{B.8}$$

Equations (B.6), (B.7), and (B.8) define an equilibrium (δ^c, q_1^c). We have the following result:

Proposition B.3. *With big data analytics and collective switching cost, data portability increases the switching threshold δ^c, that is, $d\delta^c/d\lambda > 0$ for all $\lambda \in [0,1]$, if and only if*

$$\underbrace{q_1^c v'(\lambda q_1^c)}_{Switch-Facilitating} < \underbrace{\frac{[v'(\lambda q_1^c) + \lambda q_1^c v''(\lambda q_1^c)](1-F(\delta^c))}{-U''(q_1^c)}[v_B'(q_1^c) + v'(q_1^c) - \lambda v'(\lambda q_1^c)]}_{Demand-Expansion}, \tag{CS}$$

where $-U''(q_1^c) = C''(q_1^c) - v''(q_1^c)[1+F(\delta^c)] - \lambda^2 v''(\lambda q_1^c)(1-F(\delta^c))]$.

Proof. The equilibrium (δ^c, q_1^c) satisfy

$$\begin{aligned} v(q_1^c) - v(\lambda q_1^c) + v_B(q_1^c) - \delta^c &= 0, \\ C'(q_1) - v'(q_1) - v'(q_1)F(\delta^c) - \lambda v'(\lambda q_1)(1-F(\delta^c)) &= 0. \end{aligned}$$

Similarly, we take total differentiation with respect to λ and obtain

$$-\frac{d\delta^c}{d\lambda} + \underbrace{[v_B'(q_1^c) + v'(q_1^c) - \lambda v'(\lambda q_1^c)]}_{A}\frac{dq_1^c}{d\lambda} = \underbrace{q_1^c v'(\lambda q_1^c)}_{B},$$

$$\begin{aligned} &- \underbrace{[v'(q_1^c) - \lambda v'(\lambda q_1^c)]f(\delta^c)}_{\Gamma}\tfrac{d\delta^c}{d\lambda} \\ &+ \underbrace{[C''(q_1^c) - v''(q_1^c)[1+F(\delta^c)] - \lambda^2 v''(\lambda q_1^c)(1-F(\delta^c))]}_{\Delta}\tfrac{dq_1^c}{d\lambda} \\ &= \underbrace{[v'(\lambda q_1^c) + \lambda q_1^c v''(\lambda q_1^c)](1-F(\delta^c))}_{E}. \end{aligned}$$

Thus, $d\delta^c/d\lambda > 0$ if $\Delta B < AE$, which can be rewritten as

$$q_1^c v'(\lambda q_1^c) < \frac{[v'(\lambda q_1^c) + \lambda q_1^c v''(\lambda q_1^c)](1-F(\delta^c))}{-U''(q_1^c)}[v_B'(q_1^c) + v'(q_1^c) - \lambda v'(\lambda q_1^c)].$$

□

Proposition B.3 generalises Proposition 3. Furthermore, comparing the concavity of the utility functions in Conditions (IS) and (CS) (to be more specific, the term $-v_B''(q_1^s)F(\delta^s)$ in Condition (IS)), we can see that data portability, *ceteris paribus*, is more likely to raise entry barrier under collective switching cost compared to individual switching cost, as the first period utility function is less concave in Equation (CS) and thus the *demand-expansion* effect becomes stronger. Comparing the numerators in Equations (NB) and (CS), we can see that data portability, *ceteris paribus*, is more likely to raise entry barrier under collective switching cost compared to no big data analytics, as the *demand-expansion* effect is stronger due to the presence of $v'_B(q_1^c)$ in Equation (CS). In the next section, we dig deeper into the difference between Propositions B.2 and B.3 taking into account differences between the optimal first period data provisions and those between switching thresholds under individual and collective switching costs.

A necessary condition for all Conditions (NB), (IS), and (CS) to be satisfied is $v'(\lambda q_1^c)+\lambda q_1^c v''(\lambda q_1^c) > 0$, which is equivalent to $R_v(q) = -\frac{qv''(q)}{v'(q)} < 1$. This is related to the curvature of $v(q)$ and is satisfied when $v(q)$ is not too concave.[29] The following corollary is then immediate:

Corollary B.1. *When $R_v(q) \geq 1$, data portability always facilitates entry.*

A few remarks: First, the conditions in Propositions B.2 and B.3 are sufficient and necessary for data portability to increase the entry barrier monotonically when λ increases from 0 to 1. If, however, the only available options are no data portability ($\lambda = 0$) and full data portability ($\lambda = 1$), then we only need weaker conditions for full data portability to raise entry barrier, as we only need $\delta(1) > \delta(0)$ but not $\delta(\lambda)$ being monotonically increasing in λ.

Second, whilst the conditions in Proposition B.2 and B.3 need to be satisfied for every $\lambda \in [0, 1]$, we can simplify these conditions with additional regularity assumptions $U'''(q) \leq 0$, $R'_v(q) \geq 0$, and $R_{v'}(q) \leq 2$ to facilitate meaningful comparative statics (which are explained later in Proposition B.4 and are clearly satisfied in the linear-quadratic example). Under these assumptions we only need Condition (IS) and (CS) to be satisfied at $\lambda = 1$ for the entry barrier to monotonically increase with the degree of data portability. That is, in the case of individual switching cost, we only need

$$\frac{1}{1 - R_v(q_1^s(1))} < \frac{1 - F(\delta^s(1))}{q_1^s(1)} \frac{v'_B(q_1^s(1))}{C''(q_1^s(1)) - 2v''(q_1^s(1)) - v''_B(q_1^s(1))F(\delta^s(1))},$$

where $(\delta^s(1), q_1^s(1))$ is the equilibrium corresponding to $\lambda = 1$. In the case of collective switching cost, we only need

$$\frac{1}{1 - R_v(q_1^c)} < \frac{1 - F(\delta^c(1))}{q_1^c(1)} \frac{v'_B(q_1^c(1))}{C''(q_1^c(1)) - 2v''(q_1^c(1))},$$

[29] Alternatively, it means that the coefficient of relative risk aversion is smaller than one.

where $(\delta^c(1), q_1^c(1))$ is the equilibrium corresponding to $\lambda = 1$.

B.4 When is Data Portability More Likely to Raise Entry Barrier?

We have seen that with big data service and with either individual or collective switching cost, data portability can lead to higher entry barrier for the entrant. We now discuss the conditions under which this is more likely to happen. To fix ideas, let us consider the case of collective switching cost[30] and rewrite Condition (CS) as

$$\frac{1}{1 - R_v(\lambda q_1^c)} < \frac{1 - F(\delta^c)}{q_1^c} \frac{v_B'(q_1^c) + v'(q_1^c) - \lambda v'(\lambda q_1^c)}{C''(q_1^c) - v''(q_1^c)[1 + F(\delta^c)] - \lambda^2 v''(\lambda q_1^c)(1 - F(\delta^c))}. \tag{CS'}$$

Consider first a change in parameters (e.g., a shift in the marginal cost of providing data, an increase in the stand-alone value of big data analytics, or a change in the entrant's technology) that leads to a new equilibrium (δ^{cc}, q_1^{cc}) such that $\delta^{cc} > \delta^c$ and $q_1^{cc} > q_1^c$, without changing the first order derivatives of $v(q)$ and second order derivatives of $C(q)$ and $v(q)$.[31] We can show that:

Proposition B.4. *If $d\delta^{cc}/d\lambda > 0$ for all $\lambda \in [0, 1]$, then $d\delta^c/d\lambda > 0$ for all $\lambda \in [0, 1]$ if $U'''(q) \leq 0$, $R_v'(q) \geq 0$, and $R_{v'}(q) \leq 2$, where $R_{v'}(q) = -\frac{qv'''(q)}{v''(q)}$.*

Proof. If $R_v' > 0$, the left-hand side of Condition (CS') is increasing in q_1^c. On the right-hand side, clearly, $\frac{1-F(\delta^{cc})}{q_c^{cc}} < \frac{1-F(\delta^c)}{q_c}$. Since $v_B(\cdot)$ is concave, we have $v_B'(q_1^{cc}) < v_B'(q_1^c)$. We also have the denominator $C''(q_1^c) - v''(q_1^c)[1 + F(\delta^c)] - \lambda^2 v''(\lambda q_1^c)(1 - F(\delta^c)) = -U''$, which is increasing in q_1^c. Furthermore, we have $v'(q_1^c) - \lambda v'(\lambda q_1^c)$ decreases with q_1^c. Specifically,

$$\frac{\partial v'(q_1^c) - \lambda v'(\lambda q_1^c)}{\partial q_1^c} = v''(q_1^c) - \lambda^2 v''(\lambda q_1^c),$$

which is negative as $\lambda^2 v''(\lambda q_1^c)$ is decreasing in λ, since

$$\frac{\partial \lambda^2 v''(\lambda q_1^c)}{\partial \lambda} = \lambda[2v''(\lambda q_1^c) + \lambda q_1^c v'''(\lambda q_1^c)] = \lambda v''(\lambda q_1^c)[2 - R_{v'}(\lambda q_1^c)] < 0.$$

[30] The case of individual switching cost yields similar results.

[31] This is a natural case to consider. Specifically, define $\delta(q_1)$ as the entry threshold for a given q_1 and $q_1(\delta)$ as the optimal first period data provision for a given entry threshold δ. Since both functions are upward-sloping, any change in parameters, which affects only one function or both functions in the same direction, must move the equilibrium (δ, q_1) in the same direction.

Hence, $v''(q_1^c) - \lambda^2 v''(\lambda q_1^c) < v''(q_1^c) - v''(q_1^c) = 0$. In addition, $v''(q_1^c) - \lambda^2 v''(\lambda q_1^c)$ being negative also means the denominator $-U''$ is increasing in $F(\delta^c)$. In summary, the right-hand side of Condition (CS') is decreasing in δ^c and q_1^c, together with the left-hand side being increasing in q_1^c, this means that the condition is more difficult to satisfy for $\delta^{cc} > \delta^c$ and $q_1^{cc} > q_1^c$. □

The conditions in Proposition B.4 are clearly satisfied in the linear-quadratic case, as $U''' = R_v' = R_{v'} = 0$. They are also satisfied, for instance, when $v(\cdot)$ takes the CRRA form and C''' is sufficiently large. More specifically, Proposition B.4 shows how data provision influences the impact of data portability on entry. When the level of data provision is high and thus the switching threshold is high, data portability is less likely to further raise the threshold.

Network Effects and Innovation Environment

With Proposition B.4, it is straightforward to generalise Corollary 1 and 2. First, we have

Corollary B.2. *Data portability is more likely to raise entry barrier in the case with collective switching cost compared to the case with individual switching cost, i.e., if $d\delta^s/d\lambda > 0$, then $d\delta^c/d\lambda > 0$.*

The proof is straightforward and thus omitted. Intuitively, in addition to the difference in the concavity of first period utility function as discussed in Proposition B.3, consumers also provide less data under collective switching cost (due to the fact that they do not take into account the network externality of their individual decisions), which further amplifies the entry deterrence effect of data portability as shown by Proposition B.4.

Similarly, if the entrant is more innovative in the sense of First-Order Stochastic Dominance, i.e., the entrant's technology is characterised by ϕ with $\partial F(\delta;\phi)/\partial\phi < 0$, we can show that

Corollary B.3. *Data portability is more likely to raise entry barrier in a more innovative environment. That is, for all $\phi_1 < \phi_2$, if Conditions* (CS') *is satisfied for ϕ_1, it is satisfied for ϕ_2.*

The proof is intuitive and straightforward. In Condition (CS'), keeping δ^c and q_1^c fixed, a higher ϕ lowers $F(\delta;\phi)$ and thus increases the right-hand side (by making the utility function less concave). Furthermore, since a consumer anticipates higher probability of switching in the future and is less willing to provide data in the first period, a higher ϕ lowers δ^c and q_1^c, which further increases the right-hand side and lowers the left-hand side. Thus, the Condition is easier to be satisfied for a higher ϕ; the same analysis holds for Condition (IS). This leads to the potential long-term consequence that data portability may reduce an entrant's incentives to invest in drastic innovations, and instead divert its R&D investments to incremental innovations.

C Further Discussions

C.1 Non-Separable First and Second Period Data Provision

Let us consider the collective switching cost case.[32] Suppose a consumer's second period utility is non-separable in first and second period data provision, i.e., the second period utility from data provision depends on the total amount of data. His utility of staying with the incumbent in the second period is

$$u_2^I = max_{q_2} v_I + v(q_1 + q_2) - C(q_2) + v_B(Q_1) = v_I + w^*(q_1 + q_2^I) + v_B(Q_1),$$

where $w^*(q_1 + q_2^I)$ denotes the utility obtained from the optimal second period data provision. The utility of switching to the entrant is

$$u_2^E = max_{q_2} v_I + \delta + v(\lambda q_1 + q_2) - C(q_2) = v_I + \delta + w^*(\lambda q_1 + q_2^E).$$

Thus, the switching threshold is given by

$$\tilde{\delta} = v_B(Q_1) + w^*(q_1 + q_2^I) - w^*(\lambda q_1 + q_2^E). \tag{C.9}$$

The first period utility is given by

$$\begin{aligned} U(q_1) = \quad & v_I + v(q_1) - C(q_1) \\ & + \textstyle\int_0^{\tilde{\delta}} [v_I + w^*(q_1 + q_2^I) + v_B(Q_1)] dF(\delta) + \int_{\tilde{\delta}}^{\Delta} [v_I + \delta + w^*(\lambda q_1 + q_2^E)] dF(\delta), \end{aligned}$$

and the first order condition is given by

$$C'(q_1) - v'(q_1) - v'(q_1 + q_2^I) F(\tilde{\delta}) - \lambda v'(\lambda q_1 + q_2^E)(1 - F(\tilde{\delta})) = 0. \tag{C.10}$$

We also have $Q_1 = q_1$. Thus, taking total differentiation with respect to λ, from Equation (C.9), we obtain

$$-\frac{d\tilde{\delta}}{d\lambda} + [v_B'(q_1) + v'(q_1 + q_2^I) - \lambda v'(\lambda q_1 + q_2^E)] \frac{dq_1}{d\lambda} = q_1 v'(\lambda q_1 + q_2^E).$$

From Equation (C.10), we obtain

$$\begin{gathered} -[v'(q_1 + q_2^I) - \lambda v'(\lambda q_1 + q_2^E)] f(\tilde{\delta}) \tfrac{d\tilde{\delta}}{d\lambda} \\ +[C''(q_1) - v''(q_1) - v''(q_1 + q_2^I)(1 + \tfrac{dq_2^I}{dq_1}) F(\tilde{\delta}) - \lambda^2 v''(\lambda q_1 + q_2^E)(\lambda + \tfrac{\partial q_2^E}{\partial q_1})(1 - F(\tilde{\delta}))] \tfrac{dq_1}{d\lambda} \\ = [v'(\lambda q_1 + q_2^E) + \lambda(q_1 + \tfrac{\partial q_2^E}{\partial \lambda}) v''(\lambda q_1 + q_2^E)](1 - F(\tilde{\delta})). \end{gathered}$$

[32] We can apply a similar proof to the individual switching cost case.

Following similar reasoning as the proofs of Propositions B.2 and B.3, we have $d\tilde{\delta}/d\lambda > 0$ if

$$\begin{aligned} & q_1 v'(\lambda q_1 + q_2^E) \\ & < [v'(\lambda q_1 + q_2^E) + \lambda(q_1 + \frac{\partial q_2^E}{\partial \lambda}) v''(\lambda q_1 + q_2^E)](1 - F(\tilde{\delta})) \frac{v_B'(q_1) + v'(q_1 + q_2^I) - \lambda v'(\lambda q_1 + q_2^E)}{-U''(q_1)}, \end{aligned}$$

or equivalently

$$\frac{1}{1 + \lambda(q_1 + \frac{\partial q_2^E}{\partial \lambda}) \frac{v''(\lambda q_1 + q_2^E)}{v'(\lambda q_1 + q_2^E)}} < \frac{1 - F(\tilde{\delta})}{q_1} \frac{v_B'(q_1) + v'(q_1 + q_2^I) - \lambda v'(\lambda q_1 + q_2^E)}{-U''(q_1)}.$$

A similar result to Proposition B.4 holds with the exception of replacing $R_v' > 0$ with $A_v' > 0$, where $A_v = -\frac{v''}{v'}$ is the coefficient of absolute risk aversion. Thus, in the non-separable case, all our results and comparative statics carry through.

C.2 Complementary Basic and Data Service

Consider the case of collective switching cost and the consumer's utility from basic service and data service being complementary. Specifically, in the second period, if the consumer stays with the incumbent, he obtains a utility of

$$u_2^I = v_I[v(q_1) + v(q_2)] - C(q_2) + v_B(Q_1);$$

if the consumer switches to the entrant, he obtains a utility of

$$u_2^E = (v_I + \delta)[v(\lambda q_1) + v(q_2)] - C(q_2).$$

Let $G(x) = \max_{q_2} xv(q_2) - C(q_2)$, we can rewrite the second period utilities as

$$u_2^I = v_I v(q_1) + G(v_I) + v_B(Q_1),$$

and

$$u_2^E = (v_I + \delta)v(\lambda q_1) + G(v_I + \delta).$$

Thus, given q_1, the switching threshold, δ^m, is determined by

$$(v_I + \delta^m)v(\lambda q_1) + G(v_I + \delta^m) - [v_I v(q_1) + G(v_I) + v_B(Q_1)] = 0. \tag{C.11}$$

In the first period, the consumer chooses q_1 to maximise

$$\begin{aligned} U(q_1) = \; & v_I v(q_1) - C(q_1) \\ & + \int_0^{\delta^m} v_I v(q_1) + G(v_I) + v_B(Q_1) dF(\delta) + \int_{\delta^m}^{\Delta} (v_I + \delta) v(\lambda q_1) + G(v_I + \delta) dF(\delta). \end{aligned}$$

The first order condition is given by

$$v_I v'(q_1) + \int_0^{\delta^m} v_I v'(q_1) dF(\delta) + \int_{\delta^m}^{\Delta} (v_I + \delta)\lambda v'(\lambda q_1) dF(\delta) = C'(q_1). \tag{C.12}$$

Then the equilibrium is given by Equation (C.11) and (C.12), and $Q_1 = q_1$. Similar to the above, from Equation (C.11), the *switch-facilitating* effect is given by

$$\frac{\partial \delta^m}{\partial \lambda} \propto -q_1 v'(\lambda q_1) < 0.$$

From Equation (C.11) and (C.12), the *demand-expansion* effect is given by

$$\frac{\partial q_1}{\partial \lambda} \propto [v'(\lambda q_1) + \lambda q_1 v''(\lambda q_1)] \int_{\delta^m}^{\Delta} (v_I + \delta) dF(\delta),$$

which is positive when $R_v(q) < 1$, and

$$\frac{\partial \delta^m}{\partial q_1} \propto v_I v'(q_1) + v_B'(q_1) - (v_I + \delta^m)\lambda v'(\lambda q_1),$$

a sufficient condition for the latter to be positive is when $v_B(q_1)$ is not too concave, i.e., $v_B'/v_B \geq v'/v$, which is always satisfied in the linear-quadratic case. With complementary basic and data service, there is an additional effect that lowers the entry barrier: the marginal value of data is higher at the entrant's service due to complementarity, and hence the value of the entrant's service increases faster as a consumer provides and ports more data. Note, however, that this does not overturn our results in the main model, as we can still identify conditions under which data portability monotonically raises the entry barrier, albeit stricter.[33]

We further illustrate this with the linear-quadratic example, that is, $v(q) = q$, $v_B(q) = v_B q$ and $C(q) = cq^2/2$. In the second period, a consumer switches if

$$\max_{q_2} (v_I + \delta)(\lambda q_1 + q_2) - \frac{1}{2} c q_2^2 > \max_{q_2} v_I (q_1 + q_2) - \frac{1}{2} c q_2^2 + v_B Q_1,$$

Thus, the switching threshold δ^m satisfies

$$\lambda(v_I + \delta^m) q_1 + \frac{(v_I + \delta^m)^2}{2c} = v_I q_1 + \frac{v_I^2}{2c} + v_B Q_1.$$

In the first period, the data provision q_1 is chosen to maximise the lifetime utility, given by

$$U = v_I q_1 - \frac{1}{2} c q_1^2 + \int_0^{\delta^m} [v_I q_1 + \frac{v_I^2}{2c} + v_B Q_1] dF(\delta) + \int_{\delta^m}^{\infty} [\lambda (v_I + \delta) q_1 + \frac{(v_I + \delta)^2}{2c}] dF(\delta).$$

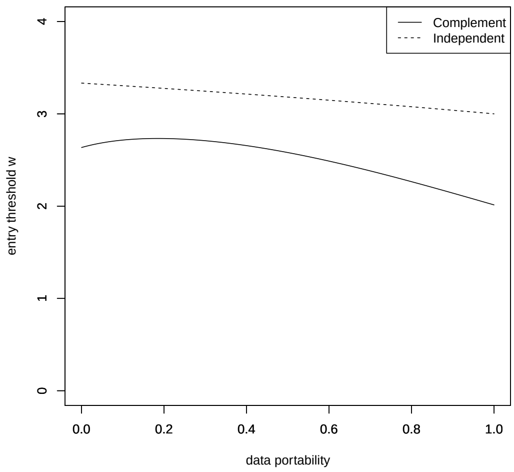
[33] Note that when the degree of data portability is small, i.e., when λ is small, the *demand-expansion* effect is always positive. Hence, we are more likely to obtain a hump-shaped relationship between the degree of data portability and the entry threshold. Yet, the main force at work remains valid.

The optimal data provision in the first period satisfies

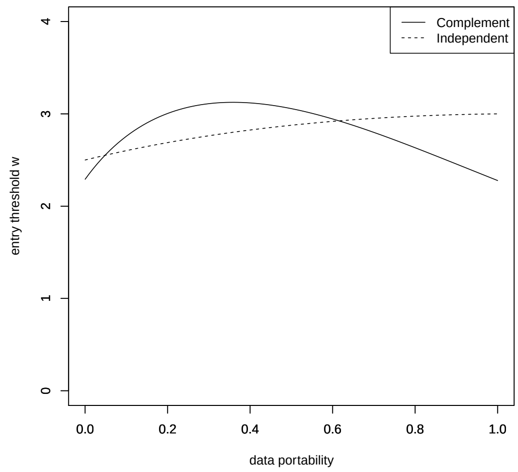
$$cq_1 = v_I + v_I F(\delta^m) + \lambda \int_{\delta^m}^{\infty} (v_I + \delta) dF(\delta).$$

Together with $Q_1 = q_1$, the equilibrium entry threshold satisfies

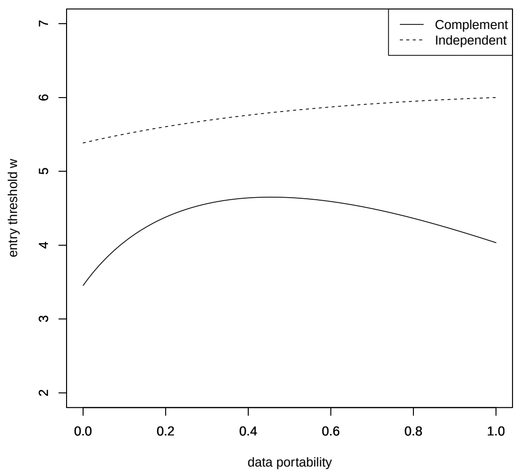
$$(v_I + \delta^m)^2 + 2(\lambda(v_I + \delta^m) - v_I - v_B)[v_I + v_I F(\delta^m) + \lambda \int_{\delta^m}^{\infty} (v_I + \delta) dF(\delta)] = v_I^2.$$



(a) $v_B = 3, \Delta = 5$



(b) $v_B = 3, \Delta = 10$



(c) $v_B = 6, \Delta = 10$

Figure 1: Impact of Data Portability on Entry Threshold ($v_I = 1, c = 2$).

Fig 1 shows the switching thresholds for different values of v_B and Δ when $F(\delta)$ is uniformly distributed on the interval $[0, \Delta]$, for both cases of independent basic and data services and complementary basic and data services. The main features remain that data portability is more likely to raise entry barrier when the value of big data is larger or when the entrant is more innovative (Δ becomes larger). It also shows that with complementary services, the condition for data portability to monotonically raise entry barrier is stricter,[34] confirming the above discussion that the *demand-expansion* effect is less harmful to the entrant with complementary services.

C.3 Multi-homing

In practice, a consumer may subscribe to multiple social networks, use multiple operating systems, or drive several cars. However, there is often an opportunity cost of time spent on these services, that is, they are still competing for consumers' screen time and attention. Hence, not only the decision about whether to multi-home, but also how much to multi-home are important. In the second period, the consumer chooses how much time to spend on each firm's service and how much data to provide to each firm, that is,

$$U_2 = \max_{t_I, q_2^I; t_E, q_2^E} t_I(v_I + q_1 + q_2^I - C(q_2^I)) + t_E(v_I + \delta + \lambda q_1 + q_2^E - C(q_2^E)) - T(t_I + t_E),$$

where t_I and q_2^I are the time spent and amount of data provided to the incumbent (t_E and q_2^E to the entrant respectively), and $T(t_I + t_E)$ is the opportunity time cost of using the services. Clearly, when the utility from one service is linear in the time spent on this service, it is not optimal for the consumer to multi-home, as it is better to spend all time on the service that generates higher value per unit of time. Thus, the model reduces to the single-homing model in our main analysis. Multi-homing becomes more meaningful if the consumer has diminishing marginal utility from one service, for instance, when

$$U_2 = \max_{t_I, q_2^I; t_E, q_2^E} s(t_I)(v_I + q_1 + q_2^I - C(q_2^I)) + s(t_E)(v_I + \delta + \lambda q_1 + q_2^E - C(q_2^E)) - T(t_I + t_E),$$

where $s(t)$ is an increasing and concave function and hence the marginal utility from time spent on one service is decreasing. In this case, the optimal time spent on the incumbent's service and the entrant's service satisfies

$$\rho = \frac{s'(t_I)}{s'(t_E)} = \frac{v_I + \delta + \lambda q_1 + w_2^*}{v_I + q_1 + w_2^*},$$

where $w_2^* = \max_{q_2} q_2 - C(q_2)$. In the first period, the consumer chooses how much time to spend (t) and how much data provide (q_1) to the incumbent, in order to maximize the

[34] For instance, one can construct an example with the uniform distribution, where data portability monotonically raises entry barrier for $\Delta = 20$ and $v_B = 40$.

total utility

$$U_1 = \begin{aligned} & max_{t,q_1} s(t)(v_I + q_1 - C(q_1)) - T(t) \\ & + s(t_I)(v_I + q_1 + q_2^I - C(q_2^I)) + s(t_E)(v_I + \delta + \lambda q_1 + q_2^E - C(q_2^E)) - T(t_I + t_E). \end{aligned}$$

Hence, the optimal level of data provision satisfies

$$s(t)(1 - C'(q_1)) + s(t_I) + \lambda s(t_E) = 0.$$

In this case, the *switching-facilitating* effect should now be interpreted as when λ increases, relatively more time is spent on the entrant's service (i.e., $\partial\rho/\partial\lambda > 0$), and the *demand-expansion* effect should be interpreted as when q_1 becomes higher, relatively less time is spent on the entrant's service (i.e., $\partial q_1/\partial\lambda > 0$ and $\partial\rho/\partial q_1 < 0$). More generally, when multi-homing is possible, instead of interpreting our results as the impact of data portability on the barrier to entry, we could interpret them as the impact of data portability on the barrier to expansion.

C.4 Consumer Surplus in a Dynamic Context

We present in this section a simple example where consumer surplus can be decreasing with data portability when we take into consideration the dynamic effects. Specifically, consider a three-period model with two overlapping generations of consumers, each of which has a mass of one. The first-generation consumers live in the first and the second periods, while the second-generation consumers live in the second and the third periods. To deliver the main insights, we use the linear-quadratic example.

In the simplest situation where second-generation consumers can freely choose which firm they want to join, the entrant in the second period always enters as it is the better option for second-generation consumers. Since there is no inter-temporal externality, a consumer in each generation faces the same problem as in our main model. That is, a first-generation consumer's problem in the first period is

$$\max U_{11} = v_I + q_{11} - c\frac{q_{11}^2}{2} + \int_0^{\delta^1} [v_I + (1+v_B)q_{11} + \frac{1}{2c}]dF(\delta_{E1}) + \int_{\delta^1}^{\Delta} [v_I + \delta + \lambda q_{11} + \frac{1}{2c}]dF(\delta_{E1}),$$

where

$$\delta^1 = (1 - \lambda + v_B)q_{11},$$

and δ_{E1} is the quality improvement offered by a potential entrant in the second period. The problem faced by a second-generation consumer is

$$\max U_{21} = v_0 + q_{21} - c\frac{q_{21}^2}{2} + \int_0^{\delta^2} [v_0 + (1+v_B)q_{21} + \frac{1}{2c}] dF(\delta_{E2}) + \int_{\delta^2}^{\Delta} [v_0 + \delta + \lambda q_{21} + \frac{1}{2c}] dF(\delta_{E2}),$$

where

$$\delta^2 = (1 - \lambda + v_B)q_{21},$$

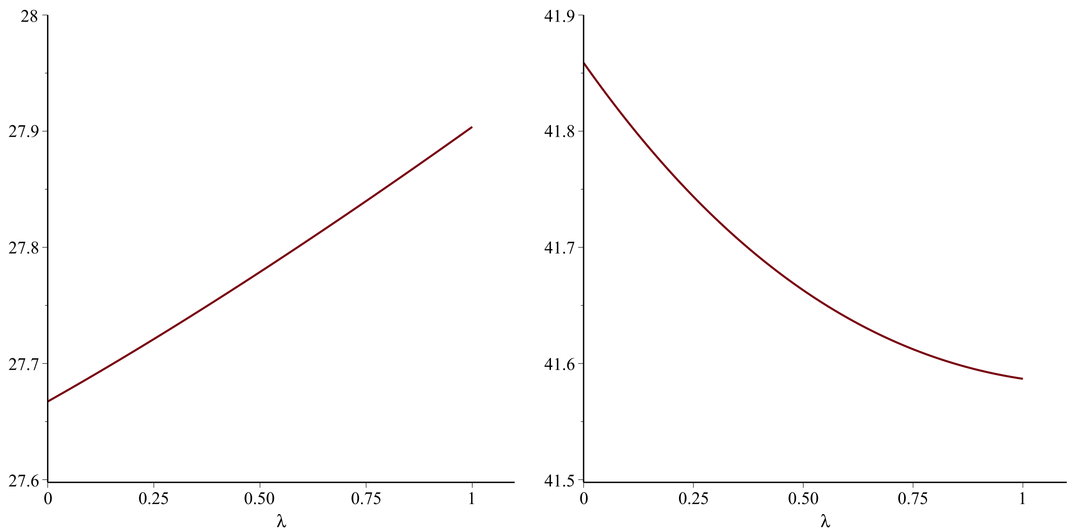
and $v_0 = v_I$ if the consumer chooses the incumbent and $v_0 = v_I + \delta_{E1}$ if he chooses the entrant in the second period (equivalent to the first period of this second-generation consumer). In addition, we assume that there is a new entrant in the third period, which can provide a quality improvement δ_{E2} over the incumbent's basic service quality in the third period, drawn from the same distribution $F(\delta)$.

However, with inter-temporal externality, the results can be quite different. Let us assume that the entrant can only enter if it attracts both generations of consumers. This captures the idea that a new entrant needs a sufficient scale in order to enter the market. It is likely to be the case when the customer base of the incumbent is already large, while new consumers only arrive in the market gradually. In such a situation, the entrant can only enter in the second period if it offers a quality improvement of at least δ^1. The utility of a second-generation consumer becomes

$$\begin{aligned} U_{21} = \; & \int_0^{\delta^1} [\max_{q_{21}} v_I + q_{21} - c\tfrac{q_{21}^2}{2} + \int_0^{\delta^2} [v_I + (1+v_B)q_{21} + \tfrac{1}{2c}] dF(\delta_{E2}) \\ & + \int_{\delta^1}^{\Delta} [v_I + \delta_{E2} + \lambda q_{21} + \tfrac{1}{2c}] dF(\delta_{E2})] dF(\delta_{E1}) \\ & + \int_{\delta^1}^{\Delta} [\max_{q_{21}} v_I + \delta_{E1} + q_{21} - c\tfrac{q_{21}^2}{2} + \int_0^{\delta^2} [v_I + \delta_{E1} + (1+v_B)q_{21} + \tfrac{1}{2c}] dF(\delta_{E2}) \\ & + \int_{\delta^2}^{\Delta} [v_I + \delta_{E1} + \delta_{E2} + \lambda q_{21} + \tfrac{1}{2c}] dF(\delta_{E2})] dF(\delta_{E1}). \end{aligned}$$

When the second-period entrant offers an insufficient quality improvement, second-generation consumers are stuck with the first-period incumbent, even though they would have preferred the entrant had they been able to act independently from the first-generation consumers. Thus, with inter-temporal externality, data portability can hurt second-generation consumers when it raises the entry barrier for the second-period entrant, i.e., when δ^1 increases with λ. Fig 2 shows an example of such a negative effect of data portability on consumer surplus.[35]

[35]The Figure is drawn for $v_I = 10, v_B = 12, \Delta = 15, c = 15$ and $F(\delta)$ being the uniform distribution on $[0, 15]$.



(a) Utility of the First Generation (b) Utility of the Second Generation

Figure 2: Impact of Data Portability on Consumer Welfare.

Thus, whereas data portability increases the utility of first-generation consumers, it decreases the utility of second-generation consumers. It is easy to extend the model beyond three periods as each generation of consumers only affects the next generation of consumers through the choice of technology (i.e., whether the entrant enters or not). Thus, a higher degree of data portability can reduce consumer welfare for all future generations by locking them in with the initial incumbent. Furthermore, if we allow the utility across generations to be related, for instance, by making the big data service value of each generation depends on data collected from all previous generations of consumers, the lock-in effect may become even stronger and further reduce consumer welfare and slow down innovation in the long run.