CrossMark

RESEARCH PAPER

Data Portability on the Internet

An Economic Analysis

Michael Wohlfarth

Abstract Data portability allows users to transfer data between competing online services. As data gets increasingly valuable for online services and users alike, the enforcement of data portability within the European Union by the General Data Protection Regulation will have important ramifications for the competition in online markets. Thus, this paper develops a game-theoretic model to examine firms' strategic reaction to data portability and to identify the ensuing market outcomes. It can be shown, among others, that although data portability is designed to protect users, they may be hurt because market entrants have an incentive to increase the amount of collected data compared to a regime without data portability. However, profits for new services and total surplus increase if the costs for implementation are not too large. This likely improves innovation and service variety. Consequently, the results provide important insights and case-specific recommendations for managers and policy makers in data-driven online markets.

Keywords Data portability · Competition between online services · Economics of IS · Switching costs · Market entry and innovation

M. Wohlfarth (✉)
University of Passau, Passau, Germany
e-mail: research@wohlfarth.org

1 Introduction

In the digital ecosystem, data is considered to be the key ingredient for many of today's revenue models, crucially determining whether a service is successful. At the same time, the protection of (personal) data, users and competition becomes increasingly important for policy makers and competition authorities. For example, the European antitrust investigations against Google attribute either to the observation that consumers might be disadvantaged or that competition and innovation is hampered (c.f., Drozdiak and Schechner 2016, for an overview of European antitrust probes against Google). In fact, personal data entered or revealed at a specific online service may lead to a lock-in effect for users as switching to competing services induces costs to re-enter the data required by the new online service (c.f., Klemperer 1987a, for related research). Hereby, (dominant) online services may benefit, but innovation and service variety might be reduced as market entry is deterred. Illustrative examples of data-induced switching costs are provided by online banking accounts (where switching leads to the necessity to re-enter recurring transferals), online mail or storage services (where switching leads to the necessity to re-enter general user information, and to re-upload files, photos, contacts or categories), or cloud computing environments (where preferences and adaptations have to be re-injected). These services suggest that a lock-in does not necessarily stem from network effects *alone*, i.e., the number of participating users or complementary provided services. Instead, as Chen and Hitt (2002) analyze empirically, there is a variety of factors (additionally) influencing a user's loyalty. We build on these observations and argue that the (amount of) already revealed (personal) data is a crucial factor for (1) online services active in data-driven markets because it

determines the service's competitive strength and thus, profitability, and also for (2) users because they might be locked-in to a certain service.

It is well known that established systems designed to lock-in users may hamper the success of new services and lead to excessive rents of incumbent firms (c.f., Katz and Shapiro 1994; Farrell and Klemperer 2007) and – eventually – to market failures. In this spirit, the European Commission has recently formulated a general "right to data portability" for personal data. Consequently, a standardized way of how information that has been actively provided can be ported from one online service to another is required (c.f., European Commission 2016b, p.45, Article 20); an issue most voluntarily provided functionalities for users to export previously revealed data do not explicitly account for (c.f., Facebook 2018; Google 2018), and an issue also highlighted by the Deputy Chief Technology Officer of the United States (c.f., Macgillivray and Shambaugh 2016). Ultimately, and especially in combination with the "right to erasure" (c.f., European Commission 2016b, p.43, Article 17), the European Commission's initiative aims to promote users' negotiation power vis-à-vis (dominant) online services by reducing lock-in effects, i.e., protecting the "fundamental rights and freedoms of natural persons" (c.f., European Commission 2016b, p.32, Article 1). However, the economic effects of such an intervention on consumer's surplus, on the amount of data online services collect from their customers, on online service's profits, and on service variety are unclear to date. Albeit the regulation is binding for all European member states since May 2018, academic analyses have so far been limited to the legal and technical dimensions of data portability. An analysis of strategic incentives, business strategies and economic outcomes is lacking, as Nobel prize laureate Jean Tirole outlined in his speech on competition and regulation of online platforms (c.f., Valero 2016.)

This paper addresses this research gap and analyses the competitive effects of a user's ability to port data from an incumbent online service or content provider (CP) to a market entrant. Hereby, we analyze the CPs' incentives (not) to promote data portability and their business strategies in data-driven markets. Additionally, we shed light on the ensuing effects on consumers as this is pivotal to the argumentation of the European Commission and the U.S. Deputy Chief Technology Officer, alike. In doing so, we develop a game-theoretic model that considers the economic effects arising from a right to data portability by considering two CPs generating revenues primarily through data revealed by users active at their platform. Thus, we abstract from any explicit revenue model (e.g., based on advertisements, or based on selling aggregated user-data to third parties), and from additional revenue streams (e.g.,

services based on a subscription model) by simply assuming that data revealed by users can be transformed into revenue. Hence, additional data has a positive effect on a CP's profits. On the other hand, revealing data bears costs (i.e., a disutility) for users: either they have some effort revealing data as such (say, the time needed to enter the data), or – more general – users give away data, to which they attribute some value to (say, privacy costs in a broader sense). Consequently, whereas collecting more data is beneficial for CPs, users experiencing a higher disutility might switch to competing CPs or even leave the market. However, users' ability to do so is impeded by established switching-costs and lock-ins. The ability to port data by means of data portability arguably lifts the established restrictions on users, but may also impact the CPs' data consumption. These effects have to be taken into account when analyzing the competitive effects.

Our obtained results show that data portability is not necessarily beneficial for users because CPs entering the market have an incentive to increase the amount of data users have to reveal. Thus, the ultimate goal to protect users is not necessarily achieved. Conversely, the CPs' incentives (not) to promote data portability are unambiguous if the costs for implementing a right to data portability are zero or comparably low: Whereas dominant CPs (incumbents) always suffer from data portability, emerging CPs (entrants) challenging incumbents are better off. However, as total surplus increases under a data portability regime, predominantly due to the arising benefits for the entrant who is able to generate higher revenues, the decision to enforce a right to data portability is far more complex than currently realized.

2 Literature Review

We refer to data portability as consumer's ability to transfer (personal) data revealed at one CP to another CP. To the best of our knowledge, the IS literature has so far not considered this concept explicitly in terms of strategic incentives, business strategies, or economic outcomes. Albeit, the technical literature demonstrated the feasibility of that concept by proposing models to conveniently port data, e.g., between cloud computing vendors. In this vein, Ranabahu and Sheth (2010) propose semantic web techniques to achieve portability and Petcu and Vasilakos (2014) inter alia highlight open standards and open application programming interfaces as technical solutions. Thus, most technical studies provide a proof of concept that data portability is technically feasible but do not explicitly discuss the possible trade-offs for the involved parties.

In light of the General Data Protection Regulation, which has become effective in May 2018, several legal

investigations have been carried out. Graef (2015) conducts a legal analysis of data portability in social networks with respect to the (European) competition law and summarizes relevant cases. Vanberg and Ünver (2017) inter alia highlight arising security issues as well as "disproportionate costs for small and medium sized companies" (Vanberg and Ünver 2017, p.14) induced by introducing a right to data portability. Swire and Lagos (2013) explicitly refer to consumer welfare and "express serious concerns about the RDP [right to data portability]" (Swire and Lagos 2013, p.338) because, (1) the problems addressed by the regulation (e.g., monopoly power through lock-ins) were legally already covered by competition law, (2) personal data could easily be exported, i.e., security problems arise, and (3) it was unclear how a common standard could be achieved if a variety of different service providers were involved. The authors conclude that "the proposed RDP appears to reduce consumer welfare" (Swire and Lagos 2013, p.379), but do not offer or discuss economic incentives or outcomes, which additionally highlights the necessity of economic backing in this context.

Moreover, this study is related to two strands of the economic literature, which will be highlighted in the following. First, as we assume users to be locked-in when using a data-intensive online service due to costs to port these data, we draw on the literature investigating the role of switching costs. The results derived from this literature show that an incumbent firm has an incentive to lower its price anticipating that an entrant enters the market (Klemperer 1989). In essence, firms thus fiercely compete in early periods to gain market shares which can then be harvested in later periods (Klemperer 1987a, b). Hence, switching costs induce softened competition in later periods which allows the remaining firms to set higher prices. Indeed, as Gehrig and Stenbacka (2004) show analytically, competing firms have an incentive to establish high switching costs. The authors show that these can be achieved by (maximum) horizontal differentiation (additionally, see Hotelling 1929; d'Aspremont et al. 1979). Within the taxonomy introduced by Ray et al. (2012), our study deals with "user-related" switching costs as they include the effort a user needs to invest to "ensure a satisfactory switch of service and to recreate or transfer features" (Ray et al. 2012, p. 199). More precisely, one may argue that within the framework provided by Ray et al. (2012), transfer costs are of particular importance to users. To demarcate our approach from previous literature related to the existence of switching costs and lock-ins, the fact that we assume data to be the considered good, which inherently determines the degree of switching costs as well as firm's profits (c.f., Sect. 3 for details) is crucial and should be highlighted. Hence, the strategy derived from the traditional switching cost literature would induce to set

lower prices in early periods (i.e., collect less data) to deter entry and gain market shares which can thereupon be harvested. This, in turn, is not necessarily the equilibrium strategy of an incumbent in a data-driven market environment, as (1) switching costs would be lower in succeeding periods and (2) profits in later periods from data already gained in early periods would be reduced. These specific aspects of the competitive environment further delineate our approach from, e.g., Caminal and Matutes (1990) who consider endogenous switching costs.

Second, our study on data portability is related to the strand of the (economic) effects stemming from interoperability. Within this strand, the literature on compatibility and standardization between different services, especially the ensuing effects of the availability of converters as considered by Farrell and Saloner (1992), should be highlighted. In their theoretical model, Farrell and Saloner show that the availability of (imperfect) converters allows users to benefit from other users using a competing technology, i.e., a converter induces benefits through compatibility. Thus, direct network effects resulting from interoperability are a central aspect of the depicted model. Another important view on interoperability is highlighted within the study conducted by Pollock (2009). Pollock evaluates the effects of controlling the possibility to convert "'software' or services' associated with one platform to run on another" assuming a two-sided market (Pollock 2009, p.155). Thus, Pollock considers interoperability being determined by indirect network effects. Additionally, the impact of the ability to control the mode of interoperability itself is investigated. Thus, the author allows the platform to directly control the costs of flow of information, i.e., the costs for interoperability. However, although interoperability plays a pivotal role in online markets, the mentioned studies do not depict the concept of data portability for several reasons. In general, interoperability should not be confounded with the portability of data (c.f., Graef 2015). Additionally, next to several technical dimensions, the central economic distinction can be seen in (1) the role of network externalities, which are not *necessarily* relevant in the context of data portability as a user's lock-in in data-driven markets is crucially influenced by the (amount of) data revealed at a certain online service and not solely by network externalities (c.f., examples provided in Sect. 1), and (2) the scope of the platform's ability to control the flow of data: since the mentioned European regulation is binding for all services alike, most existing online services are left with no possibility to strategically set the amount of data that can be ported, i.e., online services are unable to control the costs for portability.

Our proposed game-theoretic model, which will be outlined in the following section, captures the trade-offs for the involved parties and considers the specific aspects of

data-driven revenue models. We use this model to answer the following two main research questions:

RQ 1 How does a right to data portability affect the amount of data that online services collect?

RQ 2 How does a right to data portability affect consumers?

Additionally, we investigate the effects on an incumbent's and an entrant's profits, which arguably influences service variety and innovation, and investigate which regime (data portability or no data portability) is more efficient with regard to total welfare.

3 Outline of the Economic Model: Assumptions and Notation

We propose a two-stage, game-theoretic model in order to analyze the effects of introducing a right to data portability ($d = P$) vis-à-vis a regime without the possibility to port data ($d = NP$). The market environment is assumed to consist out of two content providers (CPs) and users having heterogeneous preferences over the set of content providers.

Content Providers. We consider a market with two competing, differentiated CPs ($i = A, B$) that offer substitutable services. To highlight the competitive effects of a right to data portability and to capture the implications on market entry and innovation, we consider two time periods ($t \in \{1, 2\}$) and assume that CP A is active in $t = 1$ and $t = 2$, whereas CP B enters in $t = 2$. Thus, CP A might be classified as an incumbent content provider, whereas CP B is an entrant. Although both CPs offer substitutable services and CP B enters the market in a later point in time, due to the user's preferences over the set of CPs (see explanation below), the offered services are horizontally differentiated (additionally, c.f., Irmen and Thisse 1998; Gehrig and Stenbacka 2004), i.e., users have different tastes for the services offered by the CPs. Formally, we therefore use the model proposed by Hotelling (1929) and assume that on a unit interval of length one – whereon users are uniformly distributed –, the incumbent CP A is located at $x = 0$ and the entrant CP B is located at $x = 1$ (see, e.g., Montes et al. 2018, for a similar setup). Moreover, in order to highlight the effects of introducing a right to data portability on data collection, we consider that services are free of charge, i.e., users need to reveal data and CPs are solely financed by the exploitation of this data, e.g., by showing (targeted) advertisements which is the prevalent revenue model on the internet (c.f., Dou 2004; Evans 2009; Anderson 2012) and a frequently used

assumption in the related literature (c.f., Choi and Kim 2010; Kourandi et al. 2015; Krämer et al. 2018).

Users. Users are uniformly distributed on the interval between zero and one. Consequently, users are heterogeneous in their preferences over the set of CPs. Users patronize the CP which provides them the higher utility in each period $t \in \{1, 2\}$. This utility U_i^t is determined by a CP's exogenously given base utility v_i (e.g., determined by the service's functionalities, the quality of the content, the ease-of-use), the amount of data a CP requires from users, i.e., a CP's data consumption r_A^t (which is the strategic variable of a CP and results in a disutility for users, see introductory examples stated in Sect. 1),[1] and the inherent preference of a user over the set of CPs, i.e., their tastes, which is determined by a user's location x on the unit interval. Please note that the relevance of this location can differ between different market environments. To be able to analyze this aspect formally, i.e., to account for markets with diverting characteristics, the users' preferences over the set of CPs are influenced by the parameter τ specifying the mismatch costs for users (see Sun 2012, for a similar setup). If τ is low, the users' mismatch costs are low. Thus, users preferences get relatively less important in the considered market environment and vice versa. Ultimately, it can be argued that low mismatch costs lead to a higher competitive intensity in the considered market because an user's decision which service to patronize is then predominantly determined by the CPs' qualities and data collection (c.f., "Appendix 1" which is available online via http://springerlink.com for an overview of the notation used for the model).

In period $t = 1$, CP A serves the market as monopolist. With the introduced notation, a user located at x choosing to become active at CP A derives a utility of $U_A^1(x) = v_A - \tau \cdot x - r_A^1$. See that a user's utility does not depend on the amount of revealed data. Consequently, the level of data consumption by CP A does not affect the service's quality because (1) all users need to reveal the same amount of data in order to keep programming efforts low, and (2) the principle of data minimisation manifested in the GDPR (c.f., European Commission 2016b, Article 5(1)c), makes it impossible for CP A to require unnecessary data. As a result, determining the active users in $t = 1$ is straight forward: only users deriving an utility larger or equalling zero will use the service offered by CP A, i.e., if $U_A^1(x) \geq 0$ a user is active at CP A and a user with $U_A^1(x) < 0$ does not use any service in period $t = 1$. We denote the resulting location of the indifferent user by $x^{*,d,1}$ and only

---

[1] We do not consider consumption-related benefits for users, i.e., the base utility v_i for CP i does not depend on the amount of entered data, additionally, see Sect. 6.2.

users located at $x \leq x^{*,d,1}$ are active at CP A in $t = 1$.[2] Note that the location of the indifferent user equals the market share of CP A in period $t = 1$. The strategic variable of CP A is r_A^1, i.e., setting a comparably low data consumption level (r_A^1) leads to more users being active at that CP (i.e., the market share increases all else being equal). However, the profits per user will then be lower.

In period $t = 2$, CP B enters the market. Consequently, users can now choose between two competing CPs and select the one from which they derive the higher utility. In order to investigate the competitive effects of introducing a right to data portability, we assume the market to be fully covered, i.e., at least one user can *potentially* port her data from CP A to CP B (additionally, see "Appendix 2"). The utility a user derives from staying (in case the user has been active at CP A in $t = 1$ and does not switch to the competing CP B) or becoming active at CP A in period $t = 2$ is given by

$$U_A^2(x) = \begin{cases} v_A - \tau \cdot x - r_A^2 & \text{, if } U_A^1(x) \geq 0 \\ v_A - \tau \cdot x - r_A^2 - r_A^1 & \text{, else.} \end{cases}$$

Note that r_A^2 is the strategic variable of CP A in $t = 2$. CP A is free in its decision how much data to require in that period. However, we assume that users that stay at CP A (i.e., are active at CP A in period $t = 1$ *and* $t = 2$) do not experience a disutility in $t = 2$ from data already revealed in $t = 1$. For example, if a user entered (personal) data (e.g., her name, address, date of birth, interests, or uploaded photos and documents), she does not have to re-enter, re-validate or re-upload this information. Conversely, users who were not active in $t = 1$ but decide to become active in $t = 2$ have to reveal all required data if they decide to become active in the second period. Thus, these users need to reveal data of $r_A^1 + r_A^2$. However, users may also use the competing CP B. A user located at x who becomes active at CP B in $t = 2$ derives a utility of

$$U_B^{d,2}(x)$$
$$= \begin{cases} v_B - \tau \cdot (1-x) - r_B^2 + r_A^1 & \text{, if } U_A^1(x) \geq 0 \text{ with data portability } (d = P) \\ v_B - \tau \cdot (1-x) - r_B^2 & \text{, else } (d = NP \text{ or } U_A^1(x) < 0). \end{cases}$$

The utility function $U_B^{d,2}(x)$ captures the effect that users becoming active at CP B need to enter all required data (i.e., r_B^2) either if they have not been active in $t = 1$, or if there is no ability to port already revealed data ($d = NP$). Additionally, the equation captures the effects of a right to data portability if users switch CPs: if users have been active at CP A in the first period, i.e., $U_A^1(x) \geq 0$, and are able to port already entered data to the new CP without incurring any costs (as envisaged by the European

Commission, $d = P$), they do not have to reveal this data again.[3] Based on the utility functions, the location of the indifferent user in period $t = 2$ ($x^{*,d,2}$) can be calculated. Again, the location of the indifferent user directly translates into the CPs' market shares, i.e., $x^{*,2}$ equals the market share of CP A and $1 - x^{*,2}$ equals the market share of CP B.

Content Providers' Profits. Based on the market shares given by the location of the indifferent user, CPs' payoffs can be specified by defining their profit functions. In our base model, we assume that CPs with data-driven revenue models benefit from data entered in one period also in later periods as the obtained information is still valuable to them (e.g., in terms of the ability to target ads, or tailor or customize services). However, we relax this assumption in Extension 5.3. Moreover, we do not consider any costs associated with the introduction of the right to data portability in our base model. However, we relax this assumption in Extension 5.1. Thus, for now, total profits of CP A after two periods are given by

$$\pi_A^d = \underbrace{x^{*,d,1} \cdot r_A^{d,1}}_{\pi_A^{d,1}} + \underbrace{x^{*,d,2} \cdot (r_A^{d,1} + r_A^{d,2})}_{\pi_A^{d,2}},$$

and CP B, which is only active in $t = 2$, makes total profits of

$$\pi_B^d = (1 - x^{*,d,1}) \cdot r_B^{d,2} + (x^{*,d,1} - x^{*,d,2}) \cdot ((r_B^{d,2} - r_A^{d,1}) + r_A^{d,1}),$$
$$\pi_B^d = (1 - x^{*,d,2}) \cdot r_B^{d,2}.$$

Note that we implicitly made two further assumptions. First, we assumed that CPs cannot discriminate between old, new and switching users, i.e., the amount of data a CP requires from a specific user in $t = 2$ is independent of this user's decision in $t = 1$. Thus, all users active at a CP need to reveal the same amount of data (we refer to the limitations in Sect. 6.2 for a discussion of the implications if this assumption is relaxed). Second, we assumed that all data that is transferred to CP B is valuable for the entrant. We relax this assumption in the second extension of the base model (see Extension 5.2).

Timing of the Game. To summarize, the considered two-stage game proceeds as follows:

---

[3] In "Appendix 5", we show that the entrant CP B always requires at least the amount of data CP A required in $t = 1$ if $v_B \geq v_A$. Users then only need to reveal the *net amount of required data*. If $v_B \in [15v_A/16, v_A)$, CP B sets a lower data consumption level than CP A. Then users that switch CPs derive a net benefit from a right to data portability because (1) the new service requires less data and (2) the old service has to delete already entered data due to the right to erasure which is part of the GDPR (c.f., European Commission 2016b, Article 17 and "Appendix 5").
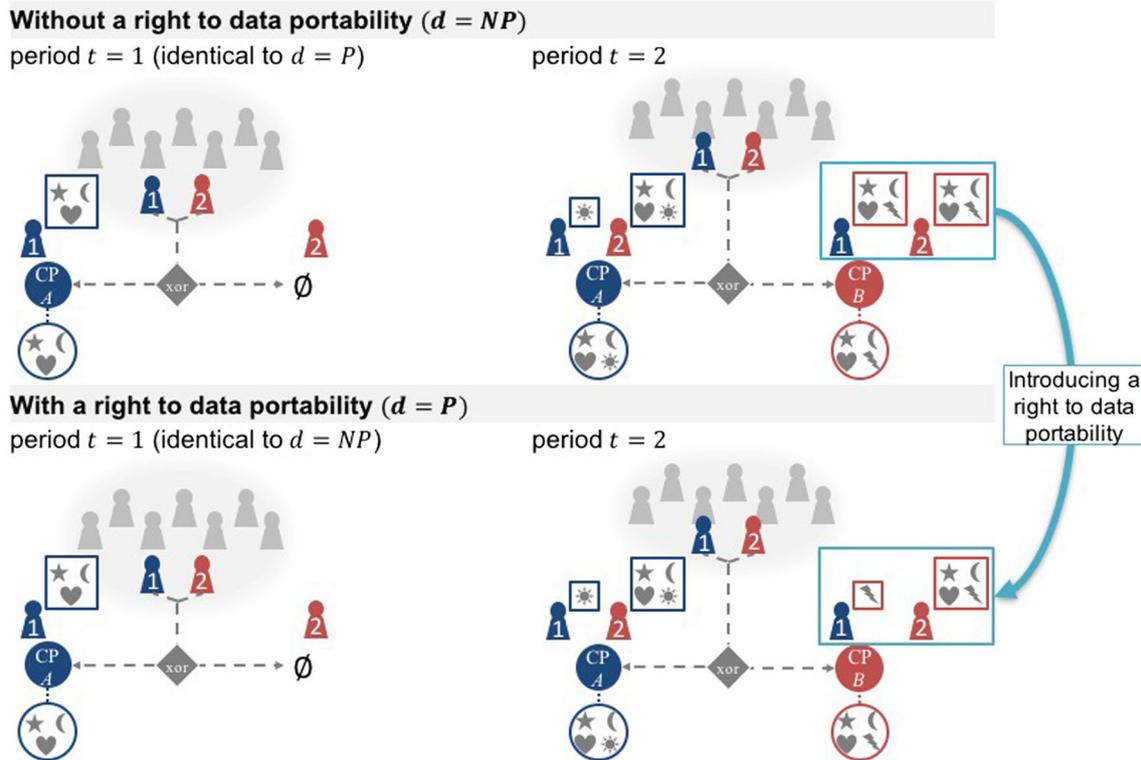
Fig. 1 Illustration of the regimes without (top) and with (bottom) a right to data portability. *Note* The effect of introducing a right to data portability is relevant in period $t = 2$ for users becoming active at the entrant CP B (see highlighted amount of data users need to reveal). Without data portability ($d = NP$), user 1 has to re-enter her already revealed data if she switches to CP B in $t = 2$. However, with data portability ($d = P$), user 1 has the ability to port her already entered data. User 2 did not enter any data in $t = 1$. Consequently, irrespective whether a right to data portability is introduced, she cannot port any data in $t = 2$

Stage 1 The incumbent CP A sets the amount of required data r_A^1 for period $t = 1$ anticipating CP B's action in period $t = 2$. Then, users decide whether to become active at CP A (if $U_A^1(x) \geq 0$).

Stage 2 Both CPs simultaneously set the amount of required data for period $t = 2$, i.e., CP A sets r_A^2 and CP B sets r_B^2. Again, users then decide at which CP they choose to become active. Under the full market coverage assumption, users in $t = 2$ are active at exactly one CP. If $U_A^2(x) \geq U_B^{d,2}(x)$, users are active at CP A and vice versa.

Figure 1 illustrates the assumed market setting. Here, squares above the user depict the (net) amount of data (illustrated by symbols) different users ($j = 1, 2$) would have to reveal in the considered period for becoming active at the respective CP. In contrast, circles underneath the CPs indicate the amount of data a CP requires. In the illustrated scenario, user 1 is active in period one, whereas user 2 becomes active only in period two. *Without* data portability (upper illustration in Fig. 1), user 1 has to re-enter the data already revealed to CP A at CP B, if she wants to switch to

CP B in the second period (thus, she needs to re-enter: star, moon and heart, and additionally needs to enter: thunderbolt). In contrast, *with* data portability (bottom illustration in Fig. 1), user 1 has the ability to port her already entered data and thus only has to enter the net amount of required data (here: thunderbolt) if she wants to switch to CP B. For user 2, who has not been active in the first period, both cases are identical, i.e., user 2 has to enter all of the CP's required data independent of the considered regime (i.e., star, moon, heart and sun to become active at CP A or star, moon, heart and thunderbolt to become active at CP B). Note that Fig. 1 only illustrates the (net) amount of data that is required by the CPs and needs to be entered by users in the respective period. A user's actual decision which CP to patronize is not illustrated in Fig. 1 because it depends (inter alia) on the base utilities.

4 Model Analysis, Results, and Discussion

We solve for the subgame perfect Nash equilibrium through backward induction beginning in Stage 2 to deduce the equilibrium amounts of required data (c.f., Sect. 4.1).

The results are successively used to analyze the effects on CPs' profits (c.f., Sect. 4.2), consumer's surplus (c.f., Sect. 4.3) and total surplus (c.f., Sect. 4.4).

In Stage 2 both CPs compete for users and revenues. Consequently, a CP's decision is affected by the decision of its competitor and the corresponding actions of users, i.e., the CPs take into account the amount of data required by the competing CP. Consequently, the payoffs of the CPs are affected by both CPs' strategic variables r_i^t. Analytically, these effects are captured by simultaneously solving and maximizing $\partial \pi_A^{d,2} / \partial r_A^2 = 0$ and $\partial \pi_B^d / \partial r_B^2 = 0$, which yields the CP's equilibrium amount of required data for period $t = 2$ (c.f., Sect. 4.1 as well as "Appendix 5" highlighting the second order conditions). In doing so, we need to calculate the location of the indifferent user in $t = 2$ by accounting for the different regimes: If users have the possibility to port their data ($d = P$) and were active in period one, the indifferent user in $t = 2$ can be calculated by solving $v_A - \tau \cdot x - r_A^2 = v_B - \tau \cdot (1 - x) - r_B^2 + r_A^1$. If users do not have the possibility to port their data ($d = NP$), but were active in period one, the indifferent user in $t = 2$ can be calculated by solving $v_A - \tau \cdot x - r_A^2 = v_B - \tau \cdot (1 - x) - r_B^2$. Technically, the indifferent user in period two might also be located right to the location of the indifferent user in period one, i.e., $U_A^1(x^{*,d,2}) < 0$. We do not explicitly analyze this case within the main analysis (see "Appendix 3" for more details). To summarize, the indifferent user in $t = 2$ is located at:

$$x^{*,d,2} = \begin{cases} -\dfrac{r_A^2 + r_A^1 - r_B^2 - \tau - v_A + v_B}{2\tau} & \text{, if } U_A^1(x^{*,d,2}) \geq 0 \quad (d = P), \\ -\dfrac{r_A^2 - r_B^2 - \tau - v_A + v_B}{2\tau} & \text{, else} \quad (d = NP). \end{cases}$$

In Stage 1 CP A serves the market as monopolist. However, it anticipates the effects on second-period profits in its decision how much data to collect. Analytically, we use the equilibrium results of Stage 2 (i.e., $r_B^{*,d,2}$ and $r_A^{*,d,2}$) to specify CP A's profits over two periods (π_A^d) and then solve and maximize $\partial \pi_A^d / \partial r_A^1 = 0$ to obtain the optimal amount of required data for CP A in period $t = 1$ (i.e., $r_A^{*,d,1}$, c.f., Sect. 4.1 as well as "Appendix 5" highlighting the second order conditions). In doing so, we need to calculate the location of the indifferent user in period $t = 1$ by solving $U_A^1 = 0$ with respect to x which leads to $x^{*,d,1} = \frac{v_A - r_A^1}{\tau}$.

4.1 Amount of Required Data by the CPs

As outlined above, to calculate the amount of required data, we maximize the CPs' profit functions considering both periods (for CP A) or only period $t = 2$ (for CP B). Successively, the equilibrium amounts of required data can be compared. Here, it can be seen that CP A requires a higher amount of data under the regime *without* data portability ($d = NP$). Interestingly, the data consumption of CP A without data portability in the first period is even higher than the monopoly data consumption $r_{Monopoly}^*$, i.e., the amount of data CP A would require without the entry of CP B:

$$r_A^{*,NP,1} = \frac{3\tau + 10v_A - v_B}{17} > \frac{v_A}{2} = r_A^{*,P,1} = r_{Monopoly}^*.$$

This highlights the effect of anticipated entry: Intuitively, CP A requires a high amount of data to generate (higher) switching costs to weaken competition in later periods (i.e., generates data-induced switching costs). The effect of weakened competition even dominates the (negative effect of) reduced period one market shares and, compared to a regular one-period monopoly, reduced profits. The observation that CP A requires an even higher amount of data than in monopoly is, at first sight, in contrast to the traditional switching cost literature. Here, anticipated entry results in price wars lowering early-period prices to gain market shares, which can thereupon be harvested in later periods (c.f., Klemperer 1989, 1995). But, within our considered setting of a data-driven market environment, lock-ins are not generated by participation alone (e.g., positive network externalities or the functionalities of a service), which can be stimulated by low prices (additionally, c.f., Extension 5.4), but additionally by a user's invested effort to enter, i.e., a user's disutility to reveal (personal) data. Thus, lock-in effects do play a pivotal role for CPs in these market environments, although the underlying rationale differs compared to traditional market environments. This is because (1) data required by a CP (i.e., "prices" set) in early periods are directly relevant to CPs' profits in later periods, and (2) the incumbent's "price setting" is (additionally) constrained by entrants in later periods. With data portability ($d = P$), the incumbent CP requires the monopoly amount of data. Because lock-in effects vanish through the users' ability to port data to the competing CP in the following period, the incumbent CP cannot benefit from establishing lock-ins anymore. Consequently, CP A maximizes its profits in the first period by requiring the same amount of data it would require in a one-period game, where it acts as monopolistic CP.

Insight 1 *Without a right to data portability, incumbent CPs anticipating the entry of a competitor have an incentive to create data-induced switching costs by increasing their data consumption to a level higher than in monopoly.*

With respect to the amount of required data in the second period, this restricting effect is also observable: the incumbent CP always requires less data if users are able to port their data, i.e.,

$$r_A^{*,NP,2} = \frac{15\tau - v_A - 5v_B}{17} > \frac{6\tau - v_A - 2v_B}{6} = r_A^{*,P,2}.$$

Conversely, evaluating optimal data collection by the entrant (CP B) reveals that the required amount of data with a right to data portability is always higher than in the case without data portability:

$$r_B^{*,NP,2} = \frac{16\tau - 9v_A + 6v_B}{17} < \tau - \frac{v_A - v_B}{3} = r_B^{*,P,2}.$$

Intuitively, CP B requires more data with data portability because users that switch from CP A experience less disutility due to the possibility to port the already entered data. Thus, these users now only reveal the net amount of required data which is lower (i.e., $r_B^2 - r_A^1 \leq r_B^2$), all else being equal, leading to higher market shares and profits for the entrant under this regime. Proposition 1 summarizes these findings:

Proposition 1 *Under a data portability regime, incumbents always require less user data, whereas entrants unambiguously increase their data consumption level.*

Next, to deduce possible business strategies for CPs (additionally, c.f., managerial implications in Sect. 6.1) and to analyze the factors influencing a CP's data consumption in equilibrium, we conduct comparative statics, i.e., analyze the effects on a CP's data consumption by changing the exogenous model parameters. First, we find that CP A's period one data consumption increases in its base utility v_A, whereas its second-period data consumption decreases in v_A, i.e., $\partial r_A^{*,d,1}/\partial v_A > 0$ and $\partial r_A^{*,d,2}/\partial v_A < 0$ irrespective of the considered regime. The negative effect on the second-period amount of required data by CP A can be explained by the incumbent's rationale to protect its market share in a competitive environment, i.e., after a competitor has entered the market: Through an increased base utility, CP A is able to require a large(r) amount of data in period one. Protecting this market share in period two (through a comparably low amount of required data in this period) dominates the positive effects arising from requiring more data in the second period. On the contrary, if its base utility is decreasing, protecting market shares does not dominate the positive effects of requiring additional data in period two. Second, an increase in CP B's base utility v_B lowers CP A's data collection: in period one to increase the share of users that are locked-in, in period two due to stronger competitive forces. Since the lock-in effect vanishes with data portability, the period one amount of required data is unaffected by v_B. In conclusion: $\partial r_A^{*,NP,1}/\partial v_B < 0$, $\partial r_A^{*,P,1}/\partial v_B = 0$, $\partial r_A^{*,d,2}/\partial v_B < 0$. Third, the mismatch costs of users (τ) have an unambiguous effect on CP A's data consumption: the higher the mismatch costs, the higher the amount of required data, i.e., $\partial r_A^{*,NP,1}/\partial \tau > 0$

and $\partial r_A^{*,d,2}/\partial \tau > 0$, because high mismatch costs reduce the competitive intensity in the market as a user's location, i.e., a user's preferences over the set of CPs, gets relatively more important. Finally, for CP B, comparative statics show that an increase in the competitor's base utility (v_A) reduces the amount of required data. In contrast to the incumbent, an increase in the own base utility (v_B) unambiguously increases the amount of required data. The effect of the mismatch costs for users on CP B's data consumption is qualitatively the same as the effect on CP A's data consumption, i.e., the higher the mismatch costs, the higher the amount of required data. Thus, it can be summarized that:

Insight 2 *A (in terms of service quality) strong competitor or low mismatch costs for users reduce a CP's amount of required data. If a CP increases its own quality, it requires more data in the first period being active.*

4.2 CPs' Profits

To analyze CPs' profits (π_i^d), we evaluate optimal profits given the just derived equilibrium amount of required data. Within the feasible parameter range (c.f., "Appendix 2"), the incumbent always suffers from data portability (i.e., $\pi_A^P \leq \pi_A^{NP}$), whereas the entrant always benefits from data portability (i.e., $\pi_B^P \geq \pi_B^{NP}$; see "Appendix 6" for analytical details). Thus, since data portability unambiguously increases an entrant's profits, service variety (and innovation) is arguably increased because entrants are more likely to enter the market due to higher profits. Hence, if the market is dominated by a single firm, data portability may be a suitable device to foster competition.

Comparative statics show that an increase in the CP's own base utility has always a positive effect on its profits. Conversely, an increase in the competitor's base utility decreases a CP's profits (i.e., $\partial \pi_i^d/\partial v_i > 0$ and $\partial \pi_i^d/\partial v_{-i} < 0$ for $i = \{A, B\}$ and $-i$ denoting the competing CP i). Interestingly, the effect of higher mismatch costs for users (i.e., an increase in τ) is ambiguous: with respect to π_A^P, π_A^{NP} and π_B^{NP}, higher mismatch costs are beneficial only if the competing CP (CP $-i$) is strong in terms of its base utility, i.e., $v_{-i} \gg v_i$; arguably because the considered CP then focuses on users which are located close to it. Otherwise, the effect of the mismatch costs τ depend on the characteristics of the considered market.[4] With regard to

---

[4] Formally, the derivative changes its sign in the feasible parameter range. The effect of an increasing τ on π_A^P is positive if $\tau > \sqrt{22v_A^2 - 12v_A v_B + 4v_B^2}/6$; the effect of an increasing τ on π_A^{NP} is positive if $\tau > \sqrt{26v_A^2 - 12v_A v_B + 4v_B^2}/6$; the effect of an increasing τ on π_B^{NP} is positive if $\tau > (6v_B - 9v_A)/16$.

π_B^P, the effect of the mismatch costs are unambiguous: the higher the mismatch costs for users, the higher the profits.

Insight 3 *A right to data portability unambiguously increases an entrant's profits arguably increasing service variety and innovation. In contrast, an incumbent always suffers under a data portability regime.*

4.3 Consumer's Surplus

To examine the effects on consumer's surplus (CS_i^d), we compare the users' utility accounting for the different regimes. With respect to users active at CP A, consumer's surplus for both periods is given by:

$$CS_A^d = \underbrace{\int_0^{x^{*,d,1}} U_A^1(x)dx}_{period\,t=1} + \underbrace{\int_0^{x^{*,d,2}} U_A^2(x)dx}_{period\,t=2}$$

Note that users active at CP B differ with regard to their utility under the regime with data portability depending on whether they have not been active in the first period (and consequently have a utility of $U_B^{NP,2}$), or whether they have been active in the first period, switch from CP A to CP B and port their data. Hence, the latter group has a lower disutility for a given amount of data required by CP B (and thus, has an utility of $U_B^{P,2}$). If data portability is not enforced, all users becoming active at CP B derive a utility of $U_B^{NP,2}$. In conclusion, consumer's surplus can be calculated by:

$$CS_B^d$$
$$= \begin{cases} \int_{x^{*,P,1}}^{x^{*,d,1}} U_B^{P,2}(x)dx + \int_{x^{*,d,1}}^1 U_B^{NP,2}(x)dx & \text{, with data portability } (d=P), \\ \int_{x^{*,NP,2}}^1 U_B^{NP,2}(x)dx & \text{, without data portability } (d=NP). \end{cases}$$

By comparing consumer's surplus in equilibrium, it can be seen that a regime *without* data portability may leave users actually better off. Thus, the sum of consumer's surplus at both CPs *can* decrease with introducing a right to data portability, i.e., $CS_{A+B}^P = CS_A^P + CS_B^P < CS_A^{NP} + CS_B^{NP} = CS_{A+B}^{NP}$ (see "Appendix 7" for analytical details). Consequently, although data portability is most commonly justified by the potential benefits for end customers (c.f., Macgillivray and Shambaugh 2016; European Commission 2016b), this goal is not necessarily achieved.

Moreover, it can be shown that (relatively) high mismatch costs for users may lead to users being worse off with a right to data portability, i.e., the consumer's surplus is reduced if the critical threshold (τ_{CS}) is exceeded. More precisely, if $\tau \geq \tau_{CS} := (174v_B - 822v_A + 17\sqrt{6658v_A^2 - 752v_Av_B + 16v_B^2})/726$, users are better off *without* a right to data portability (additionally, c.f., "Appendix 8"). Intuitively, as we have shown above, CPs

require higher amounts of data if the mismatch costs for users are high (because $\partial r_i^{*,d,t}/\partial \tau > 0$). This, in turn, increases the disutility a user derives from being active at the considered CP, which, consequently, reduces consumer's surplus (i.e., $\partial CS_{A+B}^d/\partial \tau < 0$). However, the threshold τ_{CS} is not always within the feasible parameter range: If the CPs' base utilities are relatively equal (i.e., $v_B < v_{B,CS} := 447/160 \cdot v_A$), consumers unambiguously benefit under a data portability regime. Additionally, higher base utilities always positively affect consumer's surplus (i.e., $\partial CS_{A+B}^d/\partial v_i > 0$). Proposition 2 summarizes these findings:

Proposition 2 *The possibility to port data from one online service to another online service has ambiguous effects on consumer's surplus. If both services offer a comparable service quality for users (i.e., $v_B < v_{B,CS}$), consumer's surplus always increases. However, if the entrant offers a better service (i.e., $v_B \geq v_{B,CS}$), users may suffer under a data portability regime if their mismatch costs to using a service are higher than τ_{CS}.*

Figure 2 illustrates the possible negative effect on consumer's surplus for a specific parameter constellation by showing total consumer's surplus, as well as the consumer's surplus at each CP with and without data portability for different mismatch costs.
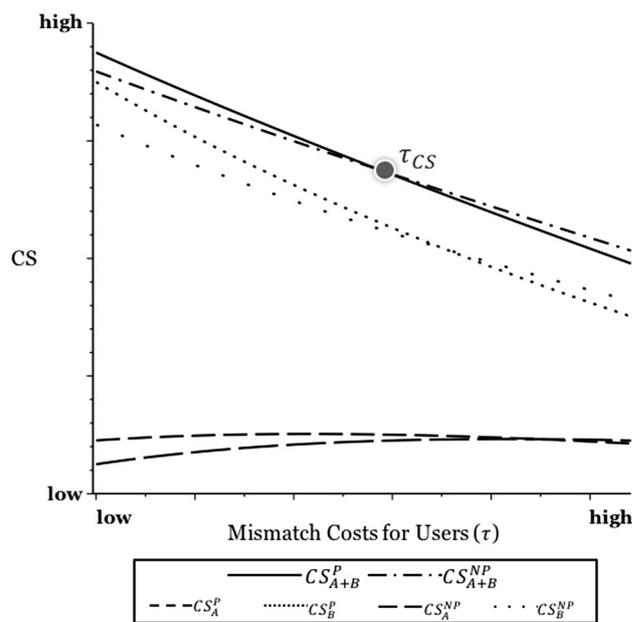

Fig. 2 Illustration of consumer's surplus for different mismatch costs. *Note* Illustration of total consumer's surplus with ($d = P$, solid line) and without data portability ($d = NP$, dash-dotted line) for $v_A = 1$ and $v_B = 4$. As $v_B > v_{B,CS}$, users are worse off if $\tau > \tau_{CS}$. Additionally, consumer's surplus at each CP i for the different regimes is illustrated [dashed (dotted) lines refer to CP A (CP B, respectively)]

4.4 Total Surplus

Finally, total surplus (TS^d) being the sum of consumer's surplus and CPs' profits, i.e.,

$$TS^d = \sum_{i=A,B} (\pi_i^d + CS_i^d)$$

is examined (see "Appendix 8" for analytical details). Within the feasible parameter range, it can be concluded that total surplus is unambiguously increasing with a right to data portability, i.e., $TS^P > TS^{NP}$. Thus, although consumers might be worse off in some cases and CP A always experiences lower profits under a regime with a right to data portability, the increased profits of CP B always outweigh these effects.

Insight 4 *Total surplus unambiguously increases with a right to data portability.*

5 Extensions

In the following, we explore four extensions to the base model, which confirm the robustness of the main insights highlighted by Proposition 1 and 2 and provide more nuanced results: Sect. 5.1 considers costs for CPs implementing a right to data portability (subscript F), Sect. 5.2 assumes that not all data that is ported to a CP is relevant to that CP (subscript ID), Sect. 5.3 considers cases where the value of collected data is diminishing over time (subscript DV), and Sect. 5.4 considers services that are characterized by network effects (subscript NWE).

5.1 Costs for Providing the Possibility to Port Data

Until now, we assumed that the possibility to port (personal) data does not incur any costs for the CPs. However, giving users the possibility to port personal data may result in additional costs such as costs for the programming effort to implement the technical functionalities. To account for such costs, we extend the model by assuming that both CPs face some exogenous costs F if a right to data portability is introduced. Consequently, the CPs' profit functions *with* a right to data portability now incorporate an additional fixed cost term F (see "Appendix 9").

The timing of the game remains unchanged. By solving for the subgame perfect Nash equilibrium, it is easy to see that the CPs' data consumption remains unchanged by

introducing (fixed) costs to implement the possibility to port data. This implies that also (1) all insights with respect to the amount of required data (c.f., Proposition 1), and (2) all insights with respect to consumer's surplus remain unchanged (c.f., Proposition 2). Consequently, users can still be worse off if a right to data portability is introduced. In contrast, CPs' profits change if a right to data portability is introduced. Obviously, CPs' profits are affected negatively by introducing costs, i.e., $\partial \pi_{i,F}^P / \partial F < 0$ with $i \in \{A, B\}$. Thus, the entrant is *not* necessarily better off if a right to data portability is introduced. Instead, the entrant is worse off (i.e., $\pi_{B,F}^P < \pi_{B,F}^{NP}$), if the fixed costs for the implementation of a functionality to port data exceed the critical threshold $\hat{F}$ (see "Appendix 9"), i.e., if

$$F > \hat{F} := \frac{(10v_A - v_B + 3\tau) \cdot (35v_B - 44v_A + 99\tau)}{5205 \cdot \tau}.$$

Thus, if the costs associated with providing the possibility to port personal data are too high, the right to data portability does not necessarily stimulate market entry or innovation as entrants may find it unprofitable to enter the market at all. Please note that this very same result is true, if we would assume that fixed costs are only relevant for entrants but not for established firms (i.e., incumbents). Moreover, total surplus may now decrease with the introduction of a right to data portability because *all* CPs as well as users can be worse off. Therefore, policy makers need to deliberately define the scope of data that can actually be ported and additionally specify the concrete mechanism of data portability in order to reduce costs. For example, in many cases the transmission of personal data should not occur *directly* between different CPs as this arguably increases implementation costs, particularly as the transmission needs to be secure in order to protect users' sensitive data.

Insight 5 *If implementing a right to data portability is associated with fixed costs F, even entrants can suffer from introducing a right to data portability if the resulting costs exceed $\hat{F}$. Then, also total surplus is likely to be reduced as all CPs are worse off and users can be worse off under a data portability regime.*

5.2 Porting Irrelevant Data

Although this paper investigates the effects of a right to data portability on two CPs providing substitutable services, these CPs may not necessarily require identical data from users becoming active at their platform. Whereas we address a benchmark case in our base model by assuming that all data that is transferred to the competing CP is valuable, we now modify our model to account for cases

---

⁴ Formally, the derivative changes its sign in the feasible parameter range. The effect of an increasing τ on π_A^P is positive if $\tau > \sqrt{22v_A^2 - 12v_A v_B + 4v_B^2}/6$; the effect of an increasing τ on π_A^{NP} is positive if $\tau > \sqrt{26v_A^2 - 12v_A v_B + 4v_B^2}/6$; the effect of an increasing τ on π_B^{NP} is positive if $\tau > (6v_B - 9v_A)/16$.

where also irrelevant data (*ID*) is ported to the entrant (CP *B*).

In doing so, we introduce the parameter $\gamma \in [0,1]$ defining the share of ported data that is (also) useful for the CP where the data is ported to (here: the entrant CP *B*). For example, a user may have entered her name, date of birth and cellphone number at CP *A* in $t = 1$ (i.e., r_A^1) and now ports this data to CP *B* in $t = 2$. However, CP *B* requires the name, date of birth and address from users becoming active at the platform (i.e., r_B^2) and cannot analyze or monetize a user's cellphone number. Consequently, only some share of the ported data is relevant to the new CP. Thus, the *net amount of required data* is not given by $r_B^2 - r_A^1$ as in the base model, but by $r_B^2 - \gamma \cdot r_A^1$. Hence, if data portability is possible, the utility of users that have been active at CP *A* in $t = 1$ and switch to CP *B* changes compared to the base model. By assuming that only a share of the ported data is useful for the new CP, a user located at x becoming active at CP *B* in $t = 2$ derives a utility of

$$U_{B,ID}^{d,2}(x)$$
$$= \begin{cases} v_B - \tau \cdot (1-x) - r_B^2 + \gamma \cdot r_A^1 & \text{, if } U_A^1(x) \geq 0 \text{ with data portability (d = P),} \\ v_B - \tau \cdot (1-x) - r_B^2 & \text{, else (d = NP or } U_A^1(x) < 0). \end{cases}$$

Consequently, with data portability, the location of the indifferent user changes in $t = 2$, which also affects CPs' profits as well as the amount of required data (c.f., "Appendix 10" for analytical details). Note that this extension is a generalization of the base model outlined above. Thus, assuming $\gamma = 0$, the results are identical to the benchmark case without data portability because none of the ported data is useful for the entrant. Conversely, assuming $\gamma = 1$, the results are identical to the benchmark case with data portability where all ported data is relevant to the entrant.

To deduce more nuanced results, we solve the game through backward induction. Due to the extreme cases already analyzed, we restrict our analysis to $\gamma \in (0,1)$. In summary, we obtain:

$$r_{A,ID}^{*,P,1} = \frac{(3\tau + v_A - v_B)\gamma - 3\tau - 10v_A + v_B}{\gamma^2 - 2\gamma - 17} \text{ with } r_A^{*,NP,1} > r_{A,ID}^{*,P,1} > r_A^{*,P,1},$$

$$r_{A,ID}^{*,P,2} = \frac{(-3\tau + 2v_A + v_B)\gamma - 15\tau + v_A + 5v_B}{\gamma^2 - 2\gamma - 17} \text{ with } r_A^{*,NP,2} > r_{A,ID}^{*,P,2} > r_A^{*,P,2},$$

$$r_{B,ID}^{*,P,2} = \frac{2\tau\gamma^2 - (4\tau + 3v_A)\gamma - 16\tau + 9v_A - 6v_B}{\gamma^2 - 2\gamma - 17} \text{ with } r_B^{*,NP,2} < r_{B,ID}^{*,P,2} < r_B^{*,P,2}.$$

It can be seen that a higher γ increases the entrant's amount of required data, i.e., an entrant CP's data consumption increases with the amount of data that is ported *and* valuable, whereas the incumbent's amount of required data is reduced (i.e., $\partial r_{A,ID}^{*,P,t}/\partial\gamma < 0$ with $t \in \{1,2\}$ and $\partial r_{B,ID}^{*,P,2}/\partial\gamma > 0$). Additionally, the incumbent's period one amount of required data now also (negatively) depends on v_B: Due to the assumption that not all data is relevant to CP *B*, CP *B*'s decision in $t = 2$

now affects CP *A*'s decision in $t = 1$. This has not been the case in the base model. In the base model, v_B does not affect the data consumption in period one, because all of the data collected by CP *A* is transferred and valuable for CP *B*. Consequently, CP *A* behaves like a one-period monopolist with respect to its data consumption irrespective of CP *B*'s decision in $t = 2$. However, Proposition 1 still continues to hold, i.e., the incumbent still requires less data and the entrant requires more data if users have the possibility to port (some share of their) personal data. Moreover, CPs' profits behave intuitively with respect to the introduced parameter γ, i.e., the incumbent's profits decrease, whereas the entrant's profits increase the more data is relevant to the entrant, i.e., $\partial \pi_{A,ID}^P/\partial\gamma < 0$ and $\partial \pi_{B,ID}^P/\partial\gamma > 0$. Consequently, the incumbent may be able protect its profits by strategically reducing the amount of explicitly stored information that can be ported with a right to data portability, e.g., by inferring information from a user's action on the website instead of requiring data to be actively entered by users (because only data provided by users may be subject to data portability, c.f., European Commission 2016b) or by requiring data from users that is only useful in combination with other data that is not subject to data portability.

Assuming that not all data is relevant to the entrant also affects consumer's surplus. However, Proposition 2 still continues to holds, i.e., if the entrant provides a better service quality, users may actually be worse off with a right to data portability. Here, it can be seen that $\gamma \in (0,1)$ can dampen the negative effects of data portability on consumer's surplus compared to the base model with data portability: If users suffer most with a right to data portability assuming $\gamma = 1$, i.e., if the mismatch costs are very high, they suffer less with $\gamma \in (0,1)$. Consequently, from a policy perspective, restricting the amount of data that can be ported may be a device to protect users. However, this necessitates that policy makers need to precisely analyze the competitive intensity of the market apriori, because restricting the amount of data that can be ported also dampens consumer's surplus in cases where users benefit from a right to data portability (additionally, c.f., "Appendix 10"). Finally, it can be shown that total surplus is always higher with a right to data portability–although only some share of the ported data is actually relevant to the entrant.

Insight 6 *If users can port their personal data from an incumbent to an entrant but only some share of this data $\gamma \in (0,1)$ is relevant to the entrant, incumbents (entrants) reduce (increase) their data consumption with a right to data portability which may lead to users being worse off compared to a regime without a right to data portability.*

5.3 Diminishing Value of Collected Data

The benchmark case analyzed in the base model assumes that the data an incumbent collected in $t = 1$ is equally important in $t = 2$, i.e., has an identical effect on profits. In the following, we relax this assumption by assuming that the value of data is diminishing (DV), i.e., the incumbent can only monetize a share $\rho \in [0, 1]$ of collected data in succeeding periods. Thus, ρ represents the share of data collected in period one that is (still) valuable for CP A in period two. Herewith, CP A's profit function changes to

$$\pi_{A,DV}^{d} = \underbrace{x^{*,d,1} \cdot r_A^{d,1}}_{\pi_{A,DV}^{d,1} = \pi_A^{d,1}} + \underbrace{x^{*,d,2} \cdot (\rho \cdot r_A^{d,1} + r_A^{d,2})}_{\pi_{A,DV}^{d,2}}.$$

It is worth mentioning that assuming $\rho = 1$ leads to the benchmark cases analyzed in Sect. 4. Thus, we concentrate on cases with $\rho < 1$. Note that users' utility functions remain unaffected by introducing ρ. Consequently, the formulas derived in the benchmark case to calculate the locations of the indifferent users can also be used for this extension. Moreover, CP B's profit function does not change compared to the base model. However, CP A now incorporates the diminishing value of the data collected in $t = 1$ in its profit function for $t = 2$. Due to solving the game through backward induction, this affects the amount of required data for *all* CPs in *each* period. For the regime with a right to data portability, we obtain:

$$r_{A,DV}^{*,P,1} = \frac{(-3\tau - v_A + v_B)\rho + 3\tau - 8v_A - v_B}{\rho^2 - 2\rho - 17} \text{ with } r_{A,DV}^{*,P,1} < r_A^{*,P,1},$$

$$r_{A,DV}^{*,P,2} = \frac{(3\tau + v_A - v_B)\rho^2 + (-3\tau + 5v_A + v_B)\rho - 18\tau - 3v_A + 6v_B}{\rho^2 - 2\rho - 17}$$
$$\text{with } r_{A,DV}^{*,P,2} > r_A^{*,P,2},$$

$$r_{B,DV}^{*,P,2} = \frac{2\tau\rho + (-4\tau + 3v_A)\rho - 16\tau + 3v_A - 6v_B}{\rho^2 - 2\rho - 17} \text{ with } r_{B,DV}^{*,P,2} > r_B^{*,P,2}.$$

and for the regime without a right to data portability:

$$r_{A,DV}^{*,NP,1} = \frac{(-3\tau - v_A + v_B)\rho - 9v_A}{\rho^2 - 18} \text{ with } r_{A,DV}^{*,NP,1} < r_A^{*,NP,1},$$

$$r_{A,DV}^{*,NP,2} = \frac{(3\tau + v_A - v_B)\rho^2 + 6v_A\rho - 18\tau - 6v_A + 6v_B}{\rho^2 - 18}$$
$$\text{with } r_{A,DV}^{*,NP,2} > r_A^{*,NP,2},$$

$$r_{B,DV}^{*,NP,2} = \frac{2\tau\rho^2 + 3v_A\rho - 18\tau + 6v_A - 6v_B}{\rho^2 - 18} \text{ with } r_{B,DV}^{*,NP,2} > r_B^{*,NP,2}.$$

See that introducing the parameter ρ has a negative impact on CP A's period one data consumption, i.e., the incumbent requires less data in $t = 1$ compared to the benchmark case. Conversely, the period two amount of required data increases with ρ, i.e., the incumbent as well as the entrant require more data in $t = 2$. In conclusion, $\partial r_{A,DV}^{*,d,1}/\partial\rho < 0$ and $\partial r_{i,DV}^{*,d,2}/\partial\rho > 0$. Intuitively, compared to the base

model, the benefits from data collected in $t = 1$ that the incumbent CP A can convey to the succeeding period is lower. This leads to a lower data consumption in period one; however, in period two, the incumbent then increases its data consumption compared to the benchmark case. This also leads to an increasing data consumption of the entrant CP B. Please note that without a right to data portability, CP A still requires at least the amount of data a monopolist would require. This corroborates our insight that incumbent firms have an incentive to generate data-induced switching costs (i.e., $r_{A,DV}^{NP,1} \geq r_{Monopoly}^{*}$). Moreover, it can easily be shown that Proposition 1 continues to hold, i.e., the incumbent reduces its data consumption with a right to data portability (i.e., $r_{A,DV}^{*,NP,t} > r_{A,DV}^{*,P,t}$) whereas the entrant increases its data consumption (i.e., $r_{B,DV}^{*,NP,2} < r_{A,DV}^{*,P,2}$). With respect to CPs' profits, it can be shown that CP A (CP B) suffers (benefits) with $\rho < 1$, i.e., $\pi_{A,DV}^{d} < \pi_A^{d}$ and $\pi_{B,DV}^{d} > \pi_B^{d}$, respectively. Comparative statics reveal that the effect of ρ on the CPs' profits is monotone, i.e., $\partial\pi_{A,DV}^{d}/\partial\rho > 0$ and $\partial\pi_{B,DV}^{d}/\partial\rho < 0$ within the feasible parameter range. Moreover, with respect to consumer's surplus, also Proposition 2 continues to hold, i.e., users can – again – be worse off with the possibility to port data (see "Appendix 11").

Insight 7 *If the value of data an incumbent collects in period one is not equally valuable in period two, the incumbent reduces its data consumption in period $t = 1$, but increases its data consumption in period $t = 2$. In contrast, the entrant unambiguously increases its data consumption compared to a scenario where the value of collected data does not change over time. However, compared to a regime without a right to data portability, the incumbent (entrant) still reduces (increases) its data consumption which can lead to users being worse off.*

5.4 The Role of Network Effects

As highlighted in the previous sections, network effects are not a precondition for online CPs to become successful and are not necessarily the (main) source for users to become locked-in. However, the utility a user derives from being active at an online service may nevertheless be affected by the number of other users active at that platform, i.e., direct network effects may exist and influence a user's decision, but also the CPs' strategies in setting the amount of required data. Intuitively, the presence of positive network effects may reduce a user's incentive to switch to an entrant CP because the derived utility from the already installed base at the incumbent may outweigh the potentially higher base utility from the joining CP – although data already entered can be ported to that joining CP with a right to data

portability. To investigate the role of network effects formally, we modify the users' utility functions and incorporate positive direct network effects (NWE). In doing so, we assume that the total number of users active at the considered CP has a positive effect on a user's utility, i.e., $U_{A,NWE}^{d,t}(x) = U_A^{d,t}(x) + \omega \cdot x^{*,d,t}$ for CP A and $U_{B,NWE}^{d,2}(x) = U_B^{d,2}(x) + \omega \cdot (1 - x^{*,d,t})$ for CP B, respectively with $\omega > 0$. By changing the utility functions, also the location of the indifferent user changes. Relying on the concept of fulfilled expectations (i.e., in equilibrium, the network size determined by the location of the indifferent user equals the expected one, additionally, c.f., Katz and Shapiro 1985), the indifferent user in period $t = 1$ is now located at $x_{NWE}^{*,d,1} = \frac{v_A - r_A}{\tau - \omega}$ and the indifferent user in period $t = 2$ is now located at:

$$x_{NWE}^{*,d,2} = \begin{cases} \dfrac{v_B + \omega - \tau - r_B^2 + r_A^1 + r_A^2 - v_A}{2(\omega - \tau)} & , \text{if } U_{A,NWE}^1(x^{*,d,2}) \geq 0 \quad (d = P), \\[2ex] \dfrac{v_B + \omega - \tau - r_B^2 + r_A^2 - v_A}{2(\omega - \tau)} & , \text{else} \quad (d = NP). \end{cases}$$

The resulting profit functions as well as our proposed two stage game remain qualitatively unchanged (additionally, c.f., "Appendix 12").

Again, we solve for the subgame perfect Nash equilibrium using backward induction and derive the period one and period two level of data consumption as shown in Sect. 18.1. Compared to the base model without incorporating network effects (c.f., Sects. 3 and 4), one can easily show that CPs never require more data, i.e., the existence of positive direct network effects has a negative impact on CPs' data consumption ($\partial r_{i,NWE}^{*,d,t}/\partial\omega < 0$). Intuitively, CP A now has the possibility to lock-in users without increasing its data consumption. This improved competitive situation also leads to CP B reducing its data consumption which is beneficial to users (see Fig. 3). However, our results with respect to CPs' data consumption highlighted in Sect. 4.1 continue to hold, i.e., $r_{A,NWE}^{*,NP,t} > r_{A,NWE}^{*,P,t}$ and $r_{B,NWE}^{*,NP,2} < r_{B,NWE}^{*,P,2}$, and consequently, Proposition 1 continues to hold.

As the CP's data consumption changes, incorporating network effects has ramifications on all players within our considered market. However, also our other results of introducing a right to data portability qualitatively remain unchanged which further corroborates the robustness of the model: The incumbent always suffers from introducing a right to data portability, the entrant is always better off, and total surplus always increases. Moreover, the effect of data portability on consumers remains ambiguous. Although consumer's surplus with a right to data portability is now higher in more cases, i.e., the intersection of both functions is shifted to the edge of the feasible parameter range (c.f., Fig. 3 for an illustration and comparison), users
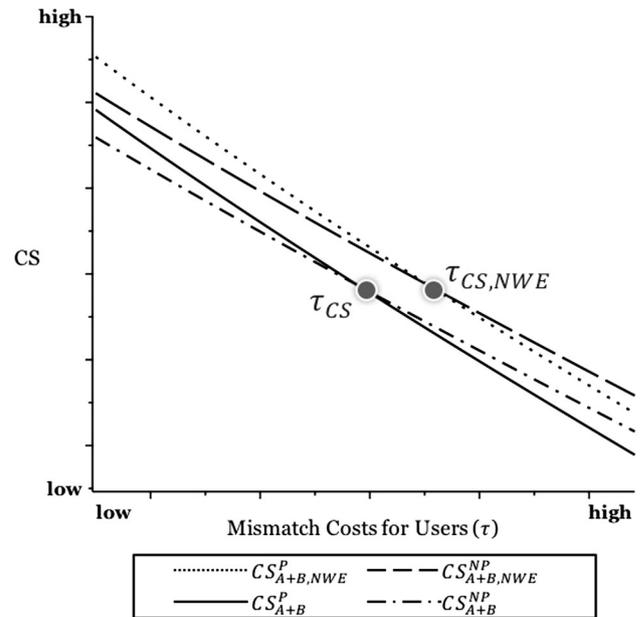

Fig. 3 Comparison: Consumer's surplus for services with and without network effects. *Note* Total consumer's surplus with and without incorporating positive direct network effects with and without a right to data portability for different users' mismatch costs using the parameters $v_A = 1, v_B = 4$ and $\omega = 0.05$. The upper (lower) two curves refer to a model with (without) incorporating network effects

nevertheless may experience a lower consumer's surplus compared to a regime without a right to data portability if their mismatch costs exceed $\tau_{CS,NWE}$, i.e., also Proposition 2 continues to hold (additionally, c.f., Sect. 18.3).

Insight 8 *If being active at a CP induces positive direct network effects for users, the CPs' level of data consumption is lower. However, introducing a right to data portability increases (reduces) an entrant's (incumbent's) level of data consumption which may lead to users being worse off compared to a regime without a right to data portability.*

6 Conclusion

Data portability allows users to transfer their data entered at a certain service to another service. Although some online services have implemented such features voluntarily, and built-in autofill features of internet browsers can reduce the effort to create new accounts, a standardized and mandatory ability for users to port (personal) data is pursued by the European Commission for all online services available in the EU's member states through the General Data Protection Regulation (European Commission 2016b). Additionally, this topic also gains momentum for non-European policy makers, as the request for information

in the United States suggests (c.f., Macgillivray and Shambaugh 2016).

Despite the importance of this issue resulting from the far-reaching implications on business strategies of online services and thus on the total economy, we are – to the best of our knowledge – the first to analyze the resulting competitive effects theoretically. In doing so, we not only shed light on current policy issues, but also highlight relevant implications on the interface of the IS, the technical and the economic realm to better understand and develop systems' value propositions. For this purpose, we propose a game-theoretic model that captures competing online services' strategic incentives and identify the feasible market outcomes together with the implications for all stakeholders.

In conclusion, we find that if the CP's costs to implement data portability are not too large, on the one hand, data portability fosters market entry, which arguably enhances service variety and innovation, but on the other hand, incumbent services unambiguously suffer from data portability. Whereas such an outcome might be desired by policy makers to alleviate concerns about dominant online services, we highlight that end users may actually suffer from a right to data portability, because new services have an incentive to increase the amount of collected data compared to a regime without data portability. However, as the total surplus increases due to higher overall profits, a decision to introduce a mandatory right to data portability invokes a complex assessment. In the following, we outline policy implications as well as strategies for services active in data-driven markets based on the obtained results and discuss avenues for future research.

6.1 Policy and Managerial Implications

From a *policy perspective*, the rationale to introduce a (general) right to data portability is clearly focused on the protection of end users (see, e.g., European Commission 2016b, Article 1). Consequently, our results imply that data portability should *not* be applied to all online services because consumer's might actually be worse off. On the other hand, considering the total economy, overreaching goals such as the Digital Single Market Strategy (DSM strategy) within the European Union (c.f., European Commission 2016a) or former-president Obama's executive order on competition from April 2016 (c.f., Obama 2016) highlight the importance of open, fair and non-discriminatory (data-driven) markets. As we show that the entrant's profits increase under data portability, a right to data portability may attribute to these goals. However, these goals are only achieved if the resulting costs (for implementation as well as administration) of a right to data portability are low. Therefore, our findings evoke the necessity for policy makers to carefully weigh whether they want to promote market entry to stimulate innovation and successively service variety, or purely focus on consumer's surplus.

If new services should be incentivized to enter the market, data portability should be enforced strictly with few exceptions. To date, the concept of data portability proposed by the European Commission solely focuses on personal data revealed by users themselves. Hence, data revealed by third persons (say, reviews for a private lift, or endorsements on professional networking sites) are excluded in the current version of the regulation. Therefore, policy makers might think of extending the scope of data that can be ported. In fact, as highlighted in the mid-term review on the implementation of the Digital Single Market Strategy, the European Commission already "subject to impact assessment, prepare[s] a legislative proposal [...] which takes into account [...] the principle of porting non-personal data" (European Commission 2017, p.11). In most cases, extending the scope of portable data would be in line with the goal of enhancing consumer's surplus. However, it has to be taken into consideration that (1) porting sensitive data (e.g., credit card numbers, tax IDs, social security numbers) bears important privacy and security risks, although users entered these data voluntarily, and (2) there are cases where users are actually worse off with a right to data portability, as we have shown throughout all of our model specifications and analyses. Our results suggest that users are likely to be worse off if base utilities are asymmetric, e.g., if the entrant has a superior value proposition providing the user a higher base utility. Arguably, entry is then beneficial for the entrant even without a right to data portability. Consequently, one may hypothetically think of a concept where data portability is only granted to some services. Although this seems possible in theory, the likeliness of success of such an approach is questionable as (1) this concept would contradict popular "neutrality regimes", which might get increasingly important on a service level (c.f., Easley et al. 2018), (2) the current political view aims at giving end users back the control of their (personal) data; independent of the considered service (c.f., European Commission 2016b), and (3) the nature of the internet with independent parties and hard-to-control data flows makes supervision costly. However, as we have shown that the negative effects of data portability on consumer's surplus can be dampened by restricting the amount of data that can be ported, this might be a possible way to facilitate market entry and to limit potential adverse effects on consumers.

From a *managerial perspective*, it has to be emphasized that incumbent services have an unambiguous incentive to inhibit the concept of data portability because their opportunity to soften competition vanishes, leading to reduced profits. In contrast, entrant services or start-ups

should promote the concept of data portability because their flexibility in setting the amount of data that is collected rises, leading to higher profits and thus, earlier profitability. If services have no possibility to influence the scope of data that can be ported, incumbents should pursue a differentiation strategy if the entrant is superior in terms of its base utility. This arguably increases a user's mismatch costs which reduces the competitiveness of the market and ultimately benefits the incumbent. For this purpose, incumbents may try to change (aspects of) their service offering (i.e., differentiate) to escape the fierce competition with the new service. In contrast, a strategy designed to imitate the competitor can be seen as an incumbent's opportunity if the entrant is relatively equal in terms of its base utility and if the mismatch costs of users are already comparably high. This might be achieved by matching all of the entrant's value propositions to reduce mismatch costs which increases competition and thus, profits (see effects of the users' mismatch costs on profits outlined in Sect. 4.2). Additionally, incumbents may try to (1) infer information from a user's browsing behavior as data that has not been actively provided by users is not covered by the right to data portability, and (2) require "proxy data" from users that is only useful for services if they are analyzed in combination with other data (that is not subject to data portability). The entrant always benefits from higher mismatch costs of users and should thus differentiate as much as possible from the incumbent, e.g., by acting as the industry's innovation leader.

6.2 Limitations and Avenues for Future Research

Finally, we wish to conclude by highlighting possible model extensions and limitations that should be taken into consideration and analyzed in future studies. First, the market environment could be changed to capture the effects of data portability on two existing, competing services. In our terminology, CP B would then already be active in period one and data can be ported from CP A to CP B and vice versa. Arguably, as the CP's flexibility in setting the amount of required data is reduced, CPs should suffer under a regime that enforces data portability. Conversely, such a market environment would be beneficial for end users. Second, the possibility to discriminate between new users and existing users might be seen as a possible model extension. However, this extension would assume that services have a non-uniform data consumption for data from different user groups, which may increase programming efforts and potentially complicates the provision of a streamlined and consistent (service) portfolio. With data portability, the entrant would then collect a relatively high amount of data from new (i.e., not switching) users and additionally maintains flexibility for the share of users that

may switch services, leading to reduced consumer's surplus. Third, we assumed that data entered once has no effect on a user's utility in succeeding periods. Whereas we believe that this is a suitable benchmark, one may argue that the disutility of already entered data only diminishes over time, i.e., the effects of trust for a certain service or the possibility of data breaches at a CP might be included into the analysis. Incorporating trust can be achieved by assuming that there is a lower (or no) disutility if the same service is used again, whereas there is some disutility if the same data is ported to another service. Fourth, we only assumed the costs of revealing personal data. However, entering (more) personal data may also lead to a higher base utility of services because the service can be better personalized to a user's needs. This effect can be introduced, e.g., by assuming that the valuation of a service is an increasing concave function of the costs. Finally, a right to data portability arguably also induces positive effects on other CPs, which supply independent or complementary services, but are not modeled within this study focusing on competing CPs. Thus, the positive effect of data portability on service variety and innovation may be stronger than assumed in this study.

Appendix 1: Notation

The notation of the game-theoretic model outlined in Sect. 3 and solved in Sect. 4 is stated according to its occurrence in the text in Table 1. Moreover, the notation introduced in Sect. 5 is presented.

Appendix 2: Thresholds for the Feasible Parameter Range

In this paper, we build on Hotelling's model of horizontal differentiation (c.f., Hotelling 1929) in order to identify the competitive effects of introducing a right to data portability. In doing so, we assume that a unit mass of users is uniformly distributed on the interval [0, 1]. By calculating market shares, which can directly be deduced from the location of the indifferent user, we formally need to ensure that the indifferent user is in all cases located within the

Table 1 Notation used in the game-theoretic model and its extension

Variable	Explanation	First occurrence
Base model (Sects. 3 and 4)		
d	Specifying the considered regime: $d = NP$ refers to a regime *without* a right to data portability, $d = P$ refers to a regime *with* a right to data portability	Section 3
i	Considered Content Providers (CPs), i.e., $i \in \{A, B\}$ and $-i$ refers to the competing CP i	Section 3
t	Time period t with $t \in \{1, 2\}$	Section 3
x	A user's location on the unit interval with $x \in [0, 1]$	Section 3
$U_i^t(x)$	Utility a user at location x derives from being active at CP i in t	Section 3
v_i	Base utility of CP i, e.g., determined by the service's functionalities, or the quality of provided content	Section 3
r_i^t	Amount of required data / data consumption, i.e., the disutlity a user derives from revealing data a CP i requires in period t from users being active at that CP	Section 3
τ	Mismatch costs for users to using a certain service (affects the users' preferences over the set of CPs)	Section 3
$x^{*,d,t}$	Location of the indifferent user assuming regime d in period t	Section 3
π_i^d	Profit of CP i under regime d	Section 3
$r_i^{*,d,t}$	Equilibrium amount of required data CP i requires under regime d from users becoming active in period t	Section 4.1
CS_i^d	Consumer's surplus of users active at CP i under regime d. CS_{A+B}^d refers to the total consumer's surplus	Section 4.3
τ_{CS}	Threshold that defines the condition for τ that leads to users being worse off with a right to data portability (if $\tau > \tau_{CS}$)	Section 4.3
$v_{B,CS}$	Threshold that defines the condition for CP B's base utility that leads to τ_{CS} being within the feasible parameter range (if $v_B > v_{B,CS}$)	Section 4.3
TS^d	Total surplus under regime d	Section 4.4
Extension 5.1 assuming fixed costs (Subscript F)		
F	Exogenous fixed costs affecting CP i's profits	Extension 5.1
$\hat{F}$	Critical threshold for F that leads to CP B being worse off with a right to data portability	Extension 5.1
Extension 5.2 assuming porting irrelevant data (subscript ID)		
γ	Share of data that is ported from CP A to CP B and that is valuable for CP B (with $\gamma \in [0, 1]$)	Extension 5.2
Extension 5.3 assuming a diminishing value of collected data (subscript DV)		
ρ	Share of data collected in period $t = 1$ that is (still) valuable for CP A in $t = 2$ (with $\rho \in [0, 1]$)	Extension 5.3
Extension 5.4 assuming network effects (subscript NWE)		
ω	Parameter defining the benefit of network effects for users, i.e., ω is multiplied with the size of the network at the considered CP (with $\omega > 0$)	Extension 5.4
$\tau_{CS,NWE}$	Threshold that defines the condition for τ that leads to users being worse off with a right to data portability assuming network effects (if $\tau > \tau_{CS,NWE}$)	Section 5.4

interval [0, 1]. Consequently, for the regime with data portability and for the regime without data portability, we require $x^{*,d,1} \leq 1, x^{*,d,2} \geq 0$ and $x^{*,d,2} \leq 1$, i.e., the CPs' market shares are always positive and do not exceed 100%. As highlighted above, we assume (1) full market coverage in $t = 2$ for analytical tractability and (2) to analyze the effects of data portability, an entrant's base utility that is large enough so that (at least) one user can potentially port its user data from CP A to CP B, i.e., $U_B^{d,2}(x^{*,d,1}) \geq 0$. These assumptions lead to several conditions and thresholds stated next.

With a right to data portability. The following thresholds for τ refer to the regime with a right to data portability.

Condition *P1*: Indifferent user in $t = 1$ within the feasible parameter range

$$\tau > th\_p\_1 := \frac{v_A}{2}.$$

Condition *P2*: Indifferent user in $t = 2$ within the feasible parameter range (market share smaller 100%)

$$\tau > th\_p\_2 := \frac{v_A - v_B}{3}.$$

Condition *P3*: Indifferent user in $t = 2$ within the feasible parameter range (market share larger 0%)

$$\tau > th\_p\_3 := \frac{-v_A + v_B}{3}.$$

Condition *P4*: Overlapping market shares (full market coverage), i.e., at least one user has to be able to port its data

$$\tau < th\_p\_4 := \frac{5v_A}{12} + \frac{v_B}{3}.$$

Without a right to data portability. The following thresholds for τ refer to the regime without a right to data portability.

Condition *NP1*: Indifferent user in $t = 1$ within the feasible parameter range

$$\tau > th\_np\_1 := \frac{7v_A + v_B}{20}.$$

Condition *NP2*: Indifferent user in $t = 2$ within the feasible parameter range (market share smaller 100%)

$$\tau > th\_np\_2 := \frac{9v_A - 6v_B}{16}.$$

Condition *NP3*: Indifferent user in $t = 2$ within the feasible parameter range (market share larger 0%)

$$\tau > th\_np\_3 := \frac{v_B}{3} - \frac{v_A}{2}.$$

Condition *NP4*: Overlapping market shares (full market coverage)

$$\tau < th\_np\_4 := \frac{5v_A}{24} + \frac{v_B}{3}.$$

Please note that these conditions restrict the *feasible parameter range* where the regime with and without data portability can be compared in. We account for these thresholds by comparing the regimes with and without data portability only in those cases where the value of τ is feasible in *both* regimes. This "lowest common denominator" delimits the feasible parameter range used for the analyses, i.e., we require

$$\tau \in [max\{th\_p\_1, th\_p\_2, th\_p\_3, th\_np\_1,$$
$$th\_np\_2, th\_np\_3\}, \quad min\{th\_p\_4, th\_np\_4\}].$$

Appendix 3: Location of the Indifferent User

In the following, we show that $U_A^1(x^{*,d,2}) \geq 0$ is satisfied in all relevant cases, i.e., the indifferent user in period two is given by $x^{*,P,2} = -(r_A^2 + r_A^1 - r_B^2 - \tau - v_A + v_B)/2\tau$ *with* data portability, and by $x^{*,NP,2} = -(r_A^2 - r_B^2 - \tau - v_A + v_B)/2\tau$ *without* data portability.

In doing so, assume $U_A^1(x^{*,d,2}) < 0$. The location of the indifferent user is then calculated by solving $v_A - \tau \cdot x - r_A^2 - r_A^1 = v_B - \tau \cdot (1 - x) - r_B^2$ which yields $x_{new}^{*,2} = x^{*,P,2} = -(r_A^2 + r_A^1 - r_B^2 - \tau - v_A + v_B)/2\tau$. Note that by assuming $U_A^1(x^{*,d,2}) < 0$, the indifferent user is located right to the indifferent user in period $t = 1$, i.e., $x_{new}^{*,2} > x^{*,d,1}$. Consequently, users do not port their data although they would be able to do so, i.e., now the case with and without data portability coincides. We use $x_{new}^{*,2}$ to specify firms' profits. Again, we solve the game through backward induction. We use the obtained equilibrium results and calculate $U_A^1(x_{new}^{*,2})$. The resulting term is only smaller than zero iff $\tau > \tau^{min} := 2v_A 3 + v_B/3$. However, we assumed that CP *B*'s base utility is large enough so that at least one user can potentially port its user data (see above). This implies that $\tau < \tau^{max} = th\_p\_4 := 5v_A 12 + v_B/3$. It can easily be seen that $\tau^{min} > \tau^{max}$. Consequently, proofing by contradiction, $U_A^1(x^{*,d,2}) \geq 0$ is always satisfied.

Appendix 4: Myopic versus Strategic Users

In the following, we show that it is irrelevant whether users are assumed to be myopic or strategic. In doing so, first, consider the regime without a right to data portability. Here, the analysis remains identical due to the two stages assumed for our game-theoretic model and the assumption that data revealed in $t = 1$ does not lead to a disutility for users in $t = 2$. Thus, users do not have any benefit in $t = 2$ if they reveal more data in $t = 1$. Furthermore, CPs have no incentive to reduce their data consumption in case users provided additional data. Second, consider the regime with a right to data portability and assume a strategic user that is willing to accept a negative utility in $t = 1$ to be able to port (more) data to CP *B* in $t = 2$. However, CP *B* would then simply increase its data consumption in $t = 2$ leading to users being worse off compared to a user that is not willing to accept a negative utility in $t = 1$. Similar, also CP *A* has no incentive to reduce its data consumption as users do not experience a further disutility from data revealed in $t = 1$. Consequently, users would also suffer with data portability if they do not switch to CP *B*. Thus, in conclusion, users would unambiguously be worse off if they decide to accept a negative utility in $t = 1$ which is why they would not be willing to accept a negative utility in the first place. Consequently, assuming strategic users would not change the model's results as users' decisions coincide.

Appendix 5: Amount of Required Data (r_i^t)

The equilibrium amount of required data is (c.f., Sect. 4.1):

$$r_A^{*,P,1} = \frac{v_A}{2}$$

$$r_A^{*,P,2} = \frac{6\tau - v_A - 2v_B}{6}$$

$$r_B^{*,P,2} = \tau - \frac{v_A - v_B}{3}$$

$$r_A^{*,NP,1} = \frac{3\tau + 10v_A - v_B}{17}$$

$$r_A^{*,NP,2} = \frac{15\tau - v_A - 5v_B}{17}$$

$$r_B^{*,NP,2} = \frac{3\tau + 10v_A - v_B}{17}$$

The second order conditions are:

$$\frac{\partial^2 \pi_A^{P,1}}{\partial (r_A^1)^2} = -2/\tau < 0$$

$$\frac{\partial^2 \pi_A^{P,2}}{\partial (r_A^2)^2} = -1/\tau < 0$$

$$\frac{\partial^2 \pi_B^{P,2}}{\partial (r_B^2)^2} = -1/\tau < 0$$

$$\frac{\partial^2 \pi_A^{NP,1}}{\partial (r_A^1)^2} = -17/9\tau < 0$$

$$\frac{\partial^2 \pi_A^{NP,2}}{\partial (r_A^2)^2} = -1/\tau < 0$$

$$\frac{\partial^2 \pi_B^{NP,2}}{\partial (r_B^2)^2} = -1/\tau < 0$$

Consequently, the equilibrium amount of required data for CP A and CP B, respectively, constitute the profit maximizing data consumption.

Moreover, it can easily be shown that the amount of data CP A requires is higher under a regime *without* data portability ($d = NP$). For the first period, the amount of required data *with* data portability can only be higher if $\tau < -v_A/2 + v_B/3$. In the second period, the amount of required data *with* data portability can only be higher if $\tau > 11v_A/12 + v_B/3$. However, both conditions violate the feasible parameter range defined in "Appendix 2". Similar, $r_B^{*,NP,2}$ can only be higher than $r_B^{*,P,2}$ iff $\tau > -10v_A/3 + v_B/3$. Again, this condition violates the feasible parameter range defined in "Appendix 2". Consequently, within the feasible parameter range $r_A^{*,NP,1} > r_A^{*,P,1}$, $r_A^{*,NP,2} > r_A^{*,P,2}$, and $r_B^{*,NP,2} < r_B^{*,P,2}$.

Next, we like to highlight the different cases that may occur with regard to CP's data consumption to provide further intuition for the utility functions stated in Sect. 3.

(case i) – *users cannot port their data* ($d = NP$). The derived utility at CP B can be calculated the same way as the derived utility for users that decided to use CP A in $t = 1$. Depending on r_B^2, more or less users are willing to switch to CP B. As we assume the market to be fully covered, users switch or become active at CP B if $U_B^{NP,2} > U_A^{NP,2}$. Please note that in this case, users need to re-enter the already revealed data because they have no possibility to port their data. From an analytical perspective, it is not relevant whether CP B requires more or less data than CP A. The decision which CP to patronize is only affected by the resulting utility which – of course – is influenced by the amount of required data set by the respective CP. The indifferent user can be derived by solving $v_B - \tau(1-x) - r_B^2 = v_A - \tau x - r_A^2$ with respect to x (c.f., Sect. 4).

(case ii)–*users can port their data* ($d = P$). The derived utility at CP B is now influenced by the amount of data CP A required from users in $t = 1$. Please note that due to the full market coverage assumption and in line with the assumptions highlighted in Sect. 3 as well as "Appendix 3", users (again) switch or become active at CP B if $U_B^{P,2} > U_A^{P,2}$. For users that have not been active at CP A in $t = 1$ (i.e., $U_A^1 < 0$), the utility function for users deciding to use CP B equals the one from the no portability case ($d = NP$, see above) because these users simply need to reveal all of the required data. For users that have been active at CP A in $t = 1$ and now switch to CP B, two sub-cases can be differentiated:

sub-case a) $r_B^2 \geq r_A^1$: Users have to reveal additional information if they become active at CP B. For example, users already revealed their name and address (r_A^1) but CP B requires their name, address and cellphone number (r_B^2). As users can port their data, they do not need to re-enter their name and address but need to (additionally) reveal their cellphone number which induces a disutility. This represents the most intuitive scenario. The resulting utility function for users that switch to CP B thus is $v_B - \tau \cdot (1-x) - r_B^2 + r_A^1$ and the indifferent user can be calculated by solving $v_B - \tau \cdot (1-x) - r_B^2 + r_A^1 = v_A - \tau \cdot x - r_A^2$ with respect to x (c.f., Sect. 4).

sub-case b) $r_B^2 < r_A^1$: Users need to reveal less data at CP B. Analytically, this case only occurs iff $v_B < v_A \wedge v_B > 15v_A/16$, i.e., $v_B \in [15v_A/16, v_A)$. In all other cases, either no feasible parameter range exists, or $r_B^2 \geq r_A^1$ (c.f., sub-case a). Consequently, in almost all cases considered in this paper, CP B requires at least the amount of data CP A required in period $t = 1$, which is why the examples and intuition provided focus on these cases. If CP B requires *less* data, users do not need to reveal additional data. Consequently, they do not

experience a disutility if they switch to CP B in $t = 2$, i.e., all data required at CP B is ported. We assume that the resulting utility function for users that switch to CP B (again) is $v_B - \tau \cdot (1 - x) - r_B^2 + r_A^1$ which is in line with the intuition of the disutility a user derives from revealing data being some kind of privacy costs (c.f., Sect. 1). Consequently, the user derives a net benefit from porting data because (1) the new service offered by CP B requires less data that does not need to be re-entered and (2) the data already provided to CP A is deleted at that CP because the European General Data Protection Regulation also encompasses a right to erasure (c.f., European Commission 2016b, Article 17), i.e., in the end, less data is disclosed to online services. The indifferent user can thus be calculated by solving $v_B - \tau \cdot (1 - x) - r_B^2 + r_A^1 = v_A - \tau \cdot x - r_A^2$ with respect to x (c.f., Sect. 4).

Appendix 6: CPs' Profits (π_i^d)

With data portability $(d = P)$, the CPs' profits are:

$$\pi_A^P = \frac{18\tau^2 + 12\tau(v_A - v_B) + 11v_A^2 - 4v_Av_B + 2v_B^2}{36\tau},$$

$$\pi_B^P = \frac{(3\tau - v_A + v_B)^2}{18\tau}.$$

Without data portability $(d = NP)$, the CPs' profits are:

$$\pi_A^{NP} = \frac{18\tau^2 + \tau(18v_A - 12v_B) + 13v_A^2 - 6v_Av_B + 2v_B^2}{34\tau},$$

$$\pi_B^{NP} = \frac{(16\tau - 9v_A + 6v_B)^2}{578\tau}.$$

To determine whether CPs are better off with data portability, we calculate the intersection of the CP's profit functions under the different regimes (i.e., π_i^P and π_i^{NP}). Although the profit functions intersect two times, both intersections are outside the feasible parameter range given by the restrictions specified in the "Appendix 2". Consequently, the effect of data portability on the incumbent and entrant is unambiguous. It can easily be shown that the incumbent (entrant) always suffers (benefits) from data portability, i.e., $\pi_A^P \leq \pi_A^{NP}$ and $\pi_B^P \geq \pi_B^{NP}$.

Appendix 7: Consumer's Surplus (CS_i^d)

With data portability $(d = P)$, consumer's surplus equals:

$$CS_A^P = \frac{-45\tau^2 + \tau(24v_A + 30v_B) + 22v_A^2 - 8v_Av_B + 5v_B^2}{72\tau},$$

$$CS_B^P = \frac{-45\tau^2 + \tau(12v_A + 6v_B) + 7v_A^2 - 4v_Av_B + 7v_B^2}{72\tau}.$$

Without data portability $(d = NP)$, consumer's surplus equals:

$$CS_A^{NP} = \frac{-1368\tau^2 + \tau(264v_A + 912v_B) + 763v_A^2 - 88v_Av_B + 152v_B^2}{2312\tau},$$

$$CS_B^{NP} = \frac{(16\tau - 9v_A + 6v_B)(80\tau - 45v_A + 38v_B)}{2312\tau}.$$

To determine whether users are better off with data portability, we calculate $CS_A^P + CS_B^P = CS_A^{NP} + CS_B^{NP}$ and reorder the result with respect to τ. This leads to two solutions labeled by τ_{CS} and $\tau_{CS,2}$. It can be shown that $\tau_{CS} := (174v_B - 822v_A + 17\sqrt{6658v_A^2 - 752v_Av_B + 16v_B^2})/726$ can be within the feasible parameter range specified in "Appendix 2", whereas $\tau_{CS,2} := (174v_B - 822v_A - 17\sqrt{6658v_A^2 - 752v_Av_B + 16v_B^2})/726$ is always outside of that feasible parameter range. Consequently, the effect of data portability on consumer's surplus is ambiguous and users may suffer from a right to data portability. Whereas the effect of data portability on consumer's surplus is positive if $\tau < \tau_{CS}$, the effect is negative if $\tau > \tau_{CS}$. Please note that τ_{CS} is not always within the feasible parameter range: if $v_B < 447v_A/160$, the intersection is always outside the feasible parameter range.

Appendix 8: Total Surplus (TS^d)

With data portability $(d = P)$, total surplus is:

$$TS^P = \frac{-18\tau^2 + 36\tau(v_A + v_B) + 55v_A^2 - 20v_Av_B + 10v_B^2}{72\tau}.$$

Without data portability $(d = NP)$, total surplus is:

$$TS^{NP} = \frac{-200\tau^2 + \tau(888v_A + 496v_B) + 783v_A^2 - 500v_Av_B + 178v_B^2}{1156\tau}.$$

All intersections of the functions are outside the feasible parameter range specified by the restrictions given in "Appendix 2". Consequently, the effect on total surplus is unambiguous. It can easily be shown that total surplus always increases with data portability, i.e., $TS^P > TS^{NP}$. Please note that this result assumes that total surplus is the unweighted sum of producer's and consumer's surplus.

Appendix 9: Fixed Costs for Data Portability (F)

By introducing fixed costs F for implementing a right to data portability, CP A's profits can be calculated by $\pi_{A,F}^{P} = \pi_{A}^{P} - F$ and CP B's profits by $\pi_{B,F}^{P} = \pi_{B}^{P} - F$, respectively. Note that the profit functions *without* a right to data portability ($d = NP$) remain unchanged because CPs do not face any additional costs if they do not have to implement such functionalities, i.e., $\pi_{A,F}^{NP} = \pi_{A}^{NP}$ and $\pi_{B,F}^{NP} = \pi_{B}^{NP}$.

We solve for the subgame perfect Nash equilibrium through backward induction. For the regime without a right to data portability, the results equal the results from the base scenario because $\pi_{B,F}^{NP} = \pi_{B}^{NP}$ (c.f., Sect. 4.1 as well as "Appendix 6"). For the regime with a right to data portability incorporating costs for the implementation, we get

$$r_{A,B}^{*,P,1} = \frac{v_A}{2},$$

$$r_{A,B}^{*,P,2} = \tau - \left(\frac{v_A + 2v_B}{6}\right),$$

$$r_{B,B}^{*,P,2} = \frac{v_B - v_A}{3} + \tau.$$

These results can be used to specify the CPs profits (c.f., Sect. 4.2).

Comparing the therewith deduced results, it can be seen that the entrant CP B now can be worse off with a right to data portability, if the fixed costs for the implementation exceed the critical threshold $\hat{F}$. This threshold can be calculated by solving $\pi_{B,F}^{P} = \pi_{B,F}^{NP}$ with respect to F, i.e., we solve

$$\pi_{B,F}^{P} = \pi_{B,F}^{NP},$$

$$\frac{9\tau^2 + (6v_B - 6v_A - 18F)\tau + (v_A - v_B)^2}{18\tau} = \frac{(16\tau - 9v_A + 6v_B)^2}{578\tau}.$$

with respect to F which specifies the critical threshold $\hat{F}$. It follows that the entrant CP B is worse off, if

$$F > \hat{F} := \frac{(10v_A - v_B + 3\tau) \cdot (35v_B - 44v_A + 99\tau)}{5205 \cdot \tau}.$$

Appendix 10: Porting Irrelevant Data (ID)

Assuming that users also port irrelevant data from CP A to CP B, a user's utility function changes to $U_{B,ID}^{d,2}(x)$ if they become active at CP B. Consequently, also the location of the indifferent user changes in period $t = 2$. Note that

CP A's utility function and the location of the indifferent user in $t = 1$ remains unchanged.

To calculate the indifferent user in $t = 2$, we (again) need to account for the different cases that may evolve. We stick to the assumption used in the base model. Thus, if users have the possibility to port their data ($d = P$ with subscript ID), the indifferent user in $t = 2$ can be calculated by solving $v_A - \tau \cdot x - r_A^2 = v_B - \tau \cdot (1 - x) - r_B^2 + \gamma \cdot r_A^1$. The indifferent user without data portability ($d = NP$) can (again) be calculated by solving $v_A - \tau \cdot x - r_A^2 = v_B - \tau \cdot (1 - x) - r_B^2 + r_A^1$. Consequently, the indifferent user in $t = 2$ is located at

$$x_{ID}^{*,d,2} = \begin{cases} -\dfrac{r_A^2 + \gamma \cdot r_A^1 - r_B^2 - \tau - v_A + v_B}{2\tau} & , \text{if } U_A^1(x_{ID}^{*,d,2}) \geq 0, d = P \\[2mm] -\dfrac{r_A^2 - r_B - \tau - v_A + v_B}{2\tau} & , \text{else}\,(U_A^1(x_{ID}^{*,d,2}) \geq 0; d = NP). \end{cases}$$

Based on the market shares given by the location of the indifferent user, the profits of the CPs can be specified. The total profits of CP A for both periods are given by

$$\pi_{A,ID}^{d} = \underbrace{x^{*,d,1} \cdot r_{A,ID}^{d,1}}_{\pi_{A,ID}^{d,1}} + \underbrace{x_{ID}^{*,d,2} \cdot (r_{A,ID}^{d,1} + r_{A,ID}^{d,2})}_{\pi_{A,ID}^{d,2}}$$

CP B, which is only active in $t = 2$, makes total profits of:

$$\pi_{B,ID}^{d} = (1 - x^{*,d,1}) \cdot r_{B,ID}^{d,2} + (x^{*,d,1} - x_{ID}^{*,d,2})$$
$$\cdot ((r_{B,ID}^{d,2} - \gamma \cdot r_{A,ID}^{d,1}) + \gamma \cdot r_{A,ID}^{d,1}),$$
$$\pi_{B,ID}^{d} = (1 - x_{ID}^{*,d,2}) \cdot r_{B,ID}^{d,2}.$$

Using the equilibrium amounts of required data stated above, we receive:

$$\pi_{A,ID}^{P} = \frac{(2\gamma - 13)v_A^2 + 6v_A(\tau - v_B/3)(\gamma - 3) - 18(\tau - v_B/3)^2}{2\tau(\gamma^2 - 2\gamma - 17)},$$

$$\pi_{B,ID}^{P} = \frac{2(\gamma^2\tau - \gamma(2\tau + 3v_A/2) - 8\tau + 9v_A/2 - 3v_B)^2}{\tau(\gamma^2 - 2\gamma - 17)^2},$$

$$\pi_{A,ID}^{NP} = \pi_A^{NP},$$

$$\pi_{B,ID}^{NP} = \pi_B^{NP}.$$

For consumer's surplus, we receive:

$$CS_{A,ID}^P = \frac{1}{8\left((\gamma^2 - 2\gamma - 17)^2\tau\right)}.$$
$$\left(4\gamma^4 v_A^2 - (24(\tau + v_A/2 - v_B/3))v_A\gamma^3\right.$$
$$+ (-113v_A^2 + (60\tau - 20v_B)v_A$$
$$+ 36(\tau - v_B/3)^2\left) * \gamma^2\right.$$
$$+ \left(150v_A^2 + (564\tau - 188v_B)v_A - 288(\tau - v_B/3)^2\gamma\right)$$
$$\left. + 763v_A^2 + (264\tau - 88v_B)v_A - 1368(\tau - v_B/3)^2\right),$$

$$CS_{B,ID}^P = \frac{1}{8\left((\gamma^2 - 2\gamma - 17)^2\tau\right)}.$$
$$((-20\tau^2 + (24v_A + 8v_B)\tau + 8v_A(v_A - v_B))\gamma^4$$
$$+ (8\tau^2 + (-96v_A + 16v_B)\tau$$
$$- 116v_A^2 + 40v_Av_B - 8v_B^2)\gamma^3$$
$$+ (600\tau^2 + (84v_A - 288v_B)\tau + 295v_A^2$$
$$- 68v_Av_B + 16v_B^2)\gamma^2$$
$$+ (-928\tau^2 + (-1020v_A + 352v_B)\tau$$
$$+ 470v_A^2 + 252v_Av_B + 16v_B^2)\gamma$$
$$- (1280(\tau + (6v_B - 9v_A)/16)\cdot$$
$$(-9v_A/16 - 19v_B/40 + \tau)),$$
$$CS_{A,ID}^{NP} = CS_A^{NP},$$
$$CS_{B,ID}^{NP} = CS_B^{NP}.$$

Users can be worse off with a right to data portability if $CS_{A+B,ID}^P < CS_{A+B,ID}^{NP}$. This occurs if

$$\tau > \tau_{CS,ID} := \frac{1}{522\gamma^3 + 1380\gamma^2 - 17394\gamma + 28560}$$
$$\cdot (17\sqrt{1594}(((\gamma^2 - 4430\gamma/767 + 6962/797)v_A^2$$
$$- 164v_Bv_A/797(\gamma^2 - 183\gamma/41 + 231/41)$$
$$+ 8v_B^2/797(\gamma - 2)^2 * (1/797))(\gamma^2 - 2\gamma - 17)^{1/2}.$$
$$+ (436v_A + 106v_B)\gamma^3 + (-2322v_A + 732v_B)\gamma^2 + (7728v_A$$
$$- 4914v_B)\gamma - 20638v_A + 7208v_B).$$

Restricting the amount of data that can be ported dampens the effect of data portability on consumer's surplus. This may lead to users suffering less if the user's mismatch costs are low. However, restricting the amount of data that can be ported also dampens the effect of data portability on consumer's surplus if users benefit with a right to data portability. Consequently, compared to a scenario with full data portability ($\gamma = 1$), consumer's surplus with $\gamma \in (0, 1)$ is lower, i.e., $CS_{A+B,ID}^P < CS_{A+B}^P$, if

$$\tau < \tau_{ID,P} := \frac{1}{30\gamma^3 + 126\gamma^2 - 882\gamma + 726}.$$
$$(-\sqrt{1690}((\gamma^2 - 2\gamma - 17)^2((\gamma^2$$
$$- 3526\gamma/845 + 3329/845)v_A^2$$
$$- 1/169(32(\gamma - 1)(\gamma - 47/20)v_Bv_A)$$
$$+ 1/845(8v_B^2(\gamma - 1)^2))^{1/2}$$
$$+ (30v_A + 6v_B)\gamma^3 + (-126v_A + 54v_B)\gamma^2$$
$$+ (270v_A - 234v_B)\gamma$$
$$- 822v_A + 174v_B).$$

Total surplus can be calculated according to the formula given in Sect. 4.4.

Appendix 11: Diminishing Value of Collected Data (DV)

Assuming that the data collected in $t = 1$ is not equally important in period $t = 2$ does not change the user's utility function or the entrant's profit function. However, the incumbent CP A's profit function changes as highlighted in Extension 5.3. This leads to a new equilibrium data consumption and subsequently, to diverting profits, consumer surplus, and total profits.

Using the equilibrium amounts of required data stated in Extension 5.3, we receive:

$$\pi_{A,DV}^P = \frac{(-2\rho - 9)v_A^2 - (6(\rho + 1))(\tau - (1/3)v_B)v_A - 18(\tau - (1/3)v_B)^2}{2(\rho^2 - 2\rho - 17)\tau},$$
$$\pi_{B,DV}^P = \frac{2(\tau\rho^2 + (-2\tau + 3v_A(1/2))\rho - 8\tau + 3v_A(1/2) - 3v_B)^2}{2(\rho^2 - 2\rho - 17))\tau},$$

$$\pi_{A,DV}^{NP} = \frac{(-2\rho - 11)v_A^2 - (6(\tau - (1/3)v_B))(\rho + 2)v_A - 18(\tau - (1/3)v_B)^2}{2(\rho^2 - 18)^2\tau},$$
$$\pi_{B,DV}^{NP} = -\frac{(2\tau\rho^2 + 3v_A\rho - 18\tau + 6v_A - 6v_B)^2}{2(\rho^2 - 18)^2\tau}.$$

For consumer's surplus, we receive:

$$CS_{A,DV}^P = \frac{1}{8(\rho^2 - 2\rho - 17)^2 \tau} \cdot$$
$$(4v_A^2 \rho^4 + 60 v_A(\tau - 2v_A(1/15) - (1/3)v_B)\rho^3$$
$$+ (7v_A^2 + (-48\tau + 16v_B)v_A$$
$$+ 252(\tau - (1/3)v_B)^2)\rho^2$$
$$+ (306 v_A^2 + (-48\tau + 16v_B)v_A$$
$$- 288(\tau - (1/3)v_B)^2)\rho + 483 v_A^2$$
$$+ (900\tau - 300 v_B)v_A$$
$$- 1584(\tau - (1/3)v_B)^2),$$

$$CS_{B,DV}^P = \frac{1}{8(\rho^2 - 2\rho - 17)^2 \tau} \cdot ((-20\tau^2 + 8\tau v_B)\rho^4$$
$$+ (80\tau^2 + (-84 v_A - 32 v_B)\tau$$
$$- 8v_A^2 + 20 v_B v_A)\rho^3$$
$$+ (168\tau^2 + 48 v_A \tau - 113 v_A^2$$
$$- 8v_A v_B - 32 v_B^2)\rho^2 + (-712\tau^2$$
$$+ (552 v_A + 208 v_B)\tau$$
$$- 62 v_A^2 - 32 v_A v_B + 40 v_B^2)\rho - 1136\tau^2$$
$$+ (-84 v_A + 32 v_B)\tau$$
$$+ 435 v_A^2 + 164 v_B v_A + 244 v_B^2),$$

$$CS_{A,DV}^{NP} = \frac{1}{8(\rho^2 - 18)^2 \tau} \cdot (4 v_A^2 \rho^4$$
$$+ 60 v_A(\tau + 2 v_A(1/15) - (1/3)v_B)\rho^3$$
$$(-5 v_A^2 + (96\tau - 32 v_B)v_A + 252(\tau - (1/3)v_B)^2)\rho^2$$
$$- (108(\tau - 5 v_A(1/3) - (1/3)v_B))v_A \rho$$
$$- (1620(-2 v_A(1/3)$$
$$- (1/3)v_B + \tau))(8 v_A(1/15) - (1/3)v_B + \tau)),$$

$$CS_{B,DV}^{NP} = -\frac{1}{2(\rho^2 - 18)^2 \tau} \cdot (5(\tau\rho^2 - 9\tau + (3/2)v_A \rho$$
$$+ 3 v_A - 3 v_B) \cdot (\tau - 2 v_B(1/5))\rho^2 + 3 v_A \rho(1/2)$$
$$- 9\tau + 3 v_A + 21 v_B(1/5)).$$

Users can be worse off with a right to data portability if $CS_{A+B,DV}^P < CS_{A+B,DV}^{NP}$. This occurs if

$$\tau > \tau_{CS,DV} := \frac{1}{12\rho^4 + 36\rho^3 - 234\rho^2 - 1080\rho + 540} \cdot$$
$$(2(((\rho^6 - 2\rho^5 - 39\rho^4$$
$$- 34\rho^3 + 741\rho^2 + 1520\rho - 1045/2)v_A^2$$
$$+ 4 v_B(\rho 4 - 5\rho^3 - 41/2\rho^2 - 95/2\rho + 25)v_A$$
$$+ 16 v_B^2(\rho - 1/2)^2)(\rho^2 - 18)^2)^{1/2} - 2 v_A \rho^5$$
$$+ (-2 v_A + 4 v_B)\rho^4 + (3 v_A + 4 v_B)\rho^3 + (74 v_A - 74 v_B)\rho^2$$
$$+ (180 v_A - 216 v_B)\rho + 540 v_A + 108 v_B)).$$

Total surplus can be calculated according to the formula given in Sect. 4.4.

Appendix 12: Network Effects (NWE)

The Amount of Required Data

As highlighted in Sect. 5.4, with network effects, a user's utility function changes. Because the location of the indifferent user changes, the corresponding profits change yielding different equilibrium amounts of required data. For CP A, the equilibrium amount of required data in $t = 1$ equals:

$$r_{A,NWE}^{*,P,1} = r_A^{*,P,1} = \frac{v_A}{2},$$
$$r_{A,NWE}^{*,NP,1} = r_A^{*,NP,1} - \frac{3\omega}{17} = \frac{3\tau + 10 v_A - v_B - 3\omega}{17}.$$

and in $t = 2$:

$$r_{A,NWE}^{*,P,2} = \tau - \omega - \frac{v_A}{6} - \frac{v_B}{3},$$
$$r_{A,NWE}^{*,NP,2} = \frac{15(\tau - \omega) - v_A - 5 v_B}{17}.$$

For CP B, the equilibrium amount of required data (in $t = 2$) equals:

$$r_{B,NWE}^{*,P,2} = r_B^{*,P,2} - \omega = \tau - \omega - \frac{v_A - v_B}{3},$$
$$r_{B,NWE}^{*,NP,2} = r_B^{*,NP,2} - \frac{16\omega}{17} = \frac{16\tau - 9 v_A + 6 v_B - 16\omega}{17}.$$

CPs' Profits

The calculation of the CPs' profits incorporating network effects qualitatively remains unchanged compared to the base model (c.f., Sect. 3 for details). Using the location of the indifferent users (c.f., Extension 5.4) and the equilibrium amount of required data (c.f., Sect. 18.1), the CPs' profits with data portability $(d = P)$ and with network effects yield:

$$\pi_{A,NWE}^P = \frac{v_A^2}{4(\tau - \omega)} + \frac{(3(\tau - \omega) + v_A - v_B)^2}{18(\tau - \omega)},$$
$$\pi_{B,NWE}^P = \frac{(3(\tau - \omega) - v_A + v_B)^2}{18\tau - 18\omega}.$$

Without data portability $(d = NP)$ and with network effects:

$$\pi_{A,NWE}^{NP} = \frac{18(\omega^2 + \tau^2) + (-36\tau - 18v_A + 12v_B)\omega + 12\tau(18v_A - 12v_B) + 13v_A^2 - 6v_A v_B + 2v_B^2}{34\tau - 34\omega},$$

$$\pi_{B,NWE}^{NP} = \frac{(16(\tau - \omega) - 9v_A + 6v_B)^2}{578\tau - 578\omega}.$$

Consumer's Surplus

For consumer's surplus, we get:

$$CS_{A,NWE}^{P} = \frac{1}{72(\tau - \omega)^2} \cdot (54\omega^3 + (-153\tau + 18v_A + 36v_B)\omega^2$$
$$+ (144\tau^2 + (-42v_A - 66v_B)\tau$$
$$- 12v_A^2 + 6v_A v_B + 6v_B^2)\omega - 45\tau^3 + (24v_A + 30v_B)\tau^2$$
$$+ (22v_A^2 - 8v_A v_B - 5v_B^2)\tau),$$

$$CS_{B,NWE}^{P} = \frac{1}{72(\tau - \omega)^2} \cdot (-45\tau^3 + (12v_A + 6v_B + 144\omega)\tau^2$$
$$+ (-153\omega^2 + (-30v_A - 6v_B)\omega$$
$$+ 7v_A^2 + 4v_A v_B + 7v_B^2)\tau$$
$$- 6\omega(-9\omega^2 - 3\omega v_A + v_A^2 + v_A v_B + v_B^2)),$$

$$CS_{A,NWE}^{NP} = \frac{1}{2313(\tau - \omega)^2} \cdot (1728\omega^3 + (-4824\tau + 108v_A + 1152v_B)\omega^2$$
$$+ (4464\tau^2 + (-372v_A - 2064v_B)\tau$$
$$- 486v_A^2 + 36v_A v_B + 192v_B^2)\omega - 1368\tau^3 + (264v_A + 912v_B)\tau^2$$
$$+ (763v_A^2 - 88v_A v_B - 152v_B^2)\tau,$$

$$CS_{B,NWE}^{NP} = -\frac{160}{289(\tau - \omega)^2} \cdot \left(6\omega^2/5 + (-11\tau/5 + 27v_A/40 + 2v_B/5)\omega\right.$$
$$+ \tau(\tau - 9v_A/16 - 19v_B/40))$$
$$\cdot (-9v_A/16 + 3v_B/8 - \omega + \tau).$$

Users are worse off with a right to data portability if $CS_{A,P} + CS_{B,P} < CS_{A,NP} + CS_{B,NP}$. The resulting threshold can be calculated by solving $CS_{A,P} + CS_{B,P} = CS_{A,NP} + CS_{B,NP}$ with respect to τ. This yields $\tau_{CS,NWE}$.

Total Surplus

Total surplus can be calculated according to the formula given in Sect. 4.4.

Comparison to the Base Model

Compared to the base model *without* considering network effects, the incumbent can benefit in terms of profits from the existence of network effects due to a higher market share in the first period. Analytically, the incumbent's profit functions with and without the existence of network effects intersect within the feasible parameter range. If the user's mismatch costs are high, the incumbent realizes higher profits *with* network effects. Formally, if $\tau > \omega/2 + \sqrt{9\omega^2 + 22v_A^2 - 8v_A v_B + 4v_B^2}/6$ (with data portability) and

if $\tau > \omega/2 + \sqrt{9\omega^2 + 26v_A^2 - 12v_A v_B + 4v_B^2}/6$ (without data portability), the incumbent realizes higher profits if network effects are considered. Conversely, the entrant always realizes lower profits. Unsurprisingly, consumers are unambiguously better off if positive direct network effects are considered.

References

Anderson SA (2012) Advertising on the internet. In: Peitz M, Waldfogel J (eds) The Oxford handbook of the digital economy. Oxford University Press, New York, pp 355–390

Caminal R, Matutes C (1990) Endogenous switching costs in a duopoly model. Int J Ind Organ 8(3):353–373

Chen PY, Hitt LM (2002) Measuring switching costs and the determinants of customer retention in Internet-enabled businesses: a study of the online brokerage industry. Inf Syst Res 13(3):255–274

Choi JP, Kim BC (2010) Net neutrality and investment incentives. RAND J Econ 47(5):1145–1150

d'Aspremont C, Gabszewicz JJ, Thisse JF (1979) On Hotelling's "Stability in competition". Econometrica 47(5):1145–1150

Dou W (2004) Will internet users pay for online content? J Advert Res 44(4):349–359

Drozdiak N, Schechner S (2016) EU files additional formal charges against Google. http://www.wsj.com/articles/google-set-to-face-more-eu-antitrust-charges-1468479516. Accessed 29 Nov 2018

Easley R, Guo H, Kraemer J (2018) From network neutrality to data neutrality: a techno-economic framework and research agenda. Inf Syst Res Forthcom 29(2):253–272

European Commission (2016a) Online platforms. https://ec.europa.eu/digital-single-market/online-platforms-digital-single-market. Accessed 29 Nov 2018

European Commission (2016b) Regulation (EU) 2016/679 (...) 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). http://data.europa.eu/eli/reg/2016/679/oj. Accessed 29 Nov 2018

European Commission (2017) Communication from the commission () on the Mid-Term Review on the implementation of the Digital Single Market Strategy (SWD(2017) 155 final). http://ec.europa.eu/newsroom/document.cfm?doc_id=44527. Accessed 29 Nov 2018

Evans D (2009) The online advertising industry: economics, evolution, and privacy. J Econ Perspect 23(3):37–60

Facebook (2018) Downloading your info. https://www.facebook.com/help/131112897028467/. Accessed 29 Nov 2018

Farrell J, Klemperer P (2007) Coordination and lock-in: competition with switching costs and network effects. In: Armstrong M,

Porter RH (eds) Handbook of industrial organization. Elsevier, Amsterdam, pp 1967–2072

Farrell J, Saloner G (1992) Converters, compatibility, and the control of interfaces. J Ind Econ 40(1):9–35

Gehrig T, Stenbacka R (2004) Differentiation-induced switching costs and poaching. J Econ Manag Strategy 13(4):635–655

Google (2018) Download your data. https://support.google.com/accounts/answer/3024190. Accessed 29 Nov 2018

Graef I (2015) Mandating portability and interoperability in online social networks: regulatory and competition law issues in the European Union. Telecommun Policy 39(6):502–514

Hotelling H (1929) Stability in competition. Econ J 39(153):41–57

Irmen A, Thisse JF (1998) Competition in multi-characteristics spaces: hotelling was almost right. J Econ Theory 78(1):76–102

Katz M, Shapiro C (1985) Network externalities, competition, and compatibility. Am Econ Rev 75(3):424–440

Katz ML, Shapiro C (1994) Systems competition and network effects. J Econ Perspect 8(2):93–115

Klemperer P (1987a) Markets with consumer switching costs. Q J Econ 102(2):375–394

Klemperer P (1987b) The competitiveness of markets with switching costs. RAND J Econ 18(1):138–150

Klemperer P (1989) Price wars caused by switching costs. Rev Econ Stud 53(3):405–420

Klemperer P (1995) Competition when consumers have switching costs: an overview with applications to industrial organization, macroeconomics, and international trade. Rev Econ Stud 62(2):515–539

Kourandi F, Krämer J, Valletti T (2015) Net neutrality, exclusivity contracts, and internet fragmentation. Inf Syst Res 26(2):320–338

Krämer J, Schnurr D, Wohlfarth M (2018) Winners, losers, and Facebook: the role of social logins in the online advertising ecosystem. Manag Sci. https://doi.org/10.1287/mnsc.2017.3012

Macgillivray A, Shambaugh J (2016) Exploring data portability. https://obamawhitehouse.archives.gov/blog/2016/09/30/exploring-data-portability. Accessed 29 Nov 2018

Montes R, Sand-Zantman W, Valletti T (2018) The value of personal information in online markets with endogenous privacy. Manag Sci. https://doi.org/10.1287/mnsc.2017.2989

Obama B (2016) Executive order—steps to increase competition and better inform consumers and workers to support continued growth of the american economy. https://obamawhitehouse.archives.gov/the-press-office/2016/04/15/executive-order-steps-increase-competition-and-better-inform-consumers. Accessed 29 Nov 2018

Petcu D, Vasilakos AV (2014) Portability in clouds: approaches and research opportunities. Scalable Comput Pract Exp 15(3):251–270

Pollock R (2009) The control of porting in platform markets. J Econ Asymmetries 6(2):155–180

Ranabahu A, Sheth A (2010) Semantics centric solutions for application and data portability in cloud computing. In: Proceedings of the international conference on cloud computing technology and science (CloudCom), pp 234–241

Ray S, Kim SS, Morris JG (2012) Research note—online users' switching costs: their nature and formation. Inf Syst Res 23(1):197–213

Sun M (2012) How does the variance of product ratings matter? Manag Sci 58(4):696–707

Swire P, Lagos Y (2013) Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique. Md Law Rev 72(2):335–380

Valero J (2016) Tirole: Brussels must level the playing field for online platforms. http://www.euractiv.com/section/digital/news/tirole-brussels-must-level-the-playing-field-for-online-platforms. Accessed 29 Nov 2018

Vanberg AD, Ünver MB (2017) The right to data portability in the GDPR and EU competition law: odd couple or dynamic duo? Eur J Law Technol 8(1):1–22

Wohlfarth M (2017) Data portability on the internet: an economic analysis. In: Proceedings of the international conference on information systems (ICIS), Seoul. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1020&context=icis2017