Internet Regulation, Two-Sided Pricing, and Sponsored Data [*]

Bruno Jullien[†]   Wilfried Sand-Zantman[‡]

December 2017

Abstract

We consider zero-rating by Internet service providers. We analyze the implications of offering sponsored data plans that allow content providers to pay for traffic on behalf of their consumers. These plans boost consumption of high-value content and decrease the networks' incentives to exclude low-value content. The welfare effect of allowing this price discrimination depends on the proportion of content targeted and the value of contents. Our analysis is extended to various cases (one-sided pricing, competing network, heterogenous cost, paid contents).

1   Introduction

The pricing of traffic on the Internet is the subject of many controversies due to contrasting views on how Internet service providers (ISPs), which manage the physical network, should treat various types of content and on their relationship with content providers. In this paper, we discuss the role of traffic management methods that allow content providers to pay for the data used by their customers. Examples of such methods include agreements such as AT&T's

Sponsored Data plan, Verizon FreeBee Data, and other *zero-rating* programs, whereby the data generated by the consumers of a provider are not counted in the subscriber's own monthly data limits.[1] Zero-rating plans are offered not only in developed countries but also in developing countries, through partnerships between large application developers and mobile operators, such as the programs Free Basics by Facebook or Google Free Zone. These types of agreements have raised concerns among net neutrality advocates and are the object of an intense debate between Internet actors and regulators.[2] The issue is whether discriminatory agreements may be justified by efficiency considerations.

To see how this problem differs from standard discriminatory pricing problems, let us begin by adopting a global perspective on the optimal pricing strategy for consumer content. The Internet can be seen as a three-party business whereby content providers and consumers use the network to trade. This trade creates some costs (mostly for the network) and some benefits both for consumers and content providers, either directly, such as when consumers pay for usage, or indirectly, such as when the presence of consumers generates ad revenues. When the presence of these consumers results in either costs or benefits, it is natural for content providers and the network to adjust their prices to deter or foster consumer usage. However, two major factors impede such adjustment. First, many content providers are free and thus cannot use prices to induce the behavior desired from consumers. Second, content providers are likely to differ in terms not only of the cost they impose on the network but also the benefits they create. The combination of a "missing price" for content and heterogeneity implies that benefits and costs will not be internalized properly by consumers or the network, which compromises the efficient use of network capacity. In this work, we derive the network's optimal pricing strategy when content providers differ in the social value they generate and cannot directly affect consumer behavior via price. We explore how network tariffs that target both consumers and content providers can be designed to alleviate the misallocation problem of network capacity and promote efficient network use.

Our main contribution is to show how ISPs can use zero-rating in order to screen contents according to their willingness to pay for traffic and to highlight two effects related to the two-sided nature of the market.[3] First, as it is typical for two-sided markets, the ISP internalizes the consumer surplus which reduces the potential adverse effect on welfare of content discrimination. Moreover, the benefits from screening content providers is passed-

---

[1]See *https://developer.att.com/sponsored-data* for AT&T's plan, and *https://www.5screensmedia.com/enterprise* or *https://www.sandvine.com/solutions/subscriber-services/sponsored data.html* for platforms intermediating such services.

[2]Regulators have declared zero rating to be anti-competitive in Canada, Chile, India, Norway, the Netherlands, and Slovenia. See the OECD's *Digital Economy Outlook* for 2015.

[3]In this paper we abstract from potential anti-trust issues.

through to consumers to an extent that depends on demand elasticity and the intensity of competition between bottleneck ISPs. Second, as some content is free, zero-rating plans may substitute for missing prices by transmitting signals to consumers about the externalities their traffic induces, improving efficiency. Indeed, the content providers choosing to pay to participate to a zero-rating program are those deriving the highest value from increasing traffic. The data allowance proposed by the ISP may however increase or decrease the consumption of lower value contents compared with a situation where zero-rating is banned. While zero-rating plans improve welfare when they screen free content from paid content, this may not be the case when they are used to screen similar types of content. The welfare consequences then depend on the distribution of the value generated by the traffic of different types of contents.

The setup we consider is such that a network intermediates the traffic between content providers and consumers. Content providers receive a benefit proportional to traffic, such as advertising revenue or direct utility for the producer. This benefit is heterogenous, with high-benefit and low-benefit content providers, but is their private information. Furthermore, the total benefits depend also on the usage level chosen by consumers. Because consumer usage increases the network's total cost, a price can be charged to one or both of the parties involved in traffic generation. When only consumers are charged, we refer to one-sided pricing, whereas when both consumers and content providers are charged, we refer to two-sided pricing. We focus on the two-sided pricing case but demonstrate in the extension that our insights apply also to the one-sided case. To match current practices, we assume that consumers pay a tariff that depends on the data allowance they choose while the content producers are charged a non-discriminatory linear price for the traffic they generate.

We also allow the network to propose a "sponsored data" plan to the content providers whereby the traffic to their content provider is removed from the consumer data allowance. By relaxing the constraint on traffic, this influences the consumption choice and raises traffic and, therefore, the content providers' revenues. Given the network's option, each content provider must trade off the volume of consumption against the cost of traffic. The high-benefit content providers will sponsor consumption to generate higher advertising revenues, while low-benefit providers will still be constrained by the data allowance. Such a zero-rating plan thus enables the network to discriminate among various content types. These practices emerge naturally as a correction for allocative inefficiencies arising from the absence of some prices (here, the price of content).

Absent the sponsored data program for the content providers, the network may decide to exclude the low-value providers to extract more from the high-value providers. Allowing zero-rating alleviates this standard monopoly trade-off. Providers have the option to increase

consumption by paying for traffic, which allows the network to differentiate consumption by content type. Despite improving efficiency compared with uniform tariff, the menu of prices results in socially suboptimal consumption levels. Indeed, the network reduces consumption levels for contents that do not choose to pay for traffic in order to raise the relative value and thus the price of sponsored data.

When zero-rating is not allowed, the network cannot extract the rent from the high-benefit content without excluding the low-benefit content. It follows that there will be more exclusion under uniform tariff than under zero-rating. Absent exclusion with uniform price, the welfare consequences of banning this discrimination are ambiguous due to the opposite effects on the two types of content. We show that a ban on zero-rating reduces welfare if the proportion of low-benefit content is large, but it raises welfare if the value of low-benefit content is not too low.

We extend the analysis in several directions. First, we consider the case of one-sided pricing. Sponsored data generates higher incremental value for adopters of the plan than under two-sided pricing, whereas the consumption under data allowance decreases. The welfare analysis is similar, but there cannot be exclusion of any content in this case.

Second, we discuss the case of elastic demand and competing networks. We show that competing networks will choose to propose the same sponsored data program as in the benchmark case, as each network still wants to maximize the value generated per consumer. We also show that elastic consumer participation argues for less regulation because the network will pass on to consumers a larger share of the gains obtained from content providers.

Third, we discuss a more general model that accommodates content providers differing in terms of demand, advertising benefits and traffic intensity. We show that the analysis extends to this case and that what matters for the adoption of sponsored data is the content providers' benefit per unit of traffic generated.

Finally, we extend the analysis to the more general case in which content can be either free or paid. In this extended setting, the network has still some incentives to propose a sponsored data program. And this program allows to increase global efficiency.

Our work is first related to the literature on two-sided markets (see Caillaud and Jullien, 2003; Rochet and Tirole, 2003, 2006; Armstrong, 2006), in that we aim to characterize the optimal pricing for each side of the market, content providers and consumers. We combine the participation model of Armstrong (2006) and the usage model of Rochet and Tirole (2003). In our basic model, the total number of agents on the consumer side is fixed and their consumption is affected only by the price and number — or more precisely the types — of content providers in the market. On the content-provider side, profit depends not only on the network charges providers must pay but also on the number of consumers and the data

allowance. The number of content providers can vary, in part because low-benefit providers may be priced out of the market by network charges.

Some contributions in the field of telecommunications have studied the sender–receiver pricing structure. This literature emphasizes the importance of "call externalities" and, hence, the social benefits associated with using positive receiver prices (Jeon, Laffont and Tirole, 2004). Hermalin and Katz (2004) develop a related idea while focusing on how best to address uncertainty over the private value of exchanging messages, and the game — namely the choice to call or to wait for a call — induced by the tariff structure. In our paper, the structure of communication is different because it is the receiver (i.e., the consumer) who always initiates the communication. Another difference is that, in our setup, only the sender learns the true benefit of this communication.

The literature on the Internet price regulation has been driven by the debate over net neutrality and the optimal way to price content providers and consumers (see, e.g., Economides and Hermalin, 2012, and Greenstein, Peitz and Valletti, 2016). One point emerging from two-sided market models is that, although laissez-faire can be shown to result in inefficient pricing, the precise nature of an intervention that would foster efficiency is unclear (Economides and Tåg, 2012). While neglecting the investment question (on this point, see, e.g., Choi and Kim, 2010; Hermalin and Katz, 2009), we direct our attention to the efficient management of current resources when the benefit of consumption is uncertain. Several recent contributions discuss the screening of congestion sensitive content by means of prices and differentiated quality layers, a key aspect of the net neutrality debate (Krämer and Wiewiorra, 2012; Reggiani and Valletti, 2016; Choi, Jeon and Kim, 2015; Bourreau, Kourandi and Valletti, 2015). Peitz and Schuett (2016) analyze moral hazard in traffic generation using a model that incorporates congestion externalities. At last, Somogyi (2017) proposes an analysis of zero-rating with a particular focus on congestion but in a setting where some contents cannot have access to zero-rating offers and exogenous consumers' prices.

Our work departs from these papers by considering consumption usage and the allocative role of consumer prices. A specific contribution is to show when zero-rating induces screening among traffic-sensitive content types and enhances efficiency.

The remainder of the paper proceeds as follows. After describing the model in Section 2, we analyze in Section 3 the outcome under uniform pricing and under zero-rating, and we discuss the consequences of welfare consequence of banning zero-rating. In Section 4, we explore the extensions mentioned above. Section 5 concludes.

2 Model

2.1 Framework

We analyze the tariff charged for traffic by a network (an ISP, in the case of the Internet) to two sides of the market: a unit mass of consumers and a unit mass of content providers. In practice, some content is delivered freely to consumers while other content is paid for. To focus on our paper's novel aspects, we simplify the analysis by assuming in the core of the paper that all content is free and consider the possibility of paid content in the extension section. Moreover, contents are non-competing and consumers visit all content providers. Consumers are homogenous with utility function $u(q)$, strictly concave on $[0, \bar{q}]$ such that $u'(0) = \bar{p} > 0$ and $u'(\bar{q}) = 0$. Alternatively, we can consider the demand for each content when consumers bear a cost p per unit of content by $q = D(p)$, with $D(\cdot) = u'^{-1}(\cdot)$ decreasing, reaching a maximum at $D(0) = \bar{q}$ and such that $D(\bar{p}) = 0$.

Transactions between content providers and consumers generate costs and benefits in addition to the utility consumers derive from usage. More precisely, each unit of content generates a cost of traffic θ, born by the network, that we normalize to $\theta = 1$. This cost may be a short-run operating cost or the long-run network's costs of expending resources to maintain service quality. We assume that consumption is positive if the consumer bears the full cost of traffic, i.e. $1 < \bar{p}$. Furthermore, each unit of consumption generates a benefit $b > 0$ for the content providers, net of the cost (if any) of distributing the content. This benefit is diverse, including the advertising revenue[4] and other gains for the content providers such as private benefits for blogs and not-for-profit organizations, the value of consumer data and the leverage of the customer base on the capital market. Content providers are heterogenous and can be of either low-benefit (LB) type ℓ, with benefits b_ℓ, or high-benefit (HB) type h, that is with benefits b_h.[5] A content is of type h with probability λ and of type ℓ with probability $1 - \lambda$. We assume that if content providers had to pay the full cost of traffic only HB content would be profitable :

$$b_\ell < 1 < b_h.$$

As a consequence, LB content providers will not operate unless consumers pay for part of the traffic cost.

It is assumed that the network cannot distinguish between the different types of content consumed; in other words, it is not allowed to practice explicit discrimination. The network

---

[4]Advertising revenue increases with consumption if the consumer time devoted to a page is increasing in consumption or if advertising is tied to consumption in some other ways.

[5]In the extension section, we will allow consumer utility and cost of traffic to depend on the content provider type.

proposes to consumers a take-it-or-leave-it offer consisting of a data allowance T, that is, a maximal level of traffic consumption, at a price P.[6] On the other side of the market, we assume that the network charges a linear price $s \geq 0$ for the traffic generated by content, so that each content provider's cost of servicing a quantity q to a consumer is sq. We also allow the network to offer sponsored data programs to content providers. This takes the form of an option removing the consumers' traffic from their data allowance T,[7] in exchange for a payment S made by the content provider to the network.[8] As the sponsored consumption is free of charge, it is predictable and equal to the maximal demand $\bar{q}$ defined above.

We consider the following timing:

1. The network proposes the data allowance T and prices P and s, as well as S if zero-rating is allowed.

2. Each content provider decides whether to be active and may also choose the sponsored data option, if available.

3. Each consumer decides whether to subscribe and then determines how to allocate the data allowance among contents.

4. Traffic is observed, and payments are made to the network.

For the main part of the paper, we focus on the case of a monopoly network with inelastic participation by consumers. Let us denote by CS the gross consumer surplus from usage. It is given by $\text{CS} = \lambda u(q_h) + (1 - \lambda) u(q_\ell)$, where q_h and q_ℓ represent the consumption of HB and LB content, respectively. Because consumers are *ex ante* identical, the network can extract the full surplus by requiring a total payment P equal to the gross consumer surplus, i.e. $P = \text{CS}$. Accordingly, the network's objective fully internalizes the consumer surplus.

Let Π denote the difference between the network's total revenue derived from content providers and the total cost of supporting the traffic. We define the *network value* as $V = \text{CS} + \Pi$. Given that $P = \text{CS}$, the network's profit is equal to the network value. Therefore it is optimal for the network to choose values of T, s and S that maximize V which we assume from now on.[9]

---

[6] Assuming a single offer (T, P) is without loss of generality. As consumers are homogenous and anticipate perfectly their traffic, we could allow for a non-linear tariff $P(T)$ without altering the analysis.

[7] For an analysis of data caps by ISPs and mobile operators, see Economides and Hermalin (2015).

[8] Notice that due to consumers ex-ante homogeneity, the ISP would not introduce zero-rating unless it is paid by content providers.

[9] We will show that network behavior leads to V also being maximized with competition between networks or elastic demand, although V does not coincide with total network profits in these cases.

As a benchmark, let us consider the socially optimal allocation with full information on the content type. This scenario corresponds to the case of a regulated network maximizing social welfare and perfectly discriminating between content types. Content with benefit b generates social welfare of $u(q) + (b-1)q$. Then social welfare is maximal at q_t^{FB}, with $t = \ell, h$, solution of

$$u'\left(q_t^{FB}\right) = \max\left\{1 - b_t, 0\right\}, \quad t \in \{\ell, h\}.$$

When the content price for a type-t content is discriminatory and equal to $s_t = b_t$, the content provider receives zero surplus. Hence, for $s_t = b_t$, the network value for each type-t content $V_t = u(q_t) - q_t + b_t q_t$ is equal to the total welfare. This implies that a network maximizing V implements the social optimum under perfect price discrimination with respect to content.

3   Uniform tariff *vs* zero-rating offers

We now investigate the network's choice of tariffs, and the impact of regulating network's pricing strategy.

3.1   Uniform tariff

We first consider the case in which all traffic is included in the data allowance and all content providers face the same pricing conditions for accessing the network. Under uniform tariff a content provider's only decision is whether to participate or not. Because content providers do not charge consumers for the good or service they offer, their profits can only be generated via the benefit per unit of traffic b. The price s charged by the network to them cannot be passed through to consumers. Therefore, given a price s, a content provider of type t stays in the market if it anticipates a nonnegative profit — that is, if $s \leq b_t$ for $t = \ell, h$. In particular, if s lies between b_ℓ and b_h, then only the HB content providers participate in the market, while all content providers participate if s is below b_ℓ.

Let us denote by $M \in \{\lambda, 1\}$ the mass of active content providers. With a data allowance T, and the same utility for every content, consumers will consume the same quantity of each content $q = T/M$. For given content participation (so for given M), choosing a data allowance amounts to choosing the consumption q of each content. Consumer's net surplus is $Mu(q) - P$. As the network can capture the full consumer surplus, the consumption level q is chosen to maximize the network value $V = Mu(q) + M(s-1)q$. This leads to q such that $u'(q) = 1 - s$ if $s < 1$ or $q = \bar{q}$ si $s \geq 1$. Framed differently, if the network chooses a price $s \leq b_\ell$ that allows all types of contents to be profitable, then the data allowance is such

that $u'(T) = 1 - s$. On the other hand, if the price s is greater than 1 (and less than b_h), only HB content providers will be on the network and the data allowance given by $T = \lambda \bar{q}$.

Therefore, the network chooses s by comparing two possible prices for content: the maximal price $s = b_\ell$ that maintains full participation of the content and the maximal price $s = b_h$ that preserves the participation of only HB content providers. The consumption levels in these two cases are respectively q_ℓ^{FB}, such that $u'(q_\ell^{\mathrm{FB}}) = 1 - b_\ell$, and $\bar{q}$. These two strategies yield the following network values:

$$V_\ell^u = u\left(q_\ell^{\mathrm{FB}}\right) + (b_\ell - 1)\, q_\ell^{\mathrm{FB}} \quad \text{when} \ \ s = b_\ell$$
$$V_h^u = \lambda\left[u\left(\bar{q}\right) + (b_h - 1)\, \bar{q}\right] \quad \text{when} \ \ s = b_h.$$

When the network chooses a price s to make both types of content providers participate, a strictly positive profit must be left to HB content providers because one cannot differentiate between contents. The network will instead choose to increase its price, and thereby keep only the HB content, when the proportion of this content is large.

Proposition 1 *Under uniform pricing, the network excludes the LB content by choosing a high price $s^u = b_h$ if and only if $\lambda > \lambda^u$. The threshold λ^u is decreasing in b_h and increasing in b_ℓ.*

Proof. See Appendix. ∎

The profit that the network obtains per unit of content is thus equal to the total value generated by the *marginal* content provider. The induced consumption is efficient for the *marginal* content and suboptimal for *inframarginal* content. As a consequence there is excessive exclusion.

3.2   Zero-rating

When choosing price s, the network faces the standard monopoly trade-off between capturing the rent of HB content providers (with high s) and avoiding the exclusion of LB content providers (with low s). One way to alleviate this trade-off consists in relying on a zero-rating policy to implement more complex tariffs. Therefore, we allow the network to offer a sponsored data option. With this option, a content provider pays a fixed fee S in order to remove its traffic from consumers' data allowance.

With the sponsored data option, consumers can adjust their consumption for each type of content. If the consumption of one type of content is removed from the data allowance, consumers will consume $\bar{q}$ of this content. If the network succeeds at inducing the LB and

the HB content providers to choose different options, then it is immediate that the HB content providers choose to sponsor data. Therefore, the data allowance only concerns the LB content. The LB content consumption, denoted q_ℓ, is then given by $q_\ell = T/(1 - \lambda)$.

We define the pricing under zero-rating as a set of price $\{s, S\}$ and a data allowance level T such that

- the LB and HB content providers are willing to participate, and

- the LB content providers choose not to sponsor the traffic, while the HB content providers choose to sponsor the traffic and pay the additional fee S.

We assume that the network offers a zero-rating program only if this allows to strictly increase its profits. We now detail how this program should be optimally structured to maximize the networks' profit.

Under zero-rating, consumers are constrained by the data allowance for one type of content only and obtain the gross surplus $\lambda u(\bar{q}) + (1 - \lambda) u(q_\ell)$. As the network captures this surplus through the price P, the network value is then equal to

$$V = \lambda \left[u(\bar{q}) + (s - 1)\bar{q} + S\right] + (1 - \lambda) \left[u(q_\ell) + (s - 1)q_\ell\right].$$

The tariff induces participation of both HB and LB provider types as long as

$$b_\ell \geq s \text{ and } b_h \bar{q} \geq s\bar{q} + S. \tag{1}$$

The following incentive compatibility conditions ensure that each content provider chooses the tariff designed for its specific type:

$$\begin{aligned} (b_\ell - s)\, q_\ell &\geq (b_\ell - s)\, \bar{q} - S \\ (b_h - s)\, \bar{q} - S &\geq (b_h - s)\, q_\ell \end{aligned} \tag{2}$$

The first condition states that a LB content providers do not gain by paying S to boost the traffic to $\bar{q}$. To interpret the second equation, we remind that all contents are perceived as equivalent by consumers. Hence, if some HB content providers choose not to sponsor the traffic, each consumer allocates her consumption uniformly across all non-sponsored contents. As content providers are non-atomistic, a single HB content provider deviating by refusing to take the option would face the same consumption q_ℓ as LB content providers.

Sponsored tariffs are thus equivalently characterized by an allocation (q_ℓ, s, S) such that conditions (1) and (2) are satisfied. The network's program is then to maximize V under these constraints.

Let us first fix the data allowance T (hence q_ℓ) and discuss the choice of tariffs s and S. First, since it is optimal to raise content prices as long as they remain compatible with the constraints, the participation constraint of the LB content providers and the incentive compatibility constraint of the HB content providers will be binding, so

$$s = b_\ell \quad \text{and} \quad S = (b_h - b_\ell)(\bar{q} - q_\ell). \tag{3}$$

Under conditions (3), the remaining constraints are satisfied because $q_\ell \leq \bar{q}$. Hence, the optimal prices are given by (3) and we obtain the following reduced program:

$$\max_{q_\ell \geq 0} \lambda \left[u(\bar{q}) - (1 - b_h)\bar{q} - (b_h - b_\ell)q_\ell \right] + (1 - \lambda) \left[u(q_\ell) - (1 - b_\ell)q_\ell \right].$$

This expression leads directly to the following statement.

Proposition 2 *The network offers a sponsored data option if and only if $1 - b_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell) < \bar{p}$. In this case the tariffs and data allowance are given by*

$$u'(q_\ell^*) = 1 - b_\ell + \frac{\lambda}{1 - \lambda}(b_h - b_\ell); \ \ s^* = b_\ell, \ S^* = (b_h - b_\ell)(\bar{q} - q_\ell^*),$$

where $q_\ell^* = T^*/(1 - \lambda)$.

Otherwise, the network sets $s^* = b_h$ with $T^* \geq \lambda\bar{q}$, and doesn't offer the sponsored data option.

Proof. The solution of the reduced program is obtained at

$$
\begin{aligned}
u'(q_\ell) &= 1 - b_\ell + \frac{\lambda}{1 - \lambda}(b_h - b_\ell) \text{ if it is less than } \bar{p} \\
q_\ell &= 0 \text{ otherwise}
\end{aligned}
$$

When $q_\ell > 0$, the prices s and S are then given by condition (3).

When $q_\ell = 0$, the conditions lead to $S + s\bar{q} = b_h\bar{q}$ which induces the same level of profit as a single price $s = b_h$. Hence, the network doesn't offer a contingent plan and simply choose a data allowance for efficient consumption of HB content. ∎

The menu of tariffs proposed by the network plays two roles. It allows the network to screen the different types of content providers, and it leads to more efficient consumption. HB content providers prefer high consumption and choose the option that removes it from the consumer data allowance. Because the benefit b_h is larger than the cost of traffic, equal to 1, this induces an efficient consumption level of the HB content. The total price paid

11

by the HB content providers is strictly less than the total benefit $b_h \bar{q}$, because the network must forego some profit to induce the HB content providers to choose the sponsored data option. Whereas the price s is simply set to capture the LB content providers' benefits surplus, the data allowance chosen is affected by two factors. First, it reflects the net cost $1 - b_\ell$ of any unit of traffic. Second, relative to efficient consumption at this net cost, it is distorted downward in order to deter the HB content providers from choosing the normal offer. Indeed, reducing T makes the sponsored data option relatively more attractive and allows the network raising S.[10]

In this setting, the network may decide to exclude the LB content providers, by choosing $T = 0$ or alternatively $s = b_h$ without zero-rating. This is the case when the quantity q_ℓ resulting from the price characterized in Proposition 2 is negative — and thus when λ is large:

$$q_\ell = 0 \iff \lambda \geq \lambda^* = \frac{\bar{p} - (1 - b_\ell)}{\bar{p} - (1 - b_h)}. \tag{4}$$

In this case, the network does not need to rely on zero-rating and may simply offer a linear price to content providers and set a large data allowance. Thus, the sponsored data option is offered only for $\lambda < \lambda^*$.

Corollary 1 *The network proposes a sponsored data option when the proportion of HB content is not too large (i.e., when $\lambda < \lambda^*$) and excludes the LB content otherwise. There is less exclusion when this option is proposed than when it is not ($\lambda^u < \lambda^*$). The threshold λ^* for exclusion is decreasing in b_h and increasing in b_ℓ.*

Proof. See Appendix. ∎

3.3   Ban on zero-rating offers

Let us now consider the welfare consequences of banning zero-rating offers. This could come from a regulatory rule such as a strict interpretation of net neutrality. As there is always exclusion of the LB content when $\lambda > \lambda^*$, we focus on the case in which the fraction of HB content is not too large. We assume that the regulator seeks to maximize total welfare. In the case of zero-rating, this welfare is

$$W^* = \lambda \left[u \left(\bar{q} \right) + \left(b_h - 1 \right) \bar{q} \right] + \left(1 - \lambda \right) \left[u \left(q_\ell^* \right) + \left(b_\ell - 1 \right) q_\ell^* \right].$$

---

[10]This bears some similarities with the analysis of capacity choices by Choi and Kim (2010). In their paper, as in ours, the network reduces the value of the basic services to the website to raise the price of the premium service.

Since under zero-rating, the consumption of HB, $q_h^* = \bar{q}$, is efficient, the main objective of regulation is to increase consumption of LB content while avoiding any decrease in the consumption of HB content.

We have shown in Section 3 that there will be more exclusion without zero rating because the network cannot accommodate the LB content and simultaneously exploit its market power on the HB content. Clearly, regulation is not desirable if it leads to exclusion. When there is no exclusion, a ban on zero-rating induces the network to raise the level of LB consumption (to q_ℓ^{FB}) and to reduce the level of HB consumption (from $\bar{q}$ to q_ℓ^{FB}). Thus, the overall effect of a ban is ambiguous.

When there is no exclusion, social welfare under the ban is given by

$$W^u = u\left(q_\ell^{\mathrm{FB}}\right) + \lambda\left(b_h - 1\right) q_\ell^{\mathrm{FB}} + (1 - \lambda)\left(b_\ell - 1\right) q_\ell^{\mathrm{FB}},$$

which we compare with W^* defined above.

Proposition 3 *If zero-rating is banned, then*

- *when $\lambda^* > \lambda > \lambda^u$, total welfare decreases;*

- *when $\lambda < \lambda^u$, total welfare decreases if λ is small enough and increases if b_ℓ is close enough to 1 (i.e., b_ℓ close to the cost of traffic).*

Proof. See Appendix ∎

When $\lambda^* > \lambda > \lambda^u$, allowing zero-rating benefits both groups of content providers. The LB content providers are now active, whereas they were excluded before. The HB content providers also benefit from the introduction of the sponsored data option because, to preserve incentive compatibility, the network must leave them some rent — which is not the case when uniform tariff is exclusionary.

When $\lambda < \lambda^u$, there is no exclusion under either regime. The effect of a ban on zero-rating is then more ambiguous because on the one hand the consumption of HB content is too low but the consumption of LB content is efficient. For λ is small, the distortion of q_ℓ (or, equivalently, of T) under zero-rating is likewise small, so the former effect dominated and the ban decreases social welfare. However, when b_ℓ is close to 1 (for given λ), the distortion of q_h under uniform tariff is small, so that the latter effect dominates and the ban can be beneficial.

In short, the introduction of zero-rating changes the consumption for both types of content. When the price $s = b_\ell$ under uniform tariff is high, it reduces significantly the consumption of LB content with little effect on the consumption of HB content, and thus decreases

social welfare. When the price s is not affected too much (λ small), the main impact of zero-rating is to increases the consumption of HB content to allow the network to decrease the rent of the HB content providers, and this is socially profitable.

In addition, one could consider an alternative regulation consisting in simply capping the amount S that content providers must pay to benefit from the sponsored data program. By setting a cap, the regulator would limit the scope for distorting the LB consumption, which is good for welfare. But this may also deter the network from offering zero-rating, at the cost of either inducing a lower consumption of HB content or the exclusion of the LB content. Therefore, one can advocate the implementation of a reasonable cap, lower than the unregulated level, but not too tight, to make sure that zero-rating is still offered.

4 Extensions

4.1 One-sided pricing

Even if our framework allows the networks to make both sides pay, the current situation is more similar to one-sided pricing, with a zero price for content, so it is worth investigating the implications of zero-rating in this setting. Note that the zero-price rule for content can be seen as one of the possible interpretation of the net neutrality regulation (see, for example, Economides and Täg (2012)). Because zero-rating relates to consumers' data allowance and not directly to traffic, there is no inconsistency in allowing it in the context of a zero price being charged to content providers for traffic. Therefore, the regulator may or may not allow it.

Let us then assume that the price s is constrained to be equal to zero, implying in particular that all content providers participate. In the absence of sponsored data program, the consumption of any content is given by q_0^u that maximizes the network value $u\left(q_0^u\right) - q_0^u$, and the data allowance is set accordingly at $T = q_0^u$. This leads to

$$V_0^u = u\left(q_0^u\right) - q_0^u \text{ with } u'\left(q_0^u\right) = 1.$$

In this situation, the network cannot tax any content and thus imposes a small consumption to consumers.

Let us suppose now that zero-rating is allowed. This means that on top of choosing a data allowance T along with a zero price ($s = 0$) for content providers, the network offers content providers the option of sponsoring consumption at a fee S. In this case, the network's pricing program is the same as in Section 3 except that the constraint $s \leq b_\ell$ is replaced by the new constraint $s = 0$. The reasoning of Proposition 2 applies with this new constraint,

which leads to the following optimal strategy for the network:

$$u'\left(q_{0\ell}^*\right) = 1 + \frac{\lambda}{1-\lambda}b_h, \quad T_0^* = (1-\lambda)\,q_{0\ell}^*, \quad S_0^* = b_h\left(\bar{q} - q_{0\ell}^*\right) \qquad \text{if } 1 + \frac{\lambda}{1-\lambda}b_h < \bar{p};$$

$$\text{and} \qquad T_0^* = (1-\lambda)\,q_{0\ell}^* = 0, \qquad S_0^* = b_h\bar{q} \qquad \text{if } 1 + \frac{\lambda}{1-\lambda}b_h \geq \bar{p}.$$

To induce the HB content providers to choose the sponsored-data option with a positive price S rather than the free standard contract, it requires the network to distort the consumption of the LB content even more than the case of two-sided pricing $(q_{0\ell}^* < q_\ell^*)$. The network value under zero-rating is then

$$V_0^* = \lambda\left(u\left(\bar{q}\right) + (b_h - 1)\,\bar{q} - b_h q_{0\ell}^*\right) + (1-\lambda)\left(u\left(q_{0\ell}^*\right) - q_{0\ell}^*\right)$$

Because the network could offer a sponsored data option while maintaining the standard data allowance, there is always a way for the network to design such an offer and increase its profit. However, the welfare implications of allowing/banning zero-rating are shown to be ambiguous.

Proposition 4 *Under one-sided pricing, the network always proposes a sponsored data option if it is allowed. A ban on zero-rating reduces total welfare when λ is large. When λ is small, the ban reduces welfare if b_ℓ is small but increases welfare if b_ℓ is close to 1 and demand is elastic enough.*

Proof. See Appendix ∎

When a sponsored data option is proposed, the consumption of HB content is unchanged, but the price of the data allowance is higher, and thus, the consumption of LB content is reduced. The attractiveness of sponsored data for content providers is naturally reduced when the network is not allowed to charge content providers for traffic. Hence, the network must reduce the non-sponsored consumption still further to maintain the sponsored data option's value. From a welfare perspective, zero-rating is better than uniform tariff when λ is large because there is efficient consumption of the HB content. When λ is small, the situation is more complex since the LB content consumption is only marginally distorted with zero-rating. However, the cost of this distortion depends on the benefit of the LB content. When this benefit is low, the consumption distortion is not excessively value-destructive, and the benefit from efficient consumption of the HB content dominates. When this benefit is high and demand is elastic enough, the low revenue due to reduced level of consumption outweighs the benefit from efficient consumption of HB content. Thus, the results obtained under two-sided pricing extend to the case of one-sided pricing.

4.2 Elastic participation and competition between networks

In the main analysis, we considered the case of a monopoly network with inelastic subscription demand. We now show that introducing demand elasticity or competition at the network level does not affect the main conclusions of our work. For this purpose, let us describe in greater detail the participation decision of the consumers.

We consider a model with an initial unit mass of consumers, a unit mass of content providers and $I \geq 1$ networks (indexed by i). Content providers are further divided into a mass λ of type h and a complementary mass $1 - \lambda$ of type ℓ. The utility of each consumer subscribing to network i and choosing a consumption profile $\{q_{ih}, q_{i\ell}\}$ and data allowance T_i at a subscription fee P_i is given by[11]

$$\lambda u(q_{ih}) + (1 - \lambda) u(q_{i\ell}) - P_i + \tilde{\varepsilon}_i$$

where the idiosyncratic shock $\tilde{\varepsilon}_i$ is a random variable that represents consumers' heterogeneity with respect to the intrinsic taste for network i. We impose no restrictions on the distribution of preference shocks, but we implicitly assume that they do not convey any information about the utility derived from consuming content.[12]

The timing of the game is unchanged, and we assume that in stage 1 each network i simultaneously makes an offer (T_i, P_i, s_i, S_i). In this slightly modified setting, content providers may deliver their content to all networks — they then pay only a variable price — whereas consumers subscribe to a single network. Moreover, we assumed that when zero-rating occurs, the offer is a fee S_i per consumer (or per mass 1 of consumers) on the network.

Let N_i denote the mass of consumers subscribing to network i, and q_i be the consumption of each content under the data allowance. Each content provider takes N_i and q_i as given and chooses whether to pay $S_i N_i$ to boost each consumer consumption to $\bar{q}$ or not. Hence, the profit of each content provider using this network is given by

$$N_i \max \left\{ (b_t - s_i) q_i, (b_t - s_i) \bar{q} - S_i \right\}.$$

In this context, the participation of content providers in network i and their choice of tariffs, as well as individual-level consumption for a given contract, are the same as before. What differs is that consumers can now choose among networks.

To unify notation between the case of uniform tariff and the case of zero-rating, we denote

---

[11]If type-t content is not available on the network, we set $q_{it} = 0$.

[12]This modeling of competition can be seen as a simplified version of the "nested discrete choice" model of demand developed in Anderson and de Palma (1992).

by q_{it} the individual consumption of type-t content on network i and by t_{it} the per-consumer payment of a type-t content provider to network i. Hence, we have

$$q_{it} = T; \ t_{it} = s_i T \qquad \text{under uniform tariff and no exclusion}$$

$$q_{ih} = T/\lambda; q_{il} = 0, t_{ih} = s_i T/\lambda, t_{i\ell} = 0 \qquad \text{under uniform tariff and exclusion}$$

$$q_{ih} = \bar{q}; \ q_{i\ell} = T/(1-\lambda), \ t_{ih} = s_i \bar{q} + S_i, \ t_{i\ell} = s_i T/(1-\lambda) \qquad \text{under zero-rating}$$

The gross consumer surplus is given by

$$\text{CS}_i = \lambda u\left(q_{ih}\right) + (1-\lambda) u\left(q_{i\ell}\right).$$

A given consumer joining network i gains $\text{CS}_i + \varepsilon_i - P_i$. Because there are several networks, the mass of consumers subscribing to network i is given by

$$N_i = \Pr\left(\text{CS}_i - P_i + \varepsilon_i \geq \max\{0, \max_{j \neq i} \text{CS}_j - P_j + \varepsilon_j\}\right).$$

where the probability Pr is related to the random variable representing consumers' heterogeneity.

The total profit of network i is then

$$N_i\left[P_i + \lambda\left(t_{ih} - q_{ih}\right) + (1-\lambda)\left(t_{i\ell} - q_{i\ell}\right)\right].$$

For any given strategy of the other networks (denoted z_{-i}), let

$$\phi_i\left(R; z_{-i}\right) = \Pr\left(R \geq \max\{0, \max_{j \neq i} CS_j - P_j + \varepsilon_j\} - \varepsilon_i\right).$$

Now, we can write the profit of network i as

$$\phi_i\left(\text{CS}_i - P_i; z_{-i}\right)\left[P_i + \lambda\left(t_{ih} - q_{ih}\right) + (1-\lambda)\left(t_{i\ell} - q_{i\ell}\right)\right].$$

Under this formulation, it is easy to see that the networks' best pricing strategy always maximizes the network value per consumer.

Proposition 5 *In any equilibrium of the game with elastic subscription demand and I networks, each network chooses a tariff structure that maximizes its value per consumer: $V_i = \lambda\left(u\left(q_{ih}\right) + t_{ih} - q_{ih}\right) + (1-\lambda)\left(u\left(q_{i\ell}\right) + t_{i\ell} - q_{i\ell}\right)$. Hence, the optimal data allowance T and tariff structure (s, S) is not affected by demand elasticity or competition, only the hook-up fee P is.*

Proof. Let $R_i = \mathrm{CS}_i - P_i$ be the expected net consumer surplus and V_i be the network value defined above. The network's profit can be written as

$$\phi_i\left(R_i; z_{-i}\right)\left[V_i - R_i\right].$$

Note that V_i is independent of the subscription fee P_i and of the other networks' strategies z_{-i}. Indeed, V_i depends solely on T and (s, S) through their effect on quantities and revenue per user from content providers. Let us fix the participation $\phi_i\left(R_i; z_{-i}\right)$ by adjusting P_i to maintain R_i constant. Then, profit maximization for a given R_i implies that the network will always choose (T_i, s_i, S_i) to maximize V_i. ∎

The value V_i is solely dependent on usage, meaning that there is a natural ordering in the pricing strategy. First, the network maximizes the value that can be shared with consumers by setting adequate data allowance and usage prices. Then, the network decides how much surplus to keep and how much to leave to consumers. Whereas the surplus R_i left to consumers (and hence the subscription fee P_i) depends on the elasticity of demand and on competition between networks, the data allowance T_i and the prices (s_i, S_i) do not. It follows that the allocation derived in the main model with a monopoly network is also the equilibrium allocation in the case of many networks competing for consumers.

As far as welfare is concerned, if total demand is fixed (inelastic consumer participation), then introducing competition at the network level does not alter our results. However, when aggregate demand is elastic, competition may increase total participation in the market. We observe that, compared with the case of inelastic demand, regulation should be more favorable to zero-rating under competition.

Corollary 2 *Zero-rating is all the more welfare-enhancing compared with uniform tariff as aggregate demand becomes more elastic.*

Proof. See Appendix. ∎

4.3 Heterogenous cost and demand

We can introduce other dimensions of heterogeneity by allowing the content providers to differ in their cost to the network θ (normalized up to now to 1) and in the demand for their product. Suppose that, per unit of consumption, content of type h generates a traffic θ_h and a benefit a_h, while content of type ℓ generates a traffic cost θ_ℓ and a benefit a_ℓ. We still rank these two types of content providers according to their respective ratios of advertising

revenue to cost and continue to assume that

$$b_\ell = \frac{a_\ell}{\theta_\ell} < 1 < b_h = \frac{a_h}{\theta_h},\ {}^{13}$$

where b_t is the benefit per unit of traffic.

We also allow the demands to differ across content types by assuming that the utility derived from the consumption Q of type-t content is $u_t(Q) = \eta_t u(Q/\eta_t)$, with $\eta_t > 0$. Although this is not essential to the argument, we assume at last that the utility function u is quadratic and thus that the demand function D is linear. Thus, the demand function for a price $p \leq \bar{p}$ of type-t content is $Q_t = \eta_t D(p)$. Therefore, η_t is a direct measure of demand intensity of type-t content. To limit the number of cases to study, we also assume that $\theta_h \eta_h > \theta_\ell \eta_\ell$. This means that the HB contents also generate high costs for the network either because the demand is high or because the load is high.

The analysis under uniform tariff is then the same as before, adjusting for the new demand. In particular, the network must choose between two prices for the content providers, $s = b_\ell$ or $s = b_h$.

We first consider the case where $s = b_\ell$. The consumer allocates then his data allowance T across the two contents. He thus solves

$$\max \lambda \eta_h u\left(\frac{Q_h}{\eta_h}\right) + (1-\lambda)\eta_\ell u\left(\frac{Q_\ell}{\eta_\ell}\right)$$
$$s.t. \lambda \theta_h Q_h + (1-\lambda)\theta_\ell Q_\ell \leq T$$

Let us denote by r the Lagrange multiplier of the data allowance constraint. This multiplier is the shadow cost of traffic faced by consumers and consumption levels are given by[14]

$$Q_t = \eta_t D(r\theta_t),$$

where r solves

$$\lambda \theta_h \eta_h D(r\theta_h) + (1-\lambda)\theta_\ell \eta_\ell D(r\theta_\ell) = T.[15]$$

The value captured by the network is then

$$V_\ell = \lambda\left(\eta_h u(Q_h/\eta_h) + (b_\ell - 1)\theta_h Q_h\right) + (1-\lambda)\left(\eta_\ell u(Q_\ell/\eta_\ell) + (b_\ell - 1)\theta_\ell Q_\ell\right).$$

---

[13]The model is also compatible with a uniform benefit $a_\ell = a_h$ per unit of consumption across types and different costs associated with consumption, in which case type h is low-cost content.

[14]The consumer maximizes $\lambda \eta_h u\left(\frac{Q_h}{\eta_h}\right) + (1-\lambda)\eta_\ell u\left(\frac{Q_\ell}{\eta_\ell}\right) + r(T - \lambda \theta_h Q_h - (1-\lambda)\theta_\ell Q_\ell)$.

[15]Notice that there is a one-to-one relationship between the data allowance and the shadow value of traffic.

Using $Q_t = \eta_t D(r\theta_t)$, the data allowance maximizing the network value will be chosen such that

$$\frac{\partial V_\ell}{\partial r} = (r + b_\ell - 1)\left\{\lambda\eta_h\theta_h^2 D'(r\theta_h) + (1-\lambda)\eta_\ell\theta_\ell^2 D'(r\theta_\ell)\right\}.$$

Thus the network chooses the data allowance $T = \lambda\theta_\ell\eta_\ell D((1-b_\ell)\theta_\ell) + (1-\lambda)\theta_h\eta_h D((1-b_\ell)\theta_h)$ that induces the shadow cost $r = 1 - b_\ell$. We then obtain the network value

$$V_\ell^u = \lambda\left(\eta_h u\left(\frac{Q_h^u}{\eta_h}\right) + (b_\ell - 1)\theta_h Q_h^u\right) + (1-\lambda)\left(\eta_\ell u\left(\frac{Q_\ell^u}{\eta_\ell}\right) + (b_\ell - 1)\theta_\ell Q_\ell^u\right)$$

$$\text{where } Q_t^u = \eta_t D((1-b_\ell)\theta_t).$$

We now consider the second case in which $s = b_h$. Here, LB content is excluded and the same reasoning shows that it is optimal for the network to induce the efficient consumption $Q_h = \eta_h\bar{q}$ with $T \geq \lambda Q_h$. The network value is

$$V_h^u = \lambda\left(\eta_h u(\bar{q}) + (b_h - 1)\theta_h\eta_h\bar{q}\right)$$

In this case again, there is a critical value λ^u such that exclusion occurs if $\lambda > \lambda^u$.

Consider now the introduction of a sponsored data option at price S with data allowance T. Following the above reasoning, let us denote by r the shadow cost of traffic under data allowance, assuming that all HB content providers and only them choose to the sponsored data option. This means that the constraint to which this shadow cost is associated is now $((1-\lambda)\theta_\ell\eta_\ell D(r\theta_\ell) \leq T$. As each content provider is infinitesimal, his individual choice doesn't affect the shadow cost r. Therefore, a type-t content provider concerned by the data allowance would face demand $Q_t = \eta_t D(r\theta_t)$. As the zero-rated consumption is $\eta_t\bar{q}$, this implies that type-h content providers opt for sponsored data if

$$(b_h - s)\theta_h\eta_h\bar{q} - S \geq (b_h - s)\theta_h\eta_h D(r\theta_h),$$

while type-ℓ content providers don't opt for it if

$$(b_\ell - s)\theta_\ell\eta_\ell\bar{q} - S \leq (b_\ell - s)\theta_\ell\eta_\ell D(r\theta_\ell) = (b_\ell - s)\frac{T}{1-\lambda}.$$

Given these constraints and the participation constraint $s \leq b_\ell$, it is easy to see that a network targeting HB content providers with the sponsored data option sets $s = b_\ell$ and $S = (b_h - b_\ell)\theta_h\eta_h(\bar{q} - D(r\theta_h))$.

From the network's perspective, the two main benefits of sponsoring are that it enables the extraction of more rent from content providers and can induce more efficient levels of

consumption. This interaction between screening on one side and incentives on the other side is a distinguishing feature of zero-rating in this extended setting. The optimal sponsored data allowance tariff T is then obtained at the value that maximizes the new objective:

$$\max_{T,Q_h,Q_\ell} \lambda\left[\eta_h u\left(\bar{q}\right) + (b_\ell - 1)\theta_h\eta_h\bar{q} + (b_h - b_\ell)\,\theta_h\left(\eta_h\bar{q} - Q_h\right)\right] + (1-\lambda)\left[\eta_\ell u\left(\frac{Q_\ell}{\eta_\ell}\right) + (b_\ell - 1)\theta_\ell Q_\ell\right]$$

$$\text{s.t. } Q_t = \eta_t D\left(r\theta_t\right),\quad t = \ell, h;\qquad (1-\lambda)\,\theta_\ell Q_\ell = T.$$

The characterization of the solution can then be extended as described next.

Proposition 6 *Suppose that content types differ in terms of costs, benefits and demands. If $\bar{p} > \theta_h\left(1 - b_\ell\right)$, there exists $\hat{\lambda} < 1$ such that the following hold:*

a) *For $\lambda < \hat{\lambda}$, the network offers zero-rating to HB content providers with*

$$s^* = b_\ell, \ T^* = (1-\lambda)\,\theta_\ell D\left(r^*\theta_\ell\right),\quad S^* = (b_h - b_\ell)\,\theta_h\eta_h\left(\bar{q} - D\left(r^*\theta_h\right)\right)$$

$$\text{where } r^* = 1 - b_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell)\frac{\eta_h}{\eta_\ell}\left(\frac{\theta_h}{\theta_\ell}\right)^2.$$

b) *If $\lambda > \hat{\lambda}$ and $\theta_h > \theta_\ell$, the network offers zero-rating to HB content providers with $s^* = b_\ell$, $T^* = (1-\lambda)\theta_\ell\eta_\ell D(\bar{p}\theta_\ell/\theta_h) > 0$ and $S^* = b_h\theta_h\eta_h\bar{q}$.*

c) *If $\lambda > \hat{\lambda}$ and $\theta_\ell \geq \theta_h$, there is exclusion of the LB content and a uniform tariff $s^* = b_h$. If $\bar{p} \leq \theta_h\left(1 - b_\ell\right)$, then $T^* = (1-\lambda)\theta_\ell\eta_\ell D\left(\theta_\ell\left(1 - b_\ell\right)\right)$ and $S^* = b_h\theta_h\eta_h\bar{q}$.*

Proof. See Appendix. ■

The new feature is that there is no exclusion if at equal consumption, the LB content induces lower traffic than the HB content. In this case, one can find a (small) data allowance that may deter consumption of HB content but not of LB content. Thus, the network need not to foreclose the LB content to gain maximal profit from the HB content providers.

As before, exclusion may be achieved by way of a uniform tariff $s = b_h$. When the proportion of HB content is below $\hat{\lambda}$ or when $\theta_\ell < \theta_h$, the network accommodates the LB content via sponsoring by HB content providers. The consumption of the LB content is nevertheless sub-optimal.

A lesson from the result is that the distinction between a small benefit and a large benefit (b_ℓ and b_h) is the relevant one to determine which content provider will choose the sponsored data option. What matters is thus the return on traffic (the revenue for each unit of traffic). In particular, the size of demand affects the prices (s, S) and the data allowance but not which content providers will sponsor data.

21

Proposition 7 *Suppose that content types differ in terms of costs, benefits and demands. If zero-rating is banned, then*

- *when $\hat{\lambda} > \lambda > \lambda^u$ or when $\lambda > \hat{\lambda}$ and $\theta_h > \theta_\ell$, total welfare decreases;*

- *when $\lambda < \lambda^u$, total welfare decreases if λ is small enough and increases if b_h is close to 1.*

Proof. see Appendix. ∎

The welfare comparison is similar to the case in which only benefits are heterogenous. A new effect is that zero-rating yields a further social benefit in that it avoids complete exclusion of the LB content when the proportion of HB content is large.

Similarly, a zero-price regulation reduces welfare if b_h is high because in that case there will be insufficient (resp. excessive) consumption of the HB (resp. LB) content.

4.4 Paid content

4.4.1 Presentation

We now investigate the network's choice of tariffs when content providers can charge positive prices to consumers. For this, we assume that each content provider is a monopoly, but the argument extends easily to the case of competition.

Whether a content provider charges a positive price or not will depend on the net benefit it obtains from consumers. Clearly, the price of content decreases with the net benefit and the content will be free for large benefits. In order to discuss this issue in a tractable and insightful way, we assume in this part that the low-benefit type receives no benefit, $b_\ell = 0$. Under this assumption, low-benefit content providers always charge a positive price.

Let us consider a tariff (T, P, s) offered by the network and accepted by the consumers. A content provider of type t charges price p_t for each piece of content, identical for all content providers of the same type. Facing the tariff and these prices, a consumer chooses a consumption pattern that maximizes

$$\lambda \left(u \left(q_h \right) - p_h q_h \right) + \left(1 - \lambda \right) \left(u \left(q_\ell \right) - p_\ell q_\ell \right) - P,$$

under the constraint

$$T \geq \lambda q_h + \left(1 - \lambda \right) q_\ell.$$

This yields demands $q_t = D(p_t + r)$, with $t = \ell, h$, where r is the Lagrange multiplier of the data allowance constraint.

Since each content provider is infinitesimal, it will take this demand as given and in particular the shadow cost r. The profit of a content provider with benefit b and price p is then $D(p+r)(p-s+b)$. Let us define $p^m(c) = \arg\max_{p\geq 0} D(p)(p-c)$ and $q^m(c)$ as the monopoly price and quantity for a cost c and demand function $D(p)$. We allow c to be negative in which case the price $p^m(c)$ may be zero. We assume that the pass-through is less than 1, that is $p'^m(c) = dp^m/dc \in (0,1)$. Then, the content's price and consumption are given by

$$\begin{cases} p_t = \max\{p^m(r+s-b_t) - r, 0\} \\ q_t = \min\{q^m(r+s-b_t), D(r)\} \end{cases} \tag{5}$$

The low-benefit (here zero-benefit) content providers always choose a positive price $p^m(r+s) > r$. For the high-benefit content providers, because $p'^m(r+s-b_h) < 1$, there exists a non-decreasing cutoff value $\hat{r}(s)$ such that $p^m(r+s-b_h) > r$ if and only if $r < \hat{r}(s)$.[16] In other words, the high-benefit content will be free of charge when the data allowance is tight. Notice that $\hat{r}(s) = 0$ if b_h is large and s is small enough that $p^m(s - b_h) = 0$.

When the data allowance constraint is binding, the shadow cost $r > 0$ is the solution of

$$T = \lambda \min\{q^m(r+s-b_h), D(r)\} + (1-\lambda)q^m(r+s). \tag{6}$$

The content prices and consumption for a given binding data allowance T and price s are then the solutions to equations (5) and (6). For a non-binding data allowance, we have $r = 0$ and $T \geq \lambda q^m(s - b_h) + (1-\lambda)q^m(s)$. Notice that choosing (T, s) is equivalent to choosing (r, s).

Let us now consider the tariffs without zero-rating. As mentioned above, the network can set the subscription fee P to extract the total consumer surplus, which leads to

$$P = \lambda(u(q_h) - p_h q_h) + (1-\lambda)(u(q_\ell) - p_\ell q_\ell)$$

The network value is then

$$\begin{aligned} V^u &= \max_{s,r} \lambda(u(q_h) - p_h q_h + (s-1)q_h) + (1-\lambda)(u(q_\ell) - p_\ell q_\ell + (s-1)q_\ell) \\ &\quad \text{s.t. } (5) \end{aligned}$$

---

[16] Notice that $0 \geq \hat{r}'(s) \geq -1$

23

Because $\partial p_t / \partial s < 1$ and $\partial q_t / \partial s < 0$, the slope

$$\frac{\partial V}{\partial s} = \lambda \left(q_h \left(1 - \frac{\partial p_h}{\partial s} \right) + (r + s - 1) \frac{\partial q_h}{\partial s} \right) + (1 - \lambda) \left(q_\ell \left(1 - \frac{\partial p_\ell}{\partial s} \right) + (r + s - 1) \frac{\partial q_\ell}{\partial s} \right)$$

is positive for $r + s \leq 1$. Thus at the optimum, still denoted r^u and s^u, we have $r^u + s^u > 1$.

Moreover, it is optimal to set $r \leq p^m (r + s - b_h)$ because on the range where $p^m (r + s - b_h) < r$ and $r + s > 1$, we have

$$V = \lambda \left(u (q_h) + (s - 1) q_h \right) + (1 - \lambda) \left(u (q_\ell) - p_\ell q_\ell + (s - 1) q_\ell \right),$$

and, using the fact that $q_h = D(r)$ and $q_\ell = q^m (r + s)$,

$$
\begin{aligned}
\left. \frac{\partial V}{\partial r} \right|_{r+s=constant} &= \lambda \frac{\partial}{\partial q_h} \left(u (q_h) - u'(q_h) q_h + (r + s - 1) q_h \right) D'(r) \\
&= \lambda \left(-u''(q_h) q_h + r + s - 1 \right) D'(r) < 0.
\end{aligned}
$$

We define $w = r + s$ as the total cost of consumption for consumers, and denote by $w^u = r^u + s^u$. The above findings show that w^u is strictly larger than 1 and that content providers charge their monopoly price:

$$
\begin{cases}
p_t^u &= p^m (w^u - b_t) - r^u, \\
q_t &= q^m (w^u - b_t).
\end{cases}
$$

We now contrast two cases. In the first one all content is paid while in the second only the low-benefit content is paid whereas the high-benefit content is free.

4.4.2 All paid content

Suppose that b_h is small enough that the high-benefit content is a paid content at the optimal tariffs. This implies that $w = r + s$ is a function $\omega(T)$ of T only, independent of s (as long as the data allowance is binding, i.e. $s < \omega(T)$), solution of

$$T = \lambda q^m (\omega(T) - b_h) + (1 - \lambda) q^m (\omega(T)).$$

The network value is then

$$
\begin{aligned}
V^u &= \max_w \lambda \left(u (q_h) - p^m (w - b_h) q_h + (w - 1) q_h \right) + (1 - \lambda) \left(u (q_\ell) - p^m (w) q_\ell + (w - 1) q_\ell \right) \\
&\quad \text{s.t. } q_t = q^m (w - b_t), \quad t = h, \ell.
\end{aligned}
$$

Given a solution w^u, the corresponding data consumption is the solution T^u of $w^u = \omega(T^u)$. As the network value V depends only on data consumption T^u and not on r and s, the optimal allocation may be implemented in various ways. The network may impose $T = T^u$ along with $s = 0$, in which case content providers do not pay and data consumption is constrained, or it may charge $s = w^u$ with no constraint on data consumption $T > T^u$.

Suppose that the network decides to introduce zero-rating and screens the types. Let (T, s, S) be the offer and r the new shadow cost of the data allowance.

Then, a content provider prefers to pay S to sponsor data if $\max_p D(p + r)(p - s + b_t)$ is smaller than $\max_p D(p)(p - s + b_t) - S$, that is if

$$\int_0^r q^m(s + x - b_t)\, dx \geq S.$$

It is then immediate that the high-benefit content has the largest benefit to take the option. If the network screens the high-benefit content provider with sponsored data, the network value is then

$$
\begin{aligned}
V &= \lambda\big(u(q_h) - p^m(s - b_h)q_h + S + (s - 1)q_h\big) + (1 - \lambda)\big(u(q_\ell) - p^m(w - b_\ell)q_\ell + (w - 1)q_\ell\big), \\
q_h &= q^m(s - b_h);\ q_\ell = q^m(w);\ S = \int_0^r q^m(s + x - b_h)\, dx.
\end{aligned}
$$

The network offers zero-rating with $S > 0$ if and only if $r > 0$. In other words, zero-rating requires that the data allowance constrains consumption. Moreover setting $r = 0$ and $s = w^u$ would implement the value maximizing allocation with uniform prices. Starting from a situation where $r = 0$ and $s = w^u$, we have

$$\left.\frac{\partial V}{\partial r}\right|_{r+s=w^u} = -\lambda\left(-q_h + (w^u - 1)D'(p^m(w^u - b_h))\right)p'^m(w^u - b_h) > 0.$$

Thus, when all content is paid, the ISP always chooses zero-rating and therefore to constrain consumption $(r > 0)$.

4.4.3 Free and paid content

We now consider the case where $\hat{r}(s) = 0$ for all relevant prices s, which ensures that the high-benefit content is free.

We first look at the uniform tariff such that each firm must pay s for any unit of load. From the above analysis we know that if the network optimal strategy leads the high-benefit content to be free, then the network doesn't restrict the data consumption $(r^u = 0)$. We

have also established that the optimal price s^u is larger than 1. More precisely it is given by the solution of

$$\lambda \bar{q} + (1 - \lambda) \left\{ [1 - p'^m(s^u)]q^m(s^u) + (s^u - 1)p'^m(s^u) D'(p^m(s^u)) \right\} = 0$$

Notice that it must be the case that $b_h > s^u$. Otherwise the high-benefit content would choose to be paid. This case then requires that $b_h > 1$.

Consider now the case where the network uses the sponsored data option to screen content providers. For given shadow cost r, the paid content provider is willing to sponsor data if

$$\int_0^r q^m(s + x)\, dx \geq S.$$

while the free content provider is willing to sponsor data if

$$(b_h - s)(\bar{q} - D(r)) \geq S$$

Recall that by assumption $\hat{r}(s) = 0$ so it must be the case that $(b_h - s)D'(x) + D(x) \leq 0$ for $x \leq r$. This implies that

$$(b_h - s)(\bar{q} - D(r)) = -\int_0^r (b_h - s)D'(x) > \int_0^r D(x)\, dx > \int_0^r q^m(s + x)\, dx.$$

Hence, as in the case of paid content, the high-benefit content providers choose to sponsor data. The program then writes as

$$
\begin{aligned}
V &= \max_{r,s} \lambda \left(u(\bar{q}) + (b_h - 1)\bar{q} - (b_h - s)D(r) \right) + (1 - \lambda) V_p(r + s) \\
\text{where} \quad V_p(w) &= u(q^m(w)) - p^m(w)q^m(w)) + (w - 1)q^m(w)
\end{aligned}
$$

so the optimality condition for the price s, still denoted s^*, is:

$$\frac{\partial V}{\partial s} = 0 \Leftrightarrow (1 - \lambda)\frac{\partial V_p}{\partial w} + \lambda D(r) = 0, \tag{7}$$

while the optimality condition for the data allowance (captured by the shadow value r) is

$$\frac{\partial V}{\partial r} = (1 - \lambda)\frac{\partial V_p}{\partial w} - \lambda(b_h - s)D'(r) = 0 \tag{8}$$

Thus, the optimal r, denoted r^*, is now positive. Moreover, as $D(r^*) < \bar{q}$, the total cost of consumption under data allowance is such that $1 < s^* + r^* < s^u$.

Indeed, sponsored data allows the network to reduce the deviation profit of the h-type content providers by restricting the data allowance and adjusting s to preserve the total cost $r + s$ of consumption of paid content. As a consequence, the network raises consumption of paid content while still imposing a data constraint so as to generate a shadow cost solution of

$$\frac{\partial V}{\partial r} = 0 \quad \Rightarrow \quad -\lambda D(r) - \lambda(b_h - s)D'(r) = 0$$

$$\Rightarrow \quad b_h - s^* = -\frac{D(r^*)}{D'(r^*)}.$$

Therefore, the network still proposes a sponsored data program in this case. This results in the same consumption of free content and a higher consumption of paid content, hence more efficiency. Moreover the paid content producers obtain a larger profit than with uniform price. The implication for the profit of the free content providers is ambiguous as they face a reduced price s but they have to pay the fee S to obtain the level of consumption $\bar{q}$.[17]

5   Conclusions

We have shown how ISP can rely on zero-rating to screen contents with high willingness to pay for traffic. In terms of allocative efficiency, this allows to direct more traffic toward more valuable contents. This is particularly relevant when content is free as market prices do not allocate consumption according to values.

Due to the two-sided nature of the activity and monetization on the consumer side of the market, the ISP internalizes the consumer surplus in its decision. Thus, potential negative impact of discrimination on the content side, if any, must be balanced with welfare benefits on the consumer side. The extent to which this ultimately benefits consumers depends on the demand elasticity for a monopoly and on the intensity of competition between bottleneck ISPs competing for consumers.

In this paper we have focused on a particular form of zero-rating, sponsored data paid by content content providers. There are other forms of zero-rating plans that would be worth considering. Zero-rating plans that are voluntarily offered by the network can be analyzed as a form of marketing and product design, with the drawback that they may raise some competitive issues on the content side and congestion issues.

Even if our understanding of the network's cost in this paper includes the management of congestion, it may be interesting to consider congestion issues more directly. To get

---

[17]Their profit is $(b_h - s^u)\bar{q}$ without zero-rating and $(b_h - s^*)D(r^*)$ with.

an idea about the potential effect, we can suppose that consumers suffer from a disutility due to congestion. As the ISP internalizes the effects on consumer surplus it will account for the negative effect of congestion and reduce data allowance accordingly, and as if the ISP was bearing the cost of this congestion directly. The analysis of uniform tariff is then similar to the analysis in Section 3.1, provided that the marginal effect of congestion on consumers' utility is not too large. In general, introducing a sponsored data option results in higher levels of traffic. Due to increasing congestion, the program becomes less attractive for the ISP (which cannot control traffic under zero-rating). Adjusting to this lower benefit from zero-rating and lower rate of adoption, the welfare analysis would remain qualitatively similar.

To conclude, this paper did not address anti-trust issues that zero-rating may raise if some large content providers compete with smaller content providers. In this case, there is a risk that ISPs favor some content providers over others, by setting tariffs that are attractive only to large content providers. This may provide a motive for capping the price that may be charged for sponsoring data to reduce barriers to entry. Another related issue that should be explored concerns vertical integration by ISPs in the content market. ISPs may be tempted to exempt consumers from data restrictions when they consume their own integrated content, thereby placing non-integrated content at a disadvantage. This is clearly an alley of research for future work.

References

[1] Anderson, S., and A., de Palma, 1992. "Multiproduct Firms: A Nested Logit Approach", *The Journal of Industrial Economics*, 40 (3), 261-276.

[2] Armstrong, M., 2006. "Competition in Two-Sided Market", *The RAND Journal of Economics*, 37 (3), 668-691.

[3] Bourreau, M., Kourandi, F. and T. Valletti, 2015, "Net Neutrality with Competing Platforms", *Journal of Industrial Economics*, 63, 30-73.

[4] Caillaud, B., and B., Jullien, 2003. "Chicken &Egg: competition among intermediation service providers", *The RAND Journal of Economics*, 34(2), 309-328.

[5] Choi, J.P., and B.-C., Kim, 2010. "Net Neutrality and Investment Incentives", *The RAND Journal of Economics*, 41(3), 446–471.

[6] Choi, J.P., Jeon D.-S. and B.-C., Kim, 2015. "Asymmetric Neutrality Regulation and Innovation at the Edges: Fixed vs. Mobile Networks", mimeo, Toulouse School of Economics.

[7] Economides, N. and B., Hermalin, 2012. "The Economics of Network Neutrality", *The RAND Journal of Economics,* 43(4), 602-629.

[8] Economides, N. and B. Hermalin, 2015. "The strategic use of download limits by a monopoly platform", *The RAND Journal of Economics*, 46(2), 297-327.

[9] Economides N. and J. Tåg, 2012. "Net Neutrality on the Internet : A two-sided Market Analysis", *Information Economics and Policy*, 24, 91–104.

[10] Greenstein S., Peitz M. and T. Valletti, 2016. "Net Neutrality: A Fast Lane to Understanding the Trade-offs", *The Journal of Economic Perspectives*, 30(2), 127-149.

[11] Hermalin, B. and M.L. Katz, 2004. "Sender or Receiver: who should pay to exchange an electronic message", *The RAND Journal of Economics*, 35(3), 423-448.

[12] Hermalin, B. and M.L. Katz, 2009. "Information and the hold-up problem", *The RAND Journal of Economics*, 40, 215-248.

[13] Jeon, D.S., Laffont, J.J., and J. Tirole, 2004."On the Receiver-Pays Principle", *The RAND Journal of Economics*, 35, 85-110.

[14] Krämer, J. and L. Wiewiorra, 2012. "Net neutrality and congestion sensitive content providers: Implications for content variety, broadband investment, and regulation", *Information Systems Research* 23(4), 1303-1321.

[15] Peitz, M. and F. Schuett, 2016. "Net neutrality and the inflation of traffic", *International Journal of Industrial Organization,* 46, 16-62.

[16] Reggiani, C. and T. Valletti, 2016. "Net neutrality and innovation at the core and at the edges", *International Journal of Industrial Organization*, 45, 16-27.

[17] Rochet J.C. and J. Tirole, 2003. "Platform Competition in Two-Sided Markets", *Journal of the European Economic Association*, 1, 990-1029.

[18] Rochet, J.C., and J. Tirole, 2006. "Two-sided market: A progress Report", *The RAND Journal of Economics*, 37(3), 645-667.

[19] Somogyi, R., 2017. "The Economics of Zero-rating and Net Neutrality", working paper CORE.

[20] Strategy Analytics Reports, 2013. *Unlimited to Tiered Data Plans: User Preference and Management Needs.*

Appendix

Proof of Proposition 1

While V_ℓ^u does not depend on λ, V_h^u is linear and strictly increasing in λ. At $\lambda = 0$, we have $V_\ell^u > V_h^u = 0$ (because q_ℓ is positive when $s = s_\ell < 1$), whereas at $\lambda = 1$, we have $V_\ell^u < V_h^u$ (because $b_\ell < b_h$). This implies that $V_\ell^u < V_h^u$ for λ above a threshold λ^u with $0 < \lambda^u < 1$.

We define the surplus $U(p) = \max_{q \geq 0} u(q) - pq$. The threshold λ^u solves

$$\lambda^u = \frac{U(1 - b_\ell)}{U(0) + (b_h - 1)\bar{q}}$$

which is decreasing in b_h and increasing in b_ℓ. ∎

Proof of Corollary 1

The first and last claims are immediate from Proposition 2 and equation (4). Let us focus on the second claim, that is, $\lambda^* > \lambda^u$.

By the convexity of $U(.)$ and $U(\bar{p}) = 0$,

$$U(1 - b_\ell) < U(0)\frac{\bar{p} - (1 - b_\ell)}{\bar{p}}$$

hence

$$\lambda^u < \frac{U(0)}{U(0) + (b_h - 1)D(0)}\frac{\bar{p} - (1 - b_\ell)}{\bar{p}}$$

Using $U(0) < \bar{p}D(0)$ (from convexity of U), we have

$$\lambda^u < \frac{\bar{p}D(0)}{\bar{p}D(0) + (b_h - 1)D(0)}\frac{\bar{p} - (1 - b_\ell)}{\bar{p}} = \frac{\bar{p} - (1 - b_\ell)}{\bar{p} + (b_h - 1)} = \lambda^*.$$

∎

Proof of Proposition 3

If $\lambda^* > \lambda > \lambda^u$, it is immediate that zero rating is better than uniform tariff, as the consumption level of HB content is the same, and there is no consumption of LB content under uniform tariff.

We focus now on the case in which $\lambda < \lambda^u$. Under zero rating, expected social welfare is given by

$$
\begin{aligned}
W^* - W^u = \ & \lambda \left[u\left(\bar{q}\right) + \left(b_h - 1\right)\bar{q} - u\left(q_\ell^{\mathrm{FB}}\right) - \left(b_h - 1\right)q_\ell^{\mathrm{FB}}\right] \\
& - \left(1 - \lambda\right)\left[u\left(q_\ell^{\mathrm{FB}}\right) + \left(b_\ell - 1\right)q_\ell^{\mathrm{FB}} - u\left(q_\ell^*\right) - \left(b_\ell - 1\right)q_\ell^*\right].
\end{aligned}
$$

At $\lambda = 0$, we have $W^* = W^u$ and $q_\ell^* = q_\ell^{\mathrm{FB}}$. Using the first-order conditions above we have the following:

$$
\frac{\partial(W^* - W^u)}{\partial \lambda} \mid_{\lambda=0} = u\left(\bar{q}\right) + \left(b_h - 1\right)\bar{q} - \left[u\left(q_\ell^u\right) + \left(b_h - 1\right)q_\ell^u\right] > 0.
$$

Therefore, at least for small values of λ, sponsored pricing dominates uniform tariff.

Suppose finally that b_ℓ is close to 1; then, q_ℓ^{FB} is close to $\bar{q}$, while $q_\ell^* < q_\ell^{\mathrm{FB}}$. Hence, $W^* - W^u < 0$. ∎.

Proof of Proposition 4

The first part follows from the fact that the profit under zero-rating is larger than the profit with a sponsored data option at price $S = b_h(\bar{q} - q_0^u)$ and $T = (1 - \lambda)\, q_0^u < \bar{q}$, since

$$
\lambda \left[u\left(\bar{q}\right) + \left(b_h - 1\right)\bar{q} - b_h q_0^u\right] + \left(1 - \lambda\right)\left[u\left(q_0^u\right) - q_0^u\right] > u\left(q_0^u\right) - q_0^u. \tag{9}
$$

As far as the welfare effects are concerned, the differential between the welfare under sponsored pricing and strict zero-price regulation is given by

$$
\begin{aligned}
\Delta = \ & \lambda \left[u\left(\bar{q}\right) + \left(b_h - 1\right)\bar{q} - u\left(q_0^u\right) - \left(b_h - 1\right)q_0^u\right] \tag{10} \\
& - \left(1 - \lambda\right)\left[u\left(q_0^u\right) + \left(b_\ell - 1\right)q_0^u - u\left(q_{0\ell}^*\right) - \left(b_\ell - 1\right)q_{0\ell}^*\right],
\end{aligned}
$$

where $q_\ell^{\mathrm{FB}} > q_0^u = D\left(1\right) > q_{0\ell}^*$. This is positive at $\lambda = 1$, hence the second part of the proposition. The last part is more intricate since the welfare differential vanishes at $\lambda = 0$ because $q_{0\ell}^*$ converges to q_0^u. Note first that

$$
\frac{\partial q_{0\ell}^*}{\partial \lambda} = D'(1 + \frac{\lambda}{1-\lambda}b_h)\frac{b_h}{(1-\lambda)^2}
$$

The slope of the welfare differential is

$$
\frac{\partial \Delta}{\partial \lambda} = \left[u\left(\bar{q}\right) + \left(b_h - 1\right)\bar{q} - \left(b_h - b_\ell\right)q_0^u\right] - \left[u\left(q_{0\ell}^*\right) + \left(b_\ell - 1\right)q_{0\ell}^*\right] + \left(\lambda b_h + \left(1 - \lambda\right)b_\ell\right)D'\left(1 + \frac{\lambda}{1-\lambda}b_h\right)\frac{b_h}{(1-\lambda)^2}
$$

Moreover,

$$\frac{\partial \Delta}{\partial \lambda} \mid_{\lambda=0} = [u(\bar{q}) + (b_h - 1)\bar{q} - u(q_0^u) - (b_h - 1)q_0^u] + b_\ell b_h D'(1) \tag{11}$$

with the first term being positive and the second negative. The second term vanishes when $b_\ell = 0$ (while q_0^u and $q_{0\ell}^*$ are not affected by b_ℓ), meaning that the slope is positive in this case. Therefore, for λ close to 0, sponsored pricing dominates zero-price when the benefit of LB firms is small.

Note also that by concavity,

$$[u(\bar{q}) + (b_h - 1)\bar{q} - u(q_0^u) - (b_h - 1)q_0^u] < b_h(\bar{q} - q_0^u).$$

Hence, the slope $\frac{\partial \Delta}{\partial \lambda} \mid_{\lambda=0}$ is negative if

$$\frac{\bar{q} - q_0^u}{q_0^u} < b_\ell \left(-\frac{D'(1)}{D(1)}\right),$$

therefore when demand is elastic and the LB content benefit is large. ∎

Proof of Corollary 2

Consider a monopoly network with elastic demand $\phi(\text{CS} - P)$. Let V^* and W^* be the network value and welfare per consumer with zero-rating. Let $\bar{V}$ and $\bar{W}$ be the network value with uniform tariff.

With network value V, the network chooses the fee P by solving

$$\max_F \phi(R)(V - R)$$
$$R = \text{CS} - P$$

The first-order condition for P then yields

$$R + \frac{\phi(R)}{\phi'(R)} = V$$

which defines R as an increasing function of V. Thus, with elastic participation, the consumers' participation $\phi(R)$ is a function $N(V)$, increasing with V.

Hence participation N^* with zero-rating is higher than participation $\bar{N}$ under a uniform price. Whenever $N^*/\bar{N} > \bar{W}/W^*$, allowing zero-rating dominates. Whether this occurs depends on the elasticity of demand. Note that V^*, W^*, $\bar{V}$ and $\bar{W}$ are independent of ϕ; thus, the ratio $N^*/\bar{N}$ increases when demand becomes more elastic. Thus, allowing zero rating will dominate for very elastic demand.

Under competition the same reasoning holds but ϕ has to be interpreted as the *equilibrium residual demand* faced by the network. ∎

Proof of Proposition 6

We proceed in three steps. First, we derive the optimal contract when the h-type provider chooses the sponsored option. Then, we show that it is not possible to make the ℓ-type provider (and only this type) choose the sponsored option. Finally, we show that it is profitable to use this contract rather than to propose the sponsored option for all types of content providers.

The optimal contract when the h-type provider chooses the sponsored option can be written as

$$V^* = \max_{T,r} \left(\begin{array}{c} \lambda \eta_h \left[u(\bar{q}) + (b_h - 1)\theta_h \bar{q} - (b_h - b_\ell)\,\theta_h D\,(r\theta_h) \right] \\ + (1 - \lambda)\,\eta_\ell \left[u\,(D\,(r\theta_\ell)) + (b_\ell - 1)\theta_\ell D\,(r\theta_\ell) \right] \end{array} \right) \tag{12}$$
$$\text{s.t.} \quad (1 - \lambda)\,\theta_\ell \eta_\ell D\,(r\theta_\ell) = T$$

In the case of a linear demand function, we have $D(p) = \bar{q}\max(1 - p/\bar{p}, 0)$. Notice that the positivity constraint on T imposes that $0 \leq r \leq \bar{p}/\theta_\ell$.

First notice that if $\bar{p} \leq \theta_h (1 - b_\ell)$, then the network can implement and capture the first-best total surplus by setting $T = (1 - \lambda)\theta_\ell \eta_\ell D\,(\theta_\ell (1 - b_\ell))$, $s = b_\ell$ and $S = b_h \theta_h \eta_h \bar{q}$. Indeed faced with this data allowance, a consumer who has the choice between LB and HB contents would consume only LB content, as $u'\,(D\,(\theta_\ell (1 - b_\ell))) = \theta_\ell (1 - b_\ell) \geq \frac{\theta_\ell}{\theta_h} u'\,(0) = \frac{\theta_\ell}{\theta_h}\bar{p}$. Hence, the only possibility for an HB content provider to have a positive demand is to buy the sponsored data option.

We now focus on the case where $\bar{p} > \theta_h (1 - b_\ell)$.
Case 1: $\theta_h \leq \theta_\ell$.
If $r < \bar{p}/\theta_\ell$, the slope of the objective V is

$$\lambda \eta_h (b_h - b_\ell)\,\theta_h^2 \frac{\bar{q}}{\bar{p}} - (1 - \lambda)\,\eta_\ell \left[r + b_\ell - 1\right]\theta_\ell^2 \frac{\bar{q}}{\bar{p}}.$$

The solution of the program is

$$r = \hat{r}\,(\lambda) \equiv 1 - b_\ell + \frac{\lambda}{1 - \lambda}\,(b_h - b_\ell)\frac{\eta_h \theta_h^2}{\eta_\ell \theta_\ell^2}$$

as long as $\hat{r}\,(\lambda) \leq \bar{p}/\theta_\ell$, which occurs for $\lambda \leq \hat{\lambda}$ where $\hat{\lambda}$ is defined by $\hat{r}\left(\hat{\lambda}\right) = \bar{p}/\theta_\ell$.

For $\lambda > \hat{\lambda}$, there is exclusion of LB content with $T^* = 0$ (or equivalently $r = \bar{p}/\theta_\ell$) and $S = b_h\theta_h\eta_h\bar{q}$.

Case 2: $\theta_h > \theta_\ell$.

In this case, if $r < \bar{p}/\theta_h$, the slope of the objective V is still

$$\lambda\eta_h\left(b_h - b_\ell\right)\theta_h^2\frac{\bar{q}}{\bar{p}} - (1 - \lambda)\,\eta_\ell\left[r + b_\ell - 1\right]\theta_\ell^2\frac{\bar{q}}{\bar{p}}.$$

If $\bar{p}/\theta_h < r \leq \bar{p}/\theta_\ell$, we have $D(r\theta_h) = 0$ and the slope of V is $-(1 - \lambda)\eta_\ell[r + b_\ell - 1]\theta_\ell^2\bar{q}/\bar{p} < 0$, because $\bar{p} > \theta_h\left(1 - b_\ell\right)$.

The solution of the program is still

$$r = \hat{r}\left(\lambda\right)$$

as long as $\hat{r}\left(\lambda\right) \leq \bar{p}/\theta_h$, which occurs for $\lambda \leq \hat{\lambda}$ where $\hat{\lambda}$ now is defined by $\hat{r}\left(\hat{\lambda}\right) = \bar{p}/\theta_h$. For $\lambda > \hat{\lambda}$ the solution is $r = \bar{p}/\theta_h$, with $D(r\theta_h) = 0$, $S = b_h\theta_h\eta_h\bar{q}$ and $T = (1 - \lambda)\theta_\ell\eta_\ell D(\bar{p}\theta_\ell/\theta_h) > 0$.

Is there a reason why the ISP would want to induce the LB content providers (and only those types) to choose the sponsored data option? This would means that the shadow value r is linked to the constraint $\lambda\theta_h\eta_h Q_h \leq T$ or equivalently that $Q_h = D(r\theta_h)$. In this case, the ICs can be written as

$$
\begin{aligned}
(b_\ell - s)\theta_\ell\eta_\ell\bar{q} - S &\geq (b_\ell - s)\theta_\ell\eta_\ell D(r\theta_\ell) \\
(b_h - s)\theta_h\eta_h D(r\theta_h) &\geq (b_h - s)\theta_h\eta_h\bar{q} - S
\end{aligned}
$$

We also have the participation constraints, with in particular the fact that $s \leq b_\ell$. The ISP's objective is then

$$
\begin{aligned}
Max_{r,s} \quad &\lambda\left[\eta_h u(D(r\theta_h)) + (s - 1)\eta_h\theta_h D(r\theta_h)\right] \\
&+ (1 - \lambda)\left[\eta_\ell u(\bar{q}) + (s - 1)\eta_\ell\theta_\ell\bar{q} + (b_\ell - s)\eta_\ell\theta_\ell(\bar{q} - D(r\theta_\ell))\right]
\end{aligned}
$$

which can be transformed into

$$
\begin{aligned}
Max_{r,s} \quad &\lambda\left[\eta_h u(D(r\theta_h)) + (s - 1)\eta_h\theta_h D(r\theta_h)\right] \\
&+ (1 - \lambda)\left[\eta_\ell u(\bar{q}) - \eta_\ell\theta_\ell\bar{q} + b_\ell\eta_\ell\theta_\ell(\bar{q} - D(r\theta_\ell)) + s\eta_\ell\theta_\ell\bar{q}D(r\theta_\ell)\right]
\end{aligned}
$$

This means that the objective is increasing is s. As s cannot exceed b_ℓ, we find $s^* = b_\ell$ and $S = 0$. But in this case, the incentive constraint of the HB content providers is violated, as long as $r > 0$. So the two incentive constraints must be binding. Let us then write the ISP's profit in this case.

$$\lambda\left[\eta_h u(D(r\theta_h)) + (s-1)\eta_h\theta_h D(r\theta_h)\right]$$
$$+\quad (1-\lambda)\left[\eta_\ell u(\bar{q}) + (s-1)\eta_\ell\theta_\ell\bar{q} + (b_\ell - s)\eta_\ell\theta_\ell(\bar{q} - D(r\theta_\ell))\right]$$

As the LB contents incentive constraint is binding, the ISP could ask this content to choose the sponsored data option. The resulting value for the network would be

$$\lambda\left[\eta_h u(\bar{q}) + (s-1)\eta_h\theta_h\bar{q} + (b_h - s)\theta_h\eta_h[\bar{q} - D(r\theta_h)]\right]$$
$$+\quad (1-\lambda)\left[\eta_\ell u(\bar{q}) + (s-1)\eta_\ell\theta_\ell\bar{q} + (b_\ell - s)\eta_\ell\theta_\ell(\bar{q} - D(r\theta_\ell))\right]$$

Focusing on the profit generated by the HB contents, one can easily check that it is higher the latter case than in the former. So any contract that would induce sponsoring only by the LB is dominated for the ISP by an offer that induces the two types of content to choose the sponsored data program.

Let us therefore look at the case where both types sponsor the traffic, by paying a fixed amount S on top of the variable price s. With only one contract, there is no incentive constraint to satisfy, only participation constraints which writes as

$$S \le (b_h - s)\theta_h\eta_h\bar{q} \text{ and } S \le (b_\ell - s)\theta_\ell\eta_\ell\bar{q}$$

Assuming that $\theta_h\eta_h \ge \theta_\ell\eta_\ell$, only the second participation constraint may be binding. The ISP's program writes as

$$\lambda\left[\eta_h u(q) + (s-1)\theta_h\eta_h\bar{q} + S\right] + (1+\lambda)\left[\eta_\ell u(\bar{q}) + (s-1)\theta_\ell\eta_\ell\bar{q} + S\right]$$

so S will chosen at its maximum. The ISP's program can then be written as

$$\lambda\left[\eta_h u(q) + s(\theta_h\eta_h - \theta_\ell\eta_\ell)\bar{q} + b_\ell\theta_\ell\eta_\ell - \theta_h\eta_h\bar{q}\right] + (1+\lambda)\left[\eta_\ell u(\bar{q}) + (b_h - 1)\theta_\ell\eta_\ell\bar{q}\right]$$

The optimal solution is to set $s = b_\ell$ and therefore $S = 0$. But this is dominated by a uniform tariff at $s = b_\ell$ and $T = \lambda\theta_h\eta_h D\left(\theta_h\left(1 - b_\ell\right)\right) + (1-\lambda)\theta_\ell\eta_\ell D\left(\theta_\ell\left(1 - b_\ell\right)\right)$. So there is no reason to propose a contract in which the sponsored option is proposed to everyone.

∎

Proof of Proposition 7

The proof is the same as the Proof of Proposition 3, adjusting for the new dimensions of heterogeneity. The welfare differential is

$$
\begin{aligned}
W^* - W^u \;=\; & \lambda \eta_h \left[u\left(\bar{q}\right) + \left(b_h - 1\right) \theta_h \bar{q} - u\left(q_\ell^{\mathrm{FB}}\right) - \left(b_h - 1\right) \theta_h q_\ell^{\mathrm{FB}} \right] \\
& - \left(1 - \lambda\right) \eta_\ell \left[u\left(q_\ell^{\mathrm{FB}}\right) + \left(b_\ell - 1\right) \theta_\ell q_\ell^{\mathrm{FB}} - u\left(q_\ell^*\right) - \left(b_\ell - 1\right) \theta_\ell q_\ell^* \right].
\end{aligned}
$$

The difference arises when $\lambda > \hat{\lambda}$ and $\theta_h > \theta_\ell$ because there is no exclusion of LB content with sponsored pricing while there is without. Therefore, it is optimal to allow zero rating.
∎