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

Bruno Jullien† Wilfried Sand-Zantman‡

June 2012, revised January 2016

Abstract

We consider a network that intermediates traffic between the consumers and the providers of free content. Two-sided pricing of consumers and content providers allows the network to extract profit and transmit information on the social value of content. Profit-maximizing tariffs give content providers the option of sponsoring the consumer traffic. We show that a cost-oriented price cap on the charge to content providers improves social welfare; in contrast, banning discrimination or imposing zero price for content providers yields suboptimal outcomes when the content is sufficiently valuable.

1 Introduction

Pricing of traffic on the Internet is the subject of intense debate between contrasting views of how the physical network operators should treat various content and of the relationship between content providers and Internet service providers (ISP) which manage the network. Efficient traffic management requires that the Web's various actors internalize the costs or benefits they impose on the ecosystem. In this paper we discuss optimal network management

*This paper had circulated before under the title "Congestion Pricing and Net Neutrality". We thank Dominik Grafenhofer, Byung-Cheol Kim, Jan Krämer, Martin Peitz, Mike Riordan, David Salant, Florian Schuett, Yossi Spiegel and Tommaso Valletti for helpful discussions and comments, as well as participants at 2012 CRESSE Conference, CREST-LEI seminar, The Future of Internet Conference (MaCCI, Mannheim), ICT 2012 Conference (ParisTech), EARIE 2013 Conference, the 2nd London IO Day and Pontifica Univerdad de Chile seminar. We gratefully acknowledge Orange, in particular Marc Lebourges, for its intellectual and financial support under the IDEI/Orange convention. Sand-Zantman also acknowledges support from the Agence Nationale de la Recherche, grant ANR-12-BSH1-0009.

†Toulouse School of Economics (CNRS and IDEI). E-mail: bruno.jullien@tse-fr.eu

‡Toulouse School of Economics (GREMAQ and IDEI) and ESSEC Business School. Manufacture des Tabacs, 21 allées de Brienne, 31000 Toulouse. E-mail: wsandz@tse-fr.eu

methods — in particular with regard to how network usage should be priced — when direct communication about those costs and benefits is difficult.

To see how this problem differs from standard pricing problems, let us start by taking 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, as when consumers pay for usage, or indirectly, as when the presence of consumers generates ad revenues. When the presence of these consumers results in either costs or benefits, it is natural to suppose that content providers and the network could adjust their respective prices to deter or foster consumer usage. Yet, there are two major factors that impede such adjustment. First, many websites are free and so cannot use prices to induce the behavior desired of consumers. Second, websites are prone to differ in terms not only of the cost they impose on the network but also of the benefits they create. Moreover these benefits will probably not be assessed correctly either by consumers or the network, which compromises the efficient use of the network capacity. In this work, we derive the network's optimal pricing strategy when websites 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 ameliorate the misallocation problem of network capacity while yielding useful information to promoting efficient network use.

Toward this end, we model a network that 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, but do not charge a retail price for content. This benefit is heterogenous, with high-benefit and low-benefit content providers, but is private information of the content providers. Furthermore, the total benefits depend also on the usage level chosen by consumers. Because consumer usage increases the network's 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; when both consumers and content providers are charged, we refer to two-sided pricing. We assume throughout that the network can charge a hookup fee to consumers[1] and in most of the paper we assume that the network is a monopoly facing consumers with inelastic participation.

In the absence of regulation, the network chooses two-sided pricing and discriminates between various content types by charging higher prices to high-benefit content providers.

[1]In most countries, fixed and mobile network operators offer consumers a menu of tariffs with different traffic allowances (i.e. nonlinear tariffs). In this case we can interpret the traffic price as the implicit price of data in the package (as in our model consumers can anticipate the total volume of traffic).

Discrimination is possible only when the network can personalize each content type's consumption level by proposing different prices to consumers. The network will then allow content providers to choose from a menu of two-sided tariffs, that are then transparent to consumers. We refer to this practice as "sponsored pricing" because the price paid by consumers falls in response to a higher price paid by content providers. This practice is similar to the attempt of AT&T's Sponsored Data program to require that content providers pay for the data used by their customers. In this scenario the data generated by consumers of a provider that subscribes to the program are not counted in the subscriber's own monthly data limits. These *zero-rating* agreements have become common in mobile telecommunications especially for video streaming, online music services, and popular contents.[2] Zero rating has raised concerns of discrimination and is the object of an intense debate among Internet actors and regulators.[3]

We show that these practices emerge naturally as a correction for allocative inefficiencies arising from the absence of some prices (here the price of content).[4] Given the network's menu, each content provider must trade off the volume of consumption against the cost of traffic. In this setup, the high-benefit content providers will sponsor consumption while other providers will prefer to reduce their costs by letting consumers pay for traffic. The different prices that result allow to transmit a signal to consumers based on the information extracted from the choice made by content providers. The mechanism improves efficiency by facilitating the transmission of information between content providers and consumers.

Despite improving efficiency, the menu of prices results in socially suboptimal consumption levels. This outcome follows because the network is a bottleneck for consumers access and so may charge excessive prices to content providers. We shall discuss some forms of regulatory intervention. First we show that imposing a price cap at unit cost on the price charged to content providers always improves welfare. We then examine the effects of preventing discrimination — that is, of forcing the network to offer a single pair of tariffs with one part to be paid by consumers and the other by content providers. Such uniform pricing prevents the network from adjusting consumption levels to the type of content. Since rents cannot be extracted from the most valuable contents without excluding the low-benefit contents, it follows that there will be more exclusion under uniform pricing than in the case of sponsored pricing. We therefore show that a ban on sponsored pricing *reduces* welfare if

[2] This is the case not only in developed countries but also in developing countries where widely used applications (e.g. Google, Facebook, Twitter, Wikipedia) sponsor zero-rating plans.

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

[4] In this paper we abstract from anti-trust issue that may arise from discriminatory practices — in particular market foreclosure. See however our conclusion on this point.

the benefits that contents generate are large enough. In addition, we show that imposing a one-sided tariff with a sponsoring option is socially dominated by a cost-oriented price cap; the same conclusion holds for a one-sided tariff without sponsoring as long as that content providers derive large benefits from consumption.

We extend the analysis in two directions. First, we discuss the case of elastic demand and competing networks. We show that, as long as the heterogeneity between ISPs is not content related, competing networks will still choose to propose the sponsored pricing program derived in the benchmark case in order to maximize their profits. We also show that elastic consumers participation argues for less regulation. Indeed, under a laissez-faire regime, the network will pass on to consumers more of its gains from content providers, which should increase consumers participation.

Second, we discuss a more general model that accommodates content providers differing in terms of both the benefit and cost they impose to the network. This added dimension of heterogeneity is yet another motive for using sponsored pricing to transmit information, since it allows consumers to factor in their traffic cost when deciding on their usage level. Therefore, welfare is increased thanks to the signaling role played by the presence of differentiated offers.

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 at characterizing the optimal pricing for each side of the market, content providers and the 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 on the market. On the content provider side, profit depends not only on the network charges it must pay but also on the number of consumers and the price that they are charged by the network. And 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 telecommunication have studied the sender–receiver pricing structure. This literature emphasizes the importance of "call externalities" and hence of 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 deal with the uncertainty about the private value of exchanging messages and the gaming — 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). One point emerging from two-sided market models is that, even though 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. By focusing on the information revelation aspect of prices, we offer a new perspective that complements previous studies on the effects of price discrimination (cf. Hermalin and Katz, 2007). Several recent contributions discuss the screening of traffic-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, 2012; Choi, Jeon and Kim, 2013; Bourreau, Kourandi and Valletti, 2015). Peitz and Schuett (2015) analyze moral hazard in traffic generation using a model that incorporates congestion externalities. Our work departs from these papers by considering consumption usage and the informational role of consumer prices. Two specific contributions are (i) showing that screening among traffic-sensitive contents can be done via different consumer prices and a single quality layer and (ii) proving the optimality of a price cap on the content side.

The rest of the paper proceeds as follows. After describing the model in Section 2, in Section 3 we derive the prices maximizing the network's profit and discuss the consequence of laissez-faire pricing for consumers and content providers. Section 4 focuses on regulation by analyzing standard price cap regulations and no-discrimination clauses. In Section 5, we propose two extensions; the first considers elastic subscription demand and competition at the network level, and the second assumes asymmetric information on the benefits and also on the costs of communication. Section 6 concludes.

2 Model

We analyze the tariff charged for traffic by a network (an ISP, in the case of the Internet) to two sides of the market: consumers and content providers. In practice, some content is delivered freely while other content is paid for. In order to focus on our paper's novel aspects, we simplify the analysis by assuming that all content is free. We assume for conciseness that contents are non-rival so that consumers visit every content provider. The expected demand for each content when consumers face a price p per unit of content is $q = D(p)$. The representative consumer utility function $U(q)$ is strictly concave with $U'(0) = \bar{p} > 0$ and $U'(\bar{q}) = 0$. Thus the demand $D(\cdot)$ is decreasing, the consumption $D(0) = \bar{q}$ of free goods is positive, and demand vanishes at price $\bar{p}$.

Any transaction between a content provider and a consumer creates costs and benefits in addition to the utility that consumers derive from usage. More precisely, each unit of content creates a cost θ to the network. Hence the consumption of q units of content costs the network an amount θq. This cost is either direct or a function related to congestion. One interpretation is that θ reflects the network's costs of expending resources to maintain the service quality. We assume that consumption is positive if priced at its true marginal cost (i.e. $\theta < \bar{p}$). At the same time, each unit of consumption generates a benefit $a > 0$ for the content providers, net of the cost (if any) of distributing the content. This benefit is diverse, including the advertising revenue[5] and other gains of the content provider 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 type ℓ or high-benefit type h; these types are characterized by respective benefits a_ℓ and a_h where $a_\ell < a_h$. We assume that any given content is type h with probability λ and thus content is of type ℓ with probability $1 - \lambda$. The type of each content is unknown to the consumers and the network, but known to the content provider. We will refer to content providers of type h as high-benefit (in short HB) and to content providers of type ℓ as low-benefit (in short LB). We shall use $b_\ell = a_\ell/\theta$ and $b_h = a_h/\theta$ to denote the respective ratios of benefits to costs. To focus on the most interesting case, we assume that one type of content could not be profitable if its provider had to pay the full cost of traffic:

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

This assumption captures the idea that only some content providers can afford the network's cost. As a consequence, some content providers will not operate unless consumers pay for part of the traffic cost.[6]

Although content providers have some private information about their benefit, the level of consumption is determined by consumers. Moreover, the network observes the ex post realization of total cost θq and can charge any side for this cost. We restrict attention to linear traffic prices — that is prices of the form $s\theta q$ and $r\theta q$ to (respectively) content providers and consumers, where the unit prices s and r are both nonnegative. However, we will allow the network to offer content providers a choice between multiple pairs of prices

[5] 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 way.

[6] The key assumption is $b_\ell < 1$. Otherwise (as we shall demonstrate), total welfare would be maximized by imposition of a content price equal to 1.

(r, s). We use Π to denote the network's (variable) profit, which is defined as the difference between the revenues from traffic and the costs of supporting that traffic.

We assume also that the network can charge a hookup fee F ex ante to consumers for subscription. The extent to which this hookup fee allows the network to capture any increase in consumer surplus due to traffic management depends on various factors, two of which include the elasticity of participation for consumers and the competition at the network level. For the most part, we focus on the case of inelastic participation, reserving the discussing of more general settings for later in the paper.

We will consider the timing as follows.

1. The network proposes the prices r and s as well as a hookup fee F.
2. Each consumer decides whether or not to subscribe.
3. Each content provider learns its type, that is the benefit that consumer's usage can generate, whether or not to be active, and may also choose a tariff.
4. Consumers observe the choice of each content provider and decide how much to consume of each provider's content.
5. Traffic is observed and payments are made to the network.

Let us denote by CS the expected consumer surplus from usage (gross of the hookup fee). Because consumers are ex ante identical and risk neutral, their subscription decision is based only on $\mathrm{CS} - F$. In this case, the network can extract the full expected surplus by setting the hookup fee equal to the consumer surplus ($F = \mathrm{CS}$) or at least can use this fee to reap part of that surplus. It is therefore optimal for the network to maximize the joint expected surplus with consumers. Accordingly, the network's objective fully internalizes the surplus of consumers. Hence we ignore the fee F and assume hereafter that the network maximizes the sum of its variable profit Π and expected consumer surplus; thus the *network value* is $V = \mathrm{CS} + \Pi$. [7] As we will show, network behavior will lead to V being maximized — even with competition between networks or elastic demand — whenever consumers do *not* have private information ex ante about their expected surplus from joining the network.

As a benchmark, we study the socially optimal prices in the case of full information on the content type and no direct transfer between content providers and consumers. This scenario corresponds to the case of a regulated network maximizing social welfare. For content of

[7] We use the term network value because, although it coincides with the network total profit in our basic model, this is not the case anymore when we introduce elastic demand and competition.

benefit a, the price perceived by consumers is $r\theta$ and so consumption is $q = D(r\theta)$. The content generates a monetary gain $a - s\theta = (b - s)\,\theta$ per unit of consumption, which means that it is offered only if $s \leq b$. Social welfare then can be written as

$$U(q) + (b\theta - \theta)q \quad \text{s.t. } s \leq b,\, U'(q) = r\theta.$$

If we ignore the feasibility constraint $r \geq 0$, then social welfare is maximal at $r = 1 - b$ and $s \leq b$. The difficulty with this solution is that optimal prices can be negative if the benefit is high enough. Once negative prices are ruled out, the optimal prices are described as follows.

Lemma 1 *Under full information, the socially (constrained) optimal allocation is obtained by charging* $r = \max\{1 - b, 0\}$ *and* $s \leq b$.

When the content price is $s = b$, content providers receive zero surplus. Hence for $s = b$, the network value $V = \text{CS} + \Pi$ is equal to the maximal total welfare. This implies that a network maximizing V implements the social optimum under full information about a.

3 Network pricing under laissez-faire

We now investigate the network's choice of tariffs under laissez-faire. Although consumer characteristics are commonly known, there is uncertainty about the content providers' benefit from any particular transaction made on the network. The network may thus seek to extract that benefit by discriminating between the two types of content. This possibility leads us to consider two price strategies for the network: uniform pricing and sponsored pricing.

Uniform pricing: The network offers a unique pair (s, r).

Under uniform pricing, a content provider's only decision is whether or not to participate and — conditionally on participation — all face the same price. Because content providers do not charge consumers for the good or service they offer, their profits can only be generated via the benefit a. The price s charged by the network to them cannot not be transferred to consumers. Therefore, a content provider participates only if $s \leq b$. Given a price s, a content provider of type t stays on 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.

With uniform pricing, 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 of allowing the network to propose tariffs that are more complex. Therefore, we now consider the possibility of the network achieving second-degree price discrimination between the two types of content providers by offering a menu of linear tariffs. We refer to this second price strategy as sponsored pricing.

Sponsored pricing: The network gives the content provider a choice between two tariffs (s_ℓ, r_ℓ) and (s_h, r_h) and the consumer is informed about which tariff the network chooses.

Sponsored pricing amounts to defining both tariffs, a base tariff ℓ and a sponsored tariff h. Content providers choose which tariff applies and that information is transmitted to consumers.[8] Note that there is no possibility of discriminating between different content providers without inducing differential consumption. Indeed, if consumers were not affected by content providers' choice, then all such content providers would invariably opt for the lower price s.[9] Yet the network may attempt to increase its profits and the value it offers to consumers by combining a higher price to content with a lower price to consumers. Content providers eager to generate high traffic (stemming from high benefits) may be willing to choose this option.

In the sponsored pricing setup, if the network succeeds at inducing the LB and the HB content providers to choose different tariffs, then consumer behavior should adapt to the tariff observed for a particular content. We thus define *sponsored pricing* as a menu $\{(s_\ell, r_\ell), (s_h, r_h)\}$ and consumption levels $\{q_\ell = D(\theta r_\ell), q_h = D(\theta r_h)\}$ such that the LB (resp. HB) content providers are willing to participate and choose tariff ℓ (resp. h), given the expected consumption of each content type.

Our definition of sponsored pricing encompasses situations in which the network chooses to exclude LB content. In fact, such exclusion should be expected when $q_\ell = 0$ (with price $r_\ell \geq \bar{p}/\theta$) and $b_\ell < s_h \leq b_h$. The reason is that, under these conditions, the LB content providers encounter zero demand and so obtain zero profit.[10] Sponsored pricing with $q_\ell = 0$ is equivalent to a uniform tariff where $(s, r) = (s_h, r_h)$, with $s_h > b_\ell$ because the LB content providers would not participate under such a tariff. Similarly a uniform tariff $s \leq b_\ell$ corresponds to sponsored pricing with $(s_\ell, r_\ell) = (s_h, r_h)$.

[8]This process requires ex ante communication. The content providers can directly inform consumers about the tariff when they visit and a standard unravelling argument shows that they will do so if the network doesn't transmit the information.

[9]It follows that keeping consumers uninformed is *not* compatible with price discrimination, because then consumption would not vary in response to content providers' choice.

[10]Under sponsored pricing, then, there is no loss of generality in assuming that all content providers participate in the mechanism.

So without loss of generality we can focus on sponsored pricing to characterize the network optimal pricing strategy. For each content type t, the consumer surplus is $\mathrm{CS}_t = U(q_t) - r_t\theta q_t$ and the network profit is $\Pi_t = (r_t + s_t - 1)\theta q_t$. The network maximizes the average joint surplus with consumers as follows

$$V = \lambda \left[U(q_h) - (1 - s_h)\theta q_h\right] + (1 - \lambda)\left[U(q_\ell) - (1 - s_\ell)\theta q_\ell\right].$$

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

$$b_\ell \geq s_\ell \text{ and } b_h \geq s_h. \tag{1}$$

As we have already argued, $q_t = 0$ is equivalent to no supply of type t content; we can therefore impose without loss of generality that condition (1) holds even when one content type is not offered. Then, the following incentive compatibility conditions ensure that each content provider chooses the tariff designed for its specific type:

$$\begin{aligned}(b_\ell - s_\ell)q_\ell &\geq (b_\ell - s_h)q_h;\\ (b_h - s_h)q_h &\geq (b_h - s_\ell)q_\ell.\end{aligned} \tag{2}$$

Sponsored tariffs are thus equivalently characterized by an allocation (q_l, q_h, s_l, s_h) such that conditions (1) and (2) are both satisfied. The network's program is then to maximize V under these two constraints. This program departs from classical textbook cases because here the transfer $s\theta q$ depends on the quantity. Nevertheless, one can follow the usual procedure for solving such programs, which we describe next.

First, since it is optimal to raise content prices as long as they remain compatible with the constraints, if follows that both the participation constraint of the LB content providers and the incentive constraint of the HB content providers will be binding, that is

$$s_\ell = b_\ell \quad \text{and} \quad (b_h - s_h)q_h = (b_h - s_\ell)q_\ell. \tag{3}$$

Second, under condition (3), all constraints are satisfied provided that $s_h \geq b_\ell$ or equivalently that $q_\ell \leq q_h$. If we replace the prices with the values given by condition (3), the reduced program can be written as a function of the quantities:

$$\max_{q_h \geq q_\ell} \lambda \left[U(q_h) - (1 - b_h)\theta q_h - (b_h - b_\ell)\theta q_\ell\right] + (1 - \lambda)\left[U(q_\ell) - (1 - b_\ell)\theta q_\ell\right].$$

This expression leads directly to the following statement.

Proposition 1 *The optimal sponsored tariffs are such that* $q_t^* = D(r_t^* \theta)$, *for* $t = \ell, h$, *where*

$$s_\ell^* = b_\ell,\ s_h^* = b_h - (b_h - b_\ell)\frac{q_\ell}{q_h};$$
$$r_\ell^* = \min\left\{1 - b_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell),\ \bar{p}\right\},\ r_h^* = 0.$$

Proof. The solution of the reduced program is obtained at

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

The condition $q_h \geq q_\ell$ holds and it determines the usage price for consumers. In the event $q_\ell = 0$, we adopt the convention that $r_\ell \theta = \bar{p}$ for clarity but any larger price would also work. The content prices are then given by condition (3). ■

The menu of tariffs proposed by the network plays two roles.[11] It allows the network to screen the different types of content providers and it leads to more efficient consumption than under uniform pricing. Consider the tariff designed for the HB content providers. Because the gains generated by these providers' traffic are higher than its cost, the network prefers high consumption and so sets a zero price for content receivers (i.e. consumers). The price s_h paid by the HB content providers is strictly less than b_h because the platform must leave some profit to induce the HB content providers to choose the right tariff. Whereas the price s_ℓ paid by the LB content providers is simply set to minimize their profit, the price r_ℓ paid by consumers to access that contents is affected by two factors. First, s_ℓ reflects the net cost of any unit of consumption; second s_ℓ is distorted so as to minimize how much profit the network must leave to HB content providers and still induce them to choose the right tariff. As it is common in the information economics literature, informational asymmetries lead the network to propose higher prices and thereby generate social costs.

In this setting, the network may decide to exclude LB content providers. This happens when the quantity q_ℓ resulting from the price characterized in Proposition 1 is negative — and thus when λ is large:

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

In this case, the network may simply rely on a uniform tariff $s = b_h$ and $r = 0$. For a

[11] The same analysis holds with $b_h < 1$ except that $r_h = 1 - b_h$.

low proportion of HB content, the network chooses to induce full participation and screens between the two content types. Thus sponsored pricing is preferred to uniform pricing for $\lambda < \lambda^*$.

Corollary 1 *Under laissez-faire, the network excludes LB contents for $\lambda > \lambda^*$; otherwise it opts for sponsoring. Total exclusion of LB content is most likely when b_h is high, θ high, and/or that b_ℓ is low.*

Proof. The claims are immediate from Proposition 1 and 4. ■

Hence it is clear that the network should never use uniform pricing without exclusion since that pricing policy is never the solution of the network's profit-maximizing program under full participation. In our two-type model, however, the network relies on uniform prices in the event of content exclusion.

Finally, from the total welfare perspective, we see that the consumption of HB content is efficient as long as the usage price charged to consumers is nonnegative. In contrast, efficient consumption of LB content would require that $r_\ell = 1 - b_\ell$, so the consumption is suboptimal. These differences provide a rationale for regulation, and one that does not depend on the network's market power on the consumer side. In fact, as it will appear clearly in the extensions, the analysis relies on the network's exclusive relationship with consumers.

4 Regulation of the content providers prices

In this section we discuss various forms of regulation and their effect on welfare. We focus on the case where this regulation concerns the content price, and we examine standard regulatory forms so as to reflect real-world practices.[12]

We assume that the regulator seeks to maximize total welfare; formally

$$W = \lambda \left[U(q_h) - (1 - b_h)\theta q_h \right] + (1 - \lambda) \left[U(q_\ell) - (1 - b_\ell)\theta q_\ell \right].$$

Let q_t^{FB} denote the (constrained) efficient consumption levels characterized by

$$q_h^{\mathrm{FB}} = D(0) \text{ and } q_\ell^{\mathrm{FB}} = D\left((1 - b_\ell)\theta\right)$$

Since q_h^* is efficient while $q_\ell^* < q_\ell^{\mathrm{FB}}$, the main regulatory concern will be to increase consumption of LB content while avoiding any decrease in the level of HB content consumption.

[12]See the online Supplementary Appendix for a discussion of optimal regulatory rules.

4.1 Cost-oriented price cap

We first consider the cost-oriented price cap regulation, as it is a simple and common form of regulation. Regulating at cost level amounts to setting a constraint of the form $s \leq 1$.

In the case of sponsored pricing, the constraints are the same as in the previous section except that $1 \geq s_h$ replaces $b_h \geq s_h$ in condition (1). Straightforward reasoning (detailed in the Appendix) establishes that the network will choose $b_\ell = s_\ell$, $q_h \geq q_\ell$. The price of the HB content now accounts for the price cap and is given by

$$s_h = \inf\left\{1, b_h - (b_h - s_\ell)\frac{q_\ell}{q_h}\right\}.$$

The reduced program then be written as

$$\max_{q_h \geq q_\ell} \lambda \left[U(q_h) - \theta q_h + \inf\{q_h, b_h q_h - (b_h - b_\ell)q_\ell\}\theta\right] + (1-\lambda)\left[U(q_\ell) - (1-b_\ell)\theta q_\ell\right]$$

The difference here when compared with the unconstrained sponsored pricing program is that, at some level, further reducing the consumption of LB content no longer reduces the rent of the HB content providers. We thus can show the following result.

Proposition 2 *A cost-oriented price cap leads to $\bar{s}_\ell = b_\ell$, $\bar{s}_h \leq 1$ and also to an efficient consumption of HB content $\bar{q}_h = D(0)$. As compared with laissez-faire, for LB content it leads to*

- *the same consumption $\bar{q}_\ell = q_\ell^*$ if $q_\ell^* \geq D(0)\frac{b_h - 1}{b_h - b_\ell}$ or to*
- *higher consumption $\bar{q}_\ell = \min\left\{D(0)\frac{b_h-1}{b_h-b_\ell}, q_\ell^{\mathrm{FB}}\right\}$ if $D(0)\frac{b_h-1}{b_h-b_\ell} \geq q_\ell^*$.*

The consumption is efficient if $D(0)\frac{b_h-1}{b_h-b_\ell} \geq q_\ell^{\mathrm{FB}}$.

Proof. See Appendix. ■

It is clear that the price cap has no effect whenever the original solution involves only content prices that are lower than one. Otherwise, the network must raise the HB content price to the cost and also raises the consumption of LB content.

The main consequence of such a price cap is to deter the network from extracting too much surplus from HB content providers. This has not only a direct effect on how surplus is shared but also an indirect and positive effect on efficiency. Indeed, since the network cannot extract too much from the HB content providers, it has less incentive than before to distort the consumption of LB content. If the efficient level of LB content consumption is low enough, then the network's pricing of HB content providers at cost is compatible with efficient consumption of LB content and so the allocation will be efficient. In short, imposing a price cap leads to increased welfare and less distortions of consumption levels.

Corollary 2 *Under sponsored pricing, a price cap at cost,* $s \leq 1$ *increases welfare.*

Proof. It follows from $\bar{q}_\ell \in \left[q_\ell^*, q_\ell^{FB}\right]$. ■

The result hinges on a seesaw effect at work for usage prices. In particular, allowing the network to charge content providers above cost would not reduce consumer prices for HB content (which is already free) although this would reduce the consumption of LB content.

4.2 No discrimination

We now consider a regulation that forces uniform pricing by prohibiting any form of price discrimination — and so, by extension, any sponsored pricing. In the debate over regulation of network pricing on the Internet, this type of regulation corresponds to one form of net neutrality that has been advocated. When restricted to uniform pricing, the network can selectively reduce the consumption of LB contents only by excluding them via a price s above b_ℓ. If we denote by $M \in \{\lambda, 1\}$ the mass of active content providers, then the network maximizes the joint surplus with consumers

$$V = M \times [U(q) + (s-1)\theta q] \quad \text{with} \quad q = D(r\theta).$$

The term in brackets captures the incentives to maximize the per-content joint surplus of the network and consumers for a given value of s. The net data cost per unit of content is $(1-s)\theta$ and internal efficiency is achieved by setting — whenever feasible — a consumer price equal to this cost. Because the participation of content providers is independent of the price charged to consumers, the network chooses

$$r = \max\{1-s, 0\}. \tag{5}$$

Given (5), the choice of the network boils down to choosing the price s that content providers will be charged. Observe that for a given participation level M, the network value increases with the price s. Therefore, the network chooses s by comparing two possible prices for content: the maximal price $s = b_\ell$ that maintains full participation with $r = 1 - b_\ell$, and the maximal price $s = b_h$ that preserves the participation of only HB contents providers with $r = 0$ (since $1 - b_h < 0$). The consumption levels in these two cases are then $q_\ell^u =$

$D\left(\left(1-b_{\ell}\right)\theta\right)$ and $q_h^u = D\left(0\right)$, which respectively yield the following network values:

$$V_{\ell}^{u} = U\left(q_{\ell}^{u}\right) + \left(b_{\ell}-1\right)\theta q_{\ell}^{u} \quad \text{when } s = b_{\ell}$$
$$V_{h}^{u} = \lambda\left[U\left(q_{h}^{u}\right) + \left(b_{h}-1\right)\theta q_{h}^{u}\right] \quad \text{when } s = b_{h}.$$

The key difference between uniform pricing and sponsored pricing is that, when both types of content providers participate, the network needs to leave a higher rent to the HB content providers because it cannot selectively reduce q_{ℓ}. Hence there is more exclusion of content under uniform than under sponsored pricing.

Proposition 3 *Under uniform pricing, the network excludes the LB contents ($s = b_h$) if and only if $\lambda > \lambda^u$. There is more exclusion than under sponsored pricing ($\lambda^u < \lambda^*$). The threshold λ^u is decreasing in b_h, and increasing in b_{ℓ}.*

Proof. See Appendix. ■

The comparative statics underlying the trade-off can be easily analyzed in terms of the relative efficiency of content types. The network value under exclusion increases with b_h and is independent of b_{ℓ}. Conversely, the network value when all content providers participate increases with b_{ℓ} and is independent of b_h. In sum: (i) exclusion occurs if b_h is large enough and/or b_{ℓ} is small enough; and (ii) the network does *not* exclude the low-benefit contents if both b_h and b_{ℓ} are close to 1.

We now turn to the welfare comparison of uniform versus sponsored pricing. On the one hand, we have shown that there will be more exclusion under uniform pricing because the network cannot accommodate LB content that should generate only low levels of consumption. On the other hand, a network that wants to attract users who consume both types of content must raise the level of LB consumption and reduce the level of HB consumption. So, as is often the case, the overall effect of a ban on price discrimination is ambiguous.

Proposition 4 *If sponsored pricing is banned, then*

- *when $\lambda > \lambda^u$, total welfare either decreases or is unaffected.*
- *when $\lambda < \lambda^u$, total welfare decreases if λ is small enough; however the effect is ambiguous for intermediate values of λ.*

Proof. See Appendix ■

A standard result in the analysis of price discrimination is that sponsored pricing yields higher welfare that uniform pricing as long as the former does not exclude LB contents. It is also widely known that the HB contents benefit from price discrimination. This benefit follows because, when allowing LB content providers to operate with a low s (as occurs for $\lambda \in [\lambda^u, \lambda^*]$), the network must leave some rents to the HB providers — which is not required when uniform prices are exclusionary.

When there is no exclusion under either regime (for $\lambda < \lambda^u$), the effect of a ban on sponsored pricing is more ambiguous. In that case, there may be too little consumption of HB content whereas the consumption of LB content increases, and welfare could either rise or fall as a result. We remark that when λ is small, the distortion of q_ℓ (or, equivalently, of r_ℓ) is likewise small and so the former effect dominates, making sponsored pricing the optimal system.

Having shown that neither a price cap nor a ban on price discrimination ensures full efficiency, we are led to wonder whether combining the two might yield a better outcome. Our next result argues to the contrary when the content benefits are large.

Corollary 3 *Under a cost-oriented price cap $s \leq 1$, sponsored pricing dominates uniform pricing if b_h is large.*

Proof. See Appendix. ■

When the benefits generated by the consumption of HB content are large enough, it is critical to ensure efficient levels of consumption. Because uniform pricing tends to reduce HB content consumption, it is all the more detrimental that b_h is large. Moreover, recall that increasing b_h tends to increase the price cap's effect, and so reduces the extent to which LB content consumption is distorted under sponsored pricing. So, when looking at the effect on the two types of contents for b_h large, sponsored pricing dominates uniform pricing.

4.3 One-sided pricing regulation

Another possible regulation consists of imposing a *zero price rule* for contents ($s = 0$ in our setting) which is another option commonly discussed in the debate about net neutrality. In this case, regulators may again decide to allow or to ban sponsored pricing.

Let us suppose first that sponsored pricing is allowed. This means that the network must first choose a base price r_ℓ for consumers along with a zero price ($s_\ell = 0$) for content providers. It may still offer content providers the option of sponsoring consumption in which case the prices become $r_h < r_\ell$ for consumers and $s_h > 0$ for content providers. Under such a regulation, the network's pricing program is the same as in Section 3 except that the constraint $s_\ell \leq b_\ell$ is replaced by $s_\ell = 0$. The reasoning of Proposition 1 applies with this new constraint so the network's optimal pricing strategy is given by

$$s_\ell = 0, \qquad r_\ell = \min\left\{1 + \frac{\lambda}{1-\lambda} b_h, \bar{p}\right\} > r_\ell^*; \tag{6}$$

$$s_h = b_h - \frac{q_\ell}{q_h} b_h, \qquad r_h = 0. \tag{7}$$

As compared with the laissez-faire case, the consumption of HB content is unchanged while the consumption of LB content is reduced (because consumer prices are higher). Total welfare and consumer welfare are then both lower than under laissez-faire, and a fortiori lower than under a cost-oriented price cap. The attractiveness of sponsored pricing is naturally reduced when the network is not allowed to charge content providers in the base tariff. Hence the network must reduce base consumption still further in order to maintain the sponsored pricing option's value.

We now turn to the case of a general ban on any price charged to content providers, which we refer to as *strict one-sided pricing*. In this case, discrimination by sponsoring is not possible and so the network sets a unique consumer price $r = 1$ (per equation 5) and the consumption of any content is given by $D(\theta)$. Under this zero-price regulation, the network cannot exclude any content and so will choose a high usage price for consumers. When the content is of HB type, a price cap at 1 performs better than the price of 0, because the former yields efficient consumption. In the case of LB content, the relative merits of price caps at 1 versus at 0 is ambiguous. We then have the following result.

Proposition 5 *Total welfare is greater under a cost-oriented price cap with sponsoring than under one-sided pricing with sponsoring. Furthermore, total welfare is higher under a cost-oriented price cap than under strict one-sided pricing if b_h is large, or if λ is either small or large enough.*

Proof. See Appendix. ■

We know from Corollary 3 that under a cost-oriented price cap it is optimal to allow sponsored pricing if b_h is large. So if the HB content is valuable enough, then the socially optimal regulation is a cap on the price charged to content providers with sponsored pricing allowed. These features would induce the optimal consumption of HB content while mitigating network incentives to distort the consumption of nonsponsored content (because such distortion would not increase the network's sponsoring revenue).

For low values of b_h, a strict zero-price regulation reduces total welfare whenever cost-orientation leads to almost efficient pricing, which can occur when λ is small. When a cost-oriented price cap is not efficient and hence leads to less consumption of LB content than under a zero-price regime, the comparison is ambiguous except when λ is large (since in that case only the HB content matters).

5 Extensions

5.1 Elastic participation and competition between networks

In the main analysis, we considered the case of a monopoly network with inelastic subscription demand. We should like to demonstrate that introducing demand elasticity or competition at the network level does not change the manner in which the variable cost is allocated between consumers and content providers — and thus does not affect the main conclusions of our work. For this purpose, we need to describe in more 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 $\{q_{ih}, q_{i\ell}\}$ is given by

$$\mathbb{E}_t \left(u(q_{it}) - \theta r_{it} q_t\right) + \varepsilon_i - F_i$$

where $\mathbb{E}$ is the expectation operator and where, for each network i: F_i is the hookup fee, ε_i is an idiosyncratic shock and r_{it} the variable price on the cost of content t. The idiosyncratic shock ε_i is a random variable that represents consumers' heterogeneity with regard to the intrinsic taste for network i. We place 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 contents.[13]

The timing of the game is unchanged and we assume that at stage 1, each network i simultaneously makes an offer $(F_i, r_{ih}, s_{ih}, r_{i\ell}, s_{i\ell})$. In this slightly modified setting, we assume that content providers may deliver their content to all networks — they pay then only a variable price — whereas consumers subscribe to a single network.

Let N_i denote the mass of consumers subscribing to network i; then the profit of each content provider using this network is given by

$$N_i(a_t - s_{it}\theta)q_{it} = N_i\theta(b_t - s_{it})q_{it}.$$

[13]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).

A content provider of type t will choose to participate in network i if $b_t \geq s_{it}$. In this context, the participation of content providers to 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.

The gross consumer surplus is given by

$$\mathrm{CS}_i = \lambda\big(U\left(q_{ih}\right) - r_{ih}\theta q_{ih}\big) + \left(1-\lambda\right)\big(U\left(q_{i\ell}\right) - r_{i\ell}\theta q_{i\ell}\big).$$

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

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

The total profit of network i is then

$$N_i\left[F_i + \lambda(r_{ih} + s_{ih} - 1)\theta q_{ih} + \left(1-\lambda\right)\left(r_{i\ell} + s_{i\ell} - 1\right)\theta q_{i\ell}\right].$$

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

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

Now we can write the profit of network i as

$$\phi_i\left(\mathrm{CS}_i - F_i; z_{-i}\right)\left[F_i + \lambda(r_{ih} + s_{ih} - 1)\theta q_{ih} + \left(1-\lambda\right)\left(r_{i\ell} + s_{i\ell} - 1\right)\theta q_{i\ell}\right].$$

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

Proposition 6 *In any equilibrium of the game with elastic subscription demand and I networks, each network chooses a tariff* (s_{it}, r_{it}), $t \in \{\ell, h\}$ *that maximizes its value per consumer:* $V_i = \lambda\left(U\left(q_{ih}\right) + \left(s_{ih} - 1\right)\theta q_{ih}\right) + \left(1-\lambda\right)\left(U\left(q_{i\ell}\right) + \left(s_{i\ell} - 1\right)\theta q_{i\ell}\right).$

Proof. The network's profit can be written as

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

where $R_i = \mathrm{CS}_i - F_i$ is the expected net consumer surplus and V_i is the network value. Note that V_i is independent of the subscription fee F_i and of the other networks' strategies z_{-i}, whereas R_i depends on F_i. Hence, the network will always choose (s_i, r_i) to maximize V_i. ■

The value V_i is solely dependent on the usage prices, so there is a natural ordering in the pricing strategy. First, the network maximizes the value that can be shared with consumers by setting adequate usage prices. Then, the network decides how much surplus to retain and how much surplus to leave to the consumers. Whereas the surplus R_i left to consumers (and hence the subscription fee F_i) depends on the elasticity of demand and on competition between networks, the prices $(r_{it}, s_{it})_{t\in\{\ell,h\}}$ do not. It follows that the prices derived in the main model with a monopoly network are the equilibrium prices also in the case of many network 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. Yet when aggregate demand is elastic, competition may increase total participation in the market. We observe that, as compared with the case of inelastic demand, the regulation of the traffic prices should be more favorable to laissez-faire under competition.

Corollary 4 *If aggregate demand is sufficiently elastic, then laissez-faire may dominate a cost-oriented price cap.*

Proof. See Appendix. ■

With elastic aggregate demand, increasing the value V_i also increases subscription demand. So if there is an optimal price cap, then it will be relatively higher and thus would strengthen the case for sponsored pricing.

5.2 Heterogenous cost

We can introduce another dimension of heterogeneity by allowing the content providers to differ in their cost to the network. We assume that consumers do not know this cost and therefore cannot incorporate this parameter into their consumption choices.[14] Even a network using "deep packet" inspection and monitoring services could not inform consumers *before* their choices are made. Because this cost is unknown, the content provider's choice of tariff is a way for it to inform consumers not only about the price r they will be paying per unit of network cost but also about the total cost of content they consume. This is another argument for the use of sponsored pricing.

More precisely, we now assume that content providers' types differ in terms of both θ and a. Content is of type h, or (θ_h, a_h) with probability λ; hence a content is of type ℓ, or (θ_ℓ, a_ℓ), with probability $1-\lambda$. Content type is known only by the content provider (i.e.,

[14]It is well documented that consumers have difficulty in correctly assessing their internet consumption (see Strategy Analytics, 2013).

not by consumers or the network). We rank these two types of content providers according to their respective ratios of advertising revenue to cost, while continuing to assume that

$$b_\ell = \frac{a_\ell}{\theta_\ell} < 1 < b_h = \frac{a_h}{\theta_h}.^{15}$$

Consumers rationally anticipate the participation of content providers and adapt their cost expectations appropriately. Because content is free, they consume $q = D\left(r\mathbb{E}\left(\theta|s\right)\right)$ of each content provider, where $\mathbb{E}\left(\theta|s\right)$ is the expected cost when the tariff is known. When cost are heterogenous, a distinction emerges between sponsored pricing and a uniform tariff.

In the uniform tariff scenario, consumers receive no signal about the cost. Conditional on the content providers participation, the average cost of data is given by

$$\mathbb{E}\left(\theta|s\right) = \begin{cases} \mathbb{E}\left(\theta\right) & \text{if } s \leq b_\ell, \\ \theta_h & \text{if } b_\ell < s \leq b_h. \end{cases}$$

Again denoting by q_t the consumption of type t content, we conclude that $q_h = q_\ell = D\left(r\mathbb{E}\left(\theta\right)\right)$ if the content price s is below b_ℓ and that $q_h = D\left(r\theta_h\right)$ and $q_\ell = 0$ if it lies between b_ℓ and b_h.

Under sponsored pricing, in contrast, consumers receive a signal of the cost. Indeed if the network succeeds in inducing the LB and the HB content providers to choose different tariffs, then consumers should realize that the average cost is different for the two tariffs. Hence, they will adapt their behavior to both the price and the cost. We therefore now redefine sponsored pricing as a menu of tariffs $\left\{\left(s_\ell, r_\ell\right), \left(s_h, r_h\right)\right\}$ such that the following two properties hold:

1. consumers anticipate that the cost is θ_ℓ for the tariff ℓ and θ_h for the tariff h, and they choose their consumption accordingly for each tariff;

2. the LB (resp. HB) content providers are willing to participate and choose tariff ℓ (resp. h), given the anticipated consumptions.

To satisfy the first condition we must have consumption levels $q_\ell = D\left(r_\ell\theta_\ell\right)$ and $q_h = D\left(r_h\theta_h\right)$ with respective tariffs ℓ and h. As a result, the information transmitted to consumers is different in the case of uniform versus sponsored pricing, since signals are more precise under the latter. A statement that will simplify the analysis is that despite their differences regarding consumer demand, the sponsored pricing strategy subsumes the uniform pricing strategy.

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

Lemma 2 *For any uniform tariff* (s, r), *there exist sponsored tariffs* (s_ℓ, r_ℓ) *and* (s_h, r_h) *that result in the same consumption levels* $\{q_\ell, q_h\}$ *and the same network value* V.

Proof. Consider a uniform tariff with $s \leq b_\ell$ and $q_\ell = q_h = D(r\theta^e)$. The same allocation can be obtained with sponsored pricing if we set $s_h = s_\ell = s$ and $r_h\theta_h = r_\ell\theta_\ell = r\mathbb{E}(\theta|s)$. Although consumers know the type of content, their consumption is the same for both types. Content providers are indifferent between the two tariffs.

Now consider a uniform tariff $b_\ell < s \leq b_h$ and $q_h = D(r\theta_h)$. The same allocation can be obtained via the sponsored tariffs $(s_h, r_h) = (s, r)$ and $r_\ell\theta_\ell \geq \bar{p}$. ■

From the network's perspective, the two main benefits of sponsoring are that it can then extract more rent from content providers and can induce more efficient levels of consumption. This interaction between screening on one side and signaling on the other side is a distinguishing feature of sponsored pricing that is not shared by a standard screening model. That being said, both the participation constraints and the incentive compatibility constraints are the same under both models. Hence the optimal sponsored pricing mechanism is obtained as before but with a new objective as follows:

$$\max_{q_h \geq q_\ell} \lambda\left[U(q_h) - (1-b_h)\theta_h q_h - (b_h - b_\ell)\theta q_\ell\right] + (1-\lambda)\left[U(q_\ell) - (1-b_\ell)\theta_\ell q_\ell\right].$$

The characterization can then be extended as described next.

Proposition 7 *When contents types differ in terms of both costs and benefits, the network's profit-maximizing sponsored tariffs are such that* $q_t^* = D(r_t^*\theta_t^*)$, *for* $t = \ell, h$, *where*

$$\begin{aligned} s_\ell^* &= b_\ell,\ r_\ell^* = \min\left\{1 - b_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell)\frac{\theta_h}{\theta_\ell}, \frac{\bar{p}}{\theta_\ell}\right\} \\ s_h^* &= b_h - \frac{q_\ell}{q_h}(b_h - b_\ell),\ r_h^* = 0. \end{aligned}$$

Proof. The solution of the reduced program is the same as before except that now the relevant cost is θ_h for q_h whereas the *virtual cost* is $(1-b_\ell)\theta_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell)\theta_h$ for the consumption of LB content. ■

Note that we always have $q_h > q_\ell$ and so, under laissez-faire, consumers correctly anticipate the cost. The solution is the same as before except that the relevant costs now differ across tariffs, which enables a more efficient allocation.

The subsequent analysis is similar to that of the main model. In particular: under laissez-faire the network excludes LB content for $\lambda > \lambda^*$, where the cutoff level for total exclusion of LB content is

$$\lambda^* = \frac{\bar{p} - (1-b_\ell)\theta_\ell}{\bar{p} - (1-b_\ell)\theta_\ell + (b_h - b_\ell)\theta_h}$$

which is decreasing in b_h and θ_ℓ, but increasing in b_ℓ and θ_h.

As before, exclusion may be achieved by way of a uniform tariff $s > b_\ell$. When the proportion of HB content is below λ^*, the network accommodates the LB content via sponsoring. The analysis of a price cap is also similar to that in the main model. Because there are always separation of types and full participation, the cost is always reflected in the consumer price and so we still have that a cost-oriented price cap enhances social welfare.

The uniform tariff case differs slightly. The net expected data cost per unit of content is $(1-s)\mathbb{E}(\theta|s)$ and internal efficiency is achieved by setting a consumer price equal to this cost whenever feasible. Since content providers' participation is independent of the consumer price, it follows that the network will choose (under the price cap) between a tariff $(s=1, r=0)$ with exclusion and a tariff $(s=b_\ell, r=(1-b_\ell))$ with full participation and consumption $D((1-b_\ell)\mathbb{E}(\theta))$. Thus the network excludes the LB content if λ is high enough that

$$\lambda U(D(0)) > U((1-b_\ell)\mathbb{E}(\theta)) + (b_\ell - 1)\mathbb{E}(\theta)D((1-b_\ell)\mathbb{E}(\theta)).$$

One difficulty with this case is that $\mathbb{E}(\theta)$ depends on λ which complicates some proofs but is not enough to overturn our results.

Proposition 8 *All the conclusions of Section 4 hold when various content differs in both the cost and benefit.*

Proof. see Appendix. ■

If we compare the case just described to the one where $\theta_h = \theta_\ell = \mathbb{E}(\theta)$, then the level of exclusion is the same with either heterogenous or homogenous costs under a cost-oriented price cap. Under laissez-faire there may be more or less exclusion depending on whether θ_h is larger or smaller than θ_ℓ. Recall that the benefit b_t is normalized by the traffic generated, so b_h may be high either because a_h is high or because θ_h is low. If costs imposed on the network are lower for the HB than for LB content, then there will be more exclusion when costs are heterogenous than when they are equal. The reverse conclusion holds when $\theta_h > \theta_\ell$.

The welfare comparison is also similar irrespective of the cost heterogeneity. A new effect is that sponsored pricing yields a further social benefit than a uniform tariff with full participation in that the consumption level can be adjusted to reflect the true cost (i.e. rather than the average cost). Hence cost heterogeneity reinforces our previous claim that price discrimination has a positive welfare effect if market power is successfully regulated by imposition of a price cap. However, this strategy is not sufficient to yield nonambiguous comparisons for intermediate values of λ. Yet we do know that ban on discrimination reduces welfare if λ is either high or low — and if the HB content is extremely valuable.

Similarly, a zero-price regulation reduces welfare — as compared with a price cap at cost — if b_h is high, because in that case there will be insufficient (resp. excessive) consumption of HB (resp. LB) content.

6 Conclusion

This paper has investigated the effect of missing content prices on the efficient pricing of transmission network services. We have shown that, when consumers control their consumption but are unaware of the induced effect on content providers and the network, a direct or indirect signal should be sent to them. In the standard setting with paid goods, this signal is sent through the price chosen by the content providers; when the goods are free, however, that approach is not feasible and so the network prices must substitute for the missing content price. In this context, our analysis has highlighted some interesting elements.

First, since networks provide unique access to consumers, it follows that even competitive networks will not choose fully efficient tariffs and instead will induce excessive exclusion of contents. This conclusion is similar to results obtained in the context of telecommunication termination charges,[16] although we focused on consumption efficiency under of adverse selection.

Second, allowing networks to propose a menu of tariffs — among which each content provider must choose — may help to mitigate the excessive exclusion of contents. By letting each content provider choose not only its own price but also the price paid by its consumers, sponsored pricing prevents the exclusion of the low value/traffic-intensive content and increases the volume of consumption of high value content.

Our analysis also suggests that some regulation is optimal. In particular, requiring that the tariff proposed to content providers not exceed a price cap reduces excessive exclusion while preserving flexibility in offers and screening possibilities. Thus we find that most of the benefits from regulation can be reaped with a price cap at cost. Imposing further restrictions — such as uniform pricing or zero price — is not desirable if the content providers' willingness to pay is high. We remark that a price cap does not constrain the price charged to consumers, which is not desirable when the network uses a two-part tariff on the consumer side.

This paper did not address anti-trust issues that sponsored data may raise, in particular the risk that Internet service providers favor some content providers over others. Note however that price cap regulation along the lines suggested here would go a long way toward alleviating those concerns, provided that all content providers have access to sponsored data

[16] These are the fees charged to one network when one of its subscribers calls a subscriber belonging to another network

programs.

A possible extension of this work would be to discuss the choice of content providers to be free or to use prices to mediate their relationship with consumers. Endogenizing this choice would certainly modify the effect of regulation and provide interesting new insights on the optimal regulation of this industry.

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., Jeon D.-S. and B.-C., Kim, 2013. "Asymmetric Neutrality Regulation and Innovation at the Edges: Fixed vs. Mobile Networks", mimeo, Toulouse School of Economics.

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

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

[8] 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.

[9] Hermalin, B. and M.L. Katz, 2007. "The Economics of product-line restrictions with an application to the network neutrality debate", *Information Economics and Policy*, 38, 215-248.

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

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

[12] Jullien, B. and W. Sand-Zantman, 2015. "Internet Regulation, Two-Sided Pricing, and Sponsored Data - Online Supplementary Appendix", IDEI.

[13] 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.

[14] Peitz, M. and F. Schuett, 2015. "Net neutrality and inflation of traffic", mimeo, University of Mannheim.

[15] Reggiani, C. and T. Valletti, 2012. "Net neutrality and innovation at the core and at the edges", Economic Discussion Paper EDP-1202, The University of Manchester.

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

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

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

Appendix

Proof of proposition 2

We consider here the case of a price cap on s_h equal to cost. In the case of sponsored pricing, the constraints can be written as

$$b_\ell \geq s_\ell \quad \text{and } 1 \geq s_h$$
$$(b_\ell - s_\ell)\, q_\ell \geq (b_\ell - s_h)\, q_h$$
$$(b_h - s_h)\, q_h \geq (b_h - s_\ell)\, q_\ell$$

As the network value is decreasing with the prices paid by contents providers, we have

$$\begin{aligned} s_\ell &= \inf\{b_\ell, b_\ell - (b_\ell - s_h)\frac{q_h}{q_\ell}\} \\ s_h &= \inf\{1, b_h - (b_h - s_\ell)\frac{q_\ell}{q_h}\} \end{aligned}$$

Suppose that $s_h \leq b_\ell < 1$. Then $s_\ell = b_\ell - (b_\ell - s_h)\frac{q_h}{q_\ell}$. It means that

$$s_h = b_h - (b_h - s_\ell)\frac{q_\ell}{q_h} = (b_h - b_\ell)\left(1 - \frac{q_\ell}{q_h}\right) + s_h,$$

which is only possible if $q_\ell = q_h$ and thus $s_h = s_\ell$.

Thus we have $s_h \geq s_\ell = b_\ell$ which requires $q_\ell \leq q_h$. The program then writes as

$$\begin{aligned} \max_{q_h \geq q_\ell} \quad & \lambda \left[U(q_h) - \theta q_h + \inf\{q_h, b_h q_h - (b_h - b_\ell) q_\ell\}\theta\right] \\ & + (1-\lambda)\left[U(q_\ell) + (b_\ell - 1)\theta q_\ell\right] \end{aligned}$$

Given that $b_h \geq 1$, optimality implies that $q_h = D(0)$. The objective is concave in q_ℓ with a kink at $q_\ell = D(0)\frac{b_h-1}{b_h-b_\ell}$. We thus find:

1. If $q_\ell^* \geq D(0)\frac{b_h-1}{b_h-b_\ell}$, the price cap is not binding, i.e. $1 > s_h^*$.
2. If $q_\ell^* \leq D(0)\frac{b_h-1}{b_h-b_\ell} \leq D((1-b_\ell)\theta)$, then the network will choose $q_\ell = D(0)\frac{b_h-1}{b_h-b_\ell}$ and $s_h = 1$.
3. If $D((1-b_\ell)\theta) \leq D(0)\frac{b_h-1}{b_h-b_\ell}$, then $q_\ell = D((1-b_\ell)\theta)$ and $s_h = 1$.

■

Proof of proposition 3

While V_ℓ^u does not depend on λ, V_h^u is linear in λ. At $\lambda = 0$ we have $V_\ell^u > V_h^u = 0$ (because demand is positive at price $(1-b_\ell)\theta$) and 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 and $0 < \lambda^u < 1$.

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

$$\lambda^u = \frac{S((1-b_\ell)\theta)}{S(0) + (b_h - 1)\theta D(0)}$$

which is decreasing in b_h and increasing in b_ℓ.

Let us show that $\lambda^* > \lambda^u$, where

$$\lambda^* = \frac{\bar{p} - (1-b_\ell)\theta}{\bar{p} + (b_h - 1)\theta}$$

By convexity of $S(.)$ and $S(\bar{p}) = 0$

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

hence

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

Using $S(0) < \bar{p}D(0)$ we have (since the RHS increases with $S(0)$)

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

■

Proof of proposition 4

If $\lambda > \lambda^u$ it is immediate that price discrimination is weakly better than uniform pricing as the consumption level of HB content is the same, and there is no consumption of LB content under uniform pricing. The superiority is strict except when $\lambda > \lambda^*$ in which case the two regimes are equivalent.

We focus now on the case where $\lambda < \lambda^u$. Under sponsored pricing, expected social welfare is given by

$$W^* = \lambda \left[U\left(q_h^*\right) + (b_h - 1)\,\theta q_h^*\right] + (1 - \lambda)\left[U\left(q_\ell^*\right) + (b_\ell - 1)\,\theta q_\ell^*\right].$$

As for the case with uniform pricing, expected social welfare is given by

$$W^u = U\left(q_\ell^u\right) + \lambda\,(b_h - 1)\,\theta q_\ell^u + (1 - \lambda)\,(b_\ell - 1)\,\theta q_\ell^u$$

where $q_\ell^u = D\left((1 - b_\ell)\,\theta\right)$

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

$$\frac{\partial (W^* - W^u)}{\partial \lambda}\,|_{\lambda=0} = U\left(q_h^*\right) + (b_h - 1)\,\theta q_h^* - \left[U\left(q_\ell^u\right) + (b_h - 1)\theta q_\ell^u\right] > 0.$$

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

Proof of corollary 3

The welfare differential in favor of sponsored pricing is:

$$\begin{aligned} &\lambda \left[U\left(q_h^*\right) + (b_h - 1)\,\theta q_h^* - U\left(q_h^u\right) - (b_h - 1)\,\theta q_h^u\right] \\ &+ (1 - \lambda)\left[U\left(\bar{q}_\ell\right) + (b_\ell - 1)\,\theta \bar{q}_\ell - U\left(q_\ell^u\right) - (b_\ell - 1)\,\theta q_\ell^u\right], \end{aligned} \tag{8}$$

This is clearly positive if there is exclusion under uniform pricing. We then focus on the other case, that is when there is no exclusion.

The first term is positive as q_h^* is efficient. The second term may be positive or negative.

If $q_\ell^{\mathrm{FB}} \leq D(0)\frac{b_h-1}{b_h-b_\ell}$ then $\bar{q}_\ell = q_\ell^{\mathrm{FB}}$ maximizes welfare, implying that the second term is non-negative. This occurs when

$$\frac{D(0)-b_\ell q_\ell^{\mathrm{FB}}}{D(0)-q_\ell^{\mathrm{FB}}} \leq b_h$$

which holds if b_h is high.

Notice that if $q_\ell^{\mathrm{FB}} > D(0)\frac{b_h-1}{b_h-b_\ell}$, the second term of (8) is negative because

$$\bar{q}_\ell = \max\left\{q_\ell^*, D(0)\frac{b_h-1}{b_h-b_\ell}\right\} < q_\ell^u = q_\ell^{\mathrm{FB}}$$

Thus the comparison is ambiguous for intermediate values of λ. ■

Proof of proposition 5

The welfare differential between a cost-orientated price cap and strict zero-price regulation is given by

$$\begin{aligned}\Delta \;=\; & \lambda\left[U\left(q_h^*\right)+\left(b_h-1\right)\theta q_h^*-U\left(q^0\right)-\left(b_h-1\right)\theta q^0\right] \\ & +(1-\lambda)\left[U\left(\bar{q}_\ell\right)+\left(b_\ell-1\right)\theta \bar{q}_\ell-U\left(q^0\right)-\left(b_\ell-1\right)\theta q^0\right],\end{aligned} \tag{9}$$

where $q^0 = D(\theta)$. This is strictly positive for λ small because then $\bar{q}_\ell$ is close to the efficient quantity q_ℓ^{FB} and for λ large as q_h^* is efficient.

As in the proof of corollary (3), if

$$\frac{D(0)-b_\ell D((1-b_\ell)\theta)}{D(0)-D((1-b_\ell)\theta)} \leq b_h \tag{10}$$

a price cap at 1 yields an efficient outcome and dominates the zero price.

Suppose that (10) is violated, then

$$\bar{q}_\ell = \max\left\{D\left(\left(1-b_\ell\right)\theta+\frac{\lambda}{1-\lambda}(b_h-b_\ell)\theta\right), D(0)\frac{b_h-1}{b_h-b_\ell}\right\}$$

is suboptimal, decreasing in λ. The second term of (9) is nonnegative if $\bar{q}_\ell \geq D(\theta)$. This is the case if

- either $\frac{D(0)-b_\ell D(\theta)}{D(0)-D(\theta)} \leq b_h$
- or $\left(1-b_\ell\right)\theta+\frac{\lambda}{1-\lambda}(b_h-b_\ell)\theta \leq \theta$ which can be written as $\lambda \geq b_\ell/b_h$. ■

Proof of corollary 4

Consider first a monopoly network with elastic demand $\phi(\text{CS} - F)$. Let V^* and W^* be network value and welfare per consumer under laissez-faire. Let $\bar{V}$ and $\bar{W}$ be network value and welfare per consumer with a cost-oriented price cap. When $q_\ell^* < D(0)\frac{b_h - 1}{b_h - b_\ell}$, we know that $V^* > \bar{V}$ and $W^* < \bar{W}$.

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

$$\max_F \phi(\text{CS} - F)(V + F - \text{CS})$$

With elastic participation the consumers' participation is $N(V)$, increasing with V, given by (under standard concavity conditions):

$$N(V) = \phi(\text{CS} - F) \text{ where } F = \text{CS} - V + \frac{\phi(\text{CS} - F)}{\phi'(\text{CS} - F)}$$

Hence participation N^* under laissez-faire is higher than participation $\bar{N}$ under price cap. Whenever $N^*/\bar{N} > \bar{W}/W^*$, laissez-faire dominates. Whether this occurs or not depends on the elasticity of demand. Notice 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 laissez-faire will dominate for very elastic demand.

The argument extends to competition provided that higher values of V result into higher aggregate demand. ■

Proof of proposition 8

Proposition 2 and corollary 2 revisited

The proof of proposition 2 is the same replacing the objective with

$$\lambda\left[U(q_h) - \theta_h q_h + \inf\{q_h, b_h q_h - (b_h - b_\ell) q_\ell\}\theta_h\right] + (1-\lambda)\left[U(q_\ell) + (b_\ell - 1)\theta_\ell q_\ell\right].$$

and $q_\ell^* = D\left((1 - b_\ell)\theta_\ell + \frac{\lambda}{1-\lambda}(b_h - b_\ell)\theta_h\right)$.

Corollary 2 is then unchanged.

Proposition 3

The proof of proposition 3 has to be adapted as follows. We now have $V_\ell^u = U(q_\ell^u) + (b_\ell - 1)\mathbb{E}(\theta) q_\ell^u$ with $q_\ell^u = D((1 - b_\ell)\mathbb{E}(\theta))$ while $V_h^u = \lambda[U(D(0)) + (b_h - 1)\theta_h D(0)]$. It is still the case that V_h^u is linear in λ but V_ℓ^u is not constant. However

$$\frac{\partial^2 V_\ell^u}{\partial \lambda^2} = -(b_\ell - 1)^2(\theta_h - \theta_\ell)^2 D'((1 - b_\ell)\mathbb{E}(\theta)) > 0.$$

Hence $V_h^u - V_\ell^u$ is concave, negative at $\lambda = 0$ and positive at $\lambda = 1$. This implies that there exists a unique threshold $\lambda^u < 1$ such that exclusion of the LB content occurs under uniform pricing if and only if $\lambda > \lambda^u$.

The proof that $\lambda^u < \lambda^*$ uses

$$\begin{aligned}
\lambda^u &= \frac{S\left((1-b_\ell)\left(\lambda^u\theta_h + (1-\lambda^u)\,\theta_\ell\right)\right)}{S\left(0\right) + \left(b_h - 1\right)\theta_h D\left(0\right)} \\
\lambda^* &= \frac{\bar{p} - (1-b_\ell)\theta_\ell}{\bar{p} - (1-b_\ell)\theta_\ell + \left(b_h - b_\ell\right)\theta_h}.
\end{aligned}$$

By convexity of $S(.)$

$$\lambda^u < \frac{\lambda^u S\left(\left(1-b_\ell\right)\theta_h\right) + \left(1-\lambda^u\right) S\left(\left(1-b_\ell\right)\theta_\ell\right)}{S\left(0\right) + \left(b_h - 1\right)\theta_h D\left(0\right)}$$

so that

$$\lambda^u < \frac{S\left(\left(1-b_\ell\right)\theta_\ell\right)}{S\left(0\right) + \left(b_h - 1\right)\theta_h D\left(0\right) + S\left(\left(1-b_\ell\right)\theta_\ell\right) - S\left(\left(1-b_\ell\right)\theta_h\right)}.$$

By convexity of $S(.)$ and $S\left(\bar{p}\right) = 0$,

$$S\left(\left(1-b_\ell\right)\theta_\ell\right) < S\left(0\right)\frac{\bar{p} - (1-b_\ell)\theta_\ell}{\bar{p}} \text{ and } S\left(\left(1-b_\ell\right)\theta_h\right) < S\left(0\right)\frac{\bar{p} - (1-b_\ell)\theta_h}{\bar{p}}$$

We thus have

$$\begin{aligned}
\lambda^u &< \frac{S\left(0\right)\frac{\bar{p}-(1-b_\ell)\theta_\ell}{\bar{p}}}{S\left(0\right) + \left(b_h - 1\right)\theta_h D\left(0\right) + S\left(0\right)\frac{\bar{p}-(1-b_\ell)\theta_\ell}{\bar{p}} - S\left(0\right)\frac{\bar{p}-(1-b_\ell)\theta_h}{\bar{p}}} \\
&< \frac{S\left(0\right)\frac{\bar{p}-(1-b_\ell)\theta_\ell}{\bar{p}}}{S\left(0\right)\left(1 + \frac{(1-b_\ell)(\theta_h-\theta_\ell)}{\bar{p}}\right) + \left(b_h - 1\right)\theta_h D\left(0\right)}.
\end{aligned}$$

Using $S\left(0\right) < \bar{p}D\left(0\right)$, we conclude that

$$\lambda^u < \frac{\bar{p} - (1-b_\ell)\theta_\ell}{\bar{p} + \left(1-b_\ell\right)\left(\theta_h - \theta_\ell\right) + \left(b_h - 1\right)\theta_h} = \lambda^*.$$

Proposition 4

The proof is the same, accounting for heterogenous cost.

Corollary 3

The proof is the same adjusting for heterogenous cost. Notice that it is more likely that banning discrimination reduces welfare because it prevents consumption from reflecting the true cost.

Proposition 5

The proof is the same with $q^0 = D\left(\mathbb{E}\left(\theta\right)\right)$. Welfare is higher for both types of content with a cost-orientated price cap than with strict zero-price regulation

- if $\frac{D(0)-b_\ell q_\ell^{\mathrm{FB}}}{D(0)-q_\ell^{\mathrm{FB}}} \leq b_h$,
- or if $q_\ell^{\mathrm{FB}} > D(0)\frac{b_h-1}{b_h-b_\ell}$ and
  - either $\frac{D(0)-b_\ell D(\mathbb{E}(\theta))}{D(0)-D(\mathbb{E}(\theta))} \leq b_h$,
  - or $\lambda \geq \frac{\mathbb{E}(\theta)-(1-b_\ell)\theta_\ell}{\mathbb{E}(\theta)+(b_h-b_\ell)\theta_h-(1-b_\ell)\theta_\ell}$. ■