NET Institute*

www.NETinst.org

Working Paper #10-25

December 2010

**The Economics of Network Neutrality**

Nicholas Economides
Stern School of Business, NYU

Benjamin Hermalin
University of California, Berkeley

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

# The Economics of Network Neutrality*

Nicholas Economides† and Benjamin E. Hermalin‡

Abstract

Pricing of Internet access has been characterized by two properties: Parties are directly billed only by the Internet service provider (ISP) through which they connect to the Internet and the ISP charges them on the basis of the amount of information transmitted rather than its content. These properties define a regime known as "network neutrality." In 2005, some large ISPs proposed that application and content providers directly pay them additional fees for accessing the ISPs' residential clients, as well as differential fees for prioritizing certain content. We analyze the private and social incentives to introduce such fees when the network is congested and more traffic implies delays. We find that network neutrality is welfare superior to bandwidth subdivision (granting or selling priority service). We also consider the welfare properties of the various regimes that have been proposed as alternatives to network neutrality. In particular, we show that the benefit of a zero-price "slow lane" is a function of the bandwidth the regulator mandates be allocated it. Extending the analysis to consider ISPs' incentives to invest in more bandwidth, we show that, under general conditions, their incentives are greatest when they can price discriminate; this investment incentive offsets to some degree the allocative distortion created by the introduction of price discrimination. *A priori*, it is ambiguous whether the offset is sufficient to justify departing from network neutrality.

*Keywords:* network neutrality, two-sided markets, Internet, monopoly, price discrimination, regulation, congestion.

*JEL Classification:* L1, D4, L12, L13, C63, D42, D43.

*The authors thank seminar participants at the Telecommunications Policy Research Conference 2010, the Stern School of Business, the University of California, Berkeley and the 38th Research Conference on Communication, Information and Internet Policy. The financial support of the Newhouse Foundation and of the Thomas and Alison Schneider Distinguished Professorship in Finance is gratefully acknowledged.

†Stern School of Business • New York University • 44 West Fourth Street • New York, NY 10012-1126 • email: neconomi@stern.nyu.edu • Also NET Institute, www.NETinst.org.

‡University of California • Department of Economics • 508–1 Evans Hall #3880 • Berkeley, CA 94720–3880 • email: hermalin@berkeley.edu.

Date: December 11, 2010

# Contents

# 1 Introduction

At issue when a network or platform facilitates the transactions of other parties are the prices it charges these parties for this facilitation. Since its commercialization in the mid-1990s, a party (website or household) connecting to the Internet pays only its direct provider of access. Other carriers on the network, even if they are subsequently involved in transporting on behalf of that party, do not collect payment from that party.

Over time, the market for connecting websites, particularly commercial sites, has become relatively competitive. At the same time, however, there is significant monopoly power in broadband residential Internet access; that is, the connections consumers and households make to their Internet Service Provider (ISP). Moreover, even when ISPs might compete for residential customers, the fact that these customers almost always connect through a single ISP (single-home) means that an end user's ISP has a monopoly on others' access to that end user through the Internet.

To date, consumers' ISPs have not directly charged websites for the content that passes over ISPs' networks into consumers' homes, so-called "last-mile service." In summer 2005, the FCC reclassified Internet services in a way that allowed for the possibility of ISPs charging content and application providers for last-mile service.[1] In fall 2005, AT&T proposed that a new fee be paid directly to it by applications and content providers whose information packets were carried by AT&T to residential customers, irrespective of where those application and content providers connected to the Internet. In particular, even application and content providers that did not connect to Internet through AT&T would be subject to being charged by AT&T. Soon other telecom and cable TV companies proposed doing the same. In short, residential-access ISPs propose to introduce fees to application and content providers (which are now zero). Additionally, they proposed to divide their residential access connection pipe into "fast access" and "slow access" lanes and charge application and content providers on the basis of the speed of access they chose. The *status quo* (zero fees and no discrimination) has been dubbed *network neutrality*, so imposing fees to the "other side" of the network or introducing price discrimination are considered departures from network neutrality.

The issue of network neutrality is controversial and complex. In October 2009, the FCC proposed rules that would impose network neutrality by law.

[1] Until the summer of 2005, telecommunication-facility-based Internet transmissions were subject to common carrier regulation that included non-discrimination requirements. Other Internet transmissions, those not telecommunication-facility-based, were not subject to common carrier regulation. Thus, DSL service was considered a common carrier service, and therefore subject to nondiscrimination provisions. Cable modem service, in contrast, was not considered common carrier service, and therefore did not have to abide by such provisions. In the summer of 2005, the Federal Communications Commission (FCC) changed the classification of Internet transmissions from "telecommunications services" to "information services." See *Nat'l Cable & Telecomm. Assn. v. Brand X Internet Services*, 125 S. Ct. 2688 (2005). This implied that there were no longer "non-discrimination" restrictions on Internet service pricing.

Broadly the proponents of network neutrality are consumer groups, the Obama administration,[2] and companies such as Google, Skype, Amazon, eBay and Microsoft, while the opponents are Cisco Systems and telecom and cable companies.

There are many advocacy papers written on the subject, but significantly less economic research. In the latter category, Economides and Tåg (2009), in a model of differentiated consumers and applications providers and monopoly or duopoly ISPs, compare a world of zero fees by ISPs to one with positive fees. They find that, for most parameter values, total surplus is higher at zero fees. With respect to the issue of price discrimination via differential quality of service, Hermalin and Katz (2007) find that restricting an ISP to a single quality level has the following effects: (a) application and content providers that would otherwise have purchased slow access are excluded from the market; (b) applications "in the middle" of the market utilize higher and more efficient speeds than otherwise; and (c) applications at the top utilize lower and less efficient speeds than otherwise. Total surplus may rise or fall, although their analysis suggests that prohibiting discrimination is likely to harm welfare. Choi and Kim (in press) also discuss prioritization under congestion using a queuing model.[3] They consider two application or content providers who are competing for eyeballs. In their setup, in equilibrium, prioritization is bought by the more efficient provider, leading to consumers switching to this provider. This improves productive efficiency but also increases utility losses from consumers who do not buy their most preferred service. In addition there are the implications of waiting costs. In their model, the net effect on welfare is ambiguous.

Our model departs from the earlier literature in a number of ways. First, we explicitly assume congestion. This means there could be a purely allocative reason to depart from network neutrality beyond any reasons stemming from the exercise of monopoly power. Second, we assume that ISPs are not producing their own content and applications for which they may seek priority *vis-à-vis* the content and applications of independent producers. This assumption abstracts away from one reason ISPs could have to violate network neutrality.[4] Additionally, unlike Choi and Kim and others, we do not take the amount of information (*e.g.,* number of packets) sent by a given application provider as fixed. Rather, the amount of content purchased by each household from any given application provider can vary. This is a critical extension insofar as it means expansion in bandwidth does not necessarily increase speed because larger bandwidth will attract more traffic. In this sense, our model allows for

[2]See www.barackobama.com/pdf/issues/technology/Fact_Sheet_Innovation_and_Technology.pdf.

[3]Like Choi and Kim, Cheng et al. (2011) and Krämer and Wiewiorra (2010) also employ a M/M/1 queuing model to study network neutrality.

[4]In particular, there is a fear that ISPs that are also content providers, such as Comcast, may engage in vertical foreclosure. The recent and ongoing battle between Comcast and Level 3 Communications (see, *e.g.,* "Comcast Fee Ignites Fight Over Videos on Internet" in the *The New York Times*, November 30, 2010) illustrates these concerns.

the effect, commonly observed with physical highways, that adding lanes does not always reduce commute times. In turn, this has important implications for the sort of second-degree price discrimination via quality distortions considered by Choi and Kim, Hermalin and Katz, and others.

In addition, like Hermalin and Katz, we consider a continuum of application providers rather than a competing duopoly pair. Although the implications of price discrimination and other departures from network neutrality for competition among application providers is important, the fact that such oligopolistic competition itself creates additional welfare distortions can make it harder to ascertain precisely what the welfare implications of departures from neutrality are *per se*. In particular, the so-called "law of the second best," can sometimes lead to conclusions that the welfare implications are less severe than they would be absent the additional distortions.

In many ways our model is most similar to Hermalin and Katz's. A key distinction, however, is that we seek to model the provision of quality (transmission speeds) from first principles, rather than in the "black box" manner of Hermalin and Katz. This proves critical for the following reason. One question is whether, absent price discrimination motives, there would be a welfare gain from offering multiple service tiers. In Hermalin and Katz, because different tiers are simply assumed to be differentially costly and there is, by implicit assumption, no tradeoff in the quality assigned one application provider with respect to the quality that can be assigned another, one arrives at the answer that the provision of differential quality levels is welfare superior to a single quality level. Hence, in their model, the only reason *not* to have multiple tiers is because it will lead to greater distortions via pricing than would neutrality (a common quality level). In contrast, here we show that, for a fixed amount of bandwidth, welfare is always greater under neutrality than under multiple tiers assuming no change in the number of application providers who connect. In particular, if the ISP's pricing to application providers can be constrained so that it provides last-mile service to *all* application providers, then welfare is maximized by having a single tier of service.

The next section of the paper presents our model. Because we wish to allow for variable content transport, the queuing approach of Choi and Kim and others is not well suited to our purposes. Instead, we model last-mile service as a "pipe" of given bandwidth. Tiering (multiple speeds) is captured by allowing the ISP to allocate portions of the bandwidth to the traffic from particular application providers. This is equivalent to the ISP giving priority via guarantees about average transmission speeds. Unlike some models, which limit the application providers to making money solely from advertising, the application providers can earn income by selling content directly to households or via advertising or some combination thereof. Although households place different values on the content of different application providers, we do assume that the advertising rates received and marginal costs of each application provider are the same. Despite the departure from reality this simplifying assumption entails, we would argue that variation in advertising rates and marginal costs is relatively small given the former are set by economy-wide competition and the latter are often *de minimis* given that most of the content costs are fixed rather than variable.

In Section 3 we prove that, ignoring ISP pricing and any dynamic effects with respect to the ISP's incentives to build bandwidth, welfare is maximized by having a single tier of service (*i.e.,* a common average transmission time). Although households value speed differently for the content from different providers, attempts at tiering are somewhat self-defeating insofar as "faster lanes" on the information superhighway will somewhat re-congest as the otherwise faster speeds will induce households to consume more content from those application providers given access to the faster lanes. This recongestion reduces the gains from tiering sufficiently that they no longer outweigh the costs of assigning some traffic to the "slow lanes."

In the three sections following Section 3, we consider what happens if the ISP can charge application providers for last-mile service. We show, *inter alia*, that permitting the ISP to charge application providers will lead it to reduce, in equilibrium, the fee it charges households to connect. It will also reduce the equilibrium amount of traffic carried and there will be less congestion, so average transmission times will improve. Nonetheless, in a static setting, pricing by the ISP that causes some application providers to shutdown is welfare reducing *vis-à-vis* a situation in which all application providers have access. Because of differences in how well the ISP captures surplus from households as opposed to application providers, we show that there are fundamental differences in certain comparative statics. In particular, whereas an increase in the consumer surplus households obtain from their trade with application providers increases the number of application providers that are connected in equilibrium, an increase in the profit application providers make from their trade with households reduces the number of application providers that are connected in equilibrium. That is, increasing the value of trade between the two ends of the market has asymmetric effects on market size.

Much of these three sections are taken with comparing linear pricing (one-tier service) to price discrimination (two-tier service). Although price discrimination will tend to lower the hookup fee the ISP can charge households, we show, given suitable conditions on the distribution of application-provider types, that the ISP always prefers discrimination to linear pricing. Similar to the findings of other work in the literature (*e.g.,* Hermalin and Katz, 2007), we find that the welfare effects of permitting discrimination when the ISP can otherwise engage in linear pricing are ambiguous. Via examples, we show either pricing regime can be welfare superior to the other.

An important issue in the debate over network neutrality has been its effect on the ISP's incentives to invest in bandwidth expansion. In Section 7, we investigate the consequences of different pricing regimes on the ISP's investment incentives. We show that an ISP engaged in linear pricing will under invest relative to the welfare-maximizing amount; hence, this is a case in which a monopoly provides too little quality. Imposing a binding price cap on the ISP further reduces its incentives to invest. Hence, network neutrality—in the sense of no charge for last-mile access—results in less bandwidth in equilibrium than allowing charging for last-mile access. Whether the dynamic inefficiency of neutrality outweighs its static efficiency is *a priori* ambiguous. We show, via example, that the static considerations can trump the dynamic considerations. Allowing the ISP to price discriminate essentially always raises its incentives to invest in bandwidth. This stands in contrast to some earlier results in the literature, which have found an ISP's investment incentives are not always enhanced by price discrimination. In part, as discussed below, the difference in results stems from our assumption that the amount of traffic is variable, not constant. As with linear pricing, there could be a tradeoff between static efficiency and dynamic efficiency. We show, via example, that static efficiency could still trump dynamic efficiency even given the greater investment incentives offered by price discrimination.

In Section 8, we consider two extensions. One, motivated by ongoing policy debates, is what happens if the ISP is obliged to offer a zero-price "slow lane" in addition to a for-fee "fast lane"? We show that if the slow lane is not too slow, such a policy must welfare dominate one in which the ISP is allowed full freedom to discriminate or one in which it is allowed to charge a fee so great as to exclude some application providers. The second extension briefly considers what happens if the households are heterogeneous in their preferences rather than, as we assume for most of the analysis, homogeneous. We show that our primary results, Propositions 1 and 2, continue to hold; that is, network neutrality is still welfare superior *ceteris paribus*.

We end with a short conclusion. Our principal conclusion is that if bandwidth expansion is not a pressing issue, then maintaining the current network neutrality could be welfare superior to ending it.

# 2 Model

## 2.1 Technology

Figure 1 shows the basic technology we have in mind. Households want to access application providers to obtain the application providers' content. This content must pass through a "pipe" controlled by a monopoly, the ISP. The pipe has a bandwidth, $B$. This should be interpreted as meaning the ISP has the capacity to have $B$ "units" of content (*e.g.,* packets) go from application providers to households per unit of time. We assume that the ISP can dedicate portions of that bandwidth or otherwise guarantee priority across groups of application providers. Hence, for instance, it can divide the bandwidth into $B_1, \dots, B_J$ $\sum_{j=1}^{J} B_j = B$ (or do the equivalent thereof via the granting of priority).[5]

We assume a continuum of application providers of measure one, indexed by $\theta \in [\theta_0, \theta_N) \subseteq \mathbb{R}_+$. (The rationale for labeling the upper limit $\theta_N$ will become clear later.) The distribution of $\theta$ is $F : [\theta_0, \theta_N) \to [0, 1]$. Assume $F(\cdot)$ is differentiable and that $F'(\theta) > 0$ for all $\theta \in (\theta_0, \theta_N)$.

[5]In particular, this formulation is equivalent to one in which an ISP guarantees average transmission speeds. To see this, if an application provider's *average* speed is $T$ then it is as if it were allocated $1/T$ of the bandwidth (average speed = total time ÷ content = (content ÷ allocated bandwidth) ÷ content).

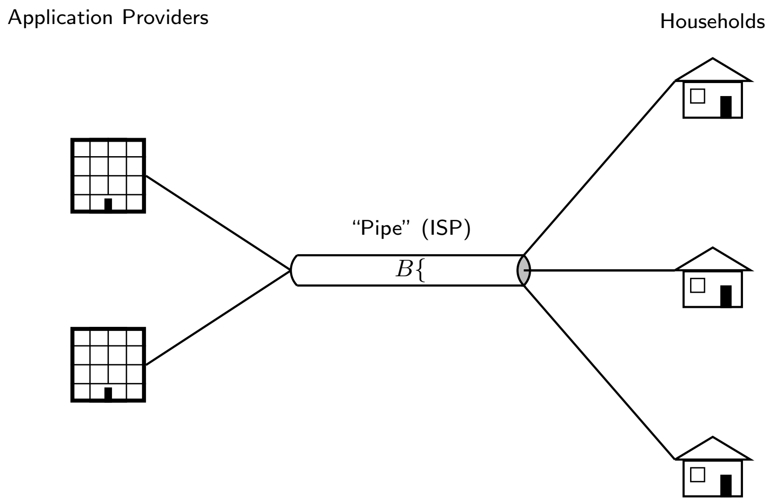


**Figure 1:** Schematic representation of technology.

Let $X(\theta)$ denote the units of content sent by application provider $\theta$. If $\mathcal{T} \subset [\theta_0, \theta_N)$ is a measurable subset of application providers with dedicated bandwidth $B_{\mathcal{T}}$, then

$$t(\mathcal{T}) \equiv \frac{\int_{\mathcal{T}} X(\theta) dF(\theta)}{B_{\mathcal{T}}}$$

is the time necessary for all content to be sent. Observe $t(\mathcal{T})$ is a measure of the congestion faced by application providers in $\mathcal{T}$ and we will treat it as such in what follows.

## 2.2 Consumers and Application Providers

Assume there is a continuum of households (consumers) of measure one. Initially, suppose that each household potentially engages in trade with each application provider. We assume households have quasi-linear utility of the form

$$U = y + \int_{\theta_0}^{\theta_N} \left( \int_0^{x(\theta)} m\Big(\frac{x\tau(\theta)}{v(\theta)}\Big) dx \right) dF(\theta) \, , \tag{1}$$

where $y$ is the numéraire good, $x(\theta)$ is the consumption of the $\theta$th provider's good (number of packets bought), and $m(\cdot)$ is the consumer's "adjusted" *marginal* utility of consumption of the $n$th provider's good, where the adjustments reflect the congestion in transmission, $\tau(\theta)$, and some value property (type) specific to the $\theta$th application provider, $v(\theta)$.[6] In practical terms, any two

[6] A more general specification would be to write $m\big(x, \tau(\theta), v(\theta)\big)$ rather than $m\big(x\tau(\theta)/v(\theta)\big)$. Unfortunately, doing so would make the model intractable. We return to

types $\theta$ and $\theta'$ such that $v(\theta) = v(\theta')$ are the same type, as there would be no way for the ISP to discriminate between them. There is no loss of generality, therefore, in assuming $v(\cdot)$ is strictly monotonic. Furthermore, because we could always redefine type through the change of variables $\tilde{\theta} = v(\theta)$, with $\tilde{\theta}$ distributed according to $F\big(v^{-1}(\cdot)\big)$, there is also no loss of generality in letting $v(\theta) = \theta$. We do so henceforth to minimize the notational burden.

If $\theta$ is in transmission group $\mathcal{T}$, then $\tau(\theta) = t(\mathcal{T})$.

Marginal utility, $m(\cdot)$, is taken to be at least twice differentiable and decreasing. We assume, further, that $m(\cdot)$ is never "too convex" in the sense that

$$m''(z) + m'(z) < 0 \tag{2}$$

for all $z \in \mathbb{R}_+$.

We assume that consumption of the application providers' goods plus any hookup fee paid the ISP never consumes a household's entire income. This and the assumption of quasi-linear utility mean that each household acquires the amount of the $\theta$th application provider's product that equates marginal utility to marginal cost (*i.e.,* price, $p$); consequently, household demand is

$$x(p, \theta) = \frac{\theta}{\tau(\theta)} m^{-1}(p) \equiv \frac{\theta}{\tau(\theta)} \omega(p)\,.$$

(Note the implicit definition of $\omega : \mathbb{R}_+ \to \mathbb{R}_+$.) Observe demand falls with congestion.

An application provider's profit is

$$\Pi(\theta) = (a + p - c)x(p, \theta) - s\,,$$

where $a$ is the advertising rate, $c$ is the cost of the content, and $s$ is a payment to the ISP. Observe, we assume a common advertising rate and cost of content. We assume, even if $a > c$, that an application provider cannot pay users to visit its site; that is, we restrict attention to $p \geq 0$.

It can be shown that assumption (2) is sufficient for $\omega(\cdot)$ to be log concave. This, in turn, is sufficient for an application provider's pricing problem,

$$\max_p (a + p - c)x(p, \theta) - s\,, \tag{3}$$

to have a unique and finite solution. Observe that solving (3) is equivalent to solving

$$\max_p (a + p - c)\omega(p)\,;$$

this point in our concluding remarks, where we note that while exploring alternative specifications is likely a reasonable project for future research, it won't eliminate the importance of the recongestion effect—the idea that fast lanes attract more traffic—which reduces the potential gains to welfare to be had from deviating from network neutrality. Moreover, the result we derive here, namely that this recongestion effect can be so strong as to make network neutrality welfare superior to its abandonment, is not reliant upon this precise specification of household utility. A task for future work is characterizing the utility functions for which the recongestion effect is dominant.

hence, the solution to (3) is independent of $\theta$. Call that solution $p^*$. For future reference define $\pi$ by

$$\pi = \big(a + p^* - c\big)\omega(p^*)\,;$$

hence,

$$\Pi(\theta) = \frac{\theta}{\tau(\theta)}\pi - s\,. \tag{4}$$

Household (consumer) surplus from trade with application provider $\theta$ is

$$cs(\theta) = \int_{p^*}^{\infty} x(p,\theta)dp = \frac{\theta}{\tau(\theta)}\int_{p^*}^{\infty}\omega(p)dp \equiv \frac{\theta}{\tau(\theta)}\sigma\,.$$

(Note the implicit definition of $\sigma$.)

Total welfare is

$$W = \int_{\theta_0}^{\theta_N} \frac{\theta}{\tau(\theta)}(\pi + \sigma)dF(\theta)\,.$$

## 3 Welfare

Suppose, first, that there is neutrality (*i.e.,* no division of the bandwidth). Hence

$$t\big([\theta_0,\theta_N)\big) = \frac{1}{B}\int_{\theta_0}^{\theta_N} \frac{\theta}{t\big([\theta_0,\theta_N]\big)}\omega(p^*)dF(\theta)\,.$$

It follows that

$$t\big([\theta_0,\theta_N)\big) = \sqrt{\frac{\mu\omega(p^*)}{B}}\,, \tag{5}$$

where $\mu$ is the average (expected) value of $\theta$.

To help keep expressions readable, define the function $\mathcal{I} : \mathbb{R}^2 \to \mathbb{R}_+$ by

$$\mathcal{I}(\theta_a,\theta_b) = \int_{\theta_a}^{\theta_b} \theta dF(\theta)\,.$$

Note $\mathcal{I}(\theta_0,\theta_N) = \mu$.

Suppose the interval of application providers were partitioned into $N$ sub-intervals, where the $n$th sub-interval is $[\theta_{n-1},\theta_n)$. Let $B_n$ denote the bandwidth allocated to the $n$th sub-interval and let $t_n$ denote the time to send the content of the application providers in the $n$th interval. Proceeding along the lines of the derivation of (5), it follows that

$$t_n = \sqrt{\frac{\mathcal{I}(\theta_{n-1},\theta_n)\omega(p^*)}{B_n}}\,. \tag{6}$$

Welfare under neutrality is, therefore,

$$W_{\text{neut}} = (\pi + \sigma)\sqrt{\frac{B\mu}{\omega(p^*)}}\,. \tag{7}$$

Welfare under a partition is

$$W_{\text{part}} = (\pi + \sigma) \sum_{n=1}^{N} \sqrt{\mathcal{I}(\theta_{n-1}, \theta_n) \frac{B_n}{\omega(p^*)}}\,. \tag{8}$$

Define $\phi_n = B_n/B$; that is, $\phi_n$ is the proportion of the bandwidth allocated to the $n$th segment. Using this notation, it follows from (7) and (8) that $W_{\text{neut}} \geq W_{\text{part}}$ if

$$\sqrt{\mu} \geq \sum_{n=1}^{N} \sqrt{\mathcal{I}(\theta_{n-1}, \theta_n)\phi_n} = \sum_{n=1}^{N} \phi_n \sqrt{\frac{\mathcal{I}(\theta_{n-1}, \theta_n)}{\phi_n}}\,. \tag{9}$$

**Proposition 1.** *Neutrality is weakly welfare superior to any division of the bandwidth in which all segments are allocated a positive portion of the bandwidth (i.e., to any division in which $\phi_n > 0$ for all $n$).*

**Proof:** Because the $\phi_n$ sum to one, the rightmost term in (9) is an expected value. By Jensen's inequality, we have (substituting the definition of $\mathcal{I}$):

$$\sum_{n=1}^{N} \phi_n \sqrt{\frac{\int_{\theta_{n-1}}^{\theta_n} \theta dF(\theta)}{\phi_n}} \leq \sqrt{\sum_{n=1}^{N} \phi_n \frac{\int_{\theta_{n-1}}^{\theta_n} \theta dF(\theta)}{\phi_n}} = \sqrt{\int_{\theta_0}^{\theta_N} \theta dF(\theta)} = \sqrt{\mu}\,.$$

So (9) holds. ∎

The only possible improvement, then, on pure neutrality would have to involve exclusion of some segment from the web. This cannot, however, improve on neutrality.

**Proposition 2.** *Neutrality is strongly welfare superior to any division that excludes some segments.*

**Proof:** Suppose some segments are excluded (*i.e.,* assigned no bandwidth). Let those segments that are *not* excluded belong to the set $\mathcal{N}$. That is, $n \in \mathcal{N}$ if and only if $\phi_n > 0$. Starting with the middle term of (9), we have

$$\begin{aligned}
\sum_{n=1}^{N} \sqrt{\mathcal{I}(\theta_{n-1}, \theta_n)\phi_n} &= \sum_{n \in \mathcal{N}} \phi_n \sqrt{\frac{\int_{\theta_{n-1}}^{\theta_n} \theta dF(\theta)}{\phi_n}} \\
&\leq \sqrt{\sum_{n \in \mathcal{N}} \phi_n \frac{\int_{\theta_{n-1}}^{\theta_n} \theta dF(\theta)}{\phi_n}} = \sqrt{\sum_{n \in \mathcal{N}} \int_{\theta_{n-1}}^{\theta_n} \theta dF(\theta)} \\
&< \sqrt{\int_{\theta_0}^{\theta_N} \theta dF(\theta)} = \sqrt{\mu}\,,
\end{aligned}$$

where the first inequality follows again from Jensen's inequality and the second from the assumption that $\theta > 0$ for all $\theta$. ■

The economics behind Propositions 1 and 2 is as follows. Consumers value speed more for content that comes from high-type application providers than for content that comes from low-type application providers. It might be tempting, therefore, to allocate more bandwidth to high-type application providers. There are two drawbacks to this, however. First, consumers buy more high-type content than low-type content; hence, a large proportion of the bandwidth would need to be allocated to the high-type application providers if speed is to increase in a meaningful way, which means a considerable worsening of congestion for low-type content. Second, allocating more bandwidth to one segment of application providers means faster traffic *holding constant their current traffic*. However, consumers respond to the faster speed by demanding more content. The recongestion effect reduces the benefit of allocating bandwidth to the high-type content. Although the same reasoning means that some of the worsened congestion from allocating less bandwidth to low-type content is alleviated by consumers reducing their consumption of low-type content, the effects don't balance out because less low-type content is consumed in the first place (*e.g.,* a 10% decrease in low-type content consumption and a 10% increase in high-type content consumption means more total consumption). When all the effects are considered, the consequence of dividing up the bandwidth are negative.

## 4 Linear Pricing

Define $t^*$ as the value on the right-hand side of equation (5). The welfare analysis of the previous section is predicated on the ISP's serving all application providers. In other words, on its setting $s$, the access price charged application providers, such that

$$\frac{\theta_0}{t^*}\pi \geq s\,, \tag{10}$$

Given that, in many settings, monopoly pricing leads to some potential customers choosing not to purchase at all, it is not necessarily reasonable to suppose that the ISP sets $s$ to satisfy (10). If (10) does not hold, then there is a welfare loss relative to the first best because the ISP is serving too few application providers (Proposition 2). Note that (10) does not require the ISP to provide the service for free; rather it requires that the hookup charge be set low enough to induce all application providers to participate.

In addition to charging the application providers, the ISP can also charge households a hookup fee, $\eta$. Given that the households are homogenous, it is feasible for the ISP to capture all consumer surplus; that is, $\eta$ equals a household's total consumer surplus.

Suppose the ISP sets $s$ via linear pricing. In equilibrium, the lowest application provider that buys, type $\underline{\theta}$, is given

$$\underline{\theta} = \frac{st_{\text{LP}}}{\pi}\,, \tag{11}$$

where $t_{\text{LP}}$ is the equilibrium time required to send all content in equilibrium under linear pricing:

$$t_{\text{LP}} = \frac{1}{B}\int_{\underline{\theta}}^{\theta_N} \frac{\theta}{t_{\text{LP}}}\omega(p^*)dF(\theta)\,. \tag{12}$$

This last expression can be rewritten as

$$t_{\text{LP}}^2 = \frac{\omega(p^*)}{B}\int_{\underline{\theta}}^{\theta_N} \theta dF(\theta)\,. \tag{13}$$

For a given price $s$, the corresponding $t_{\text{LP}}$ and $\underline{\theta}$ are the solutions to the system of equations (11) and (13).[7]

**Proposition 3.** *Under linear pricing by the* ISP*, the greater is the equilibrium hookup fee, s, application providers are charged, the smaller is (i) the hookup fee charged households; (ii) the total number of packets sent; and (iii) the time necessary to send all packets.*

**Proof:** We prove the claims in reverse order. Observe that (13) is an identity and, thus, defines $t_{\text{LP}}$ as a function of $\underline{\theta}$. Denote that function $t_{\text{LP}}(\underline{\theta})$. Differentiating (13) yields

$$2t_{\text{LP}}(\underline{\theta})t'_{\text{LP}}(\underline{\theta}) = -\frac{\omega(p^*)}{B}\underline{\theta}F'(\underline{\theta})\,. \tag{14}$$

The right-hand side is negative, hence $t'_{\text{LP}}(\underline{\theta}) < 0$.

Consider the condition for equilibrium, (11). Suppose, starting from an equilibrium, $s$ is increased. It is readily seen that $\underline{\theta}$ must increase with $s$. Because it increases, $t_{\text{LP}}$ falls. If $t_{\text{LP}}$ falls, then, given the fixed bandwidth, the number of packets sent must decrease.

A household's gross benefit is

$$\int_{\underline{\theta}}^{\theta_N} \frac{\theta}{t_{\text{LP}}}\sigma dF(\theta)\,. \tag{15}$$

Consequently, (15) equals the hookup fee, $\eta$. Differentiating (15) with respect to $\underline{\theta}$ yields:

$$\begin{aligned} -\underline{\theta}\frac{\sigma}{t_{\text{LP}}(\underline{\theta})}F'(\underline{\theta}) - \sigma\frac{\int_{\underline{\theta}}^{\theta_N}\theta dF(\theta)}{t_{\text{LP}}(\underline{\theta})^2}t'_{\text{LP}}(\underline{\theta}) \\ = -\underline{\theta}\frac{\sigma}{t_{\text{LP}}(\underline{\theta})}F'(\underline{\theta}) - \sigma\frac{B}{\omega(p^*)}t'_{\text{LP}}(\underline{\theta}) \\ = -\underline{\theta}\frac{\sigma}{t_{\text{LP}}(\underline{\theta})}F'(\underline{\theta}) + \underline{\theta}\frac{\sigma}{2t_{\text{LP}}(\underline{\theta})}F'(\underline{\theta}) < 0\,, \end{aligned} \tag{16}$$

[7] There is no reason for the ISP to price so low that the lowest-type application provider earns a surplus; hence, (11) will be an equality in equilibrium.

where the first equality follows from (13) and the second from (14). ■

An implication of this last proposition is the following:

**Corollary 1.** *Assume not all application providers are served in equilibrium. Then the imposition of a binding ceiling on the hookup fee that the* ISP *can charge application providers will cause an increase in the hookup fee charged households.*

Were all application providers otherwise served in equilibrium, then the imposition of a binding ceiling will have no effect on the hookup fee charged households.

This corollary results because reducing the hookup fee charged application providers causes more application providers to join the platform, which raises households' gross benefits. Because the ISP can capture these benefits through the hookup price it charges households, it follows that the household hookup fee will go up. Although, in equilibrium, households are, on net, no better or worse off, if households are naïve, then the ISP's promise of a lower hookup fee could create political pressure to relax any binding ceiling on the price charged application providers.

The ISP's pricing problem can be reduced to its choosing the cutoff value $\underline{\theta}$ to maximize

$$\int_{\underline{\theta}}^{\theta_N} \frac{\theta}{t_{\text{LP}}(\underline{\theta})} \sigma dF(\theta) + \underline{\theta} \frac{\pi}{t_{\text{LP}}(\underline{\theta})} \big(1 - F(\underline{\theta})\big) \,. \tag{17}$$

The first-order condition, after some algebra and using expressions (13) and (14), is equivalent to

$$-\frac{2\pi + \sigma}{2} \underline{\theta} h(\underline{\theta}) + \pi + \underline{\theta}^2 \frac{\pi h(\underline{\theta})}{2\underline{\mu}} \leq 0 \,, \tag{18}$$

where $h(\cdot)$ is the hazard rate associated with distribution $F(\cdot)$ and $\underline{\mu} = \mathbb{E}\{\theta | \theta \geq \underline{\theta}\}$. Observing that the bandwidth does not appear in (18), we can conclude:

**Proposition 4.** *Given linear pricing, the number (measure) of application providers to whom the* ISP *sells in equilibrium is independent of the* ISP*'s bandwidth.*

Intuitively, as expression (17) shows, the ISP's profit is proportional to $1/t_{\text{LP}}(\underline{\theta})$. This quantity, in turn, is multiplicatively separable in bandwidth and the volume of traffic carried (see expression (12)). This latter separability reflects the fundamental determinants of congestion: traffic and bandwidth. Hence, ultimately, bandwidth can be "factored out" of the expression for profit, which means that the determination of which application providers will be served is independent of bandwidth.

Although the number of application providers to whom the ISP sells does not vary with its bandwidth, it does vary with the base application provider profit, $\pi$, and base household surplus, $\sigma$.

**Proposition 5.** *Unless the* ISP *is already selling to all application providers, an increase in the base household surplus, $\sigma$, increases the number (measure) of application providers who the* ISP *hooks up in equilibrium. The number of application providers to whom the* ISP *sells in equilibrium is non-increasing in the level of base application provider profit, $\pi$, and strictly decreasing if the* ISP *is not selling to all application providers in equilibrium (i.e., if $\underline{\theta} > \theta_0$).*

**Proof:** From well-known comparative statics results, the first claim follows if the cross-partial derivative of (17) with respect to $\underline{\theta}$ and $\sigma$ is negative; the second if the cross-partial derivative with respect to $\underline{\theta}$ and $\pi$ is positive. The two cross-partial derivatives are, respectively,

$$\frac{\partial^2}{\partial\underline{\theta}\partial\sigma}\text{expression (17)} = -\underline{\theta}\frac{\sigma}{2t_{\text{LP}}(\underline{\theta})}F'(\underline{\theta}) < 0 \tag{19}$$

and

$$\frac{\partial^2}{\partial\underline{\theta}\partial\pi}\text{expression (17)} = \frac{1-F(\underline{\theta})}{t_{\text{LP}}(\underline{\theta})}\left(1-\underline{\theta}h(\underline{\theta})+\underline{\theta}^2h(\underline{\theta})\frac{1}{2\underline{\mu}}\right). \tag{20}$$

The derivation of the first, expression (19), is nearly identical to the derivation of (16) above. When the first-order condition (18) is an inequality, then clearly the number of application providers served is non-increasing. If the profit-maximizing $\underline{\theta}$ is interior (*i.e.,* greater than $\theta_0$), then (18) is an equality; hence,

$$0 = \pi\left(1-\underline{\theta}h(\underline{\theta})+\underline{\theta}^2h(\underline{\theta})\frac{1}{2\underline{\mu}}\right) - \frac{\sigma}{2}\underline{\theta}h(\underline{\theta}).$$

It follows that the second cross-partial derivative, expression (20), is positive. ■

The intuition behind Proposition 5 reflects the fundamental difference between the two sides, households and application providers. Households benefit the more application providers who are connected. Moreover, the increased benefit of adding more application providers is greater the greater is the consumer surplus a household gets from any one; that is, it is increasing in base surplus, $\sigma$. It follows, therefore, that the greater is $\sigma$ the more the ISP can capture by adding application providers and, hence, the lower is $\underline{\theta}$. For application providers, the opposite is true. No application provider benefits from having more application providers on the system and, in fact, each existing application provider loses because of the consequent reduction in quality (*i.e.,* the rise in congestion). Hence, for application providers, it is congestion that matters. The benefit an application provider realizes from a reduction in congestion is greater the greater its base profit. Consequently, the increase in the price, $s$, the ISP can charge due to reduced congestion is sufficiently greater with a greater $\pi$ to offset the lost sales. Another question is how do the ISP's profit, the household hookup fee, and the application provider hookup fee vary with the bandwidth.

**Proposition 6.** *An exogenous increase in bandwidth, $B$, raises* ISP *profit, household hookup fee, and the application provider hookup fee in equilibrium.*

**Proof:** Given Proposition 4, it is sufficient to consider merely how these quantities vary with $t_{\text{LP}}(\underline{\theta})$. From (11), (15), and (17), it follows, respectively, that the application provider hookup fee, the household hookup fee, and ISP profit are decreasing in $t_{\text{LP}}(\underline{\theta})$. But, from (13), $t_{\text{LP}}(\underline{\theta})$ is decreasing in bandwidth, $B$. The results follow. ■

## 5 Price Discrimination

As is well known, allowing a monopolist to engage in price discrimination can sometimes improve welfare *vis-à-vis* linear pricing. Consequently, it is conceivable that allowing the ISP to depart from network neutrality—engage in price discrimination—could improve welfare.

Whereas typically a monopolist can gain by engaging in price discrimination, it is not obvious that would be the case here. Although discrimination could be expected to allow the ISP to capture more of the application providers' surplus, there is—as can be seen from the logic behind Proposition 1—the fact that discrimination reduces household surplus, lessening the hookup fee the ISP can charge households.

To investigate discrimination, suppose the ISP can offer two services: fast, denoted by an $f$ subscript, and low-speed, denoted by an $\ell$ subscript. Let $r$ (for rate) denote an arbitrary element of $\{\ell, f\}$. The quantity $t_r$ is the time required, in equilibrium, to send all content of those application providers that subscribe to service $r$. In keeping with their names, $t_f < t_\ell$. Let $s_r$ be the subscription or hookup fee that the ISP charges application providers for service $r$. Because, otherwise, there would effectively be only one service, $s_f > s_\ell$.

In deciding which, if any, service to purchase, an application provider chooses the largest of

$$0\,,\ \frac{\theta}{t_\ell}\pi - s_\ell\,,\ \text{or}\ \frac{\theta}{t_f}\pi - s_f\,.$$

Let 0 denote "no service" and consider the services ordered $f \succ \ell \succ 0$. The following result is easily established and the proof, therefore, omitted.

**Lemma 1.** *Let $\theta > \theta'$ and $r \succ r'$. Then a $\theta'$-type application provider's preferring $r$ to $r'$ implies a $\theta$-type application provider prefers $r$ to $r'$. Conversely, a $\theta$-type application provider's preferring $r'$ to $r$ implies a $\theta'$-type application provider prefers $r'$ to $r$.*

Assuming discrimination in equilibrium, it follows from the previous lemma that there are two types of application providers, $\underline{\theta}$ and $\hat{\theta}$, with $\theta_0 \le \underline{\theta} < \hat{\theta} < \theta_N$, such that all types $\theta \ge \hat{\theta}$ purchase the $f$ service, all types $\theta \in [\underline{\theta}, \hat{\theta})$ purchase the $\ell$ service, and types $\theta < \underline{\theta}$ purchase no service. Let $\mu_r$ denote the average type of application provider conditional on its purchasing service $r$. Observe

$$s_\ell = \frac{\underline{\theta}}{t_\ell}\pi \ \text{ and }\ s_f = \left(\frac{\hat{\theta}}{t_f} - \frac{\hat{\theta}}{t_\ell}\right)\pi + s_\ell = \left(\frac{\hat{\theta}}{t_f} - \frac{\hat{\theta}}{t_\ell} + \frac{\underline{\theta}}{t_\ell}\right)\pi\,. \qquad (21)$$

In equilibrium, the times must be consistent with the volume of traffic the two segments (services) generate. Following the same logic as the derivation of (6) above, this means

$$t_\ell = \sqrt{\frac{\mathcal{I}(\underline{\theta}, \hat{\theta})\omega(p^*)}{B_\ell}} \tag{22}$$

and

$$t_f = \sqrt{\frac{\mathcal{I}(\hat{\theta}, \theta_N)\omega(p^*)}{B_f}}\,, \tag{23}$$

where $B_r$ is the bandwidth allocated service $r$. The two bandwidths together equal total bandwidth (*i.e.*, $B_f + B_\ell = B$).

Household surplus, and thus the household hookup fee, $\eta$, is

$$\begin{aligned}\int_{\underline{\theta}}^{\hat{\theta}} \frac{\theta}{t_\ell}\sigma dF(\theta) + \int_{\hat{\theta}}^{\theta_N} \frac{\theta}{t_f}\sigma dF(\theta) &= \sigma\left(\frac{\big(F(\hat{\theta}) - F(\underline{\theta})\big)\mu_\ell}{t_\ell} + \frac{\big(1 - F(\hat{\theta})\big)\mu_f}{t_f}\right) \\ &= \frac{\sigma}{\sqrt{\omega(p^*)}}\left(\sqrt{\mathcal{I}(\underline{\theta}, \hat{\theta})(B - B_f)} + \sqrt{\mathcal{I}(\hat{\theta}, \theta_N)B_f}\right), \end{aligned} \tag{24}$$

making appropriate substitutions. Observe that (24) is effectively a function of $\underline{\theta}$, $\hat{\theta}$, and $B_f$.

The ISP's return from the application providers is

$$\big(F(\hat{\theta}) - F(\underline{\theta})\big)s_\ell + \big(1 - F(\hat{\theta})\big)s_f\,. \tag{25}$$

Using expressions (21)–(23) plus the adding up constraint $B = B_\ell + B_f$, it is readily seen that (25) is also a function of $\underline{\theta}$, $\hat{\theta}$, and $B_f$.

Given plausible conditions on the distribution of application-provider types, we can show that the ISP always prefers to engage in price discrimination:

**Proposition 7.** *Assume the hazard rate associated with the distribution of application-provider types, $h(\cdot)$, satisfies the strict monotone hazard rate property (i.e., $h(\cdot)$ is increasing).*[8] *Then the* ISP *maximizes its profit by offering two classes of service (i.e., engaging in price discrimination).*

The proof, which is involved, is relegated to the appendix.

Intuitively, suppose the ISP chose to serve precisely the same set of application providers under price discrimination as it would under linear pricing (*i.e.*, $\underline{\theta}$ is the same under both pricing regimes). The ISP can always find a division of the bandwidth and a corresponding division of application-provider types $[\underline{\theta}, \hat{\theta})$ and $[\hat{\theta}, \theta_N]$ so that both market segments enjoy the same speed in equilibrium

[8]This is a property satisfied by many distributions, including the uniform.

(*i.e.,* such that $t_\ell = t_f$). Of course, such a division is profit neutral *vis-à-vis* linear pricing. Now suppose, given that bandwidth division, the ISP raised the cutoff $\hat{\theta}$ slightly. To a first-order approximation, this does not affect household surplus, so it leaves the household hookup fee unchanged. On the other hand, with respect to profit from the application providers, the usual logic of second-degree price discrimination applies: By distorting downward the quality of the low types of application providers, the ISP is able to capture more of the surplus of the high types, which raises overall profit from application providers.

## 6 Welfare Consequences of Pricing

From our earlier analysis, we know that welfare is maximized if the ISP is limited to linear pricing in which the application-provider hookup fee satisfies (10). But what if (10) is not satisfied by the profit-maximizing linear-pricing scheme? Would welfare be further reduced if the ISP were allowed to engage in discrimination or is welfare actually higher under the profit-maximizing price-discrimination scheme than under the profit-maximizing linear-pricing regime?

In what follows, we ignore "trivial" price discrimination schemes in which $t_\ell = t_f = t_{\text{LP}}$; that is, schemes that simply divide the bandwidth, but do not charge different fees to the application providers for the two segments and in which each segment is equally fast.

Given Propositions 1 and 2, if price discrimination results in no more application-provider types' being served, then discrimination cannot be better in terms of welfare than linear pricing. This yields:

**Proposition 8.** *A necessary condition for welfare under price discrimination to exceed welfare under linear pricing is that the set of application providers served in equilibrium expand.*[9]

In other words, if $\underline{\theta}_{\text{LP}}$ and $\underline{\theta}_{\text{PD}}$ are the lowest application-provider types served under linear pricing and price discrimination, respectively, then a necessary condition for welfare to be greater under price discrimination is that $\underline{\theta}_{\text{PD}} < \underline{\theta}_{\text{LP}}$. An almost immediate corollary of the proposition is, thus,

**Corollary 2.** *If the* ISP *would serve all application providers in equilibrium under linear pricing, then any non-trivial price discrimination scheme is welfare reducing.*

**Proof:** From Proposition 8, welfare cannot be greater under price discrimination. If the price discrimination scheme is non-trivial, then because square-root is a strictly concave function, expression (9) is a strict inequality by Jensen's

[9] This proposition is reminiscent of the result that a necessary condition for *third*-degree price discrimination to increase welfare is that the total amount sold increase (see, *e.g.,* Varian, 1985, 1989). This proposition is somewhat different because it concerns second-degree price discrimination. The similarity results because, in this model, second-degree price discrimination cannot improve welfare via a different distribution of quality levels *vis-à-vis* linear pricing; indeed, here, different quality levels reduce welfare. Hence, the only way in which second-degree price discrimination can enhance welfare is if it increases the volume of trade.

inequality for strictly concave functions when there is variation in what is being averaged. ■

As of this writing, ISPs in the United States do not charge application providers for last-mile service to households. Hence, the current situation is one which (10) holds and welfare, given existing bandwidth, is maximized. It follows, from the corollary, that moving away from network neutrality—holding bandwidth constant—would be welfare reducing.

On the other hand, suppose ISPs were already charging application providers for last-mile service and, as a consequence, some application providers had dropped out (*i.e.,* expression (10) failed to hold in equilibrium). By the usual logic of price discrimination, the ability to price discriminate means the monopolist has less motive to exclude types at the bottom; that is, when $\underline{\theta}_{\text{LP}} > \theta_0$, we should expect $\underline{\theta}_{\text{PD}} < \underline{\theta}_{\text{LP}}$. This effect is beneficial to welfare. *A priori*, it is not obvious whether this beneficial effect outweighs the negative effect due to misallocation (from a welfare perspective) of the bandwidth. In fact, via examples, we can show that either effect can be dominant; that is, either linear pricing or price discrimination can be welfare superior. For instance, suppose that $\theta$ is distributed uniformly on $[\theta_0, 1]$, $B = 2$, and $\omega(p^*) = \sigma = \pi = 1$. Suppose $\theta_0 = 0$. Then calculations reveal that, under linear pricing, $\underline{\theta}_{\text{LP}} \approx .457$ and welfare is approximately 1.778.[10] In contrast, under price discrimination, $\underline{\theta}_{\text{PD}} \approx .407$ and welfare is approximately 1.792 (roughly 1% greater). On the other hand, if $\theta_0 = .05$, then $\underline{\theta}_{\text{LP}} \approx .427$ and welfare under linear pricing is approximately 1.856; whereas $\underline{\theta}_{\text{PD}}$ remains .457 and welfare is approximately 1.838 (roughly 1% lower).

## 7 Dynamic Issues

The analysis to this point has treated bandwidth, $B$, as an exogenously given constant. In reality, the ISP determines $B$ via its investments in capacity. In this section, we consider the implications of various pricing regimes on the ISP's incentives to invest in bandwidth.

**Proposition 9.** *Suppose the ISP is limited to linear pricing. Then it will invest too little in bandwidth relative to the welfare-maximizing amount. Moreover, if it faces a binding cap on the hookup fee it can charge application providers, then it will invest less in bandwidth than it would absent the cap.*

**Proof:** Let $W_{\text{neut}}(B)$ equal the expression on the righthand side of (7); that is, $W_{\text{neut}}(B)$ is maximum welfare under network neutrality. Similarly, let

$$W_{\text{LP}}(B) = (\pi + \sigma)\sqrt{\frac{B\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)}{\omega(p^*)}}\,.$$

[10] A Mathematica program for constructing this and other examples is available from the authors upon request.

Because $\mu = \mathcal{I}(\theta_0, \underline{\theta}_{\text{LP}}) + \mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)$, $\theta_0 \geq 0$, and $\partial\underline{\theta}_{\text{LP}}/\partial B = 0$ (the last from Proposition 4), it follows that $W'_{\text{neut}}(B) \geq W'_{\text{LP}}(B)$; that is, the marginal social return to bandwidth is no greater under linear pricing than under neutrality.

Observe that

$$\int_{\underline{\theta}_{\text{LP}}}^{\theta_N} \Pi(\theta) dF(\theta) = \frac{\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)}{t_{\text{LP}}}\pi - s\big(1 - F(\underline{\theta}_{\text{LP}})\big)$$
$$= \pi\sqrt{\frac{B\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)}{\omega(p^*)}} - \pi\underline{\theta}_{\text{LP}}\big(1 - F(\underline{\theta}_{\text{LP}})\big)\sqrt{\frac{B}{\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)\omega(p^*)}}\,, \quad (26)$$

where the first equality follows from the definition of a type-$\theta$ application provider's net profit (expression (4)) and the second from (11) and (13). The ISP captures all surplus except that captured by the application providers; hence, its profit equals

$$\mathcal{P}(B) \equiv W_{\text{LP}}(B) - \int_{\underline{\theta}_{\text{LP}}}^{\theta_N} \Pi(\theta) dF(\theta) = W_{\text{LP}}(B) - Z\sqrt{B}\,,$$

where $Z > 0$ from (26) and because almost every application provider that operates makes a positive profit. Using the envelope theorem,

$$\mathcal{P}'(B) = W'_{\text{LP}}(B) - Z\frac{1}{2\sqrt{B}} < W'_{\text{LP}}(B) \leq W'_{\text{neut}}(B)\,. \quad (27)$$

It follows that the value of $B$ that maximizes $\mathcal{P}(B) - C(B)$, where $C(\cdot)$ is the cost-of-bandwidth function, is lower than the $B$ that maximizes $W_{\text{neut}}(B) - C(B)$.

To establish the "moreover" claim, observe that

$$\mathcal{P}(B) = \sqrt{\frac{B}{\omega(p^*)}}\left(\sigma\sqrt{\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)} + \pi\underline{\theta}_{\text{LP}}\big(1 - F(\underline{\theta}_{\text{LP}})\big)\sqrt{\frac{1}{\mathcal{I}(\underline{\theta}_{\text{LP}}, \theta_N)}}\right)\,. \quad (28)$$

From expressions (17) and (18), $\underline{\theta}_{\text{LP}}$ maximizes the expression in large parentheses. A binding cap on the hookup fee the ISP can charge application providers is equivalent to forcing the ISP to choose a lower $\underline{\theta}$. It follows, then, from (28) that the ISP's marginal profit from bandwidth expansion would be lower under a binding cap than absent such a cap. ∎

Proposition 9 indicates that there could be a fundamental distinction between efficient policy in a static setting and one in a dynamic setting. For instance, consider the example used above where $\theta$ is distributed uniformly on $[.05, 1]$. Given a fixed bandwidth, compelling the ISP to serve all application providers would raise welfare by 10%. On the other hand, letting the cost of bandwidth be $\kappa B$, it can be shown that the unrestricted ISP will build 1.649

times as much bandwidth as it would were it constrained.[11] In this example, the greater bandwidth does not fully overcome the static inefficiency, but once the dynamic consequences are considered welfare with a price cap is only 2% greater than were the ISP unrestricted.

What is the effect of permitting the ISP to price discriminate on its incentives to invest in bandwidth? As the following proposition shows, price discrimination provides stronger incentives to invest than does linear pricing.

**Proposition 10.** *If price discrimination is more profitable than linear pricing for all amounts of (total) bandwidth, $B$, then the* ISP *will invest in more bandwidth if it can engage in price discrimination than if it is limited to linear pricing.*

**Proof:** The ISP's profit under linear pricing is given by (28). Define $\phi_\ell$ and $\phi_f$ as the proportion of the bandwidth assigned the low-speed and fast segments, respectively. The ISP's profit under price discrimination can then be written as

$$\sqrt{\frac{B}{\omega(p^*)}}\Bigg(\sigma\left(\sqrt{\phi_\ell \mathcal{I}(\underline{\theta}_{\text{PD}},\hat{\theta})}+\sqrt{\phi_f \mathcal{I}(\hat{\theta},\theta_N)}\right)+\underline{\theta}_{\text{PD}}\pi\big(1-F(\underline{\theta}_{\text{PD}})\big)\sqrt{\frac{\phi_\ell}{\mathcal{I}(\underline{\theta}_{\text{PD}},\hat{\theta})}} \\ +\hat{\theta}\pi\big(1-F(\hat{\theta})\big)\Big(\sqrt{\frac{\phi_f}{\mathcal{I}(\hat{\theta},\theta_N)}}-\sqrt{\frac{\phi_\ell}{\mathcal{I}(\underline{\theta}_{\text{PD}},\hat{\theta})}}\Big)\Bigg) \quad (29)$$

(to see the derivation of (29) spelled out in detail, consider the derivation of expression (35) in the appendix). Observe, *given $B$*, the ISP's optimization program is to choose $\underline{\theta}_{\text{PD}}$, $\hat{\theta}$, and $\phi_\ell$ (with $\phi_f \equiv 1-\phi_\ell$) to maximize the expression in the largest parentheses in (29) (*i.e.,* to maximize the term that $\sqrt{B/\omega(p^*)}$ multiplies). *Note this proves that the* ISP*'s choice of segments and bandwidth allocation is* independent *of total bandwidth.* By assumption, the optimal value of (29) is greater than (28). It is readily seen, therefore, that the marginal profit from bandwidth expansion is greater given price discrimination than under linear pricing. ■

A corollary of Propositions 7 and 10 is

**Corollary 3.** *Assume the hazard rate associated with the distribution of application-provider types, $h(\cdot)$, satisfies the strict monotone hazard rate property. Then the* ISP *will invest in more bandwidth if it can engage in price discrimination than if it is limited to linear pricing.*

Note that the proof of Proposition 10 also established:

**Proposition 11.** *Given price discrimination, the number (measure) of application providers receiving low-speed service and the number (measure) receiving*

[11] For this example, the value of $\kappa$ does not matter for calculating ratios and percentage increases.

*fast service in equilibrium are independent of the* ISP*'s bandwidth. Furthermore, the proportion of total bandwidth allocated to low-speed service and the proportion allocated to fast service in equilibrium are independent of the* ISP*'s bandwidth.*

Proposition 10 reaches a somewhat different conclusion than reached by Choi and Kim (in press) about the effect allowing price discrimination has on the ISP's investment incentives. In Choi and Kim's article, the amount of content sent is independent of the ISP's capacity. Consequently, expanding capacity serves only to increase speed on both segments. Because the low-speed segment is now faster, the surplus a high-type application provider would get if it switched to low-speed service is greater, which means the amount of surplus the ISP can capture from such an application provider selling it the fast service is reduced. This effect reduces the benefit the ISP gets from increasing capacity. In essence, this is a standard story for second-degree price discrimination via quality distortions: Improving the low type's quality reduces the rent that can be captured from the high type. One caveat, though, is that the fast service is also becoming faster. This increases a high-type application provider's willingness to pay for fast service. As Choi and Kim observe, which of the relevant effects dominates is *a priori* ambiguous.

In our model, as reflected by Propositions 4 and 11, discrimination is based on the *proportion* of the bandwidth allocated to a service. In essence, an increase in total bandwidth is irrelevant for discrimination because it is essentially being absorbed by the transmission of more content.

Consider, again, the example in in which $\theta$ is distributed uniformly on $[.05, 1]$. Under those assumptions, the ISP will install 5% more bandwidth under price discrimination than under linear pricing. Despite the greater bandwidth, the static inefficiency due to price discrimination dominates slightly. Hence, welfare—even considering dynamic effects—under linear pricing is 0.7% greater than under price discrimination.

# 8 Extensions

In this section, we briefly consider two extensions of our model.

## 8.1 Zero-Price Slow Lane

A proposal that has been made in the ongoing network-neutrality debate has been to have the low-speed service priced at zero, while ISPs can charge for the fast service. This is identical to the price discrimination scheme considered above, except for the constraints $s_\ell = 0$ and $\underline{\theta} = \theta_0$. By revealed preference, such a regime would be worse for the ISP than the second-degree price discrimination scheme it would choose absent those constraints. Furthermore, if the ISP were not further constrained to allocate some bandwidth to the low-speed service, it would elect to allocate it none: In essence, given the constraints $s_\ell = 0$ and $\underline{\theta} = \theta_0$, but no constraint on $B_\ell$, the situation is identical to the situation of linear pricing considered in Section 4.

Hence, a zero-price slow lane is a distinct policy *vis-à-vis* permitting linear pricing, but prohibiting price discrimination only if the ISP is obliged to dedicate some bandwidth to the low-speed service. Let $\underline{B}_\ell$ be the minimum amount of bandwidth that must be provided the low-speed service. Given Propositions 1 and 2, the welfare-maximizing policy would be to allocate all the bandwidth to the "slow lane." Because all the relevant functions are continuous, it follows, invoking the implicit function theorem where necessary, that

**Proposition 12.** *There exists a policy, under which a sufficiently great minimum bandwidth, $\underline{B}_\ell$, to the zero-price slow lane is mandated, that welfare dominates allowing the* ISP *to freely price discriminate. If linear pricing would result in some application providers' not being served, then there exists such a policy that would also welfare dominate linear pricing.*

### 8.2 Heterogeneous Households

The analysis so far has assumed that households are homogenous. There are numerous ways in which this assumption could be relaxed, not all of which can be explored in a single paper. As one model of heterogenous consumers, let $\mathcal{A} \subseteq \mathbb{R}_+$ be a set of household types and suppose that a household of type $\alpha \in \mathcal{A}$ obtains a benefit of

$$\int_{\theta_0}^{\theta_N} \left( \int_0^{x(\theta)} \alpha m\Big(\frac{x\tau(\theta)}{\theta}\Big) dx \right) dF(\theta) \,,$$

where $x(\theta)$ is, again, the consumption of content from provider $\theta$ and $m(\cdot)$ has the same properties assumed previously. Observe that higher-type households have a greater marginal utility from content than do lower-type households.

It is readily shown that an $\alpha$-type household's demand for content from provider $\theta$ at price $p$ is

$$x(p,\theta,\alpha) = \frac{\theta}{\tau(\theta)} m^{-1}\left(\frac{p}{\alpha}\right) \,.$$

Assume each application provider is limited to linear pricing; in particular, no application provider can discriminate on the basis of household type. Let $p(\theta)$ be the price charged by application provider $\theta$. Although, as we will show shortly, all application provider's who sell will charge the same price in equilibrium (as before), for the moment we allow different prices. As a convention, set $p(\theta) = \infty$ if application provider $\theta$ is not selling in equilibrium.

A household's consumer surplus is

$$\int_{\theta_0}^{\theta_N} \left( \int_{p(\theta)}^{\infty} \frac{\theta}{\tau(\theta)} m^{-1}\left(\frac{p}{\alpha}\right) dp \right) dF(\theta) \equiv \int_{\theta_0}^{\theta_N} \frac{\theta}{\tau(\theta)} \hat{\sigma}\big(\alpha, p(\theta)\big) dF(\theta) \,. \qquad (30)$$

**Lemma 2.** *The greater a household's type, the weakly greater its consumer surplus from an application provider.*

**Proof:** It is sufficient to show that $\hat{\sigma}(\cdot, p)$ is a non-decreasing function for all $p$. Hence, it is sufficient to show that $m^{-1}(p/\alpha)$ is non-decreasing in $\alpha$ for all $p$. This follows immediately because $m(\cdot)$ and, thus, $m^{-1}(\cdot)$ are decreasing functions. ■

An immediate corollary of Lemma 2 and expression (30) is

**Corollary 4.** *If given the expected pricing of the application providers, a household of a given type wishes to hookup (is willing to pay $\eta$), then so too is any household of higher type. Conversely, if a household of given type doesn't wish to hookup, then neither does any household of lower type.*

Let $G : \mathcal{A} \to [0, 1]$ denote the distribution over household types. Assume the expectation of $\omega(p/\alpha)$ with respect to $G$ exists. Let

$$\Omega(p) = \int_{\mathcal{A}} m^{-1}\left(\frac{p}{\alpha}\right) dG(\alpha)$$

denote that expectation. Similarly, define

$$\Omega(p, \alpha) = \int_{\{z \in \mathcal{A} | z \geq \alpha\}} m^{-1}\left(\frac{p}{z}\right) dG(z)\,.$$

If $\underline{\alpha} = \inf\{\alpha \in \mathcal{A} | \alpha \text{ hooks up}\}$, then an application provider's pricing problem is

$$\max_p (a + p - c) \int_{\{\alpha \in \mathcal{A} | \alpha \geq \underline{\alpha}\}} x(p, \theta, \alpha) dG(\alpha) = \max_p (a + p - c) \frac{\theta}{\tau(\theta)} \Omega(p, \underline{\alpha})\,.$$

It follows, as was the case in maximizing (3), that the solution, $p^*(\underline{\alpha})$, is independent of $\theta$. Define $\omega^*(\underline{\alpha}) = \Omega\big(p^*(\underline{\alpha}), \underline{\alpha}\big)$. As will be seen shortly, $\omega^*(\underline{\alpha})$ plays a role similar to the one $\omega(p^*)$ played in the earlier analysis.

Define

$$\pi(\underline{\alpha}) = \big(a + p^*(\underline{\alpha}) - c\big)\omega^*(\underline{\alpha}) \quad \text{and} \quad \sigma(\underline{\alpha}) = \int_{\{\alpha \in \mathcal{A} | \alpha \geq \underline{\alpha}\}} \hat{\sigma}\big(\alpha, p^*(\underline{\alpha})\big) dG(\alpha)\,.$$

As the notation is meant to suggest, $\pi(\underline{\alpha})$ and $\sigma(\underline{\alpha})$ play roles similar to $\pi$'s and $\sigma$'s in the earlier analysis. Exploiting that similarity, it is straightforward to extend Propositions 1 and 2 as follows:

**Proposition 13.** *Hold fixed the number of households who connect (i.e., fix $\underline{\alpha}$). Then neutrality with respect to the treatment of application providers is weakly welfare superior to any division of the bandwidth in which all segments of the application providers are allocated a positive portion of the bandwidth and strongly welfare superior to any division that excludes some segments.*

Proposition 13 can be read as showing that conditional on whatever distortions might arise from the ISP's pricing to households, there is no welfare benefit to be gained from allowing the ISP to exclude or discriminate among application providers.

Redoing all the analysis of the earlier sections with heterogenous households is beyond the scope of this paper. Given many aspects of the welfare analysis were already ambiguous, the added complication of heterogeneous households will serve only to further muddy the waters.

A new question that can be asked is whether the ISP should be permitted to discriminate against households by providing different households different fractions of the bandwidth? Although it can be shown that such discrimination cannot lead to greater profits for the application providers (holding constant the households served), the effects on household welfare are ambiguous. The ambiguity arises because the ensuing change in allocation of content among households caused by bandwidth division make it impossible to reach a general prediction about overall household welfare.

## 9 Concluding Remarks

A platform in a two-sided market has the possibility of setting prices in both sides of the market. In some markets, both sides pay the platform. For example, a game platform collects both from end users, and independent games developers. In other markets, the platform finds it optimal to have one side pay it while the other side receives money from the platform. For example, a credit-card network receives money from merchants but pays money or gives other benefits to card holders; an operating system such as Windows charges a positive price to users but subsidizes software applications.

On the Internet, since its commercialization, a pricing regime known as network neutrality has prevailed: Parties pay only the ISP through which they purchase access and are charged prices that reflect the volume of information sent and received, but which do not depend on its content or origin. In 2005, certain residential ISPs (residential phone and cable TV companies) sought to deviate from network neutrality by demanding direct payments from content and applications providers if they wished to reach the ISPs' residential customers, as well as additional payments for prioritization of information packets. In October 2009, the FCC proposed to encapsulate, as law, the network neutrality tradition of the Internet, and, as of this writing, this proceeding has not ended.

This paper has analyzed the private and social benefits of allowing residential ISPs to impose such fees on the content and application providers seeking to reach their residential subscribers. The situation is different from many other two-sided markets because the existence of congestion present on the Internet does not arise in many other networks, such as in the virtual network between operating systems and applications or the payment-card networks. Unlike some models analyzing network neutrality, our model assumes no vertical integration of ISPs into content or applications. Such vertical integration would create incentives for an ISP to prioritize its own content. Absent vertical foreclosure fears and given the theoretical possibility that variable prioritization of information packets when consumers value speedier transmission more for some content than other could be beneficial given overall congestion, our model might seem "tipped" against network neutrality.

Such bias notwithstanding, we in fact find the opposite: Network neutrality is welfare superior to bandwidth subdivision (price discrimination via the use of different "lanes" of service). The key insight is that attempts to speed the content that is most time sensitive to consumers leads consumers to purchase more of that content, recongesting those lanes. This reduces the gain from such attempts and does so enough that these gains no longer outweigh the losses from slowing other content.

We investigate what form of pricing an unregulated monopoly ISP would adopt. Even though we allow a monopoly ISP to charge fixed fees, the ability of the ISP to capture the greater consumer surplus generated by network neutrality is sufficiently limited that the ISP will generally find it more profitable to engage in price discrimination that is welfare inferior to network neutrality. Hence, network neutrality will exist only if mandated via regulation. On the other hand, it is possible that the policy margin is between permitting linear pricing by the ISP that excludes some application providers and price discrimination. In this case, we show that welfare can be greater under price discrimination than linear pricing. A necessary condition for this to occur is that the set of application providers served in equilibrium is greater under discrimination than linear pricing. This provides a straightforward way to test whether a violation of equal treatment through price discrimination is desirable or not: Will it expand the set of applications providers or not?

ISPs have claimed that they need to be allowed to charge network and applications providers so that they can invest in new bandwidth. We show that claim is justified insofar as discrimination indeed increases the ISP's incentives to expand bandwidth. This is in contrast with the model of Choi and Kim (in press) that shows that the ISP can have an incentive to reduce bandwidth to increase the price difference between its "fast" and "slow" lanes. Although we find justification for the ISPs' claims in this regard, whether such investment incentives trump the static inefficiency of discrimination are, *a priori*, ambiguous.

One of the possible "solutions" discussed by parties in the FCC proceedings has been a setup under which the standard or "slow" lane stays at zero pricing while content and applications providers have the option to pay for prioritized "managed services." In our setup, this is similar to a zero-price slow lane and a positive-price fast lane. We show that the ISP would set the slow lane size to zero (*i.e.,* degrade it the point that it is effectively unusable) unless constrained by a regulator to a minimum slow lane size. We further observe that the welfare-maximizing policy would be to allocate *all* bandwidth to the slow lane. But even short of that ideal, there is a sufficient allocation of bandwidth to the slow lane that dominates unrestricted price discrimination and, provided linear pricing would not otherwise yield universal service, dominates linear pricing.

Work remains. Although we have sought to be as general as possible in our analysis, the complexity of the environment dictates that at least some restrictions on functional forms are necessary to obtain a tractable model. We would argue that the restrictions we've imposed are plausible, at least with respect to what they imply about marginal and cross-marginal effects. On the other hand, alternative assumptions, especially with respect to household (consumer) preferences, could be made and they are worth considering. The recongestion effect we identify, which has been overlooked in earlier analyses, should remain an important effect in alternative models. Nonetheless, it is worth knowing more about the properties of preferences that lead to equilibria—such as those derived here—in which the recongestion effect dominates the gains from prioritizing more time-sensitive content over less time-sensitive content. Another future direction for research is how price discrimination across households (as some ISPs do with different speed services) affects conclusions about the welfare effects of allowing or prohibiting residential ISPs to discriminating against application providers. At one level, discrimination across households simply puts households into different customer classes and the analysis of the paper could be read as analyzing application providers selling to a single such class across the total bandwidth allocated to that class. At a more general level, if the application providers cannot price discriminate across these different classes, then the analysis could be more complex and nuanced.[12]

## Appendix A: Proofs and Technical Matters

**Lemma A.1.** *Let $\Psi : [z_0, z_1] \to [0,1]$ be a differentiable distribution function, where $z_0 \geq 0$. Assume, too, that the corresponding hazard rate satisfies the strict monotone hazard rate property (i.e., the hazard rate is increasing). Then the function that maps $z$ to*

$$\frac{\int_z^{z_1} \zeta d\Psi(\zeta)}{z\big(1-\Psi(z)\big)} \tag{31}$$

*is decreasing in $z$.*

**Proof:** Expanding the numerator in (31) via integration by parts reveals (31) equals

$$\frac{z\big(1-\Psi(z)\big) + \int_z^{z_1} \big(1-\Psi(\zeta)\big)d\zeta}{z\big(1-\Psi(z)\big)} = 1 + \frac{\int_z^{z_1} \big(1-\Psi(\zeta)\big)}{z\big(1-\Psi(z)\big)}\,.$$

The result follows, therefore, if

$$\frac{\int_z^{z_1} \big(1-\Psi(\zeta)\big)d\zeta}{z\big(1-\Psi(z)\big)}$$

is decreasing in $z$; or, because log is a monotonic transformation, if

$$\log\left(\frac{\int_z^{z_1} \big(1-\Psi(\zeta)\big)d\zeta}{1-\Psi(z)}\right) - \log(z)$$

[12] If different household classes had different marginal utility functions, $m(\cdot)$, then they would have different $\omega(\cdot)$ functions, which means the application providers would wish to set different $p^*$s for different classes. If the application providers cannot do this, then the proper analysis of the overall market is no longer necessarily simply repeating the analysis of the paper for each household class.

is decreasing in $z$. Given $-\log(z)$ is decreasing in $z$, the result follows if

$$\frac{\int_z^{z_1}\big(1-\Psi(\zeta)\big)d\zeta}{1-\Psi(z)} \tag{32}$$

is decreasing in $z$. The monotone hazard rate property means that

$$\frac{\Psi'(z)}{1-\Psi(z)}\,,$$

the hazard rate, is increasing; hence,

$$\frac{-\Psi'(z)}{1-\Psi(z)} = \frac{d}{dz}\log\big(1-\Psi(z)\big)$$

is decreasing, which means $1-\Psi(z)$ is log concave. By Lemma A1 of Hermalin and Katz (2004), if $1-\Psi(z)$ is log concave, then so too is $\int_z^{z_1}\big(1-\Psi(\zeta)\big)d\zeta$. Hence,

$$\frac{-\big(1-\Psi(z)\big)}{\int_z^{z_1}\big(1-\Psi(\zeta)\big)d\zeta}$$

is a decreasing function of $z$. It follows immediately that (32) must be too. ■

**Proof of Proposition 7:** The line of argument is as follows. Let $\underline{\theta}$ be the profit-maximizing cutoff under linear pricing (*e.g.,* the solution to (18)). Consider a price-discrimination scheme with the same $\underline{\theta}$. If we can show such a scheme exists that is more profitable than linear pricing, then we're done because a profit-maximizing scheme cannot be worse than that scheme and, thus, it must be more profitable than linear pricing.

Observe we can write the ISP's profit from the application providers under such a price discrimination scheme as

$$\begin{aligned}\big(F(\hat\theta)-F(\underline{\theta})\big)s_\ell &+ \big(1-F(\hat\theta)\big)s_f \\ &= \big(1-F(\underline{\theta})\big)s_\ell + \big(1-F(\hat\theta)\big)\pi\hat\theta\left(\frac{1}{t_f}-\frac{1}{t_\ell}\right)\,, \end{aligned}\tag{33}$$

where the righthand side follows from the left given expression (21).

Profit under linear pricing, using expressions (13) and (17), is

$$\sigma\sqrt{\frac{B\mathcal{I}(\underline{\theta},\theta_N)}{\omega(p^*)}} + \underline{\theta}\pi\big(1-F(\underline{\theta})\big)\sqrt{\frac{B}{\omega(p^*)\mathcal{I}(\underline{\theta},\theta_N)}}\,. \tag{34}$$

Profit under price discrimination, using (21)–(24) and (33), is

$$\begin{aligned}\frac{\sigma}{\sqrt{\omega(p^*)}}\left(\sqrt{B_\ell\mathcal{I}(\underline{\theta},\hat\theta)}+\sqrt{B_f\mathcal{I}(\hat\theta,\theta_N)}\right) + \underline{\theta}\pi\big(1-F(\underline{\theta})\big)\sqrt{\frac{B_\ell}{\omega(p^*)\mathcal{I}(\underline{\theta},\hat\theta)}} \\ + \hat\theta\pi\big(1-F(\hat\theta)\big)\left(\sqrt{\frac{B_f}{\omega(p^*)\mathcal{I}(\hat\theta,\theta_N)}}-\sqrt{\frac{B_\ell}{\omega(p^*)\mathcal{I}(\underline{\theta},\hat\theta)}}\right)\,. \end{aligned}\tag{35}$$

Observe that a $1/\sqrt{\omega(p^*)}$ can be factored out of both expressions (34) and (35). Because our interest is in comparing these two expressions, it follows that there is no loss of generality with respect to the analysis at hand in normalizing $\omega(p^*)$ to 1, which we now do.

Fix a $B_\ell \in (0, B)$. Define $\hat{\theta} \in (\underline{\theta}, \theta_N)$ such that

$$\frac{B}{\mathcal{I}(\underline{\theta}, \theta_N)} = \frac{B_\ell}{\mathcal{I}(\underline{\theta}, \hat{\theta})} . \tag{36}$$

That such a $\hat{\theta}$ exists follows because $\mathcal{I}(\underline{\theta}, \cdot)$ is a continuous and increasing function.

Next, we show that with this $B_\ell$ and $\hat{\theta}$, the ISP profits are the same under linear pricing (*i.e.,* as given in expression (34)) and this particular price discrimination scheme (*i.e.,* as given in expression (35)). To prove this claim, observe that for such a $B_\ell$ and $\hat{\theta}$, the difference in the large parentheses in (35) is zero:

$$\frac{B_f}{\mathcal{I}(\hat{\theta}, \theta_N)} = \frac{B_\ell \left( \frac{\mathcal{I}(\underline{\theta}, \theta_N)}{\mathcal{I}(\underline{\theta}, \hat{\theta})} - 1 \right)}{\mathcal{I}(\hat{\theta}, \theta_N)} = \frac{B_\ell \big( \mathcal{I}(\underline{\theta}, \theta_N) - \mathcal{I}(\underline{\theta}, \hat{\theta}) \big)}{\mathcal{I}(\hat{\theta}, \theta_N) \mathcal{I}(\underline{\theta}, \hat{\theta})} = \frac{B_\ell}{\mathcal{I}(\underline{\theta}, \hat{\theta})} , \tag{37}$$

where the first equality follows from (36) and because $B = B_f + B_\ell$ and the last equality follows because $\mathcal{I}(x, z) - \mathcal{I}(x, y) = \mathcal{I}(y, z)$. Moreover,

$$\sqrt{B_\ell \mathcal{I}(\underline{\theta}, \hat{\theta})} + \sqrt{B_f \mathcal{I}(\hat{\theta}, \theta_N)} = \sqrt{\frac{B \mathcal{I}(\underline{\theta}, \hat{\theta})^2}{\mathcal{I}(\underline{\theta}, \theta_N)}} + \sqrt{\frac{B \mathcal{I}(\hat{\theta}, \theta_N)^2}{\mathcal{I}(\underline{\theta}, \theta_N)}}$$

$$\sqrt{\frac{B}{\mathcal{I}(\underline{\theta}, \theta_N)}} \big( \mathcal{I}(\underline{\theta}, \hat{\theta}) + \mathcal{I}(\hat{\theta}, \theta_N) \big) = \sqrt{B \mathcal{I}(\underline{\theta}, \theta_N)} , \tag{38}$$

where the first equality follows using (36) and (37) and the last because $\mathcal{I}(x, z) - \mathcal{I}(x, y) = \mathcal{I}(y, z)$. Using (36)–(38), it is readily seen that (34) and (35) are equal; that is, with this "scheme," the ISP's profit is the same as it would be under linear pricing.

Holding $B_\ell$ constant, it follows that if (i) the derivative of (35) with respect to $\hat{\theta}$ is positive and (ii) the derivative of the difference in large parentheses in (35) is also positive, then a scheme with the same $B_\ell$ but slightly larger cutoff $\hat{\theta}$ would (i) be more profitable than linear pricing and (ii) induce self-selection (*i.e.,* satisfy the requirement that $t_f < t_\ell$). The derivative of (35) with respect to $\hat{\theta}$, using (36) and (37), is

$$- \frac{1}{2} \underline{\theta} \big( 1 - F(\underline{\theta}) \big) \left( \frac{B_\ell}{\mathcal{I}(\underline{\theta}, \hat{\theta})} \right)^{3/2} \frac{\hat{\theta} F'(\hat{\theta})}{B_\ell}$$

$$+ \frac{1}{2} \hat{\theta} \big( 1 - F(\hat{\theta}) \big) \left( \left( \frac{B_f}{\mathcal{I}(\hat{\theta}, \theta_N)} \right)^{3/2} \frac{1}{B_f} + \left( \frac{B_\ell}{\mathcal{I}(\underline{\theta}, \hat{\theta})} \right)^{3/2} \frac{1}{B_\ell} \right) \hat{\theta} F'(\hat{\theta}) . \tag{39}$$

The second line of (39), which is positive, shows that condition (ii) is met (*i.e.,* $1/t_f - 1/t_\ell$ is increasing in $\hat{\theta}$). Using (37), the sign of (39) is the same as the sign of

$$-\underline{\theta}\big(1 - F(\underline{\theta})\big)\frac{1}{B_\ell} + \hat{\theta}\big(1 - F(\hat{\theta})\big)\left(\frac{1}{B_f} + \frac{1}{B_\ell}\right) . \tag{40}$$

Expression (40) is positive if

$$\underline{\theta}\big(1 - F(\underline{\theta})\big) < \hat{\theta}\big(1 - F(\hat{\theta})\big)\left(\frac{1}{\frac{\mathcal{I}(\underline{\theta},\theta_N)}{\mathcal{I}(\underline{\theta},\hat{\theta})} - 1} + 1\right) , \tag{41}$$

where (36) has been used to substitute out $B_\ell$. The righthand side of (41) can be rewritten as

$$\hat{\theta}\big(1 - F(\hat{\theta})\big)\frac{\mathcal{I}(\underline{\theta},\theta_N)}{\mathcal{I}(\hat{\theta},\theta_N)} .$$

It follows, then, that (41) holds given Lemma A.1. So condition (i) is satisfied; that is, a small increase in $\hat{\theta}$ increases profit, so there must be a price-discrimination scheme that yields the ISP greater profit than linear pricing. ■

## References

**Cheng, H. Kenneth, Subhajyoti Bandyopadhyay, and Hong Guo**, "The Debate on Net Neutrality: A Policy Perspective," *Information Systems Research*, March 2011, *22* (1). in press.

**Choi, Jay Pil and Byung-Cheol Kim**, "Net Neutrality and Investment Incentives," *RAND Journal of Economics*, in press.

**Economides, Nicholas and Joacim Tåg**, "Net Neutrality on the Internet: A Two-sided Market Analysis," 2009. Working Paper, Stern School of Business, New York University.

**Hermalin, Benjamin E. and Michael L. Katz**, "Sender or Receiver: Who Should Pay to Exchange an Electronic Message?," *RAND Journal of Economics*, Autumn 2004, *35* (3), 423–448.

**_ and _** , "The Economics of Product-Line Restrictions with an Application to the Network Neutrality Debate," *Information Economics and Policy*, 2007, *19*, 215–248.

**Krämer, Jan and Lukas Wiewiorra**, "Network Neutrality and Congestion-Sensitive Content Providers: Implications for Service Innovation, Broadband Investment and Regulation," 2010. Working Paper, Universität Karlsruhe.

**Varian, Hal R.**, "Price Discrimination and Social Welfare," *American Economic Review*, September 1985, *75* (4), 870–875.

_ , "Price Discrimination," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, 1st ed., Vol. 1, Amsterdam: Elsevier, 1989, pp. 597–654.