of the exclusionary rights. At most, holdout problems will limit somewhat the number of exclusionary rights sold, preventing the purchaser from achieving a complete monopoly. Nevertheless, they are quite unlikely to cause the strategy to fail altogether.

Suppliers may indeed realize that the purchaser is buying exclusionary rights and therefore may anticipate that, if the purchaser succeeds, unrestrained suppliers subsequently will be able to charge the rivals a higher price for their inputs.[202] A sophisticated supplier will take into account the likelihood of this higher price in calculating its opportunity cost of selling an exclusionary right to the purchaser.[203] This action may raise the cost of the rights to the purchaser and limit the number it wishes to buy. However, it will not eliminate the purchaser's demand for rights altogether. The power over price gained by the purchaser creates inherent, mutual gains to the purchaser and the supplier from the sale of the right to the purchaser, gains large enough to offset the higher price.[204] As stated earlier, this benefit to the purchaser increases with the size of its market share.[205]

---

202. This analysis does not apply to overbuying an input, only to explicit exclusionary rights. In Overbuying, all suppliers charge the higher, market-clearing input price, even to the purchaser of the exclusionary rights.

203. For example, altering somewhat the conditions set forth in note 199, *supra*, suppose one supplier anticipates that the purchaser's exclusionary rights strategy will surely succeed whether or not the supplier agrees to sell its rights. Suppose the supplier forecasts that the post-exclusion input price for remaining units of the input will rise above the current input price of $200 to $225. The supplier would make the following calculation in deciding whether to sell exclusionary rights. Assuming that a restrained supplier could still export to Samoa at a price of $190 less $5 additional transportation costs, and ignoring the possibility of a counteroffer by Klor's, the excluded firm, the supplier would require a bid from the purchaser of at least $40 in excess of the rivals' offer. Thus, a $40 increment is necessary. This $40 increment exceeds the $15 increment in the previous example in note 199, *supra*, of no counteroffers because the supplier calculates its potential revenue loss from exclusion on the basis of the anticipated post-restraint price of $225, instead of the current pre-restraint price of $200. This $40 does not include a counteroffer by Klor's because, in this case, Klor's has no incentive to make a counteroffer high enough to be acceptable to the suppliers. The suppliers would require payment of $25 per unit in exchange for the right to continue to buy at $200, for a total cost to Klor's of $225, the anticipated post-restraint price.

204. Building on the example set out in note 199, *supra*, when counteroffers are permitted, suppose that Broadway-Hale would have to pay each manufacturer $65 per unit of reduction in sales to Klor's. Assuming that before the exclusion Klor's bought 10 units from the manufacturers, Broadway-Hale's cost for totally excluding Klor's would be $650 (i.e., $65 per unit times 10 units). This cost of excluding Klor's would be offset by the higher prices Broadway-Hale could obtain for its output. For example, suppose that before the exclusion Broadway-Hale's sales were 90 units, for a pre-exclusion market share of 90%. Assume also that Broadway-Hale raises its post-exclusion price by $45, an amount less than the increase in Klor's unit costs (its $50 cost of alternative inputs), and that this price increase allows Broadway-Hale to hold its sales constant at its pre-exclusion level of 90 units. This assumption is equivalent to assuming that the market demand elasticity is approximately 0.5. In this case, Broadway-Hale's revenue will rise by $4050 (i.e., 90 units times $45 per unit), for an increase in net profit equal to $3400 (i.e., $4050-$650).

205. For example, in the hypothetical described in note 204, *supra*, if Broadway-Hale's output were higher, its gains from exclusion would be higher. *See also infra* note 208. This analysis is very similar, if not identical, to the previous analysis of counterstrategies. In the analyses of Lewis, *supra* note 198, and R. MACKAY, *supra* note 172, suppliers' anticipations of the higher post-exclusion input

HeinOnline -- 96 Yale L.J. 274 1986-1987

The second holdout problem, the supplier's incentive to capture more of those monopoly gains for itself, is more a matter of the distribution of monopoly profits among the parties than an issue of whether the exchange will take place.[206] Moreover, in the case of an input market with a large number of suppliers, competition among suppliers to sell exclusionary rights will prevent, or at least limit, this second form of holdout.[207]

Certain factors increase the likelihood that the potential purchaser can offer an amount sufficient to overcome potential holdout problems and induce the suppliers to grant exclusionary rights. First, as noted above, the larger the purchaser's market share, the greater its reward for achieving power over price, hence the greater its willingness to pay to achieve anticompetitive exclusion.[208] Second, the overall demand for the suppliers'

---

price effectively make the supply of exclusionary rights upward-sloping. That is, the more rights purchased, the higher the price the purchaser must pay. Assuming that the purchaser cannot price discriminate among suppliers, the purchaser has an incentive to act like a monopsonist, taking into account the higher price that must be paid for all rights if additional rights are bought. This higher price for rights, however, is offset by the additional gains to the purchaser arising from the higher price it will earn on its sales as a result of its enhanced market power. Because the purchase of more rights translates into a larger market share for the purchaser and a higher price received on more sales, it can afford to pay more for the rights. Indeed, as demonstrated by Mackay, the demand for rights can have an upward-sloping portion. R. MACKAY, *supra* note 172. The interaction of this supply and demand entails an equilibrium in which a purchaser with a significant initial market share gains additional, but incomplete, monopoly power. Only if the purchaser had an initial market share of zero (a "pure promoter," in Mackay's parlance) will the strategy fail altogether; otherwise, it will succeed. However, contrary to the "merger to monopoly" claims of Bork and others, Mackay emphasizes that the purchaser will be unable to achieve a complete monopoly (i.e., a market share of 100%).

206. Of course, such bargaining problems may cause a certain percentage of negotiations to fail. However, in the case of failed exclusionary rights negotiations, the purchaser often can shift to other potential sellers. Moreover, the fact that a small fraction of deals will fall through does not vitiate the proposition that a large fraction of deals to exclude rivals can succeed.

207. For example, the purchaser can avoid this type of holdout problem by committing to purchase any and all rights offered at some price or by committing to purchase some fixed number of rights from the suppliers who bid the lowest prices. In either case, competition among suppliers will prevent individual suppliers from holding out for a better deal. Of course, under this strategy, holdouts also may arise from suppliers' expectations that the purchaser will act opportunistically by violating its commitment to limit its purchases of exclusionary rights, and purchase additional rights subsequently at a higher price. Such a belief might entice some suppliers to wait for the higher price. As Mackay points out, in this event the purchaser can assuage the fears of such suppliers by offering a "most-favored-nations" provision in the agreement that would assure suppliers that higher price if more rights are purchased. *See* R. MACKAY, *supra* note 172, at 19–21. More generally, any credible commitment to purchase in a non-discriminatory fashion (e.g., employing a standard form contract with a most-favored-customers clause) can prevent individual sellers from holding out for a subsequent discriminatory high price. *See* Salop, *supra* note 51.

Warren Schwartz has suggested to us still another type of holdout problem—that suppliers may agree to exclude rivals, but then continue to sell to rivals secretly. That is, exclusionary rights agreements have enforcement problems analogous to the problem of cartel enforcement. We agree that such problems may exist in some cases. They are discussed in the analysis of Cartel Ringmaster. *See supra* Section VI.A.3. However, unlike covert, illegal cartels, exclusionary rights agreements will be enforced at low cost by contracts courts if they are not prohibited under the antitrust laws. Thus, the problems of enforcement should be less serious than in the standard cartel analysis. *See* Posner, *supra* note 103.

208. The importance of market share can be illustrated by altering the assumed sales of Broadway-Hale in the numerical example set out in note 204, *supra*. If Broadway-Hale's pre-exclusion

HeinOnline -- 96 Yale L.J. 275 1986-1987

product may be so broad that losing only those few buyers who compete with the purchaser will have negligible effects on the suppliers' revenues. That is, the suppliers may be able to sell the extra units with little or no price reduction to other buyers who do not compete with the exclusionary rights purchaser.[209]

To illustrate, these factors may have worked together to make feasible Alcoa's purchases of exclusionary rights from electric power suppliers.[210] During the early twentieth century, Alcoa was the only manufacturer of aluminum in the United States. It therefore stood to realize large gains from maintaining its power over price. If denying its potential competitors access to most, but not all, strategically located sources of electric power would have so raised these rivals' costs that it could maintain significant market power, Alcoa could have succeeded in excluding rivals and shared the gains from maintaining this market power with the utilities from which it bought exclusionary rights. The utilities may have sacrificed few, if any, electricity sales and thereby foregone few profits by their agreement because they could replace their lost sales by selling that electricity to other users outside the aluminum industry.

It thus appears that, in general, if a firm can gain power to raise price by raising its rivals' cost through exclusionary rights agreements, as we have argued will sometimes be the case, there is no reason to assume that counter-bidding by rivals or the inability to compensate suppliers will normally prevent these agreements from being executed.[211] We do not claim that rivals will never outbid purchasers for exclusionary rights or that suppliers will always find it profitable to grant them. For example, if

---

sales were 10 units (i.e., a market share of 50%) and its post-exclusion sales remain constant at 10 units, then the price increase of $45 only raises its revenue by $450 (i.e., 10 units times $45 per unit). (This price increase and quantity decrease assumes a market elasticity of 2.2, a higher elasticity than in the previous example.) Taking into account a $650 cost to purchase the exclusionary rights, the purchaser's net profit falls by $200 (i.e., $400-$650). Thus, in this second example, exclusion of Klor's would not be profitable. As these examples show, the profitability of exclusion depends on market share, the cost of the rights, the elasticity of demand and other variables. By changing the relative sizes of these variables, numerous examples of profitable complete or incomplete exclusion at far lower market shares or of unprofitable exclusion even at market shares approaching 100% can be constructed. The point of these examples is not to specify numerical thresholds for enforcement purposes, but to show how the variables interact.

209. Of course, this targeted reduction in its market is only possible when explicit exclusionary rights are sold, the markets can be separated, and arbitrage can be prevented.

210. See the previous discussion of *Alcoa*, *supra* Section IV.A.2.b.

211. Our analysis thus far assumes that the only cost borne by the purchaser of exclusionary rights is the price paid to suppliers for those rights. In some cases, however, especially those involving exclusionary use of governmental processes, exclusionary strategies also may raise the production costs of the purchaser, as well as the price paid for the rights. Profitability is more likely the greater the increase in the production costs of rivals relative to the increase in the purchaser's production costs. *See* S. SALOP & D. SCHEFFMAN, *supra* note 73; Salop, Scheffman & Schwartz, *supra* note 73. In the case of Real Foreclosure, the purchaser's relative cost increase will be smaller if its relative use of the input is less than its rivals, if it is vertically integrated into production of the input to a greater degree than its rivals, or if its purchase price is protected by bargaining power or a long term contract.

HeinOnline -- 96 Yale L.J. 276 1986-1987

there were 10,000 utilities, each of which could fully satisfy the demand of a large aluminum producer, Alcoa would have had little success with an exclusionary rights strategy. Further, single-firm monopoly suppliers will usually, but not always, be unwilling to convey exclusionary rights that give their purchasers power to raise price. That observation, however, does not undercut the central argument that suppliers who are not monopolists can benefit from such anticompetitive restraints or imply that a monopoly supplier always should be immunized from challenge for exclusionary conduct. Put another way, although exclusion is not always successful, one cannot assume that a competitive market for exclusionary rights is likely to lead to competitive product markets.[212] These conclusions alone should be sufficient to rebut assertions that antitrust law should be indifferent to claims of anticompetitive exclusion by vertical agreement or merger.

C. *Efficiency Justifications*

As explained above, exclusionary rights agreements may lower the costs of either purchasers or suppliers.[213] Indeed, costs savings are likely to be the only competitive effects of vertical restraints adopted and maintained in unconcentrated markets characterized by easy entry. Firms have incentives to seek power over price, but they also have incentives to integrate vertically when transaction costs exceed the benefits of conducting business through organized markets;[214] a priori, one cannot presume which of these incentives will explain a randomly selected exclusionary rights agreement. How, then, should antitrust policy address these dual incentives, these mirror-image potential results?

One possible response would be to treat all exclusionary rights agreements as lawful because (a) at least some are motivated by, and result only in, efficiencies, and (b) standards for assessing the probability of an-

212. This point is emphasized in the rent-seeking literature. *See* Posner, *supra* note 73; Tullock, *supra* note 73. An analogy to merger analysis might clarify the point. In markets unfettered by antitrust constraints on horizontal mergers, by merging a firm often could, in effect, buy from enough of its rivals the right to exclude them. This strategy would be profitable because the firm could share with its rivals the gains from the resulting monopoly. Similarly, a firm could acquire market power by purchasing from its rivals' suppliers the right to exclude its rivals. In neither case would the result be tolerable simply because it arose from open market deals. In short, one cannot assume that competition for the right to be the monopolist usually yields the efficient market outcome. Only if firms bid for the monopoly, not by offering a high price for the right to monopolize, but by bidding to supply high quality goods at low prices could efficiency be achieved. *See* Demsetz, *Why Regulate Utilities?*, 11 J.L. & ECON. 55 (1968); S. BORENSTEIN, ON THE EFFICIENCY OF COMPETITIVE MARKETS FOR OPERATING LICENSES (Mich. Inst. Pub. Pol'y Stud. Discussion Paper No. 226, Sept. 1985); D. SAPPINGTON & J. STIGLITZ, INFORMATION AND REGULATION (Princeton Univ. Discussion Paper, 1986).

213. *See supra* text accompanying notes 66-69.

214. *See* O. WILLIAMSON, *supra* note 152; R. BORK, *supra* note 1, at 135-37.

HeinOnline -- 96 Yale L.J. 277 1986-1987

ticompetitive effects stemming from exclusion are too complicated or unwieldy to produce trustworthy estimates. Identical arguments, however, could be made to abolish all antitrust review of mergers, joint ventures, and claims of price fixing. Such an uncritical acceptance of potentially anticompetitive conduct would be no more supportable than a rule of "inhospitality" that prohibited all potentially exclusionary restraints.[215] Unless we are to abandon all hope of rational fact-finding and discriminating analysis, both of these extreme approaches are simply too draconian. Two other approaches deserve fuller exploration.

1. *Treating Efficiencies as Irrelevant*

The dominant view of antitrust law always has been that where anticompetitive effects are probable, efficiencies are no defense.[216] Although the so-called "rule of reason" analysis takes account of efficiency claims, it does so principally by subjecting assertions of anticompetitive effects to close scrutiny when plausible efficiency arguments are offered.[217] We know of no case in which the Supreme Court concluded that the practice in question probably conferred on a firm power to raise price, but nevertheless upheld the practice on the ground that the harm to consumers was outweighed by cost savings to the firm adopting the practice.[218]

Similar treatment of exclusion claims certainly cannot be described as irrational. Suppose one knows with reasonable certainty in a particular case that an exclusionary rights agreement gave its purchaser significant power over price by raising its rivals' costs. If the agreement also enabled

---

215. *See* Easterbrook, *supra* note 49, at 4-9; *see also* R. BORK, *supra* note 1, at 136-44.

216. *See, e.g.*, FTC v. Procter & Gamble Co., 386 U.S. 568, 580 (1967) ("Possible economies cannot be used as a defense to illegality."); *cf.* Williamson, *supra* note 126.

217. *See, e.g.*, Northwest Wholesale Stationers, Inc. v. Pacific Stationery & Printing Co., 105 S. Ct. 2613, 2620-21 (1985) (refusing to apply per se rule without considerable inquiry into market conditions); Continental T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36, 51-58 (1977) (vertical restraints to be governed by rule of reason rather than per se rule because of possible procompetitive effects). Efficiency considerations can also affect choice of remedies. In two important joint venture cases, the Supreme Court chose remedies that regulated, rather than disbanded, combinations that achieved market power, probably because of the efficiencies these combinations generated. *See* Associated Press v. United States, 326 U.S. 1, 21-22 (1945); United States v. Terminal R.R. Ass'n, 224 U.S. 383, 409-13 (1912).

218. A case that might have raised this issue was Broadcast Music, Inc. v. Columbia Broadcasting Sys., 441 U.S. 1 (1979). The Court in *Broadcast Music* held the per se rule against price fixing among competitors inapplicable to blanket licensing agreements negotiated with television networks by holders of copyrights in musical compositions. The Court said the agreement among competing copyright holders escaped condemnation because large efficiencies were created (or, as the Court preferred to phrase it, because the blanket license was "a different product" with "unique characteristics"). *Id.* at 21-22. A rule of reason analysis might have required the Court to decide whether the cost-savings features of the blanket license outweighed its price-enhancing features. This issue was avoided, however, when, on remand, the Second Circuit concluded that the blanket license had no anticompetitive price enhancing effect whatsoever. Broadcast Music, Inc. v. American Soc'y of Composers, Authors & Publishers, 620 F.2d 930, 935-37 (2d Cir. 1980), *cert. denied*, 450 U.S. 970 (1981).

HeinOnline -- 96 Yale L.J. 278 1986-1987

the purchaser to reduce its costs, its concomitant acquisition of monopoly power means that consumers cannot expect to realize much, if any, of the benefits of that cost reduction; instead, only the firm's stockholders gain. In any event, efficiencies often can be achieved without exclusionary rights.

2. *Putting Efficiencies in the Balance*

The view that efficiencies are irrelevant, although not irrational, is not necessarily compelling. At least in some extreme cases, the purchaser may achieve such a substantial cost reduction that, even when it subsequently exercises its market power, the purchaser likely would set a price lower than that in effect before the agreement.[219] If this occurs, and the cost reduction could not have been achieved by any other means, consumer welfare is not harmed by the practice. Alternatively, if power over price is not certain, and in its absence the competitive price would fall, then substantial cost savings might lower the expected value of the price consumers would pay.

Further, as a normative matter, one may not wish to define the consumer welfare that antitrust law protects solely as the prices consumers pay. An increase in price, caused by the exercise of market power, harms consumers in two, arguably disparate, ways. First, there is the unambiguously harmful "deadweight" loss of sales occasioned by the shift from the lower, more competitive price. Second, and more controversial, is the "monopoly transfer," the price premium paid by those still willing to buy.

That monopoly transfer increase (or part of it) may be competed away by the erection of strategic entry barriers or by the entry of inefficient competitors. In this case, it represents an increase in deadweight loss. Alternatively, it may represent only a (possibly temporary) transfer of wealth from some people to others.[220] By treating stockholders as "honorary consumers," a consumer welfare standard could be indifferent to the wealth transfer. That view could justify a real efficiency gain to the purchasing firm or the suppliers that could lower society's costs of producing goods.[221] Moreover, the monopoly rent transfer may induce innovation and further cost savings by competitors.

On these premises, some might rationally conclude that efficiencies

---

219. Fisher, Lande & Vandaele, *Afterword: Could a Merger Lead to Both a Monopoly and a Lower Price?*, 71 CALIF. L. REV. 1697 (1983).

220. *See* Lande, *Wealth Transfers as the Original and Primary Concern of Antitrust: The Efficiency Interpretation Challenged*, 34 HASTINGS L.J. 65 (1982); Posner, *The Social Costs of Monopoly and Regulation*, 83 J. POL. ECON. 807 (1975).

221. However, production inefficiencies may also be inflicted on the purchaser's rivals. *See supra* Section IV.B.3.

HeinOnline -- 96 Yale L.J. 279 1986-1987

should be part of the calculus. Millions of dollars in cost savings might be more important than fulfilling the desires of a few consumers willing to buy at the pre-restraint price but not at the later one.

If efficiencies are put into the balance, the weight of the evidence in favor of the cost savings must be evaluated. Empirical evidence tending to prove actual cost savings should be preferred over evidence establishing only the logical possibility that such efficiencies will be realized. The empirical basis might consist of before-the-fact data that support the inference that, in the particular situation, the practice probably will generate efficiency benefits; or it might be after-the-fact evidence that the exclusionary rights agreement actually had the claimed effect.[222] Either type of empirical support would improve the reliability of the particular claim.

Balancing also raises the issue of the proper benchmark for evaluating cost-savings claims. What usually causes the anticompetitive effect in exclusionary rights agreements is a single contractual provision giving the supplier the duty to exclude rivals of the purchaser, not the entire purchase agreement for the inputs. Therefore, sometimes the cost savings can be achieved even as the anticompetitive effects are excised; the court can invalidate only the exclusionary provisions, not the entire agreement. In these cases, there would be no need for a court to balance cost savings against price increases. It follows that efficiency claims should be ignored unless the exclusionary right itself, not just the input purchase agreement in which the exclusionary right is embedded, can be shown to reduce costs, both absolutely and relative to the price increases suffered by consumers. Sometimes, however, the agreement and the exclusionary term cannot be separated. For example, Real Foreclosure in the form of Overbuying is a case in which the larger agreement is impossible to separate from the exclusionary right. It is the actual transfer of the input, not some ancillary part of the transfer, that has the exclusionary impact. For this reason, Overbuying is one of the hardest practices to evaluate. The inseparability of purchase agreement and exclusionary right also provides a strong rationale for courts to be most permissive toward this practice.

3. *The Policy Dilemma*

At the heart of this dilemma is the question of how to describe the consumer welfare that antitrust law seeks to protect. If the Sherman Act gives consumers and firms an entitlement to enter transactions that they would have made but for restraints of trade that confer monopoly power on certain other firms, then an exclusionary agreement that confers mo-

222. For examples of evidence that might be introduced, see Salop & Simons, *A Practical Guide to Merger Analysis*, 29 ANTITRUST BULL. 663 (1984).

HeinOnline -- 96 Yale L.J. 280 1986-1987

nopoly power on its purchaser by raising its rivals' costs should not be saved by efficiency claims unless these savings are so substantial that price will fall in spite of the increase in monopoly power.[223] On the other hand, if the Sherman Act places little value on the division of wealth between consumers and stockholders and instead seeks to promote a net value of efficient transactions and low production costs, and if one doubts the ability of firms to erect entry barriers or their propensity to engage in rent-seeking behavior, then the likelihood and magnitude of any cost reductions should be part of the calculus employed in judging exclusion claims.[224]

This Article is not the place to attempt to resolve that dilemma, but we can describe its contours. We would take account of very large efficiencies where the process costs of discovering and credibly measuring them are not prohibitive. Pending further refinement of methods for estimating efficiencies, the weight of precedent rather clearly lies with those who would refuse to recognize efficiencies as a justification, while, if necessary, erring in the direction of leniency in devising methods or standards for estimating the likelihood of anticompetitive effects.[225] Thus, the burden currently rests on proponents of change to describe how the magnitude of efficiencies should be measured and where the balance(s) should be struck. Both of these tasks are enormously complex and difficult. To our knowledge, no one has undertaken them. Until they are accomplished, the chances of persuading decisionmakers openly to alter the status quo seem dim.

In constructing their argument, those who would employ efficiencies as an offsetting defense in exclusionary rights cases should be prepared to accomplish the following. First, they should develop standards for estimating the magnitude of those efficiencies.[226] Second, they should develop methods to assess whether efficiencies could be obtained by less restrictive or less anticompetitive techniques. Third, they should distinguish in practice between pecuniary economies and real efficiencies, and count only the latter.[227] Fourth, they should establish standards for distinguishing among effective exclusion resulting in price increases that (a) only transfer wealth from consumers to stockholders, (b) generate activities by purchasers, such as rent-seeking and strategic entry deterrence, that siphon wealth from

223. *See* Fisher & Lande, *Efficiency Considerations in Merger Enforcement*, 71 CALIF. L. REV. 1580, 1631-32 (1983).

224. This would imply that any production inefficiencies inflicted on rivals should also be taken into account. *See supra* Section IV.B.3.

225. For example, this is precisely what the Court did in Continental T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36 (1977).

226. For example, measure the dollar benefits of reduction in free rider effects.

227. If antitrust law should be indifferent to the distribution of wealth between producers and consumers, then it should be indifferent to the wealth distribution between producers and input suppliers as well.

HeinOnline -- 96 Yale L.J. 281 1986-1987

consumers to non-productive entities, and (c) provide incentives for wealth-increasing innovation. Only after explaining how these kinds of calculations can be made or why they should be ignored can one assert, except in extreme cases, that a cost savings necessarily should be balanced against quantitative proof of likely anticompetitive effects.

Two kinds of exclusionary rights agreements nevertheless are easy to address. At one extreme, where all that is involved is the sale of a good, service, or other commodity, all of which is consumed efficiently by the purchaser (i.e., the net foreclosure rate is zero), with no additional exclusivity provisions, the arrangement might best be treated as lawful per se. Where the purchaser obtains neither a promise by the supplier not to deal with others nor more of the good than the purchaser presently can consume profitably, the odds are quite high that the arrangement will have no anticompetitive effects.[228]

At the opposite pole, where the purchaser obtains only a naked exclusionary right and the purchaser does not itself acquire inputs from the supplier, that arrangement should be presumptively suspect. Claims of efficiency in such cases are almost certain to be implausible, and there are no efficiencies stemming from input purchases to protect. Consequently, purchasers could fairly be assigned the burden of proving the absence of a reasonable likelihood of anticompetitive effect.

Remaining are cases in which some exchange of goods or services occurs, but an express exclusionary right also is conveyed. Partly because of the weight of precedent and partly because the exclusionary right itself often should not confer distinctive cost-savings, we think it would not be unreasonable to leave defendants with the burden of proving measurable, specific, countervailing efficiency justifications in specific exclusionary rights cases in which plaintiffs have proved actual or probable competitive injury.

VIII. Alternative Comprehensive Standards

As noted at the outset, courts have not treated exclusion claims as a unitary antitrust issue. Rather, they have classified exclusion cases according to the type of commercial arrangement challenged, promulgating different standards for different arrangements. Naturally, this has led most

228. Similarly, we think a strong case can be made for placing a large burden on plaintiffs in Overbuying cases. They should be required to show that the Overbuying was significantly in excess of what the defendant reasonably needed or, perhaps, that the input purchases were so large that significantly inefficient resource use would occur. This latter approach would require proof that, as a result of the inefficient resource use from Overbuying, the purchasing firm's marginal cost is, in effect, driven up to a level above the price it receives for its output. This burden would generally be very difficult for the plaintiff to carry; it is equivalent to the Areeda-Turner test for predatory pricing, as applied to input purchases. *See* S. Salop & D. Scheffman, *supra* note 73.

HeinOnline -- 96 Yale L.J. 282 1986-1987

commentators to take similarly narrow views of the issues. We are not the first, however, to suggest a more unified antitrust exclusion theory. The most complete approaches, other than ours, appear to be (a) the "output" test advocated by both Easterbrook and Posner, (b) the guidelines for vertical mergers issued in June 1984 by the Department of Justice (DOJ), and (c) the DOJ's Vertical Restraints Guidelines (VRGs) issued on January 23, 1985.

A. *The "Output" Test*

Then Professors, now Judges, Richard Posner and Frank Easterbrook have separately proposed that the legality of vertical restraints be evaluated on the basis of their effect on output.[229] The restraints would be illegal only if output declined. The test follows from the argument that consumers are injured if industry output falls and prices rise, and they are benefitted if industry output rises and prices fall. There are three main problems with this test.

First, applying the output test to exclusionary restraints easily and frequently leads to its misuse. The output test should involve changes in total industry output and was designed to evaluate only industry-wide restraints that allegedly facilitated collusion. When the test has been applied to restraints adopted by only one or a few firms, even its advocates have erroneously evaluated only the output of the firm adopting the restraint, not the entire market. For example, as Easterbrook puts it:

> The economists therefore might look at output changes in the short run. Does the firm using the challenged practice increase sales or reduce them? An increase suggests procompetitive effects, a lower effective price per unit of quality delivered. Does the firm increase its market share or lose it? Again an increase suggests procompetitive effects.[230]

He is clearly incorrect. Under this test, a firm that demolished all its rivals' plants would escape liability because its market share increased to 100 percent.[231]

---

229. Easterbrook, *Vertical Arrangements and the Rule of Reason*, 53 ANTITRUST L.J. 135 (1984) [hereinafter Easterbrook, *Vertical Arrangements*]; Easterbrook, *supra* note 49; Posner, *The Next Step in the Antitrust Treatment of Restricted Distribution: Per Se Legality*, 48 U. CHI. L. REV. 6 (1981).

230. Easterbrook, *Vertical Arrangements*, *supra* note 229, at 154.

231. It appears that Easterbrook has confused the expected effects of horizontal price collusion (in which the colluders' joint output would be expected to fall) with those of exclusionary practices that raise the costs of rivals (in which the perpetrator's output could rise). This error is so obvious that it probably should be forgiven as the fault of careless drafting. However, the error succeeded in fooling the Assistant Attorney General for Antitrust. In explaining why the purchase of exclusive (and allegedly exclusionary) rights to beer sports advertising by Miller and Anheuser-Busch could not have

HeinOnline -- 96 Yale L.J. 283 1986-1987

Second, even if aggregate *industry* output were used, the test generally still would be inconclusive. When products are differentiated, industry output is not a good proxy for aggregate consumer welfare. Welfare can fall when output rises and vice versa,[232] because the restraint can give a small benefit to a limited number of consumers at the margin while decreasing the benefits received by a large number of consumers who nonetheless find that the restrained product still is the best buy in the marketplace.[233]

Third, output is difficult to measure. If the market is growing, analysts would have to measure the output increase "but for" the restraint. Similar problems arise if costs change. In addition, for differentiated products, the proper measure of output is not clear. If "revenues" are used, price increases would show up as increases in output. Measurements of physical quantities may not capture quality differences correctly.

B. *The Justice Department's Vertical Merger Guidelines*

The DOJ's guidelines for vertical mergers[234] are broadly consistent with our approach. However, only a restricted set of anticompetitive theories is covered by the guidelines, and the thresholds are set at apparently arbitrary levels.

The guidelines ask whether the vertical merger is likely to raise entry barriers into one of the markets (the "primary" market) and thereby enhance the likelihood of tacit or express collusion in that market by requiring potential entrants simultaneously to enter the other ("secondary") market.[235] In our terminology, this concern corresponds to the possibility that a vertical merger might raise rivals' costs of obtaining needed inputs

---

been anticompetitive, Assistant Attorney General Ginsburg focused on the resulting increase in the market shares of Miller and Anheuser-Busch. He stated, "[i]f the practice restrained competition in the market, one would have seen the two firms' market shares falling rather than rising." Letter from Douglas H. Ginsburg to Honorable Howard M. Metzenbaum (Nov. 1, 1985) (citing Easterbrook, *supra* note 49). The authors were consultants to a firm that had sought to have the Antitrust Division challenge these practices.

232. *See* Comanor, *Vertical Price-Fixing, Vertical Market Restrictions, and the New Antitrust Policy*, 98 HARV. L. REV. 983, 999–1000 (1985); Katz, *Non-Uniform Pricing, Output and Welfare Under Monopoly*, 50 REV. ECON. STUD. 37 (1983); Scherer, *The Economics of Vertical Restraints*, 52 ANTITRUST L.J. 687, 706 (1983); Spence, *Monopoly, Quality and Regulation*, 6 BELL J. ECON. 417 (1975). Easterbrook responds to these arguments by stating, without any claimed economic basis, that while the test may be misleading under certain conditions, the required conditions are "rare." Easterbrook, *Vertical Arrangements*, *supra* note 229, at 154 n.34.

233. This would be implausible in a highly competitive market where there are very close substitutes for the product of the firm adopting the restraint. But a focus on these competitive conditions involves moving to a test other than the output test, a change in focus that is consistent with our analysis.

234. 1984 DOJ Merger Guidelines, *supra* note 144, § 4.2.

235. *Id.* § 4.21.

HeinOnline -- 96 Yale L.J. 284 1986-1987

(DOJ's secondary market), leading to power over price in the output market (DOJ's primary market).

According to the guidelines, barriers can only increase if, after the merger, three conditions are satisfied. First, conditions must be such that to enter the output market a firm also must enter simultaneously the input market.[236] In determining whether this two-tier entry is necessary, the guidelines consider whether the integrated firms might engage in "supply squeezes" against unintegrated rivals.[237] Second, this simultaneous two-tier entry must be more expensive.[238] In our terminology, the supply squeeze corresponds to the exclusionary act, and these two conditions would imply that the costs of (potential) rivals were raised. Third, the structure and characteristics of the output market must be conducive to non-competitive performance, so that the increase in entry barriers is likely to affect its performance.[239] This inquiry corresponds to the question of whether the merger will enhance the firm's power to raise price in the output market. The Justice Department's basic approach, as well as the variables it employs as objective measures, is fundamentally consistent with our analysis, but it differs from our proposals in two major ways.

First, the DOJ guidelines restrict their attention to vertical mergers that could create barriers to entry. They do not encompass the case in which established rivals' costs are raised significantly, although barriers to entry do not increase. This oversight is insignificant in the case where low barriers to entry imply that rivals could avoid the cost increase by integration into production of the input themselves. However, it is significant in those cases where entry barriers are already high so that rivals are dependent on established firms for their supplies of the input.

Second, the guidelines adopt an enforcement threshold that seems arbitrary. In particular, the Department concludes that supply squeezes are not sufficiently likely to necessitate two-tier entry as long as, after the merger, sufficient capacity is available in the input market to serve *two* plants of minimum efficient scale in the output market.[240] This threshold is not connected to the HHI levels used elsewhere in the guidelines.[241] Nor is it justified on other grounds.

---

236. This may not occur if there is sufficient unintegrated input capacity still available to supply entrants (in our terminology, that is, if the capacity share of unrestrained firms is large). *Id.* § 4.211. In this case, rivals' costs would not rise.

237. *Id.* § 4.211 n.31. However, they do not discuss how this evaluation would be made.

238. *Id.* § 4.212.

239. *Id.* § 4.213.

240. *Id.* § 4.211.

241. For example, the Department uses an HHI equal to 1800 as a second threshold in the input market. *Id.* § 4.213. The guidelines concerning potential entry ask, inter alia, whether at least three firms are on the verge of entering. *Id.* § 4.133; *see also* MERGERS IN THE NEW ANTITRUST ERA, *supra* note 144, at 54 n.35, 55 n.37.

HeinOnline -- 96 Yale L.J. 285 1986-1987

C. *The Justice Department's Vertical Restraints Guidelines* [242]

In many important respects, our approach dovetails with that of the VRGs. Both systems of analysis flow from dissatisfaction with the dominant doctrinal themes of prevailing case law. Both would explicitly focus attention on whether competition (as well as competitors) is harmed and would define those terms in light of microeconomic price theory. As a matter of substantive antitrust policy, both systems would require that a party asserting an exclusion claim prove, at a minimum, that entry barriers in well defined input markets permit a vertically integrated firm to raise its rivals' costs materially. Both ask whether assertedly vertical exclusionary restraints have horizontal effects. As a matter of antitrust enforcement policy, both systems borrow from the merger guidelines in applying objective measures (largely reflecting market structure) to estimate the likelihood of certain untoward economic effects (principally, output restrictions and price increases). Both systems of analysis recognize that exclusionary rights contracts often may generate efficiencies, but put the burden of proving specific efficiency justifications on challenged firms.

Notwithstanding these large and important areas of congruence, our system of analyzing exclusionary rights claims differs from that of the VRGs in several significant ways. The VRGs are concerned with collusion as well as exclusion; they also include "territorial and customer restraints" that entail no exclusion of rivals, but are troublesome only to the extent that they facilitate collusion directly among firms subject to the restraint by altering the structure of their markets.[243]

The DOJ analysis centers explicitly on whether the exclusionary rights agreement subjects rivals to an unavoidable cost increase[244] and implicitly

---

242. U.S. Dept. of Justice, Vertical Distributions Restraints Guidelines, 50 Fed. Reg. 6263 (1985) [hereinafter VRGs]. We acted as consultants on the VRGs and, in that capacity, shared an early draft of this article with the Department.

243. *Id.* § 1. We take an analogous approach in the case of Cartel Ringmaster. In the case of allegations of collusion, as we discussed earlier with regard to Cartel Ringmaster, it may be unnecessary to prove power over price. *See supra* note 136.

We bring within our analytic framework restraints that the VRGs designate as "Exclusive Dealing Arrangements." VRGs, *supra* note 242, § 1. The VRGs describe these arrangements as "requirements that a buyer deal only with a particular seller or that a seller deal only with a particular buyer or group of buyers." This definition may be read to cover not only the VRGs' limited, specific illustrations of "exclusive distributorships, sole outlet provisions, and requirements contracts," but also such otherwise unmentioned case law phenomena as refusals to deal, boycotts, and denials of essential facilities. *Id.* Exclusive dealing arrangements do not, however, include vertical mergers, which are left to the DOJ's very differently structured Vertical Merger Guidelines (even though, as shown above, they present analytically identical issues). Further, the VRGs present a completely different set of standards for analyzing the legality of tying arrangements, notwithstanding that such arrangements would appear to fit the broad definition of "exclusive dealing arrangements" and can easily be analyzed within the two-step analysis set out above. *See supra* Section IV. *Compare* VRGs, *supra* note 242, § 5 *with* DOJ Merger Guidelines, *supra* note 144, § 4.

244. VRGs, *supra* note 242, § 3.22.

HeinOnline -- 96 Yale L.J. 286 1986-1987

on whether the rivals' competitor(s) thereby gain power over price. The VRGs also seek to employ objective measures of the likelihood of these effects with a "market structure screen,"[245] a first analytical step designed "to eliminate from further consideration those restraints that, in all likelihood, have no anticompetitive effects."[246]

To apply the market structure screen, the VRGs employ two new measures, the "Vertical Restraints Index" (VRI) and the "Coverage Ratio" (CR). The VRI is calculated by squaring the market share of each firm subject to the exclusive dealing arrangement and summing the resulting values.[247] The CR is the percentage of the market employing the exclusionary arrangement.[248] Employing these measures, the VRGs consider an exclusive dealing arrangement to be a candidate for antitrust proscription only if all of the following three conditions are met: (a) "the firm employing the restraint"[249] has a market share greater than ten percent; (b) the VRI in one market is at least 1200; and (c) the CR in the other market is at least sixty percent.[250] If any one of these conditions is not satisfied, the restraint is to receive no "further consideration" because it will, "in all likelihood, have no anticompetitive effects."[251]

This "market structure screen" is incomplete in at least four respects. First, the test fails to take account of the proportion of the buyer's costs accounted for by the seller's good. This screen does not differentiate between denying paintbrush makers bristles and denying them paper clips. Second, the screen overlooks the possibility that rivals' costs may be raised via price coordination by unrestrained suppliers, what we have dubbed the Frankenstein Monster technique.[252] Nor do the VRGs deter a purchaser from buying much more of an input than it needs to satisfy its

245. *Id.* § 4.

246. *Id.* § 4.1. Restraints identified by this screening are not necessarily illegal. Such restraints are merely subjected to another series of tests. *Id.* § 4.2.

247. *Id.* § 4.1 n.25.

248. *Id.* § 4.1 n.26. To illustrate, assume that bristles for paint brushes constitute a relevant product market, that only manufacturers of paint brushes buy these bristles, on a non-exclusive basis, and that the relevant market is nationwide. Further, assume that ten firms manufacture and sell bristles, each accounting for 10% of sales; five unrelated firms manufacture and sell brushes, each accounting for 20% of all brush sales and hence 20% of bristle purchases. If two brush makers then each enter separate exclusive dealing arrangements with two bristle manufacturers, so that four bristle makers are restrained, then the bristle VRI is 400 (i.e., $10^2 + 10^2 + 10^2 + 10^2$); the bristle CR is 40% (i.e., 10 + 10 + 10 + 10); the brush VRI is 800 (i.e., $20^2 + 20^2$); and the brush CR is 40% (i.e., 20 + 20).

249. As used in this Article, the phrase "firm employing the restraint" denotes the firm that purchases an exclusionary right.

250. VRGs, *supra* note 242, § 4.1. The DOJ Guidelines state these conditions somewhat differently, but on comparison our description can be seen as equivalent.

251. *Id.*

252. *See supra* Section IV.A.3.b. In the hypothetical example, *supra* note 248, if there are barriers to entry, the exclusive dealing arrangements may enable the six remaining bristle suppliers to collude against the three brush makers who remain outside the exclusive dealing arrangements.

HeinOnline -- 96 Yale L.J. 287 1986-1987

current market share, what we have called the Overbuying method of Real Foreclosure.[253] Finally, agreements that are not deemed harmless by this screen are then subjected to another series of tests that are not set in any rational structure.

These omissions appear to be symptoms of a deeper fault. The VRGs use threshold standards that lack firm theoretical or empirical support. Moreover, they appear divorced from other DOJ antitrust enforcement policies. The merger guidelines, whatever their flaws, rest upon systematic theories linking the HHI to the likelihood of price coordination and on extensive empirical work testing the theories. If we are told that a merger between competitors that raises the HHI in their market from 1700 to 2700 should be suspect, we do not have to agree with that statement to understand what it means and the policy views upon which it rests. By comparison, the VRGs are an ipse dixit. They tell us that an exclusive dealing arrangement warrants careful antitrust analysis only if the purchasing firm accounts for ten percent of sales in its market, and either the purchaser's or the supplier's market exhibits a VRI of 1200 or more while the coverage ratio in the other market is at least sixty percent. Yet the VRGs do not explain why the numbers 10, 1200, and 60—or any others remotely resembling them—should be relevant, or why the CR and VRI are evaluated in only one market instead of both.[254] Although the indices employed—purchaser's market share, VRI, and CR—all might have theoretical claims to legitimacy,[255] those very theories suggest that the screens erected upon these indices are quite incomplete.[256]

---

253. *See supra* Section IV.A.2.b. Employing the hypothetical markets described in note 248, *supra*, suppose instead that one brush maker entered exclusive dealing arrangements with five bristle producers. Because the coverage ratio is under 60% in both markets, the arrangements fall into a VRGs safe harbor. But the net foreclosure rate is 37.5% and, unless bristles represent a very insignificant portion of the costs of making paint brushes, or there are no barriers to expansion or entry, this Overbuying may generate very large cost increases for rival brush manufacturers.

254. For example, in most cases the "leading firm proviso" of the merger guidelines would proscribe any merger of a firm whose market share is 35% with any of its competitors. 1984 DOJ Merger Guidelines, *supra* note 144, § 3.12. Yet, the VRGs permit a firm to purchase exclusionary rights from one firm with 35% of the supply market and another with 24%, despite the apparent threat of unilateral output restraints or collusion. VRGs, *supra* note 242, § 4.1.

255. *See supra* Sections IV.A., VI.A.

256. *See supra* text accompanying notes 252–53. We have tried to show that two preferable alternatives exist to the VRGs. First, one could attempt to describe systematically the necessary elements of a rigorous claim of anticompetitive exclusion and leave courts and enforcement authorities, working within this structured framework, to analyze claims in light of the particular characteristics of the markets at issue. *See supra* Section IV. If that approach leaves judges and agencies too free to introduce irrelevant factors or to reach idiosyncratic results, objective measures of the likelihood of anticompetitive effects can be adopted as threshold tests. These tests would be more firmly rooted in price theory and empirical observation and would have a stronger claim to consistency with overall antitrust policy. *See supra* Section VI.

HeinOnline -- 96 Yale L.J. 288 1986-1987

IX. CONCLUSIONS

A. *A Summary Illustration*

We have argued, throughout this Article, that courts and antitrust enforcers should adopt a unified approach to claims of anticompetitive exclusion. We have proposed a progressively more detailed approach that focuses on a narrow range of factual issues, and argued that such refinements would enhance predictability and ease of administration. We can illustrate these various conclusions about the proper antitrust response to exclusion claims by considering the classic *Klor's* case.[257]

Assume plaintiff Klor's complained of a series of agreements between Broadway-Hale and brand name appliance manufacturers from which Broadway-Hale purchased exclusionary rights, i.e., a promise that the manufacturers would not sell to Klor's. To establish an antitrust violation, Klor's should have to prove, as a threshold matter, that the agreements significantly raised its costs.[258] Regardless of the mechanism of cost-raising alleged, this involves showing both that the cost of an input increased significantly, and that this input represented such a significant fraction of costs that the net effect of the exclusion on Klor's costs was great.[259]

Turning to the next level of detail, Klor's would be required to prove that one or more of the four mechanisms firms might employ to raise rivals' costs generated its cost increase. Its costs must have been raised because (a) remaining appliance manufacturers were significantly less efficient or more expensive than those foreclosed ("Bottleneck"); or (b) remaining manufacturers had limited low-cost capacity ("Real Foreclosure"); or (c) Broadway-Hale had induced enough firms to refuse to deal with, or to practice effective substantial price discrimination against, Klor's ("Cartel Ringmaster"); or (d) remaining manufacturers were so concentrated that Klor's would incur significantly greater costs in acquiring appliances ("Frankenstein Monster").[260]

Even greater specificity can be attained, if desired, by borrowing numbers directly from the Department of Justice's merger guidelines. The four asserted theories of harm can be subjected to a threshold numerical standard, measured by objective factors, to gauge their credibility.[261] If antitrust policy should be more over-inclusive or under-inclusive in upset-

---

257. Klor's, Inc. v. Broadway-Hale Stores, 359 U.S. 207 (1959); *see supra* text accompanying notes 27-30.

258. Normally, analysis would focus on Klor's marginal costs, especially because appliances are a variable cost. For an explanation of, and some qualifications on, this statement, see *supra* note 129.

259. *See supra* Sections III, V.

260. *See supra* Section IV.A.

261. See *supra* Section VI.A.

HeinOnline -- 96 Yale L.J. 289 1986-1987

ting exclusionary rights contracts than in striking down horizontal mergers, the DOJ's numerical thresholds can be adjusted accordingly.

Even if Klor's satisfied the appropriate test for proving that the exclusionary rights agreements actually or probably materially increased its costs, Klor's should also be required to prove that Broadway-Hale thereby gained power to raise price.[262] This means Klor's should be required to prove that its (and possibly other rivals') costs have been raised significantly and either (a) that those firms (actually or potentially) in Broadway-Hale's market whose costs were not increased significantly by the vertical restraint were substantially better able to coordinate prices than were all the firms (including Klor's) prior to the exclusionary agreement(s), or (b) that those remaining (actual or potential) competitors likely would not make up for the shortfall in supply of Klor's and other excluded rivals so that Broadway-Hale and others could raise their prices unilaterally.[263]

A relatively simple, objective method for testing such an allegation would be to ask whether the Justice Department merger guidelines would have permitted Broadway-Hale to merge with Klor's (and with any other firms subjected to similar refusals to deal). Another, superior method would consider the concentration and barriers to entry for the market comprising all firms that purchased exclusionary rights and all unexcluded firms. In addition to the wide range of factors specified in the merger guidelines, one could also ask (a) whether Klor's and other excluded rivals accounted for a significant fraction of output in the market; (b) whether barriers to entry (and expansion) into the market are high, including whether barriers are created by the exclusionary right itself; and (c) whether the elasticity of demand for Broadway-Hale's product is low.[264]

To take account of the possibility that Klor's could have protected itself without resort to antitrust litigation, Klor's would also have to show that Broadway-Hale has a large market share and, if relevant, that Klor's production cost increase was substantially greater than any incurred by Broadway-Hale.[265] If Klor's met the applicable standards, Broadway-Hale would have the burden of proving specific countervailing efficiencies.[266]

Absent a showing that Klor's costs were raised and that Broadway-Hale thereby gained power to increase price, Klor's would not have a

262. *See supra* Sections III, V.
263. *See supra* Section IV.B.
264. *See supra* Section VI.B.
265. *See supra* Sections VII.A., B.
266. *See supra* Section VII.C.

HeinOnline -- 96 Yale L.J. 290 1986-1987

credible claim that Broadway-Hale's exclusionary rights agreements had any anticompetitive effects. Klor's might be harmed in the sense that it could not purchase every appliance it desired, and some consumers might be "irate" that they need to shop elsewhere, but consumers as an economic entity have lost little. After the exclusionary rights agreements, the presence of Klor's or other dealers restrains Broadway-Hale's ability to raise price just as much as before the agreements.

This approach provides greater protection against plaintiffs' filing ill-founded claims than present "foreclosure" doctrines. Judges could require complainants to state, at the outset, specifically how the exclusionary rights agreement increased rivals' costs and conferred on their competitors the power to raise price, and to explain what objective criteria lend credibility to the claim. Further, judges ought to be able to narrow the issues—and, hence, the scope of pre-trial discovery and the length of trials—by confining the parties to the theories of injury asserted by the party contesting the exclusionary rights agreement. In this fashion, the inquiry is both structured and narrowed.

B. *Looking Ahead*

Our analysis should affect prevailing antitrust law and scholarship in three ways. First, basic competition theory can demonstrate that virtually all antitrust issues spring from one of two fears: (a) that the challenged practice will enable firms engaging in it to collude (or to collude more effectively) with competitors; or (b) that the practice will permit the firm employing it to exclude actual or potential competitors who, if not excluded, could restrain the firm from raising price above the competitive level. Cases stemming from the latter fear and challenging contractual arrangements or mergers between suppliers and purchasers, or tie-in sales, should be analyzed according to the two-step inquiry we have proposed: (a) does the practice significantly raise rivals' costs; and, if so, (b) does it substantially increase the firm's power over price? The Supreme Court's opinions in *Hyde*[267] and *Northwest Stationers*[268] suggest that the Justices know or sense that antitrust needs a viable, coherent theory for evaluating exclusion claims. We believe our proposal would meet that need and satisfy the conditions implicitly established in *Sylvania*,[269] *Aspen Ski*,[270] and

267. Jefferson Parish Hosp. Dist. No. 2 v. Hyde, 466 U.S. 2 (1984); *see supra* text accompanying notes 5-7.

268. Northwest Wholesale Stationers, Inc. v. Pacific Stationery and Printing Co., 105 S. Ct. 2613 (1985); *see supra* notes 8, 10-11 and accompanying text.

269. Continental T.V., Inc. v. GTE Sylvania Inc., 433 U.S. 36, 47-59 (1977).

270. Aspen Skiing Co. v. Aspen Highlands Skiing Corp., 105 S. Ct. 2847 (1985); *see supra* notes 9-11 and accompanying text.

HeinOnline -- 96 Yale L.J. 291 1986-1987

Zenith[271] that the theory be rooted in an economic analysis of the consumer welfare goals of the antitrust statutes.

Second, applications of the two-step inquiry can be refined further. Although we believe our analysis is correct, we know questions remain. Can, and should, the calculation of the magnitude of rivals' cost increases be further refined?[272] If antitrust enforcers and judges should adopt a comprehensive set of objective factors to test exclusion claims, what specific adjustments, if any, should be made to account for the probability that efforts to exclude would fail of their own accord because rivals would undermine them or because they would prove unprofitable?[273] Can, and should, efficiency defenses be subjected to objective tests of the likelihood and magnitude of cost savings?[274] These are only some of the applied research problems that arise if one accepts this Article's theoretical premises and principal conclusions.

Third, the fundamental economic concept advanced in this Article—that raising rivals' costs sometimes can be an effective means of acquiring power over price—has potential applications outside the context of vertical agreements under Section One of the Sherman Act and Section Three of the Clayton Act, on which we have focused. We identify some of these applications to indicate how the principles developed here may have wider utility. The analysis we advance for multi-firm cases appears equally applicable to cases challenging monopolization by a single firm.[275] Indeed, the very concept of monopoly power is perhaps best expressed as the ability to engage in practices that meet our two-step test.[276] Refusals to deal—whether implemented by contract or by refusal to contract, whether the product of single-firm or multi-firm behavior—should all be brought uniformly within the umbrella of the two-step analysis of potential for anticompetitive effects, as should cases alleging that defendants obtained market power by misusing government processes.[277] The concept also can explain why, in some cases, a firm should have standing to con-

271. Matsushita Electric Indus. Co. v. Zenith Radio Corp., 106 S. Ct. 1348 (1986).

272. *See supra* Sections VI.A.1., 2.

273. *See supra* Sections VII.A., VII.B.

274. *See supra* Section VII.C.

275. In these cases, the methods of raising rivals' costs would be expanded beyond the four theories set out here, but the basic two-step analysis of injury to competitors and injury to competition would be the same.

276. As Sullivan states the classic test: "A firm has monopolized in violation of Section 2 if it has deliberately followed a course of market conduct through which it has obtained or maintained power to control price or exclude competition . . . ." L. SULLIVAN, *supra* note 2, at 29. We believe that our two-step test—asking whether a competitor has excluded its rivals by raising their costs and, if so, whether that competitor thereby gained power over price—would structure the inquiry in many single-firm monopoly cases in a fashion that would preserve the central values of the classic test and elucidate its proper meaning.

277. *See* R. BORK, *supra* note 1, at 347-64.

HeinOnline -- 96 Yale L.J. 292 1986-1987

test a merger between two of its competitors.[278] The proper measure of damages in private antitrust actions brought by competitors also may be affected by the concept we have advanced.[279]

Today, the knottiest substantive issue confronting antitrust law is the proper response to claims of exclusion. The best way out of our present state of confusion is to follow the basic roadmap described in this Article, testing such claims by analyzing critically whether the challenged practices have given firms power to raise their rivals' costs and thereby conferred on those firms power to raise prices.

278. A firm challenging a merger between two of its competitors is likely to be faced with a double-edged argument. If the merger will harm competition, it will help the challenger; if it will harm the challenger, it will help competition. *See* MERGERS IN THE NEW ANTITRUST ERA, *supra* note 144, at 218–19. Whatever the other flaws in that argument, the challenger sometimes can plausibly explain how the destruction of competition between its competitors also would harm the challenger by giving the merged firm the ability to engage in anticompetitive exclusionary practices or purchase exclusionary rights. *See, e.g.*, Christian Schmidt Brewing Co. v. G. Heileman Brewing Co., 753 F.2d 1354, 1357 (6th Cir.), *cert. dismissed*, 105 S. Ct. 1155 (1985).

279. *See* Salop & White, *Treble Damages Reform: Implications of The Georgetown Project*, 55 ANTITRUST L.J. 73, 88 (1986) and references cited therein.

HeinOnline -- 96 Yale L.J. 293 1986-1987

HeinOnline -- 96 Yale L.J. 294 1986-1987

Raising Rivals' Costs
Author(s): Steven C. Salop and David T. Scheffman
Source: *The American Economic Review*, Vol. 73, No. 2, Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association, (May, 1983), pp. 267-271
Published by: American Economic Association
Stable URL: http://www.jstor.org/stable/1816853
Accessed: 11/04/2008 16:59

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=aea.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

http://www.jstor.org

Raising Rivals' Costs

By STEVEN C. SALOP AND DAVID T. SCHEFFMAN*

Conduct that unreasonably excludes competitors from the marketplace is a concern of antitrust law. Predatory pricing doctrine focuses on conduct that lowers revenues. Alternatively, a firm can induce its rivals to exit the industry by raising their costs. Some nonprice predatory conduct can best be understood as action that raises competitors' costs.

To a predator, raising rivals' costs has obvious advantages over predatory pricing. It is better to compete against high-cost firms than low-cost ones. Thus, raising rivals' costs can be profitable even if the rival does not exit from the market. Nor is it necessary to sacrifice profits in the short run for "speculative and indeterminate" profits in the long run. A higher-cost rival quickly reduces output, allowing the predator to immediately raise price or market share. Third, unlike classical predatory pricing, cost-increasing strategies do not require a "deeper pocket" or superior access to financial resources. In contrast to pricing conduct, where the large predator loses money in the short run faster than its smaller "victim," it may be relatively inexpensive for a dominant firm to raise rivals' costs substantially. For example, a mandatory product standard may exclude rivals while being virtually costless to the predator.

These elements combine to make cost-increasing strategies more credible than predatory pricing. Because these strategies do not require a sacrifice of profits in the short run, but allow profits to be increased immediately, the would-be predator has every incentive to carry out its threats. Moreover, unlike predatory pricing, cost-increasing strategies can often be made irreversible, and thus more credible.

Legal rules governing cost-increasing conduct should differ from predatory pricing standards. Price-cost comparisons alone are insufficient because such comparisons cannot distinguish price decreases from cost increases. Moreover, in some cases concerning conduct that raises rivals' costs, courts do not need to strike the difficult balance between short-run welfare gains and long-run losses. There is often no tradeoff. Cost increases generally raise prices, not lower them.

A variety of exclusionary practices can be characterized as conduct that raises rivals' costs. In the famous *Klor's* group boycott case, for example, the alleged predator Broadway-Hale may have induced a significant number of suppliers to refuse to provide needed inputs to Klor's. If these firms were the most efficient suppliers, a boycott could have raised Klor's costs and thus placed it at a competitive disadvantage. Had Klor's been a significant competitor in the market, retail prices could have been increased. Inducing suppliers to discriminate against rivals is a less extreme variant of the same conduct. Similarly, according to Oliver Williamson's 1968 analysis of the *Pennington* case, an industrywide wage contract raised the costs of the labor-intensive competitive fringe more than it raised the costs of the more capital-intensive dominant firms.

If there are scale economies or other entry barriers in retailing, exclusive dealing arrangements can raise small rivals' costs of distribution. As emphasized in the rent-seek-

*Professor of economics, Georgetown University Law Center, and professor of economics and Director, Institute of Applied Economic Research, Concordia University, respectively. Some of the issues analyzed here were discussed in Salop (1981), and in our joint 1982 study. Judith Gelman has provided helpful comments. Financial support from the Bureau of Economics of the Federal Trade Commission is gratefully acknowledged. This paper does not reflect the views of the Commission, individual Commissioners, or staff.

ing literature, product standards and other government regulations can raise rivals' relative compliance costs. Advertising expenditures and $R\&D$ races can also be used to raise rivals' costs. For example, suppose that increased advertising expenditures initiated by the most efficient advertiser must be matched in effective intensity by less efficient rivals. An advertising strategy might be profitable even absent the demand-increasing effect of the advertising. Disadvantaging competitors can provide a benefit that exceeds its costs, if the strategy allows the dominant firm to increase price or market-share.

Though currently out of fashion with antitrust enforcers, vertical price squeezes can be viewed as conduct to raise rivals' costs. Under appropriate conditions, a dominant firm finds backward integration to be a cost-effective way to raise downstream prices.[1] If the upstream merger partner has some market power, input price increases to downstream rivals (perhaps to a level *above* the monopoly price) will raise their costs, allowing the dominant firm to increase price or output. Upstream profits are sacrificed but downstream profits rise disproportionately.

The rest of this paper provides brief diagrammatic and formal analyses of these strategies. Our results are discussed intuitively; the technical analysis is taken up elsewhere. Three conditions are discussed: profitability to the dominant firm; injury to rivals; and consumer welfare losses. These conditions are then related to analogous concepts in the antitrust law of exclusionary practices.

Consider an industry consisting of a dominant firm and a competitive fringe. In such an industry, a lower-cost dominant firm acts as price leader. Competitive fringe firms follow by collectively setting some output y on the fringe supply curve S. Because each fringe

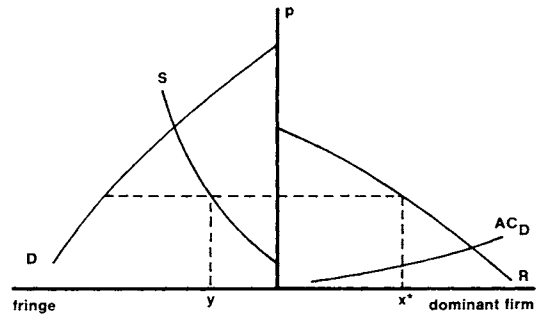

FIGURE 1

firm is small, it produces until price equals marginal cost. Indeed, for analytic simplicity, the supply curve is sometimes treated as if arising from a representative firm's marginal cost curve. At the equilibrium for such an industry, the dominant firm produces at the profit-maximizing point x^* on its residual demand curve R, as illustrated in Figure 1. The industry demand (D) and the fringe supply (marginal cost) curves are shown in the left panel. The dominant firm's residual demand (R) and average cost (AC_D) curves are pictured in the right panel. Its profits are equal to $(p - AC_D)x^*$.

Suppose the dominant firm can also select a second strategy variable to which the fringe firms must react. A general way to view this strategy is to treat the firm as selecting a "technology" for producing output or revenue. Technologies differ in cost; each fringe firm reacts by choosing a technology itself. Particular strategy variables might include product quality or advertising expenditures. Another potential instrument is the demand for necessary inputs or, alternatively, the price offered for those inputs by the dominant firm. Labor, scarce natural resources, and patentable innovations are three inputs that have concerned antitrust commentators. (Williamson, M. Maloney et al., Richard Gilbert, and Janusz Ordover-Robert Willig.) Nonprice vertical restraints like exclusive dealing and territorial restraints can also be captured in this way, because they can affect the costs of distribution. The rent-seeking literature treats cases where a firm enters the political arena in order to inflict costly regu-

[1]In the limiting case of an upstream monopolist and downstream fixed proportions (and constant returns to scale) technology, it is well known that a vertical price squeeze is unnecessary. However, few industries satisfy this structure. In other cases, vertical price squeezes can be profitable under appropriate conditions.

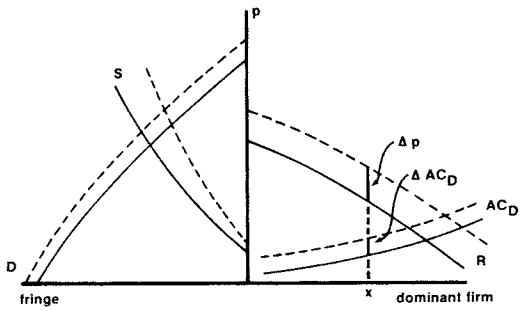

FIGURE 2

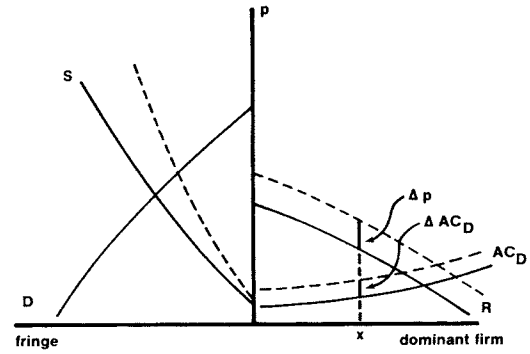

FIGURE 3

lations on its rivals, and possibly even itself (Maloney and R. McCormick).

The dominant firm's strategy may affect its own costs and market demand as well as the costs of its fringe competitors. As illustrated in Figure 2, a *sufficient* condition for a strategy to be profitable is for it to shift up the dominant firm's *residual demand curve* by more than it shifts up its *average cost curve* at the original output x^* (see equation (3) below). In this way, even if the dominant firm were to keep its output constant, the increased price-cost margin would raise its profits. Of course, the predator can generally increase its profits still further by adjusting its output.

Even if market demand is unaffected by the strategy, increases in marginal costs can reduce fringe firms' outputs and/or raise price, as illustrated in Figure 3. Sufficient increases in average costs can cause some fringe firms to exit the industry and others to forego entry. Thus, the concept of strategically erected entry barriers can be captured in this framework.

The shift in the residual demand curve depends on the elasticity of demand as well as the elasticity and shift of the fringe supply curve. The less elastic is consumer demand, the greater will be the increase in residual demand. This is because as demand elasticity falls, a given reduction in fringe supply causes a larger price rise (see equation (2) below). At the other extreme, if demand is perfectly elastic, residual demand does not increase at all.

Suppose the fringe supply curve is treated as the marginal cost curve of a representative fringe firm. Under this interpretation, the dominant firm's residual demand curve shifts up according to the increase in the fringe's marginal costs, weighted by the elasticity of the market demand curve. As illustrated in Figure 3, evaluating the profitability of the strategy requires a comparison of this price rise to the increase in the average cost of the dominant firm. Thus, in effect one must compare the effect on the *average cost* of the dominant firm relative to the *marginal cost* of the fringe, weighted by the demand elasticity.

Formally, when demand is unaffected by the strategy, the dominant firm's optimization problem is given as follows:

$$(1) \qquad Max\ px - C(x, \alpha)$$

subject to $x = D(p) - S(p, \alpha);\ \alpha \geqslant 0,$

where positive adoption of a strategy is formulated as choosing $\alpha > 0$, the dominant firm's costs $C(x, \alpha)$ are assumed to depend on its output x and the strategy α, and its residual demand consists of industry demand $D(p)$ less fringe supply $S(p, \alpha)$.

In solving this problem, a *sufficient* condition for a strategy $\alpha > 0$ to be chosen is given as follows:

$$(2) \quad 1/(1 + \sigma\varepsilon/(1-\sigma)\varepsilon_s) > \Delta AC_D/\Delta MC_F,$$

where ε and ε_s represent the elasticity of industry demand and fringe supply, respectively, σ is the market share of the dominant firm, and ΔAC_D and ΔMC_F, respectively, represent the strategy-induced *changes* in the dominant firm's average cost (C_α/x) and the representative fringe firm's marginal cost (which equals $-S_\alpha/S_p$). Differentiating the constraint in equation (1), substituting into equation (2) and rewriting, the sufficient condition becomes

$$(3) \qquad \partial p/\partial \alpha|_{x^*} > \Delta AC_D.$$

We have so far discussed only the effect of the strategy on the profitability of the dominant firm. The profitability of the fringe and consumer welfare will also be affected.

In principle, fringe output and profits may rise or fall. There are two separable effects on the fringe. First, its costs rise, lowering fringe profits. Second, because the dominant firm chooses a new output price, the profits of the fringe are further affected. These two effects generally work in opposite directions, of course, because increases in marginal cost cause price increases. For example, consider the limiting case of perfectly inelastic demand. If the fringe output is held constant, its profits fall if the strategy raises its average cost by more than the increase in fringe marginal cost. This is because, holding outputs constant, price rises by the increase in fringe marginal cost. A reduction in the fringe's output reinforces this effect whereas production increases offset the effect of the reduced profit margin.

If competitors' profits are not reduced, the strategy will obviously fail to achieve an exclusionary goal. However, if the industry is protected by entry barriers, strategies that increase the costs of fringe firms and dominant firm *equally* can still raise industry profits. In particular, if marginal costs rise by more than average costs and if demand is sufficiently inelastic, the cost increases will have a *supra*-passing-on effect, raising price by more than the increase in average cost.

Consumers' surplus is also affected by these cost-increasing strategies. Again, cost increases tend to cause price increases, which are welfare reducing. However, there may be cases in which demand and supply elasticities are increased sufficiently to cause price to fall enough to offset the welfare losses from the higher costs. In addition, in that demand (i.e. marginal consumers' surplus) is increased, consumers' surplus may rise even at a higher price. Similar results obtain for measures of aggregate economic welfare (consumers' surplus plus profits). For example, a strategy that does not raise demand, yet raises cost and price, surely lowers aggregate welfare. If demand rises, however, price, cost, and demand increases must be balanced.

For antitrust analysis, exclusionary strategies may be characterized by three conditions—profitability to the dominant firm; competitor injury; consumer welfare reduction—and their sum, the allocational efficiency (or aggregate welfare) effect. One formulation of the attempt to monopolize offense—unreasonable conduct undertaken with specific intent to monopolize that has a dangerous probability of success—can be interpreted in terms of these conditions. Long-run profitability to the dominant firm is an obvious element of intent to monopolize. Competitor injury is necessary for the conduct to have a dangerous probability of success. A strategy that reduces consumer welfare or allocational efficiency might well satisfy the unreasonableness prong of the offense.

REFERENCES

Gilbert, Richard, "Patents, Sleeping Patents and Entry Deterrence," in S. Salop, ed., *Strategy, Predation and Antitrust Analysis*, Federal Trade Commission Report, 1981.

Maloney, M. and McCormick, R., "A Positive Theory of Environmental Quality Regulation," *Journal of Law and Economics*, April 1982, *25*, 99–124.

_____, _____, **and Tollison, R.,** "Achieving Cartel Profits Through Unionization," *Southern Economic Journal*, October 1979, *42*, 628–34.

Ordover, Janusz and Willig, Robert, "An Economic Definition of Predation: Pricing and Product Innovation," *Yale Law Review*, November 1981, *91*, 8–53.

Salop, S. C. "Introduction," in his *Strategy, Predation and Antitrust Analysis*, Federal Trade Commission Report, 1981.

_____ **and Scheffman, D. T.,** "Non-Price Predation by a Dominant Firm," unpublished paper, 1982.

Williamson, Oliver, "Wage Rates as a Barrier to Entry: The Pennington Case," *Quarterly Journal of Economics*, February 1968, *85*, 85–116.

Vertical Integration and the Market for Broadcast and Cable Television Programming

Austan Goolsbee
Robert P. Gwinn Professor of Economics
University of Chicago, Graduate School of Business,
American Bar Foundation and National Bureau of Economic Research

April 2007

Acknowledgements: This research was commissioned by the Federal Communications Commission but they are not responsible for its content. The results reflect the findings of the author alone. I would like to thank Rafael Kuhn and Jennifer Li for outstanding research assistance and Robert Gertner, Matthew Gentzkow, Luis Garicano, Betsey Stevenson, Joel Waldfogel, Justin Wolfers and seminar participants at the University of Pennsylvania for helpful comments.

Summary

Historically, regulators in the United States have been concerned about the potential for vertically integrated television entertainment companies—broadcast networks and cable systems which also produce content—to exclude rival programming or to otherwise exploit market power. Successive waves of deregulation and media mergers, however, have generated a tremendous amount of vertical integration in the television industry.

This paper examines the evidence on vertical integration in television programming. It documents its prevalence and presents findings relating to whether integrated producers systematically discriminate against independent content in favor of their own content. It focuses separately on two parts of the video market: primetime broadcast programming and cable network carriage.

For broadcast networks, the vertical integration issue relates to the major networks (e.g., CBS) owning the programs they decide to broadcast during primetime (each network owner also houses at least one major television production unit). The paper documents an extensive amount of vertical integration during primetime. That alone does not answer, however, whether broadcast networks are biased against independent shows. It might be more efficient for networks to make their own shows.

Matching the data on show ownership to data from Nielsen on ratings and advertising rates, the paper devises a market test for bias by testing whether broadcast networks seem to apply an easier standard for carrying their own shows than for carrying independent programming. The results indicate that they do have a significantly lower standard for their own programs. Independent programs need to generate something like 15 to 20 percent higher advertising revenues than comparable in-house programs to get on the air. But the tougher standard does not apply to fully independent programs which are treated almost exactly like the programs that they own. The higher standard seems to apply only to independent shows created by the owners of rival broadcast networks.

For cable, the question of vertical integration relates to the decision of cable systems to carry networks that they own. The data suggest that vertical integration has been getting less prevalent over time. The previous literature has found that cable systems are more likely to carry their own networks but there is considerable debate, though, about whether this results from efficiencies or from market power.

Although the data on cable networks are not sufficient to allow direct market tests like those used on the broadcast networks, they do suggest there is little evidence that vertically integrated networks enjoy major efficiencies over independent rivals on either the revenue or the cost side.

Looking at decisions of cable systems regarding what channels to carry shows that carriage rates for vertically integrated channels are higher on systems that own the given network but only in places where there is not much competition from DBS. This suggests, potentially, a problem for an efficiency based explanation for the behavior. These results form the basis of a potential method to be used in determining when a given market faces sufficient pressure that it be treated as competitive.

1. Introduction

Historically, regulators in the United States have been concerned about the potential for vertically integrated television entertainment companies—broadcast networks and cable systems which also produce content—to exclude rival programming or to otherwise exploit market power. Successive waves of deregulation and media mergers, however, have generated a tremendous amount of vertical integration in the television industry.

This paper examines the evidence on vertical integration in television programming. It documents its prevalence and presents findings relating to whether integrated producers systematically discriminate against independent content in favor of their own content. It focuses separately on two parts of the video market: primetime broadcast programming and cable network carriage.

These two types of television have rather distinct kinds of vertical relationships. For broadcast networks, the concern is over the networks owning the shows they broadcast and the possibility that they might exclude independent programming. Explicit legal restrictions on the ability of the networks to take financial stakes in the shows (known as the financial interest and syndication or fin-syn rules) limited the networks' ability to vertically integrate before 1995, when the rules were abandoned. At that point, however, there was an immediate increase in ownership stakes by the broadcast networks and, consequently, in the concern over what such integration implies for the ability of independent programming to get on the air.

For cable television, the vertical integration concerns the distribution through cable systems (and now through DBS systems) of cable networks (e.g., MTV).

Approaching 90% of the country subscribes to some form of multi-channel video today. These distribution firms frequently also own some of the networks that they broadcast over their distribution systems—like Time Warner owning the Turner networks it carries on its cable systems. There are many public concerns that a vertically integrated media company might make life difficult for independent cable network operators and try to promote their own networks, instead.

This paper will document and explore the vertical relationships in these two parts of the television programming marketplace.

At the outset, though, it is vital to consider the difference between the *existence* of vertical integration in television programming and the *rationale* for it. To the extent there is an existing literature examining some of these questions, it tends to have a hard time answering the the nagging question of why such vertical relationships exist.

One view holds that vertical integration and foreclosing/self-promoting behavior is a strategic move on the part of powerful monopolies and is anti-competitive in nature. The other view, espoused by opponents of regulating such relationships, argue that vertical integration comes about because it is more efficient, that a combined entity is better able to create shows or networks that people will watch or to save money in producing the shows or in some other way generate a synergy. In this view, if it is hard for an independent to get its programming on the air, that is only natural. It need not reflect monopoly power. Indeed, there are some analysts that argue it almost never makes strategic sense for a company with market power in the distribution system to favor its own product in a way that was anti-competitive (the so called 'Chicago School' view).

This basic dichotomy oversimplifies the situation, of course. There are many other, more subtle explanations for vertical integration and for self-promoting behavior. An interested reader should consult survey discussions on the topic such as Motta (2004) or Rey and Tirole (2005). But the discussion of both segments of the television market will start from the premise that when asking about the fate of independent programming in the market, we need to look beyond simply the basic statistics of how many there are to see if there is evidence that the distributors are acting in an anti-competitive way.

One premise underlying the economics of vertical integration is the idea of competition. If the self-promoting activities are strategic in nature and result from a monopsonist exploiting their position, then as that market power falls, their ability to act nefariously should be reduced. There has been considerable discussion since the deregulation of the 1990s that almost every aspect of the television business has become much more competitive than it was in years past. For broadcast networks, there have been new entrants and there has been a major shift of viewers toward cable networks. For cable systems, there is now significant direct competition from Direct Broadcast Satellite (DBS) providers.

With this rise of competition as the backdrop, this study will reexamine the evidence regarding vertical integration in primetime broadcast television and on cable/DBS systems.

2. TELEVISION PROGRAMMING ON BROADCAST NETWORKS

2.1 Historical Background

Starting in 1970, the financial interest and syndication rules issued by the FCC explicitly restricted the ability of major networks to take ownership stakes in various types of programming. With the rise of cable networks, the expansion in the number of broadcast networks, and the skepticism over whether it would really be in the interest of broadcast networks to exclude rivals' programming, the fin-syn rules were abandoned.

Not surprisingly, there was an immediate increase in the vertical ownership of broadcast programming. Caves (2005) documents that the networks' full or partial ownerships stakes in the shows they broadcast rose by two thirds from 1995 to 1998—from 28% to 46%.

This occurred in an era of considerable consolidation of media companies into major conglomerates that combine multiple types of media and make vertical ties much more likely. A company like Disney now owns film studios, several TV production entities, a major broadcast network, and radio stations not to mention retail stores, theme parks and many other properties. The owners of the other major broadcast networks—Newscorp, CBS Corporation, GE, and Time-Warner—each have a massive number of different media businesses involved in content creation and distribution.

There is, without question, a considerable interest in the operation of the primetime programming choices of the major broadcast networks. It is important to note at the outset, however, that the importance of this group has become much smaller over time and that the degree of market power the big three networks once enjoyed in the market for television programming has substantially diminished from its peak. As Caves (2005) observes, in the 1970s the big three networks accounted for around 90 percent of

the prime-time viewing in the U.S. and already by 1996-1997, it was less than half and has continued declining since then.

That said, they still remain extremely powerful television entities with the biggest potential markets and it is worthwhile examining their behavior relating to the shows they own and the shows they do not own.

2.2 Data

To examine the evolution of prime-time programming on broadcast networks, this paper turns to several data sources. First, the *International Television and Video Almanacs* , 2000-2005 (Quigley, 2000-2005) gives the ownership for major prime-time shows each year from the 2000-2001 season to the 2004-2005 season.

The owners are matched to networks from publicly available industry sources in cases where they are not obvious. Clearly when an NBC show's owner is NBC/Universal, it is vertically integrated. There are other content creators, however, like Spelling Television or Greenblatt-Janollari whose vertical integration is not as immediately clear (CBS and Fox, respectively).

The data do not give the actual percentages owned by the various groups—many shows have more than one owner—so the results will not examine the impact of minority versus majority ownership. Following the previous literature, the regression results below will not examine news programs, sports programs, movies or special programs.

Second, I match the Almanac data to data from the *Nielsen Television Index Report*. These data give ratings and advertising rates (CPM) as well as the demographics of the shows' viewers. These two data sources yield a sample of matches that tends to

have a good representation of the major shows, though relatively light representation of very short-lived (less than a full season) shows. All told there are more than 200 different shows with between one and five years of data each.

I also take a completely independent sample of ownership using a different methodology and data source to get more current information. For a single two week period in late February and early March of 2007, this takes all shows scheduled to be aired that week and uses the Internet Movie Database Television site to examine ownership. I do not have Nielsen data to match to the IMDB sample, just the ownership data. Data in Kagan World Media (2005) provides a third useful consistency check on the *Almanac* data.

2.3 Basic Evidence on Vertical Integration of Broadcast Networks and Shows

The summary statistics for the extent of vertical integration in primetime television are presented in tables 1 through 4.

Table 1 takes the matched data sample from the Almanac and the Nielsen Index for the seasons 2000-2001 through 2004-2005 and documents the share of shows on a given network that were produced by that network, i.e., the share that are vertically integrated. Note these percentages are the percentage of self-owned *shows* (as opposed to the amount of time).[1]

There was fairly substantial variation across the networks in this period. CBS was the most vertically integrated with, in most years, around 60-85% of CBS shows having

[1] And it is the share of shows on these networks that also satisfy the previously listed criteria such as being in the *Television Almanac* and matching to the *Nielsen Television Index*. In practice this excludes much of the short-lived programming. This measure also does not take account of the fact that some programs run more frequently or are longer than other programs. All shows are treated the same for computational purposes here.

at least partial ownership from a Viacom owned production company (e.g., Paramount). NBC was the least, averaging less than 40% in most years (their purchase of Universal took place in the last season of the table causing the share to jump substantially).

Looking across primetime as a whole, the share of vertically integrated shows did not change much over these five years. The share hovers in the 55-59% range except for the one 65% year in 2002-03.

For an alternative calculation of this concept, table 2 presents computations from Kagan World Media's (2005) *The Economics of Primetime Television* which provide the share of total minutes during primetime that were of vertically integrated content. This excludes sports, movies and special programming and is based on the debut schedule. Note this is a *time* based measure rather than a *show* based measure so a one hour show counts for double a half-hour show, and so on. Despite being different measures of integration, the ownership shares are fairly consistent across the two samples. The Kagan data also shows CBS to be the most integrated and NBC to be the least.

Table 3 averages over the entire *Almanac* sample to compute an average for each network and then looks at the shows that are not vertically integrated. The share of truly independent programming—programming that is not owned by any company that also owns a broadcast network—is fairly constant across the networks at around 18%. The rest of the shows, around one quarter, are shows made by one broadcast network affiliated company but shown on a different network (like the show *The O.C.* which was made by Warner Brothers but broadcast on Fox). In this sample, Fox had the lowest use of other network created programs in the sample (less than 7%) while the WB had the highest (over 40%).

Finally, table 4 updates the numbers to 2007 using the information from the Internet Movie Database Television site for a two week period in late February/early March 2007. This table again computes the share of *shows* (not time) on primetime television that were vertically integrated for each broadcast network.

One change from the earlier period to 2007 is that the two smaller networks—UPN and the WB—merged to form a single network, the CW. Since the CW is partly owned by Time-Warner and CBS (itself spun off from Viacom with a number of production properties in late 2005), the chance that content on the combined network would be produced by some unit within either of those two companies is obviously higher than for either UPN or the WB alone.

The data show that the CW network has the highest vertical integration, at over 80% and all the other networks are in the 61-69% range. The average across all shows was just short of 68% and the share of true independent programming only around 13%. These may indicate an increase in the vertical integration share since the earlier data of tables 1 and 3 but not necessarily. The sampling methodologies differ so they may not be directly comparable. Tables 1 and 3 look only at shows that appear in the *Television Almanac* and can be matched to the Nielsen report. Table 4 takes a specific two week period and asks who owns each of the shows broadcast during that time.

However one chooses to measure vertical integration, however, the data document that a large fraction, typically the majority of what is on any broadcast network during primetime, was made "in house."

A different way to look at the question is to think about the shares for the production side rather than the broadcast/distribution side. The top panel of tables 4 and

5 look at what share of the primetime shows owned/produced by the company that owns a broadcast network actually end up being broadcast on that company's network. Table 4 looks at the 2000-2001 season through the 2004-2005 season. Table 5 looks at the period in early 2007.

In the longer period, more than 85% of the primetime television programming created by Disney (e.g., Touchstone Television, Buena Vista Television, ABC and so on) appears on ABC. Since Viacom owns both CBS and part of UPN, its row presents two alternative measures. Almost 58% of Viacom owned shows appear on CBS. Around 77% of Viacom owned shows appear on either CBS or UPN.

A relevant reference point for comparison is presented in the lower panel of each table. This gives the ultimate broadcast network for independently produced programs—those with no affiliation with a network company at all. In both tables, the distribution of these independent shows is fairly evenly spread across the networks. Compared to these independent shows, it is very clear that the output from in-house television production is overwhelmingly more likely to show up on its own network than the independents' is.

Taken together, tables 1-6 suggest that primetime broadcast television is a heavily vertically integrated endeavor and one can see that the life of an independent producer of programming is likely to be rather difficult.

The prevalence of this behavior, however, is not the same as explaining why it exists. While some industry participants argue that the data indicate powerful conglomerates are exerting monopsony power or otherwise acting in an anti-competitive manner, others argue that the shows are vertically integrated because that is the most efficient or most natural way to structure the business (see the discussion in Caves, 2005).

Trying to answer *why* there is so much vertical integration is a great deal more difficult than documenting its existence. There are some subtle ways to test, however, using readily available data.

2.4 Testing for Network Bias Against Independent Shows

The question at hand is whether the broadcast networks discriminate against shows they do not have an ownership stake in. Certainly the industry anecdotes discussed in Caves (2005) suggest that network executives have been known to issue what sound very much like ultimatums—"give us an ownership stake or we will not pick up this show," and so on. The issue also comes up in the work of Barth (2003) on renewal rates of vertically integrated shows.

But the efficiency view and the discrimination view, while both predicting vertical integration, differ significantly in their predictions for other parts of the market very much along the lines of the market based tests for discrimination in labor economics first outlined by Nobel Laureate Gary Becker (1971; 1993).

If the efficiency view is what drives things, then the vertically integrated shows should be better, in some sense, than the comparable independent shows. If discrimination, then the vertically integrated shows should be worse.

This is an important point to understand. If the broadcast network is systematically favoring its own shows over independent ones, then they should be holding the independent shows to a *higher* standard and so, for the same time slot and the same observable characteristics, we would expect to see broadcast network's integrated

shows performing *worse* than independents' do—because the marginal independent shows never make it on the air while the marginal vertically integrated ones do.

The key for actually conducting this test is that we have data at the level of the broadcast network and we can run a regression trying to explain their behavior. The network trying to maximize its direct profit should, on the margin, choose between their own shows and the shows of others such that the impact on profit is the same. I will interpret this as being the decision of whether the observable measures of profit seem to differ on the margin depending on who made the show, controlling for the time slot and year of the show.

In this spirit, table 7 turns to the Nielsen data on all of the shows from the 2000-2005 seasons and presents a series of multivariate regressions to test for evidence of discrimination against independent programming. There are around 500 show-year observations.

Each regression includes the dummy variables for the time slot (i.e., day x start time), the network and the year as well as a dummy variable for whether the show is vertically integrated with the network. The standard errors are clustered at the level of the show since there is likely to be serial correlation across time for each show. The regressions ask whether, on the margin, the vertically integrated shows on the air do worse than the non-integrated shows.

Column (1) starts by looking at the log number of households watching the show. The results suggest that on the margin, there is little evidence that vertically integrated shows differ from non-vertically integrated shows in the same time slot. Just as many people watch one as watch the other. There is no evidence of bias here.

But the number of viewers isn't what a broadcast network ought to be maximizing if they are trying to make money. Column (2) looks at a potentially more appropriate measure—the show's total advertising revenue. This is computed as the log of advertising CPM rates (for a 30 second advertisement) times the number of households viewing the show. Without a measure of costs, this measure of revenue is the closest thing we have to short-run profits. These do not take account of the potential for syndication, as will be relevant below. These regressions purposely exclude various types of programs where cost differences would be particularly problematic such as news programs which are almost always vertically integrated but also have a lower cost of production than scripted programs thus could potentially skew the results.

The results in column (2) rather clearly point to the vertically integrated shows performing worse than the independent ones. The independent shows are being held to a higher standard. The integrated shows in the same time slot and the same season must have 16% greater advertising revenues to get on the air. Notice that the result in column (1) suggests this is being driven entirely by the rate advertisers are paying for the shows. The number of viewers was not affected.

Column (3) repeats the regression of column (2) but adds a series of demographic variables about the nature of the audience (from the Nielsen data)—the share of male viewers, the share of eight different age groups, the share that is black, the share that owns a VCR, the share that has each of three different household sizes, the share that has each of three different educational attainment levels and has a high income. The results show that the higher standard that the independent shows are held to is not simply a reflection of the audiences being systematically different for independent shows relative

to vertically integrated ones. Even with these controls for the type of audience, the advertising revenues on the margin are significantly lower for the vertically integrated shows, consistent with them being held to a lower standard than the independents.

Caves (2005) has noted that advertisers value more than just the size and even the demographics of the audience. As he says, advertisers pay significantly higher advertising costs than a shows' observable audience characteristics would predict when the shows are considered to be of higher quality (presumably because advertisers believe that associating their brands with the better shows will lead to more purchases). So it is not unheard of to find advertising differences across shows with the same demogrpahics and audience sizes.

Column (4) breaks out the independent/non-vertically integrated programs into two distinct groups. The first comprises truly independent programming—primetime shows created and owned by entities that are not affiliated with any broadcast network. The other comprises shows created by a company that owns a rival broadcast network.

The results from this regression show that the coefficient for the fully independent programs is, in fact, almost identical to the coefficient on the vertically integrated shows (the comparison group is the rival network created shows). On the margin the vertically integrated shows have 25% less advertising revenue and the fully independent programs 23% less than independent programs made by rival networks. The networks appear to be applying a higher advertising standard *only* to the shows created by their rival networks, not by fully independent entities. They appear to treat the true independents the same way they treat their in-house product.

This result is especially interesting in that it may suggest that while the difficulty of getting independent shows on the air is great, that may be more attributable to the high risk and high costs of producing programming and the escalation of such costs over time than to the strategic squeezing of independent producers by the broadcast networks. It also suggests that a cost based efficiency explanation—that networks apply a lower revenue standard to their own programs because they can make them more cheaply (either from eliminating double marginalization considerations or from better synergies of some kind)—probably will not suffice since the truly independent creators of content should not enjoy those efficiencies.

This is not to say that the picture is rosy for the true independents, by any means. There is no information in these data about syndication profits from shows and the differential coefficients of the fully independent producers relative to the rival networks may simply reflect that the rival networks have more bargaining power over syndication revenue than the fully independent producers do. In other words, if a broadcaster can't get part of the syndication profits from the show's producer, they may require that show to generate higher advertising revenue to put it on the air. That would look like the pattern found in these results but would also mean that independent producers are in weak bargaining positions.

Unfortunately, there is not publicly available information about the syndication contracts of shows and who gets paid what. One suggestive piece of evidence comes from column (5) where the impact of vertical integration and full independence are allowed to vary for shows that have been on the air for six years or less (where syndication is not certain) versus those that have been on the air for longer than six years

(where syndication is very likely if not already in existence). The results show that for the young programs the standards for what gets on the air are more similar (the coefficient on the young programs is derived by adding the two coefficients together)—about 13% lower for the vertically integrated shows and 12% for the fully independent ones relative to rival network's shows—than they are for the older programs whose differentials are around 40 to 50%. The fact that the different standard for outside network programming varies by the age of the show does not necessarily imply that it is driven by syndication profits but it is consistent with that idea.

Taken together the evidence on broadcast television indicates a high degree of vertical integration across all the networks and a small role for true independent producers. It does not, however, document discrimination of broadcast networks against such programming in favor of their own shows. There is only evidence of this kind of bias against the shows created by other rival broadcast networks. Such programs need to earn significantly higher advertising revenues than a comparable independent or in house program to get on the air.

3. TELEVISION PROGRAMMING ON CABLE NETWORKS

3.1 Background

The other major outlets for television programming are, of course, cable and satellite networks. These networks have increased in importance and quantity over time and now account for the majority of television watched in the country. The last FCC

competition report indicates there were more than 530 such networks in 2005 (FCC, 2006).

Cable networks have a different type of vertical integration issue than do broadcast networks. Namely, there are several major owners of cable systems such as Time-Warner that also own several of the cable networks that appear on their cable systems.

This type of vertical integration between content and distribution has generated a history of both regulatory concern at the FCC and an academic literature studying the subject. At various times the FCC has had restrictions regarding vertically integrated cable systems such as restrictions on what share of channels can be owned by the parent company and on the terms that carriage deals can take for internal versus external networks. Issues of vertical integration have also been important in FCC merger evaluations such as Time-Warner/Turner and Newscorp/DirecTV.

Some of these issues of vertical integration were more problematic when cable systems had low channel capacity and the system owner's choice of networks was extremely binding. With the advent of satellite and of digital cable, though, the channel capacity has increased significantly (as has the number of available networks) so the marginal channel not getting on the air is a much more niche network than in earlier years. Nevertheless, the issue remains of substantial public interest.

The existing literature on the impact of vertical integration in cable television has typically looked at the cable system level and the decision of whether to carry a network or not. The literature has somewhat consistently found that cable systems are more likely to carry their own networks than are similar non-vertically integrated systems. Most of

this research was conducted in the context of low channel capacity systems and little competition (unlike today).

Important contributions from this literature include work using data from the 1980s and early 1990s such as Waterman and Weiss (1996, 1997), Chipty (2001), Ford and Jackson (1997), and the work discussed in Owen and Wildman (1992). There is also a more recent literature using more recent data including Wegener and Caves (2003), Kang (2005), Chen and Waterman (2006), and Suzuki (2006). Some of these papers have been explicitly concerned with the vertical integration of premium/pay cable networks such as HBO while others have dealt with basic cable networks like CNN.

One of the important points raised in the discussion of this previous literature is quite similar to the discussion above about the broadcast networks. Before one can decide whether self-carriage is harmful, one needs to address the issue of whether the existence of vertical integration is driven by efficiencies. This is not easy to do.

Some previous work has produced evidence that efficiencies may be important. Ford and Jackson (1997) claim that vertically integrated networks save money on programming costs. Chipty (2001) and Suzuki (2006) present evidence that vertical integration can lower prices or, even when they raise prices, increase subscribership so that, on net, the vertical integration increases consumer welfare.

Others have suggested a more strategic motivation, even in the more contemporary period. Chen and Waterman (2006), for example, argue that even after the supposed rise of competition from DBS, cable networks are still engaged in a new form of exclusion whereby they relegate competitors' networks to digital tiers where fewer people subscribe.

The ideal way to address this issue would have been to apply a market based test like the one used above for the broadcast networks. This would entail looking at each system and the amount of revenue coming in to that system from the marginal channels they choose to carry and testing whether they apply a different threshold to their own channels. Unfortunately, the data are not detailed enough to do that test. There is network level data on the performance of channels nationally and there is system level information about what networks a system carries but there is not system level data on network performance.

So the evidence on the subject will be, by its nature, more suggestive than the evidence available on the broadcast networks. Nevertheless, it provides some interesting facts about the industry.

3.2 The Extent and Evolution of Vertical Integration Among Cable Networks

There have been several major shifts in the amount of vertical integration in the industry over the years as well as in the market for multiple system operators (MSOs) and DBS providers. The concentration, on a national basis, of the top companies has grown over time with the considerable consolidation of cable and the rapid growth of DBS. On a market-by-market basis, however, it has been the opposite. Each market has gone from a virtual monopoly for the local cable franchise to a market where the cable franchise shares the market with the two major DBS providers.

As the MSOs have shifted hands, the vertical integration of leading networks has changed, as well. For example, when TCI was the leading MSO, they also owned a large number of networks. They sold their cable systems to AT&T who divested most of the

networks. Newscorp has always owned a large number of networks and been independent but then in the mid 2000s bought DirecTV, suddenly making all of their networks vertically integrated.

Tables 8A and 8B show that of the top 15 networks as measured by the size of their prime time audience, the share of vertically integrated networks has been falling over time, from eight in 1997 to four in 2005. It has been falling in the wider market, as well. As table 9 indicates, according to the FCC reports, the number of cable networks has expanded dramatically in recent years but the expansion has been concentrated mainly among independent networks not vertically integrated ones. The number of networks increased by 359 in the ten years from 1996 to 2005. Independent networks made up 311 of that 359, vertically integrated networks only 48. The share of networks identified by the FCC as being vertically integrated has basically been cut in half over this period—from almost 40% in 1996 to just over 20% in 2005.

So the initial evidence in cable looks fairly different that it does for broadcast network shows in that it shows vertical integration to be on the decline. It is important to remember, however, that in the world of cable networks, there are very few small scale, independent operators. In this part of the market, independents are, themselves, giant media companies like Disney and Viacom. It is difficult to find a single major cable network owned by someone other than a major media conglomerate.

That said, it is still worth trying to understand why vertically integrated systems tend to be more likely to carry their own channels than independent cable systems and whether this can be attributed to market power.

3.3 The Performance of Cable Networks: In Search of Efficiency Explanations

The first data I examine are the national level information on cable network performance.

With efficiency gains to vertical integration, the benefits of having more vertically integrated subscribers should show up on either the revenue side or the cost side. Networks with larger vertically integrated potential subscriber bases should either be better at generating revenue (e.g., by creating programs to fit market demand or by seeding and popularizing new networks) or else they should be better at creating successful content more cheaply (e.g., by eliminating double marginalizing monopoly markups or by keeping costs down with economies of scale). These have been the types of efficiency claims made in the previous literature.

The data on 120 leading cable networks in Kagan World Media's (2006) *The Economics of Basic Cable Networks 2007* yields a sample of networks for a ten year period from 1997 to 2006 that includes information on the number of subscribers, the total programming expenses, the total revenues for each channel and then the total number of viewers (average 24 hour ratings) for a smaller set of established networks. It is not a balanced panel in that some of the networks enter after 1997.

I match to this sample to information taken from the FCC's Annual Video Competition reports and their data on which networks were vertically integrated in the given year and the size of those vertically integrated systems (in number of subscribers). The measure of vertical integration used in the regressions is the ratio of potential vertical subscribers to current subscribers for the network, i.e., the total number of subscribers to

cable systems owned by the network's parent divided by the total number of subscribers to the channel at the start of the year.

There is a lot of variation in this measure across channels and across time. For a channel like ESPN, which is owned by the non-vertically integrated Disney, this would be zero in all years. For a channel like CNN which is owned by Time-Warner, this will be the ratio of Time-Warner cable subscribers to CNN's total subscriber base in each year. For a channel like F/X, whose owner Newscorp became vertically integrated late in the sample, or BET which was owned by TCI (vertically integrated) and then sold to Viacom (not vertically integrated), the ratio will be zero in some years and positive in others.

The basic methodology will be to regress the performance of the network—measured in subscribers, revenues, expenses, and so on—on the extent of vertical integration for the network (with network and year dummies as controls).

Column (1) of Table 10A shows that doing this regression on the log number of subscribers suggests that there is a very small *negative* effect of vertical integration on the number of subscribers a channel has. So when a channel goes from being independent to being integrated, it loses subscribers. Column (2) looks at the subscriber growth rate rather than the number of subscribers and here finds a small positive coefficient. Becoming vertically integrated leads the subscriber growth rate to increase by a small amount.

Column (3) looks at the subset of networks where the Kagan data contains information on the number of viewers as well as the number of subscribers. It shows

that, conditional on the number of subscribers, the number of viewers actually watching the channel falls when it becomes vertically integrated.

Columns (4) to (6) look at the impact of becoming vertically integrated on the amount of license revenue the network gets from the distribution systems, the amount of advertising revenue they bring in and the amount they spend on programming (each regression controls for the number of subscribers since that influences the dependent variable directly). Here, again, there is very little evidence that vertical integration of a channel has any noticeably beneficial impact on revenues or costs. The same network performs exactly as well before and after it is vertically integrated.[2]

Some in the previous literature have argued that the efficiencies of vertical integration flow only to start-up networks, not to well established ones. So Table 10B presents the same regressions of Table 10A but restricts the sample only to networks started since 1997. Among these younger networks, there are no major differences with the results for all networks. There is no evidence that when new networks become vertically integrated it increases subscribers or changes their subscriber growth rates. The number of viewers, conditional on how many subscribers there are, actually seems to fall (though here it is not significant), and there is no evidence that license revenue, advertising revenue or programming cost adjust in a beneficial way when a network vertically integrates.

[2] An endogeneity concern with these regressions would be if the vertical ownership changes for a given network were endogenous to the performance of that network (i.e., if vertical owners sell networks whose performance is poor, say) but most of these ownership changes take place with the acquisition or divestment of massive media conglomerates. There is only a small amount of trading of individual networks between owners.

So collectively, the network level data does not suggest much of an impact of vertical integration on the performance of basic cable networks. Perhaps that helps explain why the degree of vertical integration has been falling over time.

3.4 Carriage Rates for Growing Networks in the Face of DBS competition: The Search for Evidence on Propensities to Self-Carry

Next, I revisit the decisions of cable systems on what networks to carry. This has been the main subject of the previous literature. Those papers have, rather consistently found that vertically integrated systems are more likely to carry their own networks but have disagreed over whether this is due to efficiency reasons or strategic foreclosure reasons.

The regressions will use the same system level data used in previous work (but updated to the present) which is Warren Publishing's Television and Cable Factbook 2007.[3] This is the same data source used in much of the literature on the cable industry such as Goolsbee and Petrin (2004), Crawford (2000), Chipty (2001), and others.

I restrict the sample to cable systems for whom the number of subscribers is not missing and that have at least 5000 subscribers. Excluding the smaller systems reduces the number of observations substantially but hardly affects the total subscribership across the entire sample (because big systems predominate). This sample restriction yields around 1,400 different cable systems. To this sample of systems, I match demographic information from the census for the main county the cable system operates in.

[3] The actual dataset used here was first edited by Greg Crawford who kindly provided it.

The basic methodology is to do a Probit regression of the system's {0,1} decision to carry a network as a function of whether the system's parent owns the channel, demographic variables, and characteristics of the system. The results will differ from the previous literature, however, by adding an important variable to indicate the amount of competition facing the cable system which is the share of DBS subscribers in the television market.

Most previous work looked at system decisions at a time when they faced little competition in distribution. For the anti-competitive explanation of their propensity to carry their own networks, as the amount of competition rises in the market, their ability to get away with self-promotion should be reduced. If they promote their own, inferior channels, they should start losing customers to the unaffiliated DBS providers whose only interest is in giving the customers exactly the networks they want. In practical terms, the positive coefficient on vertical integration in the regression of carriage rates should fall when the market's DBS share is higher.

Alternatively, the efficiency explanation of self-carriage does not suggest the relationship should fall with higher DBS penetration. That argument holds that the cable system really *ought* to be vertically integrated with the channels in question because they enhance the system's ability to either deliver better content or to invest in relationship specific assets or avoid double marginalization, or many others. How much competition they face is not relevant.

Before looking in detail though, it is important to note that the historic literature on vertical integration and the carriage decision no longer applies to most of the major vertically integrated networks because all of them are carried on virtually all major cable

systems. This is readily apparent when looking at the top part of Table 11 which lists six of the biggest vertically integrated networks–AMC, CNN, TBS, TNT, and Discovery. Each one has carriage rates close to 100% in this sample for both affiliated and non-affiliated cable systems. Clearly there is little scope for strategic behavior when every system carries has enough capacity to carry all the major channels (although the work of Chen and Waterman, 2006 does show that there may still be interesting decisions regarding what networks get carried on the digital versus the analog tier).

A similar problem faces fledgling networks whose take up is extremely low on both types of systems. So the results here will focus on wholly or partially vertically integrated basic cable networks identified in FCC (2006) that have carriage rates between 5% and 90% in the sample of cable systems. This was 16 networks, though data sample problems forced dropping four of these.[4] The sample includes BBC America, Boomerang, CNN International, FitTV, FUSE, G4, PBSKids, The Science Channel, Style, The Travel Channel, TVOne, and WE.

The lower panel of table 11 gives the carriage rates for these channels among the systems they are vertically integrated with and the ones they are not. In many cases, the integrated systems are more likely to carry their own channels. But in many cases, the opposite is true.

These raw sample means do not, however, answer whether the vertical systems favor their own networks. If all the Time-Warner cable systems are in locations with digital capacity and a large number of intensive TV watchers, for example, then they will

[4] Likely because of name matching problems, the share of systems in the data reporting carriage of CourtTV, the Independent Film Channel, DiscoveryHealth and the Outdoor Life Network were so different from the subscriber counts in the *Economics of Basic Cable Networks* that I excluded them.

be more likely to carry the Time-Warner (and other) channels for those reasons but it will look like it is due to vertical integration.

What is needed is a multivariate regression framework that estimates the impact of vertical integration on the probability of carrying the network while controlling for the impact of demographic factors in the system's area and for the system's characteristics.

The results will follow the previous literature on carriage decisions by including as demographic variables information from the U.S. Census for the county the system operates in. The variables are: population density, the population growth rate, the share of the population that is Hispanic, the share of the population that is under 18, the share of the population that is over 65, the average household size, the log of median income, and the share of people that own their home.

For variables indicating the quality and/or capacity of the cable system the regressions include the share of the cable system's plant that is fiber, a dummy for whether the system is two-way capable, and a dummy for whether it is an analog only system (i.e., no digital service is offered). Adding channel capacity did not change the results but is missing from a large number of the system level observations and thus dramatically reduced the sample. For that reason it is not included in the regressions. The standard errors are clustered at the level of the system owner since the decisions of the multiple systems owned by the same MSO are likely not to be truly independent.

The regressions also add the amount of DBS competition in the Designated Market Area. As mentioned above, if the self-carriage rates for vertically integrated cable systems decline with the amount of competition from DBS, this could be a sign that

the positive relationship does not have an efficiency explanation but is perhaps driven by market power considerations.

In each case, the regression includes the DBS share interacted with the vertical integration dummy (which should be negative if more DBS competition reduces the positive self-carriage rates) and also the DBS share on its own in case the kinds of places that have a lot of DBS subscribers differ in their unobservable tastes for the channel in some systematic way.

Of the eleven networks listed in the tables 12A to 12K, nine show significant evidence that having an ownership in a network makes systems significantly more likely to carry it (and one that is positive but not significant). In this sense, the findings very much comport with the previous literature. The interesting addition is that for nine of the eleven networks, the higher is the DBS share in the local market, the more attenuated that relationship becomes. For those nine, the interaction of vertical integration with the DBS share has a significant negative coefficient.[5] This evidence suggests, perhaps, an explanation rooted in competitive pressures rather than efficiencies.

In each of these cases, it is also possible to ask at what level of DBS market share would the vertical integration effect on carriage rates be reduced to zero. This is computed simply by dividing the one coefficient by the other (and taking absolute values). The DBS share is only a crude measure of competition but certainly if one believed the coefficients were entirely documenting the way competition reduces market

[5] For the one network not listed in the tables, FUSE, the probit showed a significant positive coefficient on vertical integration and a significant negative on the interaction with DBS but something in the data lead the standard errors to be absurdly small and the coefficients absurdly large. A linear probability model of carriage showed a very similar positive coefficient on vertical integration and a negative on the vertical integration interacted with the DBS share found for the other networks. Rather than try to figure out what feature of the data caused this strange probit result, it is simply excluded from the presentation.

power, this would give something like a threshold test for how large the competition share must be for the cable system to lose its ability to promote its own channels. This is listed for each network at the bottom of the table. The DBS share that makes the vertical integration effect equal to zero averages around 20 to 25 percent and in all but one ranges from 7 to 42.

This kind of calculation is obviously meant only to be suggestive. But applied with better data to more narrowly defined markets, this type of approach might be able to provide an empirical basis for the threshold-type exemptions often used by the FCC and other regulatory agencies where certain markets or firms are exempted from regulation when they have been deemed to be "competitive."

Thus for at least a subset of the networks there is evidence consistent with the view that DBS competition reigns in the ability of cable systems to use a vertically integrated position to promote their own channels. And the evidence at the network level gives little evidence that that vertically integrated networks attract more subscribers, grow faster, raise more ad revenue or licensing fees or have lower programming costs.

4 Conclusion

This study has documented the extent of vertical integration in television programming and tried to explore some of the apparent motivations for it.

In broadcast television, the results show that vertical integration of networks into the shows they broadcast is large and rising in numerous, independent data sets. Looking at the decisions, the networks appear to treat independent programs differently than they treat their own programs—requiring them to have higher advertising revenues to get on

the air. This rule of thumb does not seem to apply to fully independent programs, which are treated exactly as the integrated programs are. The higher standard seems to be reserved for shows owned by rival broadcast networks.

Among cable systems, the evidence suggests that vertical integration is a shrinking part of the business as the number of networks in existence continues to expand robustly while the number of vertically integrated networks has grown slowly. There is not sufficient data to conduct a market test of discrimination like the one for broadcast network shows but the data that do exist do not reveal any clear evidence of efficiencies for channels that have a larger vertically integrated component. Further, the evidence on self-carriage rates show that vertically integrated cable systems are more likely to carry their own channels except in places where there is sufficient competition from DBS. Both of these facts may be consistent with the view that the efficiency gains from vertical integration today are relatively small and that competition has taken away the ability of cable systems to take as much strategic advantage as they may have once done.

TABLE 1: SHARE OF SHOWS THAT ARE OWNED BY THE NETWORK BY YEAR AND BROADCAST NETWORK

	TOTAL	ABC	CBS	FOX	NBC	UPN	WB
2000-01	**55.2%**	54.2%	68.2%	91.4%	40.0%	54.5%	31.6%
2001-02	**54.4%**	50.0%	80.0%	64.7%	35.2%	50.0%	35.7%
2002-03	**65.1%**	64.3%	85.7%	76.9%	38.5%	66.7%	46.1%
2003-04	**59.4%**	60.0%	70.8%	76.9%	33.3%	55.6%	58.3%
2004-05	**56.8%**	66.7%	57.1%	61.5%	46.7%	44.4%	58.3%

Source: *International Television Almanac* and *Nielsen Television Index*, computed as described in text.

TABLE 2: SHARE OF TIME IN DEBUT PRIMETIME SCHEUDLE COMPRISED OF SHOWS OWNED BY THE NETWORKS

Network	ABC	CBS	FOX	NBC	UPN	WB
2001-02	50.0%	81.8%	70.%	40.9%	40.0%	30.8%
2002-03	53.8%	77.3%	73.3%	35.0%	53.8%	43.2%
2003-04	52.3%	77.3%	60.0%	38.6%	40.0%	69.2%
2004-05	54.5%	75.0%	56.7%	56.8%	30.0%	65.4%

Source: Kagan, *Economics of Prime Time Television 2005*

TABLE 3: OWNERSHIP OF PRIMETIME SHOWS BY BROADCAST NETWORK FOR 2000-2001 TO 2004-2005 SEASONS

	TOTAL	ABC	CBS	FOX	NBC	UPN	WB
Vertically Integrated Shows	**58.0%**	57.8%	72.2%	74.8%	38.6%	54.2%	44.3%
Shows Owned by Other Networks	**23.5%**	24.5%	11.1%	6.8%	37.3%	25.0%	41.4%
Fully Independent Shows	**18.5%**	17.6%	16.7%	18.4%	24.1%	20.8%	14.3%

Source: *International Television Almanac* and *Nielsen Television Index*, computed as described in text.

TABLE 4: OWNERSHIP OF PRIMETIME SHOWS BY BROADCAST NETWORK FOR FEBRUARY/MARCH 2007

	TOTAL	ABC	CBS	FOX	NBC	CW
Vertically Integrated Shows	**67.9%**	66.7%	68.4%	66.7%	61.1%	81.8%
Shows Owned by Other Networks	**19.3%**	22.2%	21.1%	8.3%	27.8%	9.1%
Fully Independent Shows	**12.8%**	11.1%	10.5%	25.0%	11.1%	9.1%

Source: IMDB.com (2007), computed as described in text.

TABLE 5: NETWORK DISTRIBUTION FOR PROGRAMMING, BY OWNER 2000-2001 THROUGH 2004-2005 SEASONS

Programming Owned By:	**Share broadcast on own network**
DISNEY/ABC	86.2%
VIACOM/CBS	57.8% (77.0% with UPN)
NEWSCORP/FOX	58.7%
GE/NBC	96.8%
TIME-WARNER/WB	38.0%
Fully Independent Programs:	**Share broadcast by network**
ABC	16.3%
CBS	21.1%
FOX	15.8%
NBC	23.4%
UPN	11.7%
WB	11.7%

Source: Author's calculation using Television Almanac and Nielsen Television Index as described in the text.

TABLE 6: NETWORK DISTRIBUTION FOR PROGRAMMING, BY OWNER FEBRUARY/MARCH 2007

Programming Owned By:	**Share broadcast on own network**
DISNEY/ABC	70.6%
CBS CORP./CBS	68.4% (84.1% with CW)
NEWSCORP/FOX	50.7%
GE/NBC	91.7%
TIME-WARNER & CBS CORP./CW	26.5% (38.9% for Time-Warner alone)
Fully Independent Programs:	**Share broadcast by network**
ABC	20.0%
CBS	20.0%
FOX	30.0%
NBC	20.0%
CW	10.0%

Source: Author's calculation using IMDB.com (2007) as described in text.

TABLE 7:
TESTING FOR NETWORK BIAS AGAINST INDEPENDENT PROGRAMMING

	(1)	(2)	(3)	(4)	(5)
Dependent Var. (in logs)	Households	Ad Rev.	Ad Rev.	Ad Rev.	Ad Rev.
Vertically Integrated	-.021 (.034)	-.164* (.074)	-.204* (.085)	-.254** (.095)	-.514** (.192)
Fully Independent				-.227+ (.133)	-.379* (.191)
Vertically Integ. x 0-6 yrs					.379+ (.206)
Fully Independent x 0-6 yrs					.258 (.216)
Other Audience Controls			Male% Age Group% Black% VCR% HHSize% Educ.%		
Dummies	Year Network Time Slot	Year Network Time Slot	Year Network Time Slot	Year Network Time Slot	Year Network Time Slot
Observations	514	495	313	495	495
R2	.833	.811	.882	.814	.829

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: Each of these is a regression at the program level with the dependent variable being the log of the variable listed at the top of the column. The total advertising revenue is computed as the CPM advertising rates times the number of viewing households. Both numbers come from Nielsen. The sample covers primetime programs from the 2000-2001 season to the 2004-2005 season as described in the text. The standard errors in parentheses are corrected for serial correlation at the level of the program because there is correlation across time for the same show. Each column includes dummies for year, network and time slot (i.e., day x time period). Column three also includes the share of the audience for the show that is male, that is part of the eight different age groups, that is black, that owns a VCR, that lives in three different household sizes, and that has each of three different levels of educational attainment as a high income. Column (5) looks at shows of age 0 to 6 only.

TABLE 8A:
VERTICAL OWNERSHIP OF NETWORKS BY PRIMETIME RATINGS, 1997

1 TNT	Time Warner (100%)
2 Nickelodeon/Nick at Nite	
3 TBS	Time Warner (100%)
4 USA Network	
5 Lifetime Television	
6 Arts & Entertainment (A&E)	
7 ESPN	
8 The Discovery Channel	TCI (49%), Cox (24.5%)
9 The Cartoon Network	Time Warner (100%)
10 The Family Channel	
11 TNN (The Nashville Network)	Time Warner (100%)
12 CNN	Time Warner (100%)
13 Sci-Fi Channel	
14 The Learning Channel	TCI (49%), Cox (24.5%)
15 fX	TCI (50%)

Source: FCC (1997)

TABLE 8B:
VERTICAL OWNERSHIP OF NETWORKS BY PRIMETIME RATINGS, 2005

1 Nickelodeon	
2 TNT	Time Warner (100%)
3 Nick at Nite	
4 USA Network	
5 Disney	
6 Lifetime	
7 Toon Disney	
8 TBS	Time Warner (100%)
9 Spike TV	
10 Fox News Channel	Newscorp (100%)
11 History Channel	
12 ESPN	
13 MTV	
14 Discovery Channel	Cox (24.5%) Advance (24.5%)
15 Sci Fi Channel	

Source: FCC (2006)

TABLE 9: NUMBER OF CABLE NETWORKS

Year	National Networks	Vertically Integrated	Independent Networks
1996	172	68	104
1997	245	95	150
2004	388	89	299
2005	531	116	415

Source: FCC (1997) and FCC (2006)

TABLE 10A: INTEGRATION AND CABLE NETWORK PERFORMANCE ALL NETWORKS, 1997-2006

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Var. (in logs)	Subscribers	Δ Subs	Viewers	License Rev.	Ad Rev.	Program. Costs
Vertical Integration Ratio	-.053** (.005)	.026** (.006)	-.548* (.238)	-.099* (.040)	.003 (.011)	.002 (.007)
Ln (Subscribers)			1.180** (.163)	1.114** (.087)	1.177** (.085)	.505** (.059)
Dummies	Year Network	Year Network	Year Network	Year Network	Year Network	Year Network
Obs.	996	926	388	869	865	996
R2	.866	.610	.986	.963	.972	.962

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are regressions where the dependent variable is the log of the variable (or the change in the log of the variable) listed at the top of the column as described in the text. The sample includes networks from 1997-2006. The vertical integration ratio is the ratio of the number of subscribers to systems vertically integrated with the network divided by the network's subscriber base.

TABLE 10B: INTEGRATION AND CABLE NETWORK PERFORMANCE NEW NETWORKS, 1997-2006

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Var. (in logs)	Subscribers	ΔSubs.	Viewers	License Rev.	Ad Rev.	Program. Costs
Vertical Integration Ratio	.017 (.034)	-.007 (.044)	-.382 (.893)	-.109* (.045)	-.009 (.029)	.019 (.033)
Ln (Subscribers)			1.020** (.232)	.809** (.153)	.812** (.175)	.466** (.082)
Dummies	Year Network	Year Network	Year Network	Year Network	Year Network	Year Network
Obs.	336	276	44	245	268	336
R2	.887	.700	.951	.941	.942	.933

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are regressions where the dependent variable is the log of the variable (or the change in the log of the variable) listed at the top of the column as described in the text. The sample includes networks from 1997-2006. The vertical integration ratio is the ratio of the number of subscribers to systems vertically integrated with the network divided by the network's subscriber base.

TABLE 11:
CARRIAGE RATES FOR INTEGRATED NETWORKS BY SYSTEM TYPE

Network	System Owns Network	System Does Not Own Network
AMC	98.7%	98.4%
CNN	99.9%	99.8%
Discovery	100%	99.8%
TBS	100%	97.2%
TNT	100%	99.5%
Boomerang	43.4%	13.2%
BBC America	89.3%	38.2%
CNN International	17.7%	5.1%
FitTV	24.5%	45.9%
FUSE	57.6%	60.8%
G4	83.7%	93.9%
PBS Kids	2.1%	8.8%
Science Channel	4.1%	15.6%
Style	5.4%	6.1%
Travel Channel	97.4%	79.7%
TVOne	7.2%	9.0%
WE	97.2%	71.2%

Notes: This table gives the share of cable systems in the sample that carry the network given in the column to the left depending on whether the system is vertically integrated with the network or not.

TABLE 12A: CARRIAGE DECISION: BBC AMERICA

	(1)
Vertically Integrated	**2.733****
	(0.719)
Vertically Integrated x DBS%	**-0.066***
	(0.028)
DBS%	-0.017
	(0.015)
Fiber Share of Total Plant	-0.013
	(0.065)
Analog System	-3.569**
	(0.378)
System Two-Way Capable	-0.246
	(0.359)
Population Density	-0.035
	(0.023)
Population Growth Rate	0.010+
	(0.005)
Hispanic%	0.017+
	(0.009)
Age <18 %	0.036
	(0.073)
Age 65+ %	-0.061
	(0.038)
Black %	-0.005
	(0.008)
Population per HH	-2.380+
	(1.420)
Ln (Income)	-0.150
	(0.701)
Homeowner %	0.007
	(0.011)
Observations	1407
DBS Share for VI Neutrality	41.3%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12B: CARRIAGE DECISION: BOOMERANG

Variable	(1)
Vertically Integrated	**1.407****
	(0.474)
Vertically Integrated x DBS%	**-0.039***
	(0.018)
DBS%	0.003
	(0.014)
Fiber Share of Total Plant	0.295**
	(0.076)
Analog System	-1.191*
	(0.473)
System Two-Way Capable	0.480
	(0.365)
Population Density	-0.043
	(0.046)
Population Growth Rate	0.007*
	(0.003)
Hispanic%	-0.010+
	(0.005)
Age <18 %	0.052
	(0.066)
Age 65+ %	-0.015
	(0.022)
Black %	-0.007
	(0.005)
Population per HH	-0.037
	(0.980)
Ln (Income)	-0.790*
	(0.397)
Homeowner %	-0.010
	(0.006)
Observations	1407
DBS Share for VI Neutrality	36.0%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12C: CARRIAGE DECISION: CNN INTERNATIONAL

Variable	(1)
Vertically Integrated	**2.279****
	(0.374)
Vertically Integrated x DBS%	**-0.066****
	(0.013)
DBS%	0.006
	(0.011)
Fiber Share of Total Plant	-0.014
	(0.102)
Analog System	-0.477
	(0.579)
System Two-Way Capable	
Population Density	0.001
	(0.019)
Population Growth Rate	0.003
	(0.004)
Hispanic%	-0.012*
	(0.006)
Age <18 %	0.057
	(0.048)
Age 65+ %	-0.049*
	(0.023)
Black %	-0.007
	(0.006)
Population per HH	-1.296**
	(0.478)
Ln (Income)	-0.343
	(0.704)
Homeowner %	-0.004
	(0.014)
Observations	1339
DBS Share for VI Neutrality	34.6%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero. The two-way capable variable equal to one predicted carriage perfectly for a number of observations so the estimation dropped that variable and the observations from the regression, resulting in the lower number of observations.

TABLE 12D: CARRIAGE DECISION: FIT TV

Variable	(1)
Vertically Integrated	**1.414****
	(0.524)
Vertically Integrated x DBS%	**-0.074***
	(0.032)
DBS%	-0.002
	(0.010)
Fiber Share of Total Plant	-0.109*
	(0.050)
Analog System	-2.249**
	(0.448)
System Two-Way Capable	-0.317
	(0.300)
Population Density	0.008
	(0.020)
Population Growth Rate	-0.003
	(0.003)
Hispanic%	-0.004
	(0.007)
Age <18 %	-0.032
	(0.026)
Age 65+ %	0.021
	(0.018)
Black %	-0.015*
	(0.007)
Population per HH	1.284*
	(0.552)
Ln (Income)	-0.072
	(0.381)
Homeowner %	-0.010
	(0.012)
Observations	1339
DBS Share for VI Neutrality	19.1%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero. The two-way capable variable equal to one predicted carriage perfectly for a number of observations so the estimation dropped that variable and the observations from the regression, resulting in the lower

TABLE 12E: CARRIAGE DECISION: G4

Variable	(1)
Vertically Integrated	**0.151**
	(0.276)
Vertically Integrated x DBS%	**-0.021+**
	(0.011)
DBS%	-0.005
	(0.011)
Fiber Share of Total Plant	0.107+
	(0.065)
Analog System	-2.692
	(0.224)
System Two-Way Capable	0.308
	(0.292)
Population Density	-0.037*
	(0.015)
Population Growth Rate	-0.006
	(0.005)
Hispanic%	0.001
	(0.007)
Age <18 %	-0.019
	(0.038)
Age 65+ %	-0.011
	(0.017)
Black %	-0.002
	(0.008)
Population per HH	0.141
	(0.340)
Ln (Income)	0.559
	(0.427)
Homeowner %	-0.005
	(0.009)
Observations	1407
DBS Share for VI Neutrality	7.2%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12F: CARRIAGE DECISION: PBS KIDS

Variable	(1)
Vertically Integrated	**2.043****
	(0.689)
Vertically Integrated x DBS%	**-0.156****
	(0.028)
DBS%	-0.015
	(0.020)
Fiber Share of Total Plant	0.380*
	(0.190)
Analog System	-0.602
	(0.560)
System Two-Way Capable	
Population Density	-0.035
	(0.033)
Population Growth Rate	0.016**
	(0.005)
Hispanic%	-0.035
	(0.018)
Age <18 %	-0.015
	(0.054)
Age 65+ %	0.078*
	(0.033)
Black %	-0.025*
	(0.011)
Population per HH	0.897
	(0.979)
Ln (Income)	0.969+
	(0.551)
Homeowner %	-0.049*
	(0.020)
Observations	1339
DBS Share for VI Neutrality	13.2%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero. The two-way capable variable equal to one predicted carriage perfectly for a number of observations so the estimation dropped that variable and the observations from the regression, resulting in the lower number of observations.

TABLE 12G: CARRIAGE DECISION: SCIENCE CHANNEL

Variable	(1)
Vertically Integrated	**3.533****
	(0.359)
Vertically Integrated x DBS%	**-0.194****
	(0.017)
DBS%	-0.014
	(0.016)
Fiber Share of Total Plant	-0.101*
	(0.047)
Analog System	-1.337**
	(0.429)
System Two-Way Capable	-0.615*
	(0.286)
Population Density	0.035
	(0.025)
Population Growth Rate	-0.003
	(0.004)
Hispanic%	0.013+
	(0.007)
Age <18 %	-0.119**
	(0.045)
Age 65+ %	0.033
	(0.023)
Black %	0.009
	(0.006)
Population per HH	1.749**
	(0.676)
Ln (Income)	0.913*
	(0.424)
Homeowner %	0.017*
	(0.007)
Observations	1407
DBS Share for VI Neutrality	18.2%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12H: CARRIAGE DECISION: STYLE

Variable	(1)
Vertically Integrated	**-0.035**
	(0.269)
Vertically Integrated x DBS%	**-0.009**
	(0.008)
DBS%	-0.006
	(0.006)
Fiber Share of Total Plant	-0.072
	(0.112)
Analog System	-1.107
	(0.570)
System Two-Way Capable	-0.680*
	(0.293)
Population Density	-0.018
	(0.018)
Population Growth Rate	-0.001
	(0.003)
Hispanic%	0.014*
	(0.007)
Age <18 %	-0.047
	(0.039)
Age 65+ %	-0.018
	(0.016)
Black %	0.004
	(0.006)
Population per HH	-0.109
	(0.557)
Ln (Income)	-0.168
	(0.240)
Homeowner %	0.005
	(0.006)
Observations	1407
DBS Share for VI Neutrality	0%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12I: CARRIAGE DECISION: TRAVEL CHANNEL

Variable	(1)
Vertically Integrated	**0.752+**
	(0.390)
Vertically Integrated x DBS%	**0.002**
	(0.008)
DBS%	-0.015*
	(0.007)
Fiber Share of Total Plant	0.140**
	(0.051)
Analog System	-1.156**
	(0.247)
System Two-Way Capable	0.508+
	(0.265)
Population Density	-0.014
	(0.016)
Population Growth Rate	0.002
	(0.003)
Hispanic%	-0.005
	(0.008)
Age <18 %	-0.004
	(0.037)
Age 65+ %	-0.020
	(0.013)
Black %	-0.012
	(0.008)
Population per HH	-0.040
	(0.451)
Ln (Income)	-0.315
	(0.251)
Homeowner %	-0.009
	(0.008)
Observations	1407
DBS Share for VI Neutrality	--

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

TABLE 12J: CARRIAGE DECISION: TV ONE

Variable	(1)
Vertically Integrated	**1.583****
	(0.497)
Vertically Integrated x DBS%	**-0.061****
	(0.019)
DBS%	-0.010
	(0.017)
Fiber Share of Total Plant	0.278+
	(0.144)
Analog System	-0.500
	(0.505)
System Two-Way Capable	
Population Density	-0.006
	(0.015)
Population Growth Rate	0.010**
	(0.003)
Hispanic%	0.017+
	(0.010)
Age <18 %	0.061
	(0.048)
Age 65+ %	0.021
	(0.051)
Black %	0.021**
	(0.004)
Population per HH	-1.394
	(0.859)
Ln (Income)	0.923*
	(0.426)
Homeowner %	-0.001
	(0.011)
Observations	1339
DBS Share for VI Neutrality	26.0%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero. The two-way capable variable equal to one predicted carriage perfectly for a number of observations so the estimation dropped that variable and the observations from the regression, resulting in the lower number of observations.

TABLE 12K: CARRIAGE DECISION: WE

Variable	(1)
Vertically Integrated	**7.317****
	(0.593)
Vertically Integrated x DBS%	**-0.412****
	(0.043)
DBS%	-0.006
	(0.014)
Fiber Share of Total Plant	-0.016
	(0.080)
Analog System	-2.168**
	(0.427)
System Two-Way Capable	-0.328
	(0.273)
Population Density	0.018
	(0.025)
Population Growth Rate	0.003
	(0.004)
Hispanic%	0.001
	(0.008)
Age <18 %	-0.056
	(0.043)
Age 65+ %	0.000
	(0.018)
Black %	-0.009
	(0.009)
Population per HH	0.930+
	(0.507)
Ln (Income)	-0.215
	(0.321)
Homeowner %	-0.004
	(0.019)
Observations	1407
DBS Share for VI Neutrality	17.8%

** Significant at 1% level * Significant at 5% level + Significant at 10% level

Notes: These are the results from a probit regression of the carriage decision for the cable systems in the sample as described in the text. Standard errors in parentheses are clustered at the MSO level to allow for correlation in the carriage decisions across systems with the same owner. The share at the bottom of the table gives the level of DBS penetration that would reduce the sum of the two vertical integration coefficients to zero.

BIBLIOGRAPHY

Barth, David (2003), Essays on the economics of **television** program supply, Ph.D. Thesis, Northwestern University

Becker, Gary S (1971), *The Economics of Discrimination.* Second edition. Chicago: University of Chicago Press. (August).

Becker, Gary S. (1993), "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101(3): 385-409.

Caves, Richard (2005), *Switching Channels*, Harvard University Press (Cambridge, Mass.).

Chen, D., and D. Waterman (2006), "Vertical Ownership in Cable Television: A New Study of Program Network Carriage and Positioning," Working paper for presentation at the TPRC annual conference, Sept. 29-Oct. 1, 2006, Washington, D.C.

Chipty, Tasneem (2001). Vertical integration, market foreclosure, and consumer welfare in the
cable television industry, The American Economic Review, 91 (3), 428-453.

Crawford, Gregory (2000), "The Impact of the 1992 Cable Act on Household Demand and Welfare," The RAND Journal of Economics, Vol. 31, No. 3, 422-449. Autumn, 2000.

Federal Communications Commission (Various), *In the Matter of Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming*, 2nd through 12th annual reports, 1995 to 2006.

Ford, G. S. and J. D. Jackson (1997). "Horizontal Concentration and Vertical Integration in the Cable Television Industry." *Review of Industrial Organization* 12(4): 501-518.

Goolsbee, Austan and Amil Petrin (2004), "The consumer gains from direct broadcast satellites and the competition with cable TV," *Econometrica* 72 (2): 351-81

Kagan World Media (2006). *Economics of Basic Cable Networks 2007*. Carmel, CA: Kagan World Media.

Kagan World Media (2005). *Economics of Primetime Television 2005*. Carmel, CA: Kagan World Media, 2005.

Kang, Jun-Seok (2005), "Reciprocal Carriage of Vertically Integrated Cable Networks: An Empirical Study," working paper, Indiana University, August**.**

Motta, M. (2004). Competition Policy: Theory and Practice. Cambridge University Press, Cambridge, UK.

Owen, B. M. and S. S. Wildman (1992). *Video Economics*. Cambridge, Massachusetts: Harvard University Press.

Quigley Publishing Company (Various), *International Television and Video Almanacs, 2000 – 2005*. New York.

Rey, P. and Tirole, J. (2005). A primer on foreclosure. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization, volume 3*. Elsevier Science, B.V., Amsterdam, forthcoming.

Suzuki, Ayako (2006), "Vertical Integration in the U.S. Cable Industry" Discussion Paper 675, Osaka University.

Waterman, D. and Weiss (1996), “The Effects of Vertical Integration Between Cable Television Systems and Pay Cable Networks.” *Journal of Econometrics* 72: 357-395.

Waterman, D. and Weiss (1997), *Vertical Integration in Cable Television*, MIT Press (Cambridge, Mass.).

Wegener, Richard and Richard Caves (2003), "Cable Networks and Cable Operators: Ownership Stakes and Decisions to Carry," Working Paper, Harvard University.

Journal of Economic Perspectives—Volume 23, Number 3—Summer 2009—Pages 61–76

Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality

Robin S. Lee and Tim Wu

The "net neutrality" debate, as it has emerged over the last five years, is a social, political and economic debate over the public information network known as the Internet and the duties of its private carriers, which include telephone and cable companies and other Internet service providers (ISPs). In the early 2000s, questions surrounding the rights of Internet carriers to block certain network attachments and control access to emergent applications or content providers led to a call to protect "network neutrality" (Wu, 2003).[1] The debate raises familiar questions for students of travel or communications infrastructure: for hundreds of years, courts and governments have struggled over the duties that carriers like ferries, railroads, or telephone companies owe to the public by virtue of their necessity to economic and social welfare. While the net neutrality debate has many aspects, in this paper we focus on one crucial issue: the de facto rule

[1] In March 2005, the Federal Communications Commission fined an Internet service provider, Madison River, for blocking an Internet telephony service. In December 2006, the FCC imposed net neutrality rules on AT&T as a condition of its merger with Bell South, forcing it to commit itself "not to provide or to sell to Internet content, application, or service providers, including those affiliated with AT&T/BellSouth, any service that privileges, degrades or prioritizes any packet transmitted over AT&T/BellSouth's wireline broadband Internet access service based on its source, ownership or destination" (Reardon, 2006). In July 2008, the FCC criticized cable firm Comcast for deliberately interfering with transmissions occurring through BitTorrent peer-to-peer software, which allows large and/or popular files to be downloaded simultaneously by many users, and required that such interference cease.

■ *Robin S. Lee is Assistant Professor of Economics, Stern School of Business, New York University, New York City, New York. At the time this paper was written, he was Research Scientist, Yahoo! Research (Microeconomics), New York City, New York. Tim Wu is Professor of Law, Columbia University Law School, New York City, New York, and a Fellow at the New America Foundation, Washington, D.C. Their e-mail addresses are* ⟨*rslee@stern.nyu.edu*⟩ *and* ⟨*wu@pobox.com*⟩.

prohibiting consumers' ISPs from charging fees to content providers for access to their customer base.

The Internet can be understood as an information network that intermediates between different groups of agents comprising *users* and *content providers,* with the critical understanding that users can also act as content providers (for the purposes of discussion, we will use the term "content" loosely, referring to all types of media, applications, retailers, and services available online). Since the Internet's spread through academia in the 1980s and mass popularity in the 1990s, it has maintained a pricing structure that is unique among information networks: users and content providers typically pay ISPs *access* fees—fixed fees to get on the Internet at all—and *usage* fees—variable fees paid based on time or bandwidth usage; however, there have not generally been any additional charges for one user of the network to reach another user or content provider. For example, content providers such as Google and Wikipedia, while paying for their own Internet access and usage, do not directly pay the ISPs of users they reach.

This began to change in the mid-2000s as certain Internet service providers began to discuss their desire to charge certain users of the Internet—notably, large application or content providers—additional fees to reach their subscribers. For example, AT&T, an ISP, might demand that content providers such as Google and Wikipedia pay AT&T to access AT&T's customers; failure to comply would result in AT&T blocking traffic from those sites to its customers (and also preventing its own customers from reaching those sites). Following usage in the telephone system, we refer to these fees as *termination* fees: AT&T would charge content providers a fee to deliver their packets much like it charges other telephone networks a fee to "terminate" their calls.

Today, through historical practice, there exists a de facto ban on termination fees—also referred to as a "zero-price" rule (Hemphill, 2008)—which forbids an Internet service provider from charging an additional fee to a content provider who wishes to reach that ISP's customers. The question is whether this zero-pricing structure should be preserved, or whether carriers should be allowed to charge termination fees and engage in other practices that have the effect of requiring payment to reach users. This paper begins with a defense of the de facto zero-price rule currently in existence. We point out that the Internet, as an intermediary between users and content providers, exhibits pricing dynamics similar to other intermediaries in "two-sided markets." In particular, we posit that the Internet's absence of payments from content creators to users' ISPs facilitates the entry of content creators. In that respect, the rule provides an alternative implementation of the policy goals provided by the intellectual property system and achieves functions similar to copyright and patent law. The rule also helps avoid the problems of Internet fragmentation, in which content providers who do not reach agreements with ISPs cannot access all customers, and consumers on a single ISP are foreclosed from accessing their content.

We then consider some of the main arguments against restricting termination fees. Although we concede that limited instances may justify certain deviations from this norm, ultimately we argue that a zero-price rule has helped make the Internet distinctive from other networks in terms of its level of creativity and social usefulness.

Zero-Pricing and Net Neutrality

The general practices specified by the phrase "network neutrality" emerged not as a closely considered policy decision, but as a consequence of how the Internet was designed and how it spread. We claim, however, that an economic case for the pricing rules inherent in a neutral network can be found in the theory of two-sided markets. This theory suggests the de facto ban on termination fees may be interpreted as a policy that provides a subsidy to content creation and provision. For a number of reasons, this subsidy appears to have been one of the forces generating the enormous wave of innovation in services and information in the last few years as well as spawning new forms of creative activity.

Internet Pricing Contrasted with Telephone and Cable Television Networks

Currently, Internet users pay access and usage fees to their service provider and then can reach any other user who is similarly connected to the Internet. The overall network does not, by its own design, distinguish between content providers and users. Consequently, content providers—who may also be users—are also able to reach an audience consisting of every single Internet user. These norms and expectations, which have created a de facto ban on termination fees, stands in sharp contrast to what is standard practice on other important information networks, like the telephone and cable networks.

One reason for the differences between networks is rooted in history. The Internet was conceived by various visionaries, particularly the Department of Defense researchers J. C. R. Licklider and Robert W. Taylor, as a "network of networks" or an "intergalatic network" that would make it possible for users of any single computer network to reach users on any other network (Licklider and Taylor, 1968). In its original, noncommercial setting, fees to access the network were paid by universities, government, and research departments. There was no motive to charge termination fees to content providers. Government regulations designating the Internet a noncommercial network also discouraged any such fees. On the technological side, Internet protocols were designed to create a network that could be universally used by different parties with very little centralized knowledge as to who was reaching whom; as a result there was little power to track or bill for reaching certain entities. In the early years, the complexity of trying to incorporate billing capabilities might have doomed the project.

In contrast, telephone and cable television networks were designed from the outset as commercial networks, where payments were the prerequisite to connectivity at all. The Bell telephone system was, from its origins, extremely careful about who would be allowed to connect to Bell customers. In the two historically "competitive" periods in wire-line telephony—from the 1890s to the 1910s and from the 1980s through the early 2000s—calling a Bell customer generally meant paying a termination fee to the owner of the local switch. Since the 1970s, termination fees on the telephone system have been regulated based on fears that the Bell companies would use their "termination monopoly"—their exclusive access to

customers—to charge exorbitant rates and bankrupt any would-be competitor. Cable television networks are priced in a similar, but in some ways opposite, manner. From the beginning, cable companies were faced with demands for fees to access the content they needed from broadcasters, copyright owners, or other sources of content. As opposed to charging to reach their users, cable companies currently pay content providers for the right to carry content on their cable networks.

The Consequences of Pricing Decisions in Two-Sided Markets

The historical reasons why the Internet has developed without termination fees does not address the issue of whether the principle is economically justified. One potentially powerful justification emerges from the economic theory of two-sided markets.

Generally, the Internet, as a network (or network of networks), can be seen as a market or intermediary that facilitates the interaction of two main groups: users and content providers. Other well-studied examples of two-sided markets include payment systems, such as credit cards or online services; hardware-software markets like videogames or operating systems; retail marketplaces such as bazaars, shopping malls, or auction houses; matching markets, such as nightclubs or job sites; and advertising exchanges, such as online advertising platforms as well as commercial telephone books.

One prevalent aspect of two-sided markets is the presence of "network effects" (Farrell and Saloner, 1985; Katz and Shapiro, 1986; Liebowitz and Margolis, 1994), where the value of a service generally increases in the number of users. Most obviously, the network is more useful to users the more people there are to e-mail, videoconference, or instant message. There are also indirect network effects in that users benefit from more content providers, and these providers also benefit from having more users to reach. "Negative" network effects are also possible: for example, in networks with congestion, the value of using a network may decrease with additional users.

The recent literature on two-sided markets has studied the pricing decisions of a network provider or platform intermediary (Rochet and Tirole, 2003, 2006; Caillaud and Julien, 2003; Armstrong, 2006; Hagiu, 2006) in the presence of such network effects and has shown that in settings with transactions costs or limits on side-payments that are allowable between agents, the division of prices charged between sides of the market matters greatly. When there are such frictions, charging fees to one side as opposed to the other—even if the sum of prices across sides is the same—can affect who uses the network, overall transaction volume, and ultimately the efficiency of the market.

Theory predicts that the optimal pricing decision for a social planner as well as for a monopolist platform provider might involve "subsidizing" one side of the market over the other(s)—that is, possibly charging one side below the marginal cost of providing service. A canonical example of this phenomena are credit card companies. Credit cards companies do not charge both sides of the market—merchants and consumers—equally. Rather, they charge a transaction fee to

merchants and typically reward consumers for using their cards with frequent flier miles or other benefits. Because using a credit card is often not only "free" for a consumer, but in fact subsidized, the effect is to encourage consumers to use credit cards more than they would if fees were charged to both sides of the market. The division of pricing matters in this case partially because merchants are often prohibited from charging consumers different prices depending on whether the method of payment is by cash or credit; in other words, since merchants are unable to "unwind" these fees, pricing has a "nonneutral" effect.

If one side is subsidized at the optimum, theory predicts that the side receiving the subsidy will be the side that either has a higher elasticity of demand with respect to price, or the side that exhibits stronger cross-side network effects—that is, an additional agent on one side of the market increases utility on the other side by more than an agent on the other side would for the original side. Another way of interpreting this result is that since demand and demand elasticities for one side of the market change when the number of agents on the other side increases, the optimal prices charged by a platform may in fact be lower than those predicted if the platform ignored these network effects and assumed demand was fixed. All of these factors lead to more complex pricing dynamics than in traditional one-sided markets, which if ignored can lead to misguided regulatory or antitrust policies (Evans, 2003; Wright, 2004a).

In contrast, if there were no transaction costs or limitations on side-payments, the division of pricing between two sides of a market would not influence transaction volume; instead, any division would merely be a transfer between different sides of the market. In these standard settings, any distinction between how the two sides of the market are charged is unnecessary. Such is the case with a value-added tax, as whether firms pay a value-added tax during the stages of production or consumers pay a sales tax at the point of purchase does not affect transaction volume. Indeed, Rochet and Tirole (2006) use the nonneutrality of price structure—that is, for a given price level, the allocation of prices across sides influences transaction volume—as their definition of what makes a market two-sided.

These insights provided by the literature on two-sided markets naturally apply to the Internet.[2] First, as discussed, network effects are prevalent: the decision by users to subscribe to Internet access depends on which content providers are online; similarly, the decision of a content provider to invest in the creation of new content and applications depends on the number of users that can be reached. Consequently, the willingness to pay (and elasticity of demand) of an agent varies according to usage by others.

Secondly, there are practical limits on the side-payments that can be easily conducted between different sides of the Internet. Due to informational asymmetries, access or subscription charges to content significantly reduce usage as con-

[2] Others have explored this connection: for example, Economides and Tåg (2007) offer an application of the two-sided market framework to modeling the Internet, while Hurwitz (2006) and Hemphill (2008) also make the connection in their policy papers.

sumers find it difficult to determine the quality of on-line content before they have used it and thus often choose not to use it at all rather than to pay an up-front fee. Moreover, payments for Internet services in general have a variety of frictions: limited consumer access to credit cards or electronic payment systems; issues with security and fraud; problems of pricing micro-transactions for very limited access; and costs of monitoring whether purchased electronic content is being used only for agreed-upon purposes. Widespread adoption of usage-based pricing by content providers is in many cases thus either impractical or inefficient; it is not surprising, then, that many content providers have chosen to rely on advertising revenues and to provide content at no charge to the user. The presence of these myriad challenges preventing content providers from passing potential termination fees on to consumers thus implies that the division of pricing will be nonneutral. This tempers the "indirect extraction" argument (raised for example by Hemphill, 2008) that a ban on termination fees would be unraveled via side-payments.

Subsidizing Content Creation and Invention

The literature on two-sided markets combined with the economic realities of the Internet suggests that the ban on termination fees serves important economic and potentially social functions.[3] First, the rule provides a direct subsidy for the production of content and inventions. Second, it also cheapens market entry by making it easier for users to switch and become content providers themselves. Vinton Cerf (2006), a co-designer of the TCP/IP protocol, captures both these points: "Because the network is neutral, the creators of new Internet content and services need not seek permission from carriers or pay special fees to be seen online. As a result, we have seen an array of unpredictable new offerings . . . [E]n-trepreneurs need not worry about getting permission for their inventions will [sic] reach the end users . . . This is a direct contrast to closed networks like the cable video system, where network owners control what the consumer can see or do."

Providing subsidies for the production of creative works and innovation is a typical goal of many government programs, including the copyright and patent laws and institutions such as the National Institute of Health or National Endowment of the Arts. Such subsidies are most often justified in economic terms as a remedy for market failure in the production of creative works or new inventions (Landes and Posner, 1989). Since both creative works and inventions have characteristics of a public good, such intervention may be necessary to avoid underproduction.

The pricing structure of the Internet can be seen as an alternative means of subsidizing creativity and innovation (as discussed in Benkler, 1999). As economic analysis suggests, setting a preferable price or ruling out certain types of fees for content providers may encourage creation of content or new inventions that would not otherwise occur. As Lessig and McChesney (2006) note, "more than 60 percent of Web content is created by regular people, not corporations," and over 100 million

[3] Other effects of banning payments from content providers to Internet intermediaries (that is, using a "bill-and-keep" system) can be found in DeGraba (2000), Hemphill (2008), and Wright (2004b).

blogs have so far been documented (*Technorati,* 2008). The Internet, as a platform, has spawned thousands of new firms and millions of sites, from mass content projects such as Wikipedia to search indexers and content aggregators such as Google and Yahoo!.

In addition to the traditional justifications for subsidizing creativity and invention, there are special reasons that subsidizing in this context might be useful. The Internet content and applications market can be understood as a "hit-driven industry," where hit products like those from Google, Yahoo!, Amazon, eBay, and others create massive spillover value for users. But extreme uncertainty plagues the creation of distinctly new content, and for every hit product there have been numerous failures, with names like Go.com, Pets.com, and Webvan.com. Given that the returns to content production are skewed and the expected value of a new online venture is low, sufficiently low costs of entry may have been and may continue to be crucial.

For similar reasons, the lack of termination fees has also been critical to the "long-tail" model of revenue generation: many Internet businesses rely on an extremely diverse product range that caters to individual niche markets (Anderson, 2006). A diverse collection of websites that yield small value individually but high value when considered as a group might not exist if faced with higher operating costs. Even a more targeted or asymmetric scheme of only levying fees on "successful" or large content providers still has the effect of depressing content creation as it reduces the potential gains to innovation for small entrepreneurs with the dream of making it big.

Of course, for a given price level, subsidizing content comes at the expense of *not* subsidizing users, and subsidizing users could also lead to greater consumer adoption of broadband. It is an open question whether, in subsidizing content, the welfare gains from the invention of the next "killer app" or the addition of new content offset the price reductions consumers might otherwise enjoy or the benefit of expanding service to new users. It may prove useful for economic theory to further illuminate and explore this tradeoff. However, given the possibility of users acting as content providers, a more accurate description of a ban on payments from content providers to Internet intermediaries is this: it is a subsidy to the creative and entrepreneurial at the expense of the passive and consumptive.

The Danger of Fragmentation

Setting aside concerns over incentives for creativity, there is another important concern about allowing Internet service providers to charge fees to content producers: it would almost certainly result in service providers "competing" for content, as seen in other platform industries, by charging different fees and bargaining on exclusive arrangements with content providers. In turn, such bilateral agreements would inevitably lead to fragmentation—where certain content would only be available on certain service providers—and hence multiple "Internets." For example, cable television is a fragmented network in this sense: not only do users

of cable television face a menu of prices for different numbers of channels, but in addition, certain channels are available only in some geographic areas.

Despite arising in equilibrium, these arrangements need not be efficient in networked industries with externalities and an incomplete contracting space (for example, Segal, 1999; Farrell and Weiser, 2003; Whinston, 2006). Potential welfare losses could also be significant, as consumers would find themselves foreclosed from accessing content available only on rival service providers, and content providers would find themselves unable to reach certain segments of the population captive to service providers with whom no agreement had been reached.[4]

Such arrangements would also be anathema to the principle of universality subscribed to by the designers of the Internet. One visionary of the Internet, Tim Berners-Lee (2008), put it this way: "It is of the utmost importance that, if I connect to the Internet, and you connect to the Internet, that we can then run any Internet application we want, without discrimination as to who we are or what we are doing. We pay for connection to the Net as though it were a cloud which magically delivers our packets. We may pay for a higher or a lower quality of service. We may pay for a service which has the characteristics of being good for video, or quality audio. But we each pay to connect to the Net, but no one can pay for exclusive access to me." Economists might rephrase this principle of universality in a language of positive network externalities and avoiding fragmentation, but the fundamental social value remains the same.

The Transactions Cost Argument for a Zero-Price

While the two-sided markets analysis might seem to justify setting different prices between agents, a crucial question remains: Why a price of zero?

There is a crucial difference between a low price and a zero-price: the ban on termination fees eliminates an entire class of transaction costs. Similar to the way in which developers can write an application for Microsoft Windows without bargaining with Microsoft at all, content providers can reach all consumers without having to negotiate individually or to pay separate fees to every service provider. For example, assuming Internet access, content providers do not have to negotiate with any service provider to get their initial production started. Transaction costs, of course, can be overcome, but their presence or absence matters.

The most obvious beneficiary of the absence of termination fees and related

[4] An argument is sometimes made that allowing exclusive arrangements might help new intermediaries to enter network markets (Lee, 2009)—in this case, the argument would be if new Internet service providers could offer exclusive content, it would be easier for them to differentiate themselves from existing providers and gain market share. However, we argue that this argument should not be given great weight in the context of Internet service providers. First, this literature also has noted that a standardization of network platforms often improves welfare (Farrell and Saloner, 1985; David and Greenstein, 1990); in this case, the benefit of standardization arises from maintaining the Internet as the sole platform and ISPs as simply conduits to the greater network. This argument is strengthened given there is no sole ISP, and hence no monopoly rents extracted upon standardization. In addition, just because net neutrality may prevent ISPs from competing on content, they may still compete in other ways: for example, they can differentiate themselves on quality of service.

transaction costs are media forms like blogs; there are millions of content providers for blogs, which are of highly variable quality. If each content provider somehow had to negotiate to gain access to users, the transaction costs alone might endanger their existence in the first place (Heller, 2008). Furthermore, social media sites such as MySpace and Facebook may not have been able to launch before the viability of their sites had been established. For new entrepreneurs or content providers, it has been unnecessary to reach agreements with every carrier to maximize the number of users and contributors accessible—and hence the value of their content—from the start. If content providers had to run a gauntlet of fees before being widely available, many business models would not have been feasible and many content providers may not have entered.

We note the lack of fees for providing Internet content lies in sharp contrast to other networks such as cable television, which involve intensive negotiation over prices for a channel's inclusion in a cable package. In these contexts, such frictions pose significant difficulties for new content in acquiring distribution and obtaining an audience. In a sense, the transaction costs induced by this bargaining often favors the established, well-financed, or overtly commercial at the expense of niche, specialized, and unproven; as we have argued, content exhibiting the latter characteristics may very well have defined much of the Internet's value and worth.

What about Content Providers Charging Service Providers?

We have not yet discussed the possibility of content providers levying fees on Internet service providers. Should these types of reverse charges be allowed? After all, if subsidization is appropriate, why stop at a zero-price and why not have ISPs pay for content? Such a regime would be similar to what is done in the cable industry, where content providers like CNN and ESPN charge cable operators a per-subscriber fee for the right to carry their content. In the Internet context, some content providers have experimented with such fees; for example, ESPN charges service providers for the right to access its ESPN360.com broadband site, as opposed to charging individual users directly (*Wall Street Journal,* 2006).

For reasons similar to those outlined above, we do not think such arrangements are desirable. The use of discriminatory termination fees, even if negative for some content providers, may still lead to positive fees levied on others. But even if asymmetric regulation would be possible—banning one direction of fees—it would still be the case that allowing for any nonzero-pricing would introduce a new class of fees to the Internet and substantially increase transaction costs, favoring some types of content providers over others. Furthermore, such fees also exacerbate the problem of fragmentation and consumer foreclosure: if some Internet service providers did not wish to pay for certain content, it would be difficult to force them to do so.

At the same time, direct regulation for this particular direction of fee payments may not be necessary, for several reasons. First, any content provider that makes a decision to charge fees will necessarily internalize the effect of having fewer consumers that can access its site; as will be discussed later, internalization is not the

case with Internet service providers levying termination fees since a collective action problem is introduced. Second, unless consumers were willing either to switch service providers or to terminate Internet access altogether in response to a particular content provider's availability, a content provider would not be able to extract rents from an ISP; it is difficult to imagine very many (if any) content providers that fit this criteria. Under these conditions, it is unlikely that a significant number of content providers (if any at all) would find it desirable to charge service providers for access. Even ESPN has shown signs of weakening its policy: it now allows anyone with a .edu or .mil domain to access its broadband site, where previously only 20 million subscribers of the ISPs that had signed agreements with ESPN had access (*New York Times,* 2008).

Other Questions and Concerns

Mandating net neutrality via government regulation or even maintaining the current de facto prohibition on termination fees has been controversial. Here, we address a number of questions and concerns that have been raised.

Won't Internet Service Providers Set Appropriate Fees on Their Own?

If setting prices to subsidize content providers produces socially desirable outcomes, wouldn't service providers simply do so? Wouldn't an Internet service provider internalize the externalities across groups and subsidize the "right" side on its own?

One possible answer is "yes" and that this is already happening. Since the early 2000s, despite some early stated interest in charging termination fees to content providers, no Internet service provider has actually tried to do so. Although there have been a few attempts by service providers to limit access to certain types of content, there are relatively few examples. It is unclear, however, whether that behavior is motivated by conscious behavior and internalization of externalities, or by the existing threat of regulation which functions as a form of enforcement.

At the same time, it seems implausible that Internet service providers have appropriate incentives to price according to the social optimum. First, two-sided market theory models predict even a monopolist provider does not subsidize the "right" sides as much as a social planner would, as profit-maximizing prices are higher than those imposed in the social optimum (Armstrong, 2006). Second, and perhaps more interestingly, the fact that a customer will often have a different ISP than a content provider means there will be strong incentives to charge fees even if zero-prices were socially optimal. Furthermore, because the value of content is shared not only by consumers subscribed to a particular ISP but across all consumers, ISPs do not completely internalize the impact of charging termination fees to content providers.

To see this point more clearly, consider the following: assume there are three Internet service providers for users—A, B, and C—and consider A's decision to

charge fees to all content providers wishing to reach its own customers. Service provider A would gain revenue and account for the possibility that such fees would lead to potentially less content (via depressed investment and innovation) for its own users. However, A does not completely internalize how its action would negatively influence content production for the entire Internet, as it does not account for the effect of a reduction in content on users of B and C. Similarly, B and C's incentives to increase termination fees would also be misaligned with the social optimum, since they also fail to completely internalize the effects of their own pricing decisions on users of other ISPs.

In effect, service providers face a prisoner's dilemma: it might be individually optimal for one provider to defect and charge positive fees to content providers, although if all content providers charged such fees, the outcome would be worse than had all providers refrained from doing so. In this sense, the existing *de facto* practice of zero-pricing for content providers on the Internet can be understood as a solution to this collective action problem. Given the temptation to defect, regulation in support of net neutrality—or the threat of such regulation—can play a useful role in maintaining a cooperative solution.[5]

Could Charging Content Providers Help Pay for Network Development or Reduce Web Congestion?

Allowing for termination fees could generate more revenue for service providers; hence, some argue that relaxing the zero-price rule will provide carriers with greater incentives to invest in existing infrastructure. Hemphill (2008) writes of an implicit tradeoff in which "not only content innovation but also infrastructure innovation must be taken into account, and that subsidizing content development necessarily must be at the expense of network development." If so, perhaps the Internet has now reached a stage of maturity in which, even if innovation in content provision is still desirable, other objectives have become more important?

Edward Whitacre, former CEO of AT&T, made a similar claim (Business Week, 2005): "Now what they [content providers] would like to do is use my pipes free, but I ain't going to let them do that because we have spent this capital and we have to have a return on it. So there's going to have to be some mechanism for these people who use these pipes to pay for the portion they're using. Why should they be allowed to use my pipes? The Internet can't be free in that sense, because we and the cable companies have made an investment and for a Google or Yahoo! or Vonage or anybody to expect to use these pipes [for] free is nuts!"

There are two flawed assumptions in this line of argument. First, it is unclear

[5] Although allowing Internet service providers to form a cooperative may ameliorate the coordination problem, there still is the misalignment of monopolist incentives from the social optimum; furthermore, and perhaps more importantly, the existence of cooperatives may be unpalatable for collusive possibilities that may be encouraged (for example, the credit card industry, which utilizes a cooperative structure, has been the subject of antitrust scrutiny and litigation, including *United States v. Visa U.S.A., Inc. et al.* (344 F.3d 229 [2d Cir. 2003], *cert. denied,* 543 U.S. 811 [2004]).

that the ability to charge termination fees would, in fact, increase incentives to upgrade existing infrastructure. Termination fees may provide a way to increase profits of Internet service providers regardless of whether they upgrade their pipes; the impact on the *marginal* incentive to invest is indeterminate.[6] Instead of investing in faster or more reliable service, firms could also pay out a (greater) dividend, undertake other projects, or even invest in increasing its returns on existing content by making it scarce and exclusive. As Wu argued (Wu and Yoo, 2007): "If you can generate revenue by charging content providers to reach customers, as opposed to charging for bandwidth, something happens. The incentives become mixed, as the provider gains an incentive to maintain a level of scarcity, and thereby maximize gatekeeper revenue. So I don't agree . . . that [termination] fees will necessarily spark more last-mile innovation."[7]

The second problem, which speaks directly to Whitacre's claim, is that the absence of termination fees does not imply Internet service providers are not properly compensated for the use of their "pipes." Indeed, ISPs, including AT&T, are actually paid twice. First, any network that transmits a packet through AT&T's network provides payment to AT&T determined by negotiated peering or transit agreements (and that network was also compensated when it received the original packet from another network, content provider, or end-user). Second, AT&T already charges consumers for access, and nothing in the current net neutrality regime prevents charging higher prices to consumers who utilize more bandwidth or demand faster service. In a sense, claims that content providers should pay for use of a consumer's ISP when consumers already pay is comparable to the Postal Service demanding a recipient pay to receive a package for which a sender has already paid postage.

Similar reasoning can be used to counter arguments which claim that termination fees might serve to reduce "web congestion." Content providers, the argument goes, design their applications without taking into account the marginal cost of higher bandwidth usage, and hence "overuse" bandwidth in their designs; that is, these content providers design applications that "spend" bandwidth with wild abandon. However, as discussed, content providers are already forced to take into account the costs of bandwidth usage: those which utilize an Internet service provider must pay the access and usage fees to make the content available in the first place, and those which are connected directly to other network providers pay fees for egress traffic based on existing peering and transit arrangements. The less

[6] In other words, in a hypothetical market already served by at least one Internet service provider where termination fees are allowed, upgrading pipes would not necessarily lead to an increase in profits. For example, in markets with at least two ISPs, the gains to investment may be competed away. However, termination fees may encourage network expansion to consumers not served by any ISP. In these instances, there may be a tradeoff between expanding Internet access to new markets on the one hand and incentivizing content creation and avoiding Internet fragmentation through a zero-price rule on the other. At the same time, by reducing content creation and availability, termination fees could also lower consumer willingness-to-pay.

[7] The "last-mile" refers to the technologies and processes (for example, coaxial, wireless, fiber) that connect an ISP to its customers.

bandwidth a content provider uses, the less it pays. Furthermore, Internet service providers can, and in certain circumstances do, charge end-users for the amount of bandwidth used regardless of what content is consumed; thus, content providers that utilize massive amounts of bandwidth will also face and internalize lower demand by consumers.

Would Favoring Some Content Improve Internet Service?

In this section, we consider the possibility that allowing networks to favor some content over others could improve service overall. If hypothetically a network could recognize and prioritize packets more sensitive to delay, like video packets, over packets that are insensitive, like email, the network would in theory function better. Similarly, if packets could be transmitted over shorter distances, less congestion would occur. This reasoning has been the basis for many proposals for "quality of service" enhancements to the Internet since at least the 1990s, and has motivated the rise of content distribution networks and caching services, such as Akamai, which "mirror" content across servers located around the world and thus allow providers to pay for improved delivery of content. Indeed, network management and quality of service inherently requires some form of packet discrimination or content co-location, and are practices with which we do not necessarily take issue (Wu, 2003); in our view, they may be palatable as long as payment is not demanded from content providers by Internet service providers as a requirement for service.[8]

Many global schemes for prioritizing some packets of information over others have so far failed because of a collective action problem inherent in their design. The Internet is comprised of hundreds of Internet service providers and millions of content providers worldwide. So far, agreement on standards to prioritize traffic on the shared network has been impossible to reach, as has any agreement to honor any standards for prioritization. There is an obvious incentive to label every packet as a "high priority" packet on the assumption that everyone else will do the same.

However, it is an open question whether using prices could somehow overcome this problem. One proposal that has been raised to address network congestion in the last mile would be to create a *tiered structure* for consumer ISP traffic: allow *all* content to travel *freely,* but at the same time allow individual Internet service providers to create a "preferred" service for traffic, or a "fast-lane," for a fee that does not depend on the identity of the content provider. In our view, this approach has the advantage over termination fees by allowing anyone access to faster service—incentivizing content providers to only label traffic "priority" if the cost was deemed worthwhile—while not foreclosing those who still opt for the "free" Internet. As a result, we do not feel as

[8] We believe caching agreements with Internet service providers and services provided by distribution networks do not impinge on content provision because these services are available to all content providers and content providers that do not use these services are still accessible by consumers. Although these services do improve the performance of certain content providers vis-à-vis others, the services are only worthwhile for content providers with significant traffic and bandwidth demands; for new entrants with low bandwidth requirements, such services provide little benefit and are a nonissue.

though a zero-pricing rule should prohibit this particular implementation, as here content providers are not *forced* to pay a termination fee to access users.

Yet although this particular solution may have desirable benefits, we raise a warning: unless sufficient bandwidth and quality of service can be guaranteed for the "free" Internet, there is a risk that such tiering will serve to sidestep the *de facto* prohibition on termination fees. For example, a priced-priority system could simply become itself a *de facto* fee charged for all content providers if the "free" Internet was of sufficiently poor quality and consumers shifted their usage behavior accordingly. In other words, even if ISPs were prevented from charging discriminatory fees to content providers (thereby reducing bargaining frictions and the potential for fragmentation), tiering still could result in transfers from content providers to Internet service providers. As argued previously, this might dampen the introduction of new content and services and eliminate the subsidy for content innovation currently provided by net neutrality.

Should the Net Neutrality Argument Be Applied to Other Networks?

We have argued the Internet's history may have created a built-in subsidy for competitive entry and creativity. There is, however, an open question of whether the principle should be extended to other networks. While a full treatment of this question is well beyond the scope of this paper, we do wish to be careful about generalizing our arguments to other industries: in particular, several institutional details may differentiate the Internet from attempts to implement a zero-pricing rule in other networks. For example, one important consequence of the Internet's universal design is that the bandwidth used by any one content provider is dynamic and proportional to its popularity: only content that is visited or popular consumes common resources, whereas sites that are never accessed utilize zero network bandwidth. Consequently, there is effectively no opportunity cost of subsidizing new content and lowering the barrier to entry, since other content is not precluded from existing or reaching users. In contrast, in media networks such as radio or cable television, each station uses a fixed amount of bandwidth or spectrum regardless of its popularity; similarly, new products sold in stores consume physical space and inventory even if no one purchases them. Thus, even if subsidizing content may be desirable, the scarcity of airtime, spectrum, or shelf space may very well render zero-pricing unappealing and undesirable in other industries.

Concluding Remarks

At its broadest, the net neutrality debate in the United States and around the world is a reincarnation of an age-old debate about the duties of firms that supply infrastructure services essential to the economy, or—in the old common law phrase—firms "affected with the public interest." In the nineteenth century, trains and canals were the focus of this debate; in the twentieth century, it was the telephone and the electric systems; and in the twenty-first century, the Internet has seized center stage.

This paper has highlighted a potential benefit of the zero-pricing aspect of net neutrality, which prevents Internet service providers from levying termination fees on content providers. The theory of two-sided markets provides an underlying rationalization for how this practice can subsidize the creation of new content and spur innovation while avoiding fragmentation of the Internet. Several open questions remain, including how close the optimal subsidy for content creation is to a zero-price rule, and to what extent welfare gains from increased content production due to a zero-price may be offset by potentially higher access or usage fees charged to consumers.

At the same time, a more fundamental question that underlies this paper is what, if anything, sets the Internet apart from other networks, past and present? This question suggests a much broader agenda for research: namely, understanding in a more parsimonious manner how different pricing rules and other features of information networks affect their influence. We have mentioned two differences—a *de facto* ban on termination fees, and a rough proportionality between content popularity and bandwidth usage—that set the Internet apart from the other networks of our time, including telephone, cable TV, and broadcasting networks. Yet while this much may be clear, we do not have anything close to a full vocabulary for understanding the different choices implicit in the designs of different networks. And we have an even weaker understanding of what the larger effects of such choices will be. Although in this paper we have isolated one interesting effect—namely, than a ban on termination fees can be used to encourage market entry by creators and innovators—this point is far from a full understanding of networks and their larger effects on society and the world.

■ *We thank the editors and Scott Hemphill for their helpful comments.*

References

Anderson, Chris. 2006. *The Long Tail.* New York: Hyperion.

Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *RAND Journal of Economics,* 37(3): 668–91.

Benkler, Yochai. 1999. "Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain." *New York University Law Review,* 74(2): 354–446.

Berners-Lee, Tim. 2008. "Neutrality of the Net." http://dig.csail.mit.edu/breadcrumbs/node/132 (accessed on November 11, 2008).

Business Week. 2005. "Online Extra: At SBC, It's All About 'Scale and Scope.'" November 7. http://www.businessweek.com/magazine/content/05_45/b3958092.htm.

Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken & Egg: Competition among Intermediation Service Providers." *RAND Journal of Economics,* 34(2): 309–28.

Cerf, Vinton G. 2006. Prepared Statement, U.S. Senate Committee on Commerce, Science, and Transportation Hearing on "Network Neutrality." http:// commerce.senate.gov/pdf/cerf-020706.pdf.

David, Paul, and Shane Greenstein. 1990.

"The Economics of Compatibility Standards: An Introduction to Recent Research." *Economics of Innovation and New Technology,* 1(1–2): 3–41.

DeGraba, Patrick. 2000. "Bill and Keep at the Central Office as the Efficient Interconnection Regime." OPP Working Paper No. 33, Office of Plans and Policy, Federal Communications Commission, Washington DC.

Economides, Nicholas, and Joacim Tåg. 2007. "Net Neutrality on the Internet: A Two-sided Market Analysis." NET Institute Working Paper No. 07-45.

Evans, David S. 2003. "Some Empirical Aspects of Multi-sided Platform Industries." *Review of Network Economics,* 2(3):191–209.

Farrell, Joseph, and Garth Saloner. 1985. "Standardization, Compatibility, and Innovation." *The RAND Journal of Economics,* 16(1): 70–83.

Farrell, Joseph, and Philip J. Weiser. 2003. "Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age." *Harvard Journal of Law & Technology,* 17(1): 85–134.

Hagiu, Andrei. 2006. "Pricing and Commitment by Two-Sided Platforms." *RAND Journal of Economics,* 37(3): 720–37.

Heller, Michael A. 2008. *The Gridlock Economy.* New York: Basic Books.

Hemphill, C. Scott. 2008. "Network Neutrality and the False Promise of Zero-Price Regulation." *Yale Journal on Regulation,* 25(2): 135–180.

Hurwitz, Justin. 2006. "Neighbor Billing and Network Neutrality." *Virginia Journal of Law and Technology,* 11(9): 1–33.

Katz, Michael, and Carl Shapiro. 1986. "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy,* 94(4): 822–41.

Landes, William M., and Richard A. Posner. 1989. "An Economic Analysis of Copyright Law." *Journal of Legal Studies,* 18(2): 325–63.

Lee, Robin S. 2009. "Vertical Integration and Exclusivity in Platform and Two-Sided Markets." http://pages.stern.nyu.edu/~rslee/papers/VIExclusivity.pdf.

Lessig, Lawrence, and Robert W. McChesney. 2006. "No Tolls on the Internet." *Washington Post,* June 8, A23.

Licklider, Joseph C. R., and Robert W. Taylor. 1968. "The Computer as a Communications Device." *Science and Technology: For the Technical Men in Management,* April, no. 76, pp. 21–31. (Reprinted in 1990 in *In Memoriam: J. C. R. Licklider: 1915–1990,* Report 61, pp. 21–41. Palo Alto, CA: Systems Research Center, Digital Equipment Corporation.)

Liebowitz, S. J., and Stephen E. Margolis. 1994. "Network Externalities: An Uncommon Tragedy." *Journal of Economic Perspectives,* 8(2): 133–50.

The New York Times. 2008. "ESPN to Offer Sports Events on the Web Free to Some." February 4. http://www.nytimes.com/2008/02/04/business/media/04espn.html (accessed November 10, 2008).

Posner, Richard. 1998. *Economic Analysis of Law.* 5th ed., pp. 36–50. New York: Aspen.

Reardon, Marguerite. 2006. "FCC approves AT&T–BellSouth merger." CNET News, December 29. http://news.cnet.com/FCC-approves-ATT-BellSouth-merger/2100-1036_3-6146369.html.

Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association,* 1(4): 990–1029.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report." *RAND Journal of Economics,* 37(3): 645–67.

Segal, Ilya. 1999. "Contracting with Externalities." *Quarterly Journal of Economics,* 144(2): 337–88.

Technorati. 2008. "State of the Blogosphere." http://technorati.com/blogging/state-of-the-blogosphere (accessed on November 12, 2008).

The Wall Street Journal. 2006. "ESPN Charges Net Providers For Right to Offer Broadband Web Site." August 1. http://online.wsj.com/article/SB115439535367922979.html.

Whinston, Michael D. 2006. *Lectures on Antitrust Economics.* Cambridge: MIT Press.

Wright, Julian. 2004a. "One-sided Logic in Two-sided Markets." *Review of Network Economics,* 3(1): 44–64.

Wright, Julian. 2004b. "Pricing Access to Internet Service Providers." *Information Economics and Policy,* 16(3): 459–73.

Wu, Tim. 2003. "Network Neutrality, Broadband Discrimination." *Journal of Telecommunications and High Technology Law,* vol. 2, pp. 141–79.

Wu, Tim, and Christopher Yoo. 2007. "Keeping the Internet Neutral?: Tim Wu and Christopher Yoo Debate." *Federal Communications Law Journal,* 59(3): 575–92.

GLIMMERS AND SIGNS OF INNOVATIVE HEALTH IN THE COMMERCIAL INTERNET

SHANE GREENSTEIN*

INTRODUCTION 25
I. THE VALUE CHAIN FOR THE INTERNET 28
  A. *Broadband* 30
  B. *Platforms* 34
  C. *Contractual Incompleteness* 39
II. COMMERCIAL BEHAVIOR AND THE INTERNET 41
  A. *Economic Experiments* 42
  B. *Vigorous Standards Competition* 46
  C. *Inventive Entrepreneurialism* 55
  D. *Absence of Unilateral Bargaining* 63
III. HEALTHY INNOVATIVE COMPETITION FROM DOMINANT FIRMS 74

INTRODUCTION

What are the signs of healthy behavior in an innovative industry?

This seemingly simple question isn't so simple to answer in a quickly evolving industry such as the Internet. Commercial behavior resides inside a complex value chain, which is a set of interrelated activities that produces a final product for end users. No single firm controls the value chain, and the quality, price, and user experience arise from the complex interactions between those participants. Moreover, over time many parts of this value chain have undergone innovative improvements, and no reasonable observer expects those improvements to cease tomorrow.

There is no agreement about which criteria observers and policy makers should use to assess the performance of the commercial Internet.

* Elinor and Wendell Hobbs Professor, Kellogg School of Management, Northwestern University. I wish to thank audiences for questions and comments. I thank also Jonathan Baker, Tim Bresnahan, Ed Felten, Greg Rosston, Jim Speta, Scott Stern, Phil Weiser, and the editors for their suggestions. This essay is part of a larger project about the commercial Internet. I received funding from the Ewing Marion Kaufman Foundation, The Searle Center, and the Dean's office at the Kellogg School of Management. None are responsible for the errors or opinions expressed herein.

Ever since the commercial Internet first emerged, there have been arguments about how to best organize its value chain to achieve maximum value for the most users. Disagreements have not diminished with time. If anything, this debate has grown shrill as the number of commercial interests and business commentators have grown.

This essay makes a novel contribution to this topic. It identifies patterns of healthy commercial behavior indicative of an innovative industry, and illustrates how to observe signs of such behavior in information technology markets, such as the Internet. Stated broadly, the essay identifies healthy behavior that correlates with desirable market-wide outcomes, such as improvement in products, lower prices, new capabilities, or other innovations that lead to productivity improvements among business users.

This essay highlights four signs of the healthy innovative behavior:

- economic experiments
- vigorous standards competition
- entrepreneurial invention
- the absence of unilateral bargaining

Unlike most prior writing in this area, the essay is not motivated by any normative proposal for governing the Internet value chain, such as net neutrality or reasonable network management, or any specific proposal for legal or regulatory reform. To be sure, the reasoning in the essay will have some implications for some aspects of these proposals, but that is not its primary purpose.

This essay is written in the spirit of aspirations to develop a "third way" for addressing infrastructure policy issues in the Internet. A third way seeks to nurture innovation by avoiding lengthy and protracted fights in agency hearings and courtrooms—avoiding events that sustain uncertainty about the value of commercial investments, sometimes for years at a time. Such sustained uncertainty damages the interests of *every* industry participant in a fast moving market, both users and suppliers.

A third way would rely on dispute resolution mechanisms that operate much faster, such as negotiations and guidelines.[1] This third way would employ arbitration and administrative resolution to disputes, avoiding the slow tools of regulatory command and control. It would avoid, in particular, an extremely damaging event that is all too common in regulatory processes for telecommunications in the United States—the slow and sometimes discursive processes associated with regulatory ping-pong between federal agency decisions and court-ordered remedies and

1. *See* Phil Weiser, *Institutional Design, FCC Reform, & the Hidden Side of the Administrative State* (Univ. Colo. Law Legal Studies, Research Paper No. 09-01 2009).

appeals.

If such "a third way" emerges, it also will aspire to reduce uncertainty. How does it do that? Such a process aspires to be predictable, saving all parties the trouble of adjudication in any but the rarest circumstances. To achieve predictability, the administrators will publish transparent guidelines for all relevant participants.

That is where this essay makes a contribution. Guidelines necessarily require a conceptual framework and benchmark for recognizing innovative behavior. The benchmark must help regulators quickly recognize when a market action does or does not contribute to a healthy innovative outcome. This essay proposes a framework for building such a benchmark.

At present, the closest any policy statements get to such a benchmark in the United States are the four Internet principles issued by the Federal Communication Commission (FCC). In their most recent restatement by the outgoing chairman of the FCC, Kevin Martin, the four principals are:

> Consumers are entitled to access the lawful Internet content of their choice; Consumers are entitled to run applications and services of their choice, subject to the needs of law enforcement; Consumers are entitled to connect their choice of legal devices that do not harm the network; Consumers are entitled to competition among network providers, application and service providers, and content providers.[2]

These principles are intended to signal the direction of future policy without committing the agency to specific actions. As noted by many observers, the principles aspire to contain both generality and flexibility in the face of inevitable change in the industry.[3] Yet, that also explains what I regard as their primary drawback. They are rather open-ended and curt in comparison to the efforts of other federal agencies to offer policy guidelines.

In my view, that curtness undermines their ability to reduce uncertainty by signalling what a federal regulator regards as healthy and unhealthy innovative behavior. They also fail to reduce regulatory delay

---

2. Appropriate Framework for Broadband Access to the Internet over Wireline Facilities, *Policy Statement*, 20 FCC Rcd. 14,986 (2005), *available at* http://www.publicknowledge.org/pdf/FCC-05-151A1.pdf/ [hereinafter *Policy Statement*]; *cf. Net Neutrality FCC / FTC,* CYBERTELECOM, *available at* http://www.cybertelecom.org/ci/neutralfcc.htm (past statements by ex-Chairman Michael Powell).

3. Footnote 15 of the *Policy Statement, supra* note 2, states, "Accordingly, we are not adopting rules in this policy statement. The principles we adopt are subject to reasonable network management," begging the question, "What is the definition of reasonable network management?" *See, e.g.,* Isen.blog, http://www.isen.com/blog/2005/08/how-martins-fcc-is-different-from.html (Aug. 7, 2005, 17:07 EST); *Net Neutrality FCC / FTC, supra* note 2.

because they are inviting regulatory hearings followed by court interpretation, triggering the usual damaging ping-pong.

One comparison with a standard benchmark of competition policy in the United States can illustrate why I perceive the four principles as open-ended and curt. In 1968 the Department of Justice and Federal Trade Commission first issued a set of very detailed merger guidelines, revising them most recently in 1997, and issuing extensive commentary on them again in 2006.[4] While the guidelines do not commit the DOJ or FTC to specific actions in specific mergers, these have become a benchmark for firms and agencies, helping firms anticipate likely DOJ and FTC responses to proposed mergers. This makes the process more predictable, which helps all parties plan, and it reduces negotiation costs for all participants.[5]

By comparison, do the four principles provide a similar level of guidance? It is not even close. The four principles cover only a narrow range of actions. There have been only a few examples to illustrate how the FCC intends to employ these principles, involving Madison River and Comcast. There are many plausible circumstances not covered, and in which the principles do not help market participants forecast whether their own decisions will generate close regulatory scrutiny or not. Such open-endedness seems particularly damaging for innovative behavior because, said simply, there are few indications about when commissioners and staff will view innovative behavior as healthy or not.

These concerns motivate focusing on identifying the behavioral signs of innovative health. I perceive there would be a gain for policy from clarifying benchmarks that any observer, even querulous lawyers on opposite sides of a policy issue, could use to assess the state of health of an innovative market, such as the Internet.

In Section I, I review the broad motivation behind the essay's core question. Section II provides an analysis of the four signs of innovative health. Section III discusses some implications of this approach for events involving dominant firms in which the FCC did or did not apply the four principles, such as disputes involving Comcast, and another between Sprint and Cogent.

I. THE VALUE CHAIN FOR THE INTERNET

The complexity and evolution of the Internet's value chain

4. DEP'T OF JUSTICE, ANTITRUST DIVISION, MERGER ENFORCEMENT GUIDELINES, http://www.usdoj.gov/atr/hmerger.htm.

5. DEP'T OF JUSTICE AND FTC, HORIZONTAL MERGER GUIDELINES, http://www.usdoj.gov/atr/public/guidelines/hmg.htm. *See also*, DEP'T OF JUSTICE AND FTC, COMMENTARY ON THE HORIZONTAL MERGER GUIDELINES, http://www.ftc.gov/os/2006/03/CommentaryontheHorizontalMergerGuidelinesMarch2006.pdf.

motivates the core question behind this essay. It is worthwhile to understand this motivation in some depth. The structure of the Internet value chain has evolved in a direction that will give rise to numerous policy issues into the foreseeable future.

The value chain for Internet services appears to be perpetually in transition. To paraphrase the economist, Bruce Owen, the players have only reached the fifth inning of a nine-inning ball game and there is no rain delay in sight.[6] That evolution raises a challenge for any regulatory framework: it makes it quite difficult to assess the general factors encouraging behavior that leads to innovative outcomes.

Indeed, ever since the Internet commercialized many of its participants have maintained a strong sense about their exceptional nature, as if innovation within the existing value chain for the Internet defied established archetypes of innovation. For example, the Internet did not arise as a consequence of one single breakthrough invention from one single genius, à la Edison and the light bulb.

That view raises a rather deep economic question about whether innovation within the Internet can be assessed with the same economic concepts used elsewhere in innovative markets, such as computing. This essay will largely argue that it can be.

The truth about the early development of the commercial Internet is less exciting than this attitude of exceptionalism would suggest. It involved a vastly dispersed set of actors. The Internet developed slowly and through a rather mundane process, accumulating capabilities over time from an enormous number of contributors. As such, it fits an archetype that scholars of innovation label as "Collective Invention."[7] For example, the creation, refinement, and improvement of e-mail prior to 1990 involved contributions from more than fifty different people over two decades, and that application was one new application among many.[8]

More specifically, the Internet initially accumulated capabilities over time in a government project hidden from mainstream view. Technical success generated interest and use, spread technology among researchers, and gained economic value by growing capabilities in a community that did not recognize its economic value for non-researchers.[9]

---

6. Bruce Owen, *Broadband Mysteries*, *in* BROADBAND: SHOULD WE REGULATE HIGH-SPEED INTERNET ACCESS? 9–38 (Robert W. Crandall & James H. Alleman eds., Am. Enter. Inst. Press 2003).

7. *See*, *e.g.*, Robert C. Allen, *Collective Invention*, 4 J. ECON. BEHAVIOR & ORG. 1, 1–24 (1983); Peter B. Meyer, *Episodes of Collective Invention* (U.S. Bureau of Labor Statistics, Working Paper No. 368, 2003), *available at* http://ssrn.com/abstract=466880.

8. Craig Partridge, *The Technical Development of Internet E-mail*, 30 IEEE ANNALS OF THE HISTORY OF COMPUTING 3-29 (2008).

9. *See, e.g.*, JANET ABBATE, INVENTING THE INTERNET (MIT Press 2000); Shane Greenstein, *Wild Ducks and Inconspicuous Accumulation: Innovation in the Government-Sponsored Internet* (Kellogg Sch. of Mgmt., Working Paper, March 2009), *available at*

Once commercialized, the Internet began to accumulate more capabilities and functions, as a range of firms began to use pieces of the Internet to enhance services provided to paying customers. Over time, "the Internet" became a label for not only the Internet, but also for all the applications that accumulated around the Internet, used pieces of the Internet, and commercialized new functions for the Internet, which cumulatively delivered an enormous array of services to a wide range of users.

Three factors in particular altered the discussion about the value chain in the last decade. First, the predominant access mode for the Internet changed. Second, several leading businesses organized several different platforms to alter the potential value chains for users and developers. Third, the predominant contractual framework for governing transactions was never completed.

Each one of these factors raises further questions about the presence of market power and its distortion on innovative outcomes. Each factor also raises questions about the ability of a savvy observer to assess the innovative health of the Internet.

I describe each of these factors in turn and explore why they motivate the core question of this essay.

A. Broadband

In the 1990s the model Internet Service Provider (ISP) was a dial-up charging $20 a month on average.[10] By the turn of the millennium this industry had generated over $10 billion in revenue,[11] which was quite impressive for an economic activity so young. At a broad level, however, it supported only applications that could tolerate some delay in the delivery of data. That restriction on the value of output rendered moot many arguments about how to best govern the value chain. Subsequent developments brought those arguments to the forefront.[12]

---

http://www.kellogg.northwestern.edu/faculty/greenstein/images/research.html.

10. Tom Downes & Shane Greenstein, *Universal Access and Local Commercial Internet Markets*, 31 RES. POLICY 1035–1052 (2002); Tom Downes & Shane Greenstein, *Understanding why Universal Service Obligations May be Unnecessary: The Private Development of Local Internet Access Markets*, 62 J. URBAN ECON. 2–26 (2007); Shane Greenstein, *Innovation and the Evolution of Market Structure for Internet Access in the United States*, *in* THE INTERNET AND AMERICAN BUSINESS 47 (William Aspray & Paul E. Ceruzzi eds., 2008) [hereinafter *Evolution of Market Structure for Internet Access*]; Shane Greenstein, *Building and Developing the Virtual World: The Commercial Internet Access Market*, 48 J. INDUS. ECON. 4 (2000).

11. Shane Greenstein & Ryan McDevitt, *The Broadband Bonus: Accounting for Broadband Internet's Impact on U.S. GDP.* (NBER, Working Paper No. 14758, 2009), *available at* http://www.nber.org/papers/w14758.

12. *See* JONATHAN E. NUECHTERLEIN & PHILIP J. WEISER, DIGITAL CROSSROADS: AMERICAN TELECOMMUNICATIONS POLICY IN THE INTERNET AGE (2005); *see also*,

The predominant mode of access changed in a short period. In September 2001, approximately 45 million U.S. households accessed the Internet through a dial-up connection, whereas only 10 million used a broadband connection.[13] By March 2006, a sharply contrasting picture emerged: approximately 47 million households (and growing) had broadband connections, whereas 34 million (and declining) used dial-up.[14] According to the latest survey of the Pew Internet and American Life Project, in April, 2009, less than 10% of U.S. households had dial-up Internet connections, and 63% of U.S. households had broadband.[15]

Consistent with the increasing adoption of broadband by households and its higher monthly prices on average, the total revenue in access markets grew. So, too, did the fraction of revenue going to broadband.[16]

Simple economic factors determined the growing trend to broadband internet serve. Dial-up became available first and diffused to more than half of U.S. households. Thereafter broadband emerged as a higher quality and more expensive alternative, albeit one available in only a few places and from a limited set of providers, if any. Over time, however, broadband became more reliable and more widely available, which enabled many households to upgrade their Internet service.

Today, most urban households face a duopoly of wire-line choice: (1) an offering from a local cable franchise, and (2) an offering from a local telephone company. In some locations, they also may face options for wireless providers, which potentially may convert the duopoly into a more competitive supply. In many suburban areas (less dense settings) households face that duopoly or only one wire-line provider. To the contrary, one wire-line provider services households in many rural settings or isolated small cities, where households lack alternatives to

---

Evolution of Market Structure for Internet Access, *supra* note 10, at 47–104 (Those arguments had antecedents in the open access movement, but became reformulated as broadband diffused. They were reformulated principally in the form of the "net neutrality" movement. As noted earlier, the FCC policy's ambiguity about the meaning of 'reasonable network management' left open many issues.); *see e.g.*, George Ou, *A Policy Maker's Guide to Network Management*, THE INFO. TECH. AND INNOVATION FOUND. (2008) (a review and analysis of various definitions and their implications), *available at* http://www.itif.org/files/Network_Management.pdf.

13. NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION (NTIA), A NATION ONLINE: ENTERING THE BROADBAND AGE (2004), *available at* http://www.ntia.doc.gov/reports.html.

14. John Horrigan, *Home Broadband Adoption*, PEW INTERNET & AM. LIFE PROJECT (2007) *available at* http://www.pewinternet.org/Reports/2007/Home-Broadband-Adoption-2007/ Data-Memo/Findings.aspx?r=1.

15. John Horrigan, *Home Broadband Adoption*, PEW INTERNET & AM. LIFE PROJECT (2009) *available at* http://www.pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx?r=1.

16. Greenstein & McDevitt, *supra* note 11.

dial-up internet service except through a satellite provider and/or other wireless ISP.

As a cause for both celebration and concern, broadband firms inhabit a position of monopoly or duopoly in a key part of the value chain. On the one hand, broadband's position reflects the ascendency of a superior product and service in replacing dial-up, an economic improvement over the near past. On the other hand, it raises concerns about the presence of market power and the incentives to make future improvements.

At a broad level, most texts in standard industrial economics stress the issues with this situation.[17] While society benefits from giving incentives to firms to create superior products and services, rewarding firms with monopoly power comes at a cost to society, presuming firms with high market share possess market power. Such firms may face weaker incentives to innovate than firms in any more competitive market structure.[18] Net neutrality advocates also have expressed a related concern that the retail market power will be used to shape the incentives of others in the value chain in adverse ways.[19]

Broadband's ascendency into the majority of households gave rise to another issue because it enabled a range of applications to blossom. Generally speaking, four types of rather different uses share the same capacity: (1) browsing and e-mail, which tend to employ low bandwidth and tolerate delay; (2) video downloading, which can employ high bandwidth and can tolerate some delay; (3) voice-over IP and video-talk, which tend to employ high bandwidth and whose quality declines with delay; and (4) peer-to-peer applications, which tend to use high bandwidth for sustained periods of time, and can tolerate delay, but, in some applications (e.g., Bit-Torrent) can impose delay on others.[20]

That range of uses and applications today also raises cheers and concerns. The Internet has evolved from a mere e-mail network for technically skilled users during its first decade into an e-mail or instant

---

17. *See* KIP VISCUSI, JOSEPH HARRINGTON & JOHN VERNON, ECONOMICS OF REGULATION AND ANTITRUST 4 (2005); *see also* DENNIS CARTON & JEFFREY PERLOFF, MODERN INDUSTRIAL ORGANIZATION (2004).

18. *See* Richard Gilbert, *Looking for Mr. Schumpeter: Where are we in the Competition-Innovation Debate*, 6 INNOVATION POL'Y & ECON. 159–215 (2006); *see also*, Jonathan Baker, *Beyond Schumpeter versus Arrow: How Antitrust Fosters Innovation*, 74 ANTITRUST L.J. 575, 575–602 (2007). While this broad point is generally accepted, there is considerable debate surrounding many aspects related to its general applicability and about what policy can/should do to foster competitive incentives aimed at raising innovation incentives.

19. *See e.g.*, LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBER SPACE (1999); LAWRENCE LESSIG, CODE, VERSION 2.0 (2006); Tim Wu & Christopher Yoo, *Keeping the Internet Neutral?: Tim Wu and Christopher Yoo Debate*, 59 FED. COMM. L.J. 575, 581 (2007); Ou, *supra* note 12.

20. *See* Ou, *supra* note 12.

messaging communications network for some, a gaming network for others, a source of news for others, and a distribution channel for video and musical entertainment for others. For others still, it is the principal media for engaging with geographically dispersed communities of friends.

While that diversity of applications wrings additional productivity out of the same capital supporting the network, it comes with a potential drawback: the use of one application affects the productivity of another. In part this is due to capacity constraints at bottleneck positions in the network; there are few backbone pathways to support browsing in isolated positions. Contributing to these constraints are geographically localized negative externalities (e.g., many modern peer-to-peer applications employ all available bandwidth, diminishing the quality of other applications in the same cable network that cannot tolerate delay).

The market for Internet access could become more complex over time. Options vary in speed, quality, and price. There have been data services from the major cellular carriers (e.g., Verizon, AT&T, and others) for several years, particularly for e-mail delivery to laptops. The most popular mechanism in the recent past was a simple device for delivery of e-mail (e.g., a BlackBerry). More complex devices have gained popularity (e.g., iPhones and smart phones), and these have download speeds that begin to approach the low end of wire-line broadband speeds.

Technological optimists forecast even faster download speeds from next generation wireless carriers (e.g., WiMax or LTE). There is still considerable uncertainty about how many of these services the market will support, about what price and sales levels will prevail, and, accordingly, what scale of deployment these prices and sales levels will support.

The pace and level of change suggest that the provision of Internet access has not stopped evolving, nor will they soon. In the best of all worlds the prior gains are permanent and the most worrisome concerns are temporary.

Why does this evolution pose a quandary for a regulator? It is not worrisome if the multiplicity of access choices erodes market power of any individual actor. It is worrisome if some actors retain market power, and use it to discourage innovations that do not serve their interests. The questions are central to any innovation policy for the Internet. What relevance will market power have to innovation policy in the Internet? Does limiting the distortions of market power provide justification for government intervention? If so, what type of action, and what are its limitations?

More to the point, in a setting where market power might or might not be present, and might or might not be employed for purposes that

run afoul of policy sensibility, private actors gain little insight into the thinking of public policy makers who publically commit to only four sentences. Surely private firms benefit from knowing how to anticipate the norms and standards employed by regulators to recognize the signs of healthy and unhealthy behavior in a situation that is changing so much.

B. Platforms

Well-designed standards and platforms hold one of the keys to the successful accumulation of functionality over time. Consider this brief overview about how platforms have changed over time, and how those changes altered the ability of a savvy observer to assess the innovative health of the Internet.

By way of background, typical use of Internet-related services requires successful execution of a set of technically interrelated activities coming from many independent firms. The failure or reduction in performance of any of these activities can lead to inferior outcomes for many users. Focusing solely on such technical action, however, misses a key dimension of how firms address the challenges. Even the simplest of activities in this value chain, such as sending e-mail, involves many participants, and efficient delivery of services depends on advanced agreement about how their business activities will interrelate. To reduce the uncertainty about how such services interoperate, commercial firms take one of two approaches: either they negotiate arrangements in advance with all relevant participants, or, if that fails, they do it all themselves.

In the parlance of business language, firms either negotiate standards with others so the task performs smoothly, or they offer a platform that accomplishes the task. Platforms are a standard bundle of components and designs around which vendors build services. Platform strategies played an important role in computing before the commercialization of the Internet.[21] Many firms naturally organized their strategic approach for commercial opportunities on the Internet with similar approaches.

---

21. *See e.g.*, Timothy Bresnahan & Shane Greenstein, *Technical Progress and Co-Invention in Computing and In the Use of Computers*, BROOKINGS PAPERS ON ECON. ACTIVITY: MICROECONOMICS 1–78 (1997); Timothy Bresnahan & Shane Greenstein, *Technological Competition and the Structure of the Computer Industry*, J. INDUS. ECON. 1–40 (1999); *see also*, ANNABELLE GAWER & M.A. CUSUMANO, PLATFORM LEADERSHIP: HOW INTEL, MICROSOFT AND CICSO DRIVE INDUSTRY INNOVATION (2002); Annabelle Gawer & R. Henderson, *Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel*, 16 J.ECON. & MGMT. STRATEGY 1, 1–34 (2007); DAVID EVANS, ANDREI HAGIU, & RICHARD SCHMALENSEE, INVISIBLE ENGINES: HOW SOFTWARE PLATFORMS DRIVE INNOVATION AND TRANSFORM INDUSTRIES (2007); ANNABELLE GAWER, PLATFORMS, MARKETS AND INNOVATION (Edward Elgar ed., forthcoming Dec. 2009).

Ever since the emergence of the Internet, several leading businesses organized different platforms to alter the potential value chains for users and developers. There are proprietary platforms, open source platforms, and business platforms, all of which interoperate to provide services and some of which compete at the same time.

As with the rise of broadband, the rise of platforms on the Internet is a source of both celebration and consternation. While platforms perform functions that firms and/or users value, their presence usually suggests that some firms/users are better off with them than without. At the same time, large or dominant platform leaders (usually) possess market power, thereby raising questions about whether those firms use their discretion in ways that lead to more innovation.

The list of important platforms today is long. To illustrate this observation, I highlight two proprietary platform providers, Microsoft and Intel, and one non-proprietary platform, open source communities.

Perhaps the best known of the commercial platform providers is Microsoft, which develops and sells an operating system branded as Windows. It organizes the computing platform around the personal computer, as well as many Intel-based servers. To produce and deliver this product Microsoft engages with a multiplicity of actors, users (e.g., businesses and households), original equipment manufacturers (OEMs, *e.g.*, Dell, HP, and others), and application developers (e.g., software vendors). The operating system allows all of them to interact with one another for more efficient delivery of services.

Microsoft's platform strategy for the Internet over the last decade has been shaped by its lucrative position selling Windows for PCs and for server functions. This has led the firm to offer a mix of supporting functionality for the Internet. For example, in the early 1990s it offered TCP/IP compatibility in Windows as means to enhance the features of its networking software. In the mid 1990s it offered a browser, partly as a gateway towards developing a broader array of Web services, and partly for defensive purposes, because it matched browsers offered by others, notably Netscape at that time. Microsoft eventually won a rather confrontational war with Netscape for market share, and continues to hold a leading position in browser usage.

Microsoft has not, however, had as much success in other aspects of its commercial Internet ventures. Despite considerable resource commitments, its MSN division has never yielded enviable success. Its attempt to build an advertising-supported set of applications—including a recent attempt to buy Yahoo![22]—also has not yielded big advances.

22. Peter Henserson & Braden Reddall, *TIMELINE: Microsoft Attempt to Buy Yahoo!*, REUTERS, May 4, 2007, *available at* http://www.reuters.com/article/topNews/idUSGOR47298420080504.

Despite a leading position in enterprise computing, it has not yet found a successful transition to cloud computing applications, such as experienced by salesforce.com, for example. Only its investments in Xbox now generate revenues in excess of operating costs,[23] as well as a significant amount of Internet gaming traffic, but it will be considerable time before it generates enough profit to recoup the billions of dollars in losses spent developing the platform in the first place, if ever.

Intel is another prominent platform provider whose strategy arose from its lucrative position in PC markets. Intel's historical platform strategy had some similarities to Microsoft's. It too stands at the middle of a large ecosystem, interacting with a range of firms, providing leadership that drives towards the standard hardware design and specification used in most desk top computers, lap tops, and net books. Its behavior also differs from Microsoft's for a simple reason; Intel interacts much more with hardware than software firms. While Intel offered the most widely used microprocessor for personal computers, it feared losing leadership in new and growing markets for integrated circuits, especially processors. It developed a faster microprocessor and invested heavily in creating demand for platforms that used it. The latter motivated Intel to invest in a wide range of activities, some of them far afield from microprocessor manufacturing.

For example, Intel designed an input-output bus for PCs, even though, until that point, it had never been in that business. Intel also designed PC motherboards and virtually gave away the design to others, as a way to foster improvements that aided its microprocessors.[24] Intel helped design and sponsor USB and corresponding USB standards, including funding the testing for conformance.[25] It also branched into sponsoring a Wi-Fi standard for laptops under the Centrino label, helping to design further upgrades to the underlying technical standard, which was designed by IEEE committee 802.11, and helping to fund conformance-testing organizations as well.[26] More recently, it has invested heavily in designing and supporting another 802 wireless standard, known as Wi-Max. In addition, Intel has worked hard to develop a position as a microprocessor provider for standard designs of

23. Erick Schonfeld, *Microsoft Lost Nearly $500 Million on the Web Last Quarter*, TECHCRUNCH, January 22, 2009, *available at* http://www.techcrunch.com/2009/01/22/microsoft-lost-nearly-500-million-on-the-web-last-quarter/.

24. GAWER & CUSUMANO, *supra* note 21.

25. Jeffrey K. MacKie-Mason & Janet S. Netz, *Manipulating Interface Standards as an Anticompetitive Strategy*, *in* STANDARDS AND PUBLIC POLICY 231 (Shane Greenstein & Victor Stango eds., 2007); Intel.com, Intel Helped Make It Easier to Connect Devices to PCs, http://www.intel.com/standards/case/case_usb.htm.

26. Shane Greenstein, *Economic Experiments and Neutrality in Internet Access*, 8 INNOVATION POL'Y & ECON. 59 (2006).

smart phone devices.

Another organizational form for developing an interrelated platform of services involves the use of open source institutions, that is, employing some variation on the General Public License (GPL) for code or a Creative Commons license for copyrighted material. While intellectual property often receives the most attention, it is not the key factor for most commercial firms. Open source differs sharply from platforms organized by Microsoft or Intel in the responsibility and activities of management by raising transparency for developers about the features of the code and its evolution. In some organizations, open source has an additional function: it substitutes participatory/collective decision making for unilateral decision making at a single firm.

In some respects, the open source movement is not new at all as an institution for platform development and support. Transparency and wide participation have played a role in the development of key protocols and standards for the Internet, known as TCP/IP, which are employed by most Internet users. These are maintained by the Internet Engineering Task Force (IETF), who maintains a set of fully documented and accessible processes for making documented code available. It invites wide participation in the design of new protocols and standards. It was not called "open source" when it started, but the processes strongly resemble the modern transparent processes with wide participation (more below).[27]

Another important platform emerges from the Web standards maintained by the World Wide Web Consortium (W3C). It, too, has a transparent process, but it employs a different model for decision making and participation. The W3C requires firms to pay for their membership, and Tim Berners-Lee and his staff retain some authority to make decisions unilaterally after consultations with the membership.

A better known example of these open source platforms is Linux. The changes to this open source project hint at how commercialization and open source have both recently changed. Linux began as a volunteer project by Linus Torvalds, but today has firm support for a consortium operated by Torvalds. This consortium supports a range of businesses operated by many firms, including IBM, Red Hat, and others.

More broadly, open source platforms now appear in many commercial ventures on the Internet. A range of other business models have emerged for platform development around open source, including businesses organized by MySQL (for databases) and Mozilla and Webkit (for the Firebox, Safari, and Chrome browsers). The same could be said

27. Scott Bradner, *The Internet Engineering Task Force*, *in* OPEN SOURCES: VOICES FROM THE OPEN SOURCE REVOLUTION 47 (Chris DiBona, Sam Ockman & Mark Stone eds., 1999).

for a range of Web 2.0 efforts, such as Facebook, YouTube, and Flickr, which do not use the GPL but instead employ licenses designed by the Creative Commons.

Taking a "snapshot" of the structure today, we can see the infrastructure supports a rather complex value chain involving the interoperability of many different commercial platforms. This is an enormous evolution: the present arrangement looks nothing like the Internet of the early 1990s, when it first commercialized.

Today many observers believe Google has the most effective platform on the Internet.[28] Its search engine is the most popular in English,[29] as well as in many other languages. That supports a very lucrative ad-placement business. Many other firms also expend considerable resources optimizing their web pages to appear high on Google's search results, so, like any important platform, Google's actions have become central to the economic prosperity of others.[30] Some observers believe this will only continue, as its popularity will allow Google to develop a range of products supporting its search business.

Other prominent platforms include those provided by Cisco (networking equipment), Research In Motion (BlackBerry), Apple (iPhone, iPod), Yahoo! (search, news, mail), Oracle (enterprise databases), E-Bay (auctions), as well as many others. These examples are only a few among many prominent commercial platforms shaping development of the Internet. It is necessarily a short list and may not have the relevant platforms for policy in the near future. Each one of these platforms deserves a longer description, and the reader should be clear that the absence of that here is due to space constraints, not their lack of importance.

Platforms add an additional layer of decision making to the provision of services. That comes with a benefit, to be sure. It lowers coordination costs, and it can smooth transactions between participants with long term relationships. But platforms also come with some strings attached. Once they exist the firms with commercial interests in their continuance will take action to make sure they do not easily go away. Growth tends to agglomerate the successful platforms, but they also

28. RANDALL STROSS, PLANET GOOGLE: ONE COMPANY'S AUDACIOUS PLAN TO ORGANIZE EVERYTHING WE KNOW (2008); Sarah Lacy, *The New Bulls-Eye on Google*, TECHCRUNCH, Feb. 18, 2009, http://www.techcrunch.com/2009/02/18/the-new-bulls-eye-on-google/.

29. Erick Schonfeld, *March ComScore Search Numbers Offer a Sign of Hope for Google*, TECHCRUNCH, Apr. 14, 2009, http://www.techcrunch.com/2009/04/14/march-comscore-search-numbers-offer-a-sign-of-hope-for-google/ (reporting that the 2009 estimates from ComScore place Google at 63.7% of all searches done in the U.S., which is over 9 billion searches).

30. STROSS, *supra* note 28.

stand in the way of complementary entry which holds the potential to oppose the commercial interests of the present platform leaders.

Aphoristically, the Internet has been called a "network of networks" since it first began to diffuse to the general public. Yet, distilling the Internet to that aphorism is misleading; it does not reflect how commercial behavior shaped the evolution of technology in the last decade and a half. Leading firms and their business partners view the commercial Internet through the same lens they view activities in the rest of computing. For them, the commercial Internet is a "network of platforms."

In sum, firms do not make investment decisions aimed in general directions. Rather, they make investments aimed at advancing their own platform strategies. Firms do not merely defend themselves against entry by a new competitor. They develop sophisticated approaches to find out which other platforms may pose a threat to their existing profitable businesses. Platforms are a central strategic determinant of the direction a firm takes.

Which of these behaviors will not raise alarms and which will? This ongoing evolution of platforms poses a thorny question to regulators: how can they recognize signs of healthy and unhealthy platform behavior in an innovative industry when platforms play such an important role? Once again, the questions are central to any innovation policy for the Internet. Once again, surely private firms benefit from knowing how to anticipate the norms and standards employed to recognize the signs of healthy and unhealthy behavior in markets where most dominant firms employ platform strategies.

C. Contractual Incompleteness

Contractual incompleteness has become a central feature of the Internet value chain. Incompleteness refers to the absence of contracts governing regular transactions or, if such contracts exist, to contracts that lack fully specified terms for all contingencies. The maturation of the Internet value chain has not yet diminished this incompleteness much, and there are no signs of change.

To many economists, such an observation is only a philosophical statement. In this essay it is also an observation with pragmatic relevance to innovation policy. It provides both justification for government intervention, as well as a limitation to it. This justification is quite distinct from the two already discussed, both of which stress the role of market power.

Contractual incompleteness arises for many reasons. The Internet involves an extraordinarily large number of parties, which renders multilateral negotiations impractical. There are so many players, in part,

because the value chain supports an extraordinarily multi-purpose network, as earlier noted. Said simply, today many parties take action and their actions influence one another. There is just no practical way to get all these participants—or even their representatives—in the same room at the same time to work out a deal by horse-trading one set of economic concerns for another.

For example, even if one set of Skype users might be willing to pay another set of Bit-Torrent users to change their behavior, there is no practical way to get them all in the same room at the same time to negotiate and sign that deal. That incompleteness might further motivate another market participant, for example, an ISP, to take further action, though I will defer that discussion until later.

Incompleteness also arises where all parties may recognize the potential for technical change to generate new applications that alter circumstances, requiring renegotiation of prior contracts whose terms are no longer relevant. Yet, many pairs of parties in this setting may fail to come to agreement for numerous reasons. Even if the recognition exists, the parties may fail to negotiate a solution due to a lack of the type of trust and mutual assumptions that usually support renegotiating commercial transactions in the face of such contractual incompleteness.

Most interesting, contractual incompleteness inhibits negotiations, as it may be impossible to consummate a deal. The relevant party may not even exist yet (if they will be entrepreneurial start-ups) and, thus, lack representation in even a basic form, such as a trade-group or related commercial organization.[31]

Legal ambiguities for innovative activities also can play a role. While contractual obligations govern some of the routine activities, it may be more difficult to erect similar obligations for new activities. For example, contracts govern the handoff of data from one backbone carrier to another, or from one Web application to an edge-caching site, such as Akamai's, or to a content-delivery network, such as Amazon's. In contrast, a looser contractual foundation governs another set of interrelated activities. For instance, when an advertising-sponsored Web application sends data to a user, the ISP delivers it without alteration, because participants await legal rulings. YouTube was founded in an era when there were multiple plausible definitions for a precise, legal, and safe harbor for copyrighted material for user-supplied video. These definitions still remain ambiguous, though court cases continue to refine them into a tighter domain.

---

31. *See* Bruce M. Owen & Gregory L. Rosston, *Local Broadband Access: Primum Non Nocere or Primum Processi? A Property Rights Approach*, *in* NET NEUTRALITY OR NET NEUTERING, SHOULD BROADBAND SERVICES BE REGULATED? 163 (Bruce M. Owen & Gregory L. Rosston eds., 2006).

Between contractual incompleteness and legal ambiguities, an efficient bargaining solution—a so-called Coasian bargaining solution—fails to arise. Indeed, such failure is endemic to the setting. The very thing that makes the Internet economically successful—the accumulation of innovation that supports a wide set of applications for many participants, including entrepreneurs—gives rise to conditions that make it harder for Coasian solutions to arise.

The lack of a Coasian bargaining solution can provide an economic justification for a potential role for government regulators in specific circumstances: to settle disputes when many participants have a stake in the solution but private parties fail to account for these externalities; or, related, to define "default" terms of commercial relationships that many partake in, when the default remains undefined; or, related, to mandate terms of standards employed by participants in the value chain when they otherwise cannot or do not come to such standards on their own.

Note, however, this argument implies a limitation on that role. It covers only those activities that firms could not already settle themselves through contracting, those without externalities, or those which necessarily involve unanticipated circumstances. To be sure, however, that is not necessarily a substantial limitation if it involves participants who are not even in a market yet, such as entrepreneurs.

Once again, another limitation on decision making also is implied. The arguments for intervention presume the existence of a well-developed set of insights about how to recognize a problem in the Internet value chain.[32] As it turns out, some arguments against intervening also presume a problem can be recognized.[33] That too motivates looking more closely at how such recognition takes place.

In short, the evolution of the Internet value chain gives rise to many of the conditions that stand in the way of a Coasian agreement. That also implies that the evolution necessarily stands in the way of making an assessment about whether the situation merits intervention or not. Once again, the questions are central to any innovation policy for the Internet. Once again, private firms benefit from knowing how to anticipate the norms and standards employed to recognize failure of a Coasian solution, i.e., the signs of healthy and unhealthy behavior in an innovative industry such as this.

II. COMMERCIAL BEHAVIOR AND THE INTERNET

Four signs of innovative behavior are examined here: economic

32. *See* LESSIG, *supra* note 19; Wu & Yoo, *supra* note 19; Ou, *supra* note 12.

33. *See* Wu & Yoo, *supra* note 19; Ou, *supra* note 12; *see* Bruce M. Owen, *Antecedents to Net Neutrality*, REGULATION, Fall 2007, at 14.

experiments, standards competition, entrepreneurial invention, and the absence of one-sided bargaining. All four play a role in the accumulation of new functionality in services on the Internet, and all four could continue to play a role in the future if the structure enables them to.

A. Economic Experiments

An economic experiment is a market-oriented action designed to help a firm learn or resolve uncertainty about an unknown economic factor. Usually such lessons cannot be learned in a laboratory or controlled environment, either because they involve learning about the nuances of market demand or learning about sets of procedures for providing new services at a lower cost.[34]

Economic experiments vary in purpose. Some experiments focus on learning about the profitability of incremental changes in business processes, whereas others seek to learn about the restructuring of organizations and the profitability that may result from the simultaneous alteration of many processes or about the profitability of restructuring the relationship among many organizations within an industry.

Internet markets have been full of economic experiments in the last fifteen years. That was especially so in the latter part of the 1990s, when firms took a wide variety of bets to learn about unknown aspects of customer demand and the costs for meeting them using Web technologies. These experiments covered all parts of the value chain for delivering services—Internet access, client-server platforms, contracting among business partners, and so on. Carriers conducted them and so did content providers.[35]

To be sure, not all experiments work out. Indeed, if the learning occurs as part of a risky business venture, many of them should not. And, accordingly, history is littered with illustrations. In Internet application markets some of these firms survived (e.g., Google, Amazon, E-Bay), and some of these did not (e.g., WebVan, Pets.com). So it goes.

Against that backdrop it is pleasing to see that recent behavior looks similar. Some firms involved in the Web 2.0 movement (Facebook, Friendster, Digg, and others) and this decade's frontier businesses (Salesforce.com and YouTube, for example) will make it, while others

---

34. *Economic experiments* pertain to any market experience that alters knowledge about the market value of a good or service. Nathan Rosenberg, *Economic Experiments*, *in* EXPLORING THE BLACK BOX: TECHNOLOGY, ECONOMICS, AND HISTORY 87 (Nathan Rosenberg ed., 1994); Scott Stern, *Economic Experiments: The Role of Entrepreneurship in Economic Prosperity*, *in* UNDERSTANDING ENTREPRENEURSHIP: A RESEARCH AND POLICY REPORT 16 (2005). Firms engage in economic experiments to reduce uncertainties about market value.

35. *See* Greenstein, *supra* note 26 (examining the role of economic experiments in the evolution of Internet access).

won't. For example, most VCs today are convinced that there will be little further entry of new businesses into Web 2.0 and the future will involve exit of many of the entrants of the last few years. So it goes again. As long as many firms are trying to learn, then the industry looks healthy in this respect.

While some experiments do not succeed, many do. A successful business continues to operate and brings new goods and services to users. Indeed, while many can recall the failures of the dot-com boom, it should be pointed out that success rates for new firms during this era were comparatively high, leaving a long string of very valuable activities in place.[36] In addition, many of the lessons learned endure, handed out as free advice from one manager to the next, benefiting a new generation of businesses.

Note how this assessment differs from the common approach and orientation of Wall Street analysts.[37] By definition, economic experiments are risky learning exercises, designed to teach a firm (or set of firms or set of VCs) about something unknown but relevant to the value chain for delivering services. It is not unusual to observe a little messiness, and there is no particular reason to anticipate the learning to yield immediate profitability. In fact, the learning is usually expensive and the benefits come later, so immediate profitability is rare.

Wall Street's short run values typically do not reward experimentation, regardless of the potential long term gains from such lessons. Consider FiOS, Verizon's program to bring fiber to residences. Many technologists think Verizon is late to the party and many stock analysts remain skeptical about the potential for large financial returns from FiOS.[38] Many analysts also remain skeptical about whether FiOS will generate steady returns, much like a utility's revenue.[39]

Using this emphasis on economic experiments, however, one might

---

36. Brent Goldfarb & David A. Kirsch, *Small Ideas, Big Ideas, Bad Ideas, Good Ideas: "Get Big Fast" and Dot Com Venture Creation*, *in* THE INTERNET AND AMERICAN BUSINESS 259 (William Aspray & Paul E. Ceruzzi eds., 2008); Brent D. Goldfarb, David Kirsch & Michael D. Pfarrer, *Searching for Ghosts: Business Survival, Unmeasured Entrepreneurial Activity and Private Equity Investment in the Dot-Com Era* (Robert H. Smith School, Research Paper No. RHS 06-027, 2005), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=825687.

37. I am using "Wall Street" as a generic term for a common style of analysis that emphasizes only firm profits of a single firm in the short term, neglecting the collective progress of a community of market participants whose activities ultimately shape user experience.

38. Saul Hansell, *Verizon's FiOS: A Smart Bet or a Big Mistake?*, N.Y. TIMES, Aug. 18, 2008, at C1; Saul Hansell, *A Bear Speaks: Why Verizon's Pricey FiOS Bet Won't Pay Off*, N.Y. TIMES, Aug. 19, 2008; Craig Moffett, *Network Upgrades Are for Ninnies*, BROADBAND DSL REPORTS, Aug. 19, 2008, http://www.dslreports.com/shownews/97086; DSLReports.com, Op-Ed, *Investor: Fios Is Doomed*, BROADBAND DSL REPORTS, Dec. 15, 2006, http://www.dslreports.com/shownews/80296.

39. Hansell, *supra* note 38; Moffett, *supra* note 38.

say, "Good for Verizon." That is precisely the type of disagreement that should arise if a firm's management undertakes a risky economic experiment. Profitable today or not, pushing the envelope will teach Verizon's management, as well as competitors like AT&T, quite a lot, and maybe that will help it lower costs or develop better targeted marketing next year.

Economic experiments also depart from another Wall Street bias for assessing the progress of innovation in terms of one firm's profitability. For purposes of public policy, it is often a poor idea to focus on one firm's success or failure to assess the benefits of learning. It is often more sensible to take a view of economic experiments taking place in an entire market, focusing instead on whether a community of suppliers or users are benefiting in the long run.

For example, this was the most insightful way to understand the earliest commercial experience with Wi-Fi, or IEEE standard 802.11b, between 1999 and 2001. During this early period of diffusion, many firms and users learned about the value of the short-range data transmission. The technology was defined, but the business case was not. Examining any single firm's experience would have yielded a rather pessimistic assessment, which was a distinctly uninformative way to understand what was happening.

While homes and enterprises explored the gains from installing wireless routers, so, too, did a completely unanticipated set of actors: coffee shops, cafés, and other hot spots. At the time all actors were trying to learn about which implementations created value and which did not. Lessons were shared in many public forums. It was a collective economic experiment, and it was generally beneficial for many users, though it was hard to identify any particular firm for whom it was super.[40]

What is an example of unhealthy experimentation? Here's one: Microsoft's lack of new releases for Internet Explorer 6.0 at the start of this decade. Microsoft deployed little new for five years, spending most of its energy and time responding to every new call for security patches, as well as dealing with the publicity nightmare that came with having its product panned so widely by so many technical experts.[41] After spending so much money to win the dominant position on browsers from

40. *See Evolution of Market Structure for Internet Access, supra* note 10.

41. *See* Martin LaMonica, *Gates Admits IE Failings, Looks to an AJAX Future*, ZDNET, Mar. 21, 2006, http://news.zdnet.co.uk/internet/0,1000000097,39258532,00.htm; *Internet Explorer 6 with Windows XP SP2*, N.Y. TIMES, http://nytimes.com.com/browsers/internet-explorer-6-with/4505-3514_7-31214886.html; CNET Reviews, http://reviews.cnet.com/4520-3514_7-5020542-1.html (last visited Apr. 2009); WinPlanet Windows Software Reviews and Downloads, http://cws.internet.com/file/11714.htm (last visited Apr. 2009); Software Informer, http://internet-explorer.software.informer.com/6.0/ (last visited Apr. 2009).

Netscape, this outcome was unnecessary, as there was no lack of capability or resources. It came from a company famous for its disciplined approach to a "three-version strategy," which deliberately takes a loss on an early version of a product in order to learn from economic experiments. It was as if all such capabilities were forgotten.

The Internet Explorer example illustrates the potential costs and benefits from having only one party conduct experiments. As to the costs, if that party has reasons not to conduct experiments, then it leaves users and third-party programmers with no alternatives. If new ideas have no channel into that one party, then all its partners and users lose from the foregone opportunity. Indeed, according to one observer, in Microsoft's case, this outcome partly resulted from the absence of market discipline after the collapse of the coalition built around Netscape, which permitted an especially bitter internal struggle for strategic direction for Internet services to permeate Microsoft's decisions, to the detriment of other firm goals, such as product development.[42] As to the benefits, only the appearance of Firefox a few years ago seemed to rouse Microsoft's managers and programmers from their internal squabbles to focus on making progress users could measure. I am pleased these days to see more activity, reversing past trends. There appear to be more new experiments coming out of the WebKit community (e.g., Safari, Chrome), as well as from Opera and others. Accordingly, some of the good ideas from these new initiatives have found their ways into the design of later releases of Internet Explorer.

The orientation of communications policy towards protecting or nurturing economic experiments has varied over time for two principle reasons. First, and broadly stated, advocates for policies to nurture experimentation generally bear a high burden of proof in public discourse, as they must argue about a future that has not yet occurred. They must argue that change in a policy will give rise to experimental behavior that has not yet arisen (or will diminish), while their opponents argue that such experimental behavior has no connection to policy.

For example, the recent debates about the need for a "Carterfone" policy in wireless technology divide precisely on these lines. One side argues that a change in policy will bring about more experimentation and the other argues that present policy encourages experimentation that

42. DAVID BANK, BREAKING WINDOWS: HOW BILL GATES FUMBLED THE FUTURE OF MICROSOFT (2001) (recounting the internal debates and fights leading up to the browser wars and beyond them); *see also* Timothy Bresnahan, Shane Greenstein & Rebecca Henderson, *Schumpeterian Competition within Computing Markets and Diseconomies of Scope* (Kellogg Sch. of Mgmt., Working Paper, 2008), *available at* http://www.kellogg.northwestern.edu/faculty/greenstein/images/research.html (providing one interpretation of these events, as a by-product of the costs of organizational diseconomies of scope between the Internet and Windows business).

would diminish if the policy changed.[43] Both points are logical, but cannot be proven without trying one policy that precludes the other.

Second, while some policies nurture the blossoming of economic experiments, that role may be unapparent until after the experiments blossom.[44] Such themes run throughout review of FCC intervention in the Internet's growth, for example. Many nurturing policies, such as policies for third party access providers, became established for reasons connected to historical events unrelated to the Internet.[45] Some, such as the policies that resulted from the *Computer Inquiries,* were in place for reasons connected to their role encouraging new entry in information technology equipment markets, but nobody had the Internet specifically in mind, and had they done so, policy makers may have made different choices.[46] In either case, such unintended consequences from prior policies make it difficult to give forward-looking advice.

I want to acknowledge these difficulties, and then restate the reason it is essential to nurture economic experiments in spite of the challenges. Said succinctly, nobody wants to see some of the Internet's biggest firms turn into Microsoft's browser division, sitting on its laurels with a buggy piece of software, slowly making upgrades, lacking any competitor to push it outward, and fighting an internal corporate fight at its own leisure, to the exclusion of other concerns. Experiments are a sign of progress; lack of them is a sign of stagnation.

B. Vigorous Standards Competition

Bleeding-edge technologies often cannot deploy on a wide scale without some routines or processes, and/or coordination of activities across many firms. Thus, the ratification of new standards generally acts as a leading indicator of impending technological progress and serves as another sign of a healthy innovative industry. While new standards and upgrades to existing standards may not arrive at a regular rate, a slow pace for development or a slow arrival of new standards should set off alarms.[47]

---

43. *See* Tim Wu, *Wireless Carterfone*, 1 INT'L J. COMM. 389 (2007); Christopher Yoo, *Network Neutrality, Consumer, and Innovation*, 25 U. CHI. LEGAL F. 179 (2008).

44. *See* Jason Oxman, *The FCC and the Unregulation of the Internet* (Office of Plans & Policy, Working Paper No. 31, 1999), *available at* http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp31.pdf; Greenstein, *supra* note 26.

45. Oxman, *supra* note 44.

46. Robert Cannon, *Where Internet Service Providers and Telephone Companies Compete: A Guide to the Computer Inquiries, Enhanced Service Providers, and Information Service Providers*, *in* COMMUNICATIONS POLICY IN TRANSITION, THE INTERNET AND BEYOND 3 (Benjamin M. Compaine & Shane Greenstein eds., 2001).

47. Lest this sound far-fetched, these ideas have been operationalized by frontier researchers. For example, this is the spirit of the exercise undertaken by Simcoe, in his study of new standards developed at the IETF. His principle finding concerns the slowing pace of

To be sure, this benchmark is particularly challenging to put into practice, because some standards are more important than others. The protocols known as TCP/IP have played a central role for decades, for example, and any alteration to them receives considerable attention, deservedly more attention than other standards. The same is so for protocols which govern the Web, as well as those standards that govern upgrades to Ethernet.

Those examples are a bit misleading, however, as they give a false sense of certainty to the enterprise of designing standards. As it turns out, there are often multiple solutions to the same problem. That may be due to differences of opinion about what the true problem is, or about how to best solve it. It shows up as different proposals for "standards." As it happens, many proposals for standards often do not get deployed or put into widespread use. In other words, new standards frequently get deployed in environments where their ultimate success remains uncertain long after development.[48]

Consider the following illustration. The deployment of Wi-Fi was far from assured. The release of 802.11b in early 1999, which eventually become widely deployed, came less than two years after the first beta release of a standard for 802.11 in 1997.[49] The first release contained multiple problems that simple field experiments revealed, generating two later descendants, given the labels "a" and "b." For numerous reasons "b" got deployed first in 1999. Though a fixed version of "a" came soon after, its availability did not determine deployment. It never deployed as widely to equipment firms. Most had largely already started to deploy "b."[50]

Another, more current example, and one more representative of the complexity and uncertainty pervasive in a standards fight, can be found in the market for "unified communications." These are a series of standard designs for making the e-mail, voice-mail, and other communications applications work more seamlessly with each other. Both Microsoft and IBM have begun to address an enterprise's communications processes by

development. Timothy Simcoe, *Delay and De Jure Standardization: Exploring the Slowdown in Internet Standards Development*, *in* STANDARDS AND PUBLIC POLICY 260 (Shane Greenstein & Victor Stango eds., 2007); Tim S. Simcoe, *Standard Setting Committees* (U. Toronto Joseph L. Rotman Sch. Mgmt., Working Paper, 2008). *See also* Iain M. Cockburn & Megan MacGarvie, *Entry, Exit, and Patenting in the Software Industry* (Nat'l Bureau of Econ. Research, Working Paper No. W12563, 2006); Iain M. Cockburn & Megan MacGarvie, *Patents, Thickets, and the Financing of Early-Stage Firms: Evidence from the Software Industry* (Nat'l Bureau of Econ. Research, Working Paper No. W13644, 2007) (investigating his principle finding concerns the slowing pace of development).

48. See URS VON BURG, THE TRIUMPH OF ETHERNET: TECHNOLOGICAL COMMUNITIES AND THE BATTLE FOR THE LAN STANDARD (2001) (documenting a similar case involving the early Ethernet standards).

49. *A Brief History of WiFi*, THE ECONOMIST, Jun. 10, 2004, at 26; Greenstein, *supra* note 26.

50. Greenstein, *supra* note 26.

offering distinct solutions.

These solutions do not arise out of the ether. They involve the integration of scores of standards into a platform upon which users customize their unique needs. Both IBM and Microsoft have had some success in developing and selling their solutions, but both have a long way to go towards an ideal. As in the above example, we can see that often no single firm can resolve a problem for every circumstance. Moreover, users may differ in whether they favor one solution or another.

Here is another illustration concerning one of the most interesting recent developments—the emergence of platforms at the edges of wireless networks. Microsoft has invested in organizing developer networks for wireless phones within its CE environment, using its platform experience as a guide. Meanwhile, Apple exported to the iPhone its experience organizing multiple providers of applications on its iPod platform and its Mac platform. Google's effort with the Android represents one alternative method for organizing the platform, and Nokia's recent efforts to develop its own music services and mapping services another. Research In Motion, the maker of BlackBerry, has organized yet another approach.

Once again, this competition among distinct platforms, with standards embedded in the platform that may not be explicit, can be interpreted as competition between bundles of standards. It is far from apparent which design offers the right solution for most users. Even the most sagacious observer cannot forecast how this competition will evolve in the next three years. In light of that intractable uncertainty, the availability of many options benefits users.

This is not to say, however, that standards competition only multiplies options. Occasionally such competition comes with substantial and durable costs. For example, there may be multiple problems that require distinct solutions, but these are inconsistent with one another, i.e., choosing one precludes another. In the face of uncertainty about the value of various alternatives or their technical efficacy, premature commitment to one standard can impose significant costs on later users. For example, Internet insiders will recognize that this is the present issue hindering different QoS proposals. Some serve to aid one goal but deter another.[51]

In short, because standards are extraordinarily important and valuable in introducing innovation to the value chain, their development and rollout anticipates new services and inventive activity. There also are often multiple solutions to similar problems, so competition between

51. See Ou, *supra* note 12.

standards proxies for multiple solutions for users.

Similar to the observation about economic experiments, this argument is headed towards a seemingly counter-intuitive observation: while this activity may be confusing to all but an insider, this messiness is a sign of good health. This may seem an especially surprising conclusion to any participant in standards processes. Any reasonably thorough case study of the processes behind the design of a standard will emphasize the frustration, confusion, and utter plethora of loose ends, even with ample funding and a functional certification process. Most participants in standards committees come out of the experience with nothing good to say about it.

All this is true, but somewhat irrelevant. More to the point, standards competition beats the alternative.

Think of one of Winston Churchill's famous quotes: "It has been said that democracy is the worst form of government except all those other forms that have been tried from time to time."[52] Similarly, there is only one saving grace for existing standards processes: standards designed in the absence of competition are usually much worse. A monopolist's tendency towards orderly, infrequent, and simplified standards leads an industry down about as unhealthy an innovative path as it can go.

If a firm with market power designs a new standard it will face strong incentives to roll it out slowly to protect a firm from cannibalizing monopoly rents. Sometimes this strategy is obvious. Consider this illustration from ancient history: in the days when IBM controlled a large part of the mainframe market it could not bring itself to abandon Extended Binary Coded Decimal Interchange Code (EBCDIC), its standardized proprietary language, or, for that matter, to help others migrate up from EBCDIC to the many other superior languages available. Despite plenty of improvements IBM could have made, its managers refused to deploy them, preferring instead to exploit locked-in users.[53]

Monopolies also face strong incentives to have a "quiet life," to paraphrase Sir John Hicks.[54] That is, monopolies may exert less effort when they choose standards, or design them to castrate user choices in such a way that leads to less inconvenience for the monopolist at the expense of the user (e.g., trimming product line breadth, or trimming away complex attributes of the product). It is less succinct an observation

52. ROBERT RHODES JAMES, 7 WINSTON S. CHURCHILL: HIS COMPLETE SPEECHES 1897–1963, at 7566 (Robert Rhodes James ed., Simon & Schuster 1983) (1974).

53. Gerald Brock, *Competition, Standards and Self-Regulation in the Computer Industry*, *in* REGULATING THE PRODUCT: QUALITY AND VARIETY 91 (Richard Caves & Marc Roberts eds., 1975).

54. John Hicks, *Annual Survey of Economic Theory: The Theory of Monopoly*, 3 ECONOMETRICA 1, 8 (1935).

than Hicks' might have preferred, but here it is: a self-interested firm faces strong incentives not to dissipate its profits (through interrupting its quiet life) if doing so serves customers in ways that do not generate additional revenue.

That may seem rather abstract, but consider this illustration from the good old days of the AT&T monopoly over residential customer premise equipment. Until the mid-1970s, most households faced a limited menu of (over-engineered and excessively rigid) choices for handset designs. Well engineered or not, there were too few choices in comparison to what a competitive market would have done. Eliminating the monopoly hold over designs led to more than one provider and over time showed just how badly the monopoly had done.

With multiple providers, each provider of customer premise equipment matched the offerings of its nearest rivals. In a short time the heated and urgent competitive behavior familiar to consumer electronics eventually overtook the market, leading to a plethora of choices at a range of prices. Compared with the choices found in just Target or Walmart today, it is remarkable that anyone in the past thought such a limited choice was a good idea.[55]

In other words, in the absence of restraining limitations on discretion, monopolies design selfish standards. An antidote to the selfish standards of monopolies is competition between standards. Indeed, it may be the best antidote. I say that even though very precise economic reasoning suggests no such conjecture can ever hold under all circumstances.[56]

An intriguing counter-example raised against this proposition is the rise of Global System for Mobile communications (GSM) in Europe, the first digital cellular standard to be put into wide use, the deployment of which led to a blossoming of designs for the European handset market. On the surface this experience seems to suggest that government-mandated standards (in a seemingly monopoly position) can sufficiently

55. Indeed, at one time, vocal and powerful participants *did* publically agree to limit customer premise equipment from third parties. From the time these debates first arose, AT&T's lawyers advanced a general argument. This stressed the potential harm an unauthorized attachment might cause to the network. The FCC initially accepted this argument in *Hush-A-Phone Corp. v. United States*, 238 F.2d 266, 269 (D.C. Cir. 1956), which the DC District Court subsequently reversed. Thereafter, and in many related cases, the FCC showed skepticism towards arguments related to potential harm from unauthorized attachments, though AT&T did persist in advancing them. *See* CYBERTELECOM FEDERAL INTERNET LAW & POLICY: AN EDUCATIONAL PROJECT, CUSTOMER PREMISES EQUIPMENT PART 68: CPE, http://www.cybertelecom.org/ci/cpe.htm.

56. See Joseph Ferrell, *Should Competition Policy Favor Compatibility?*, *in* STANDARDS AND PUBLIC POLICY 372 (Shane Greenstein & Victor Stango eds., 2006) and Joseph Ferrell & Phil Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Information Age*, 17 HARV. J.L. & TECH. 85 (2003), for a review of these arguments.

nurture a competitive equipment industry.

The surface view is extremely misleading, and misinterprets the actual sequence of events. Upon close examination the rise of GSM is not a counter example at all. It rather supports the proposition that competition generates a variety of designs and variety helps. That is, GSM's design should be interpreted in light of the competition taking place between equipment firms at the time.

By the time the GSM standard was designed and proposed, most participants in the United States anticipated the beginning of competition between CDMA and TDMA, which would come at the expense of analog systems, supported by a flourishing equipment industry led by the U.S.-based Motorola. European equipment firms were secondary in commercial leadership. In this context, had the European regulators continued to adopt digital standards that already existed in the U.S. cellular market or coordinated their efforts with their U.S. counterparts,[57] they would have adopted either CDMA, TDMA, or both, or left the choice up to the market participants, as the US did. In any of those choices Motorola's existing advantages would have continued to have some effect, as it had considerable experience supplying for the U.S. and European market. Instead, adopting a new standard, such as GSM, wiped the technological slate clean, giving all firms—American, European, and Asian—a new opportunity at the new market. It did not wipe away all incumbency advantages, which, arguably, still arose from experience with distribution and branding. It only eliminated the advantage that came from familiarity with analog technologies or designs.

As it turned out, Motorola was quite late in organizing its products for GSM based equipment and lost considerable market share to Nokia and others during the initial rollout of GSM. Seen in this light, GSM was the product of healthy competition between standards, catching an incumbent flat-footed. Moreover, Europe today benefits from the 3G that came about only because CDMA had the chance to develop in the US, a byproduct of economic experiments in different countries.

Back to the main point: competition between standards also tends to beat monopolies because it makes the design process more transparent. *Transparent* processes are those in which policies let participants know what change is imminent. It informs others openly

57. Histories of these events exist in various places. *See, e.g.*, GSMWorld.com, History, http://www.gsmworld.com/about-us/history.htm (last visited Apr. 2009); Privateline Telecommunications Expertise, GSM History (Jan. 15, 2006), http://www.privateline.com/mt_gsmhistory/02_gsm_history/; Ravings, Rantings, etc.: Clancy's Blog, GSM—history, services, architecture, http://blogs.ibibo.com/takeonlife/gsm-history-services-architechture (Feb. 6, 2007).

and vocally. In other words, these are processes participants' actions make known—sometimes well in advance—when their change will diminish the returns on others' innovative investments.

Such transparency is one of the reasons why standards processes have become a leading indicator of the imminent release of bleeding edge technologies. Interested parties monitor the designs (because they can), and know that their near rivals do the same (because the data is available to anyone). All those parties plan to match each other along the dimension of the standard and differentiate along the dimensions in which each has competitive advantage. Competition ensues once the standard is upgraded from its beta to an endorsed and official standard.

Transparency is a feature found quite frequently in sponsored open source projects, but it is not unique to that setting. It can be found in standards processes. It is thought to have great importance in interdependent value chains. Other firms will not make long-term investments if they cannot understand at a fine level of detail how their software must interact with another's.

Open source observers find that transparency can lead to more participatory decision making for standards.[58] *Participatory* processes are those in which sponsoring organizations invite comment, discussion, and input from others affected by their actions. Such organizations solicit input through public forums, e-mail lists, blogs, community sites, and a range of other activities.

Standards organizations vary considerably in their policies for encouraging or discouraging participation. For example, some organizations require fees, some require participants to meet certain technical qualifications, and others will allow any observer to attend, though not vote.

Wide participation is also found quite frequently in open source projects, particularly those without sponsorship. Wikipedia, though non-profit, is perhaps the best-known example of an online project that encourages wide participation. The Firefox browser community has quite diverse participation from numerous corners. So, too, does Linux. In both the latter cases, most participants are quite technically skilled.

However, wide participation is probably the least common attribute among standards consortia sponsored by commercial private firms. Most managers prefer to retain decision-making authority, guarding investment decisions in the name of stockholders. There is concern that giving up such discretion risks having participants take investment in directions that do not serve firm interests.[59]

---

58. Joel West & Siobhan O'Mahoney, *The Role of Participation Architecture in Growing Open Sourced Communities*, 15 INDUS. AND INNOVATION 145 (Apr. 2008).

59. *Id.*

Accommodating wide participation normally comes at a cost, such as slower decision making and more onerous managerial challenges coming to consensus. That is one reason why Tim Berners-Lee established the W3C with a less participatory structure than found in the IETF, where he had personally experienced the drawbacks of slow decision making when he first tried to standardize the core inventions behind the World Wide Web.[60]

Competition between standards is not what Wall Street analysis values. By definition, competing standards raise the risks for those with stakes in past standards (which might become obsolete) or it raises risks for those who will face competitors or entrepreneurial entrants employing new standards. Once again, it is not unusual to observe a little messiness, and there is no particular reason to anticipate the new standards to yield immediate profitability.

Contemporary Internet infrastructure contains signs of vigorous standards competition. For example, Wi-Max and LTE vie today for next generation wireless data markets. One or both technologies, as implemented and deployed by commercial firms, might very well turn out to be an unprofitable flop, but until we know that for sure, they provide potential competition for the community of firms and researchers interested in developing high-speed data transmission in the near future. That fuels a sense of urgency and gets the government bureaucracies behind wireless telephony to move quickly when they otherwise might not have. The threat can be sufficient to generate earlier investment than later investment.

While competition among standards tends to broadly yield good outcomes for all users and firms, there is an important exception. From time to time the rollout of a new standard involves a "coalition" of firms who have signed up for one design, opposing another "coalition" who has signed up for another. Such coalitions emerged in the HD DVD versus Blu-Ray fight, the Wi-Fi versus Home-RF fights, and the 56K Flex versus X2 fights, which are among many examples from the last decade. In fights between coalitions, the battle is good for everyone, except, perhaps, those in the losing coalition of a big standards battle. The losing coalition may expend considerable resources for which its members do not gain returns.

Though more complex, a similar dynamic exists in competition between organizations for control of governance over standards. Competition between groups, organizations, and communities is a good thing in general, even if specific participants lose out on occasion.

60. See TIMOTHY BERNERS-LEE & MARK FISCHETTI, WEAVING THE WEB: THE ORIGINAL DESIGN AND ULTIMATE DESTINY OF THE WORLD WIDE WEB (2000), for a description of this frustration.

Competition between organizations is different than competition between designs, however: competition between organizations may involve competition between alternative designs of standards, but it often involves competition between commitments to different processes in the future for upgrading the standards as well. The latter involves choices between commitments by specific communities of managers, technologists, and/or sponsoring firms.

The history of the Internet itself provides the best illustration of this lesson.[61] Development of TCP/IP as a foundation for a national network occurred in the presence of an alternative process and model for the same activity, organized by the International Organization for Standardization (ISO). Competition between organizations led to development of data exchange standards sooner than otherwise would have occurred had the ISO made the decision all by itself. The reverse is also true. Competition from ISO generated urgency within the communities of the Internet Architecture Board (IAB) and eventually the IETF to organize their myriad ideas and implement them quickly.

Looking more closely, this competition stressed more than merely different designs, which illustrates why competition between groups is not perfectly analogous to competition between technologies. Throughout the latter part of the 1980s there were two processes for determining standards. One process existed at the ISO, and it emphasized committee consensus in advance of deployment, with committees comprising representatives of all major stakeholders. Another existed among the descendants of the Defense Advanced Research Projects Agency (DARPA), organized around activities at the IAB, who established the IETF midway through the decade. Their process stressed bottom-up suggestions and demonstrating workable solutions before adoption.[62]

As it turned out, a bottom-up process centered in the United States made considerably more pragmatic progress, but even that was due to more than just its bottom-up nature. Even from its earliest days, IETF leadership did its best to aid the process it governed. First, it tried to provide editorial guidance and support for the entire process. That resulted in remarkably clear and comprehensive documentation (particularly from some contributors who were not practiced at clarity and thoroughness).

Second, the IETF also helped coordinate and sponsor "plugfests"

61. ABBATE, *supra* note 9.

62. *See* Andrew L. Russell, *'Rough Consensus and Running Code' and the Internet-OSI Standards War*, 28 IEEE ANNALS OF THE HISTORY OF COMPUTING, July–Sept. 2006, at 48; *cf.* William J. Drake, *The Internet Religious War*, 17 TELECOMM. POL'Y 643 (Dec. 1993); ABATTE, *supra* note 9.

where vendors could test their interoperability of actual implementations. In principle, these fests were used to verify the existence of "running code" before advancing a proposal for an RFC to a final draft. Those efforts provided the administrative glue to accumulate technical suggestions from many disparate corners.[63]

In these examples we see the benefits of messy clashes between organizations over their domains of expertise and even over the proper processes for making technical progress. It infuses decisions with a healthy tension concerning multiple options. It might be irritating for the participants involved, but the sniping results from a healthy diversity of opinion in the face of opportunity.

In sum, these clashes beat any outcome likely to arise in the presence of monopoly provision of standards.

Just as with unfettered experimentation, competition between standards and between the organizations that sponsor them yields a benefit. It may lead to innovative entrants, or it may enhance the products of one particular firm. It forces incumbents to react, or, even better yet, anticipate the entrant and innovate in advance. This fosters incentives to lower prices and to sponsor more innovative products sooner, thereby benefiting users.

C. Inventive Entrepreneurialism

Entrepreneurial initiatives involve an organization in a risky and challenging business in pursuit of a new economic opportunity. These firms are the "participant" that makes the first intrepid attempts at deploying, distributing, or servicing a new good to a wide range of customers with the intent of making a profit. Small start-ups take entrepreneurial action and so do large firms. Sometimes small businesses that take such risks are bought by large organizations, such as Cisco, IBM, or Microsoft. Sometimes small start-ups go public and grow into large firms themselves.

While the addition of more and more entrepreneurs (after some point) does not always make a situation better, their complete absence is a sign of poor innovative health. The presence of entrepreneurs provides the simplest benchmark.

It might be tempting to use the presence of start-ups funded by venture capitalists or angel investors as a measure of the presence of entrepreneurship. That is not precise or even accurate in today's markets. To be clear, while most start-ups involve entrepreneurs, not all entrepreneurs must have venture funding. Entrepreneurship also arises inside small divisions of corporations, or stems from corporate funding of

63. *See* Bradner, *supra* note 27.

spin-offs and other corporate ventures. In other words, not all innovation comes from start-ups and venture capitalists.

Indeed, there are good reasons why entrepreneurship inside large firms does not resemble entrepreneurial actions undertaken by small firms. If large firms go after the same business opportunities as small and medium firms, it is not surprising that they will take heterogeneous approaches to the same opportunity. For example, small firms may have the advantages of dexterity and surprise, while the large have the advantages of established brands, distribution channels, and strong feedback networks with existing users. The large firm will tailor its actions to its advantages.

The following is also true: both VC-funded entrepreneurs and all other kinds tend to be present at the same time in the same settings chasing the same opportunities. And so the low points are most informative: the absence of any start-ups is a pretty reliable signal of hostile environment for innovative, entrepreneurial young firms.[64]

Recent history reinforces this point. The increasing presence of entrepreneurs in communications markets has been one of the sweetest developments in the last two decades. It has brought rapid change to many sub-markets. Today we take for granted our access to e-mail, instant messaging, IP-enabled video conferencing, picture sharing, amateur-video sharing, online mapping, accessible hosted CRM applications from any location, mobile push e-mail, and a host of other utilities that no non-technical individual can understand. In virtually every case of radical change the events did not arise solely from the actions of incumbent firms with existing businesses. At some point, entrepreneurial actions got involved.

Three benefits are affiliated with the presence of a variety of entrepreneurs in comparison to their complete absence. Entrepreneurs have incentives to differentiate from incumbent firms who over-commit to one technological forecast about direction of change.[65] A related benefit has to do with overcoming inadequacies in establish organizations. Even if established firms have *incentives* to pursue a portfolio of technical directions, they may fail to act on them due to the

64. In this sense the argument here overlaps with that found in ROBERT W. FARLIE, KAUFFMAN INDEX OF ENTREPRENEURIAL ACTIVITY 1996–2007 (2008), www.kauffman.org/pdf/KIEA_041408.pdf, which measures entrepreneurial activity across different locations in the US, using self-reported decisions to open a self-owned business from the Current Population Survey.

65. This theme runs throughout a large range of studies. *See* Rebecca Henderson, *Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry*, 24 RAND J. ECON. 248 (1993); Rebecca Henderson & Kim Clark, *Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms*, 35 ADMIN. SCI. Q. 1 (1990).

absence of internal champions for a new technical direction, or the protection of rents flowing to internal champions.[66] Once again, entrepreneurs view such situations as opportunities. Third, when it is unclear which of several technical directions is most valuable, society gains from pursuing a variety of the least cost alternatives. Entrepreneurship can foster investments from distinct firms with different cost structures, each of them facing heterogenous incentives to invest in the technology.

The best historical illustration of these ideas comes from the development of the Internet itself, as it transitioned from its academic origins into a commercial service.[67] Executives at many established firms, such as AT&T and IBM, simply *did not* invest in operations nurturing any commercial future for TCP/IP services, even into the early 1990s. Some entrepreneurs viewed that as an opportunity and acted according to their vision. Thus, the initial growth of the commercial Internet involved a mix of firms from a variety of backgrounds. They shared a vision that the Internet would grow.

Some, such as PSINet and UUNet, were entrepreneurial descendents from the NSFNet. Others, such as Netscape, involved personnel from university research laboratories and executives from prior entrepreneurial commercial ventures. Others, such as those at BBN and MCI, were entrepreneurial actors inside large enterprises, who came from quite distinct backgrounds and interests. Others still were small Internet Service Providers, descendents from the bulletin board industry, who saw opportunities to establish a new service for local customer base.[68]

Because entrepreneurs often are the first to perform an economic experiment with a newly designed standard, a market with thriving entrepreneurial activity often results from the same factors that encourage a healthy amount of economic experimentation and standards competition. Yet, other factors matter, too. Entrepreneurial activity also can increase and decrease for distinct reasons.

Three additional factors play a role in encouraging entrepreneurship: low development costs, fast speed to commercialization, and strong appropriability conditions as defined by the eco-system. These are important to recognize because pragmatic policy can shape these factors.

Development costs and speed to commercialization refer to two attributes of every young firm's experience, i.e., the expense before

66. *See* CLAYTON M. CHRISTENSEN, INNOVATORS DILEMMA, WHEN NEW TECHNOLOGIES CAUSE GREAT FIRMS TO FAIL (1997).

67. *See, e.g.,* ABBATE, *supra* note 9; Greenstein, *supra* note 26.

68. Greenstein, *supra* note 26.

shipping the first major release and the amount of time it takes. Both of these are usually measured from the time a young firm first gets its funding, or its founders assign full-time managerial responsibility to somebody for the development of a product, whichever comes first.

To give a sense of scale to this discussion in Internet markets consider this example: Netscape was founded in April 1994 and sought to ship its first beta browser in four months. In fact, it took slightly longer. The first beta browser shipped in November 1994, its official product in February 1995. That effort involved several million dollars for a few months of development work and initial distribution.[69] That was at the high end of software development costs. A typical application firm in the late 1990s was expected to burn through several million dollars in a couple years. A typical software firm was expected to launch its first product in less than year, perhaps more if the product was particularly complex.

In comparison most examples of young software firms from this decade are astoundingly inexpensive until they scale up. Using open source software, modern startups have tended to work just as fast or faster, and with considerably less expense. For example, YouTube went from founding to first service in less than three months, entirely financed on the credit cards of one of the founders.[70] They did not bring in millions of dollars of working capital from any venture capitalist—in this case, Sequoia Capital—until they needed to scale their server equipment and support personnel to accommodate their spectacular growth.

That is not an isolated example. In general, it is quite common for the software firms of the Web 2.0 movement to burn no more than a few hundred thousand dollars a year and operate with less than a couple dozen employees. Many programmers with Web 2.0 startups boast about their ability to survive on "ramen profitability"—just enough revenue to buy ramen noodles for a couple founders for a while until it finds a service with wide appeal.[71] Even after funding, many firms can

69. MICHAEL CUSUMANO & DAVID YOFFIE, COMPETING ON INTERNET TIME, LESSONS FROM NETSCAPE AND ITS BATTLE WITH MICROSOFT (1998).

70. David Greising, *YouTube Founder Rides Video Clips to Dot-Com Riches*, CHICAGO TRIBUNE, Oct. 15, 2006; Scott Woolley, *Video Fixation*, FORBES, Oct. 16, 2006.

71. This phrase is due to Paul Graham. The earliest use appears to be from an essay dated May 2005, in an essay "Hiring is obsolete." He states,

> Like everything else in technology, the cost of starting a startup has decreased dramatically. Now it's so low that it has disappeared into the noise. The main cost of starting a Web-based startup is food and rent. Which means it doesn't cost much more to start a company than to be a total slacker. You can probably start a startup on ten thousand dollars of seed funding, if you're prepared to live on ramen. The less it costs to start a company, the less you need the permission of investors to do it. So a lot of people will be able to start companies now who never could have before.

The phrase "ramen-profitability" emerged from those origins, and has diffused into wider use.

accomplish amazing tasks with few permanent staff. I have toured numerous start-ups that operate with less than ten employees, and they intend to stay that way until they get their product into mass markets, at which point they will expand to less than a few dozen.

In other words, modern Web start-ups generally face low development costs, and anticipate a small scale for a long period of their earliest development, prior to scaling for a mass market. They all dream of reaching a mass market quickly, to be sure. If they do not it is not their end. They can survive and experiment for a long period.

Generalizations about the level of entrepreneurship that signal a healthy level of such activity are hard to make, not surprisingly. Some determinants of development costs and speed are outside the control of any participant. Those need to be distinguished from determinants of development costs and speed within the control of some participants.

Here is an example of determinants outside the control of entrepreneurs. Level3 entered the backbone market in the late 1990s at high expense, burning through hundreds of millions of dollars a year (maybe billions) while it built thousands of miles of new lines for its national network.[72] While Level3 initially was able to receive top dollar in revenue for its new all-IP infrastructure, its example was not followed by any other entrant.

Indeed, once the contract prices fell for backbone services in 2000-01,[73] no large new entry was observed in the backbone market except Cogent (which began service in 1999). Cogent largely did not build its own network. Instead, it put together its network from the assets of previously bankrupt firms,[74] vaguely reminiscent of how Cornell

---

See Josh Quittner, *The New Internet Startup Boom: Get Rich Slow*, TIME, Apr. 9, 2009.

72. The company's own web site boasts of something similar. It says,

> During 1998, Level 3 raised $14 billion and was called the 'best funded start-up in history.' The company constructed 19,600 route miles, and built the world's first continuously upgradeable network fully optimized for internet protocol (IP). Over the next few years, explosive demand for bandwidth fueled growth in sales. By the end of 2000, Level 3 provided service to 2,700 customers.

Level 3 Communications, A Network Built to Support the Silicon Economics Cycle, http://www.level3.com/index.cfm?pageID=245 (last visited Apr. 2009). Other web sites aimed at investors say something similar but in less generous language. For example, as recounted on fundinguniverse.com, Level 3's executives "secured rights of way from railroad companies and ordered legions of employees to dig trenches and lay cable, efforts that quickly exhausted billions of dollars in capital." Fundinguniverse.com, Level 3 Communications, http://www.fundinguniverse.com/company-histories/Level-3-Communications-Inc-Company-History.html (last visited Apr. 2009).

73. Greg Rosston, *The Rise and fall of Third Party High Speed Access*, 23 INFO. ECON. & POL'Y 21, 29 (2009).

74. On their company history web page they state:

> Although debuting at the height of the telecom industry, Cogent soon found vast market wealth eradicated and many other ISPs thrown into a state of turmoil. In a survival of the fittest competition, Cogent became the consolidator in a

assembled Western Union back in his day. The basic economics of entry suggest that new building is not justified when prices fall, capacity goes unused, and growth can be achieved through restructuring. Simply stated, the costs are too high to merit building a new frontier start-up, no matter how good they are. (Constructing a new firm from existing assets bought at fire-sale prices, apparently, is a different matter.)

Here is an example of determinants within the control of some participants. For the last decade Intel has released prototype designs for the inside of the PC and endorsed specific implementations. That action has reduced the costs of designing some components and speeded the development of others. It has fueled considerable entrepreneurial activity.[75]

Selective withholding of information also can serve strategic purposes that delay entrepreneurial competition. Intel was accused of actions, in particular of withholding technical information from other participants in a quid pro quo for licensing of its intellectual property, which generated an FTC investigation.[76] Outsiders frequently accused Microsoft of using its position to make its own life easier, such as documenting for Microsoft's use but not necessarily for any others', and not documenting code so the company could alter it to its advantage.[77]

Three aspects of these types of allegations deserve notice. First, they are extremely difficult to prove in court (at least in ways that lawyers and judges find satisfying). Second, once leveled, these allegations take on a life of their own, and continue on in many distorted forms in (on-line)

> consolidating market. Over three brief years, Cogent completed 13 acquisitions of other flailing providers. Whether it was an entire company or just select assets, Cogent was able to acquire valuable network assets, customers, peering relationships and building access agreements for pennies on the dollar.

Cogentco.com, Cogent Communication History, http://www.cogentco.com/us/about_history.php (last visited Apr. 2009).

75. *See, e.g.*, GAWER & CUSUMANO, *supra* note 21; Gawer & Henderson, *supra* note 21.

76. *In re* Intel Corp., 128 F.T.C. 213 (1999).

77. One developer suggests the following:

> Why [not document part of the internal subsystem for Win32], one might ask? Well, the official reasoning is that it allows Microsoft to tune and modify the system call layer at will, improving performance and adding features without being forced to provide backward compatibility application binary interfaces . . . . The more nefarious reasoning is that it allows Microsoft applications to cheat, and call directly into the undocumented Win32 subsystem system call interface to provide services that competing applications cannot. Several Microsoft applications were subsequently discovered to be doing just that, of course . . . . These days, this is less of a problem, as there are several books that document this system call layer . . . . But it left a nasty taste in the mouths of many early Windows NT developers (myself included).

Jeremy Allison, *A Tale of Two Standards, in* OPEN SOURCES 2.0: THE CONTINUING EVOLUTION 47 (Chris DiBona, Danese Cooper & Mark Stone eds., O'Reilly Media, Inc. 2006) (2005).

communities that mistrust the leading firm who stands accused. Third, such allegations usually presume the dominant firm could have acted differently, i.e., in a manner more considerate to the interests of other entrepreneurs, without much cost.

Internet insiders will recognize a familiar outline in the debate over Comcast's throttling of Bit-Torrent traffic. Many accusations have been leveled at Comcast for throttling traffic. When the behavior was first discovered, it unleashed more than a bit of torrent of speculation in online discussion groups about the extent of the action and the competitive motivation.[78] Not only did the lack of advance notification leave many parties suspicious about Comcast's policies, but it left observers puzzled about why it never dawned on management to issue a press release before taking action. Comcast's secrecy fueled rumors, and it came across as unnecessarily inconsiderate to users, entrepreneurs, and other on-line participants.

The final attribute of every young firm's experience is something called a*ppropriability conditions*. It refers to the ease with which entrepreneurs can retain exclusive rights over their inventions or other unique assets, usually through one of several strategies to prevent imitation: secrecy, patents, copyright, first mover advantages, or some combination of those. If conditions are weak, then entrepreneurs expect to lose quickly their unique advantages to others

Appropriability conditions are controversial for reasons related to the discussion about withholding information. They partly depend on the tenor of interfirm relationships in the competitive ecosystem constructed by leading incumbent firms. This factor has received attention by many others, so this discussion will remain brief.

For example, some incumbent firms, such as Cisco, have made it very clear that they intend to purchase other small start-ups who reach the frontier in an area Cisco considers important. That is regarded as a quite inviting setting for start-ups. Similarly, during its heyday in the late 1990s, AOL was known for its willingness to sign a reasonable deal with just about any start-up who had a service to offer for its platform. Venture capitalists also took note of these positions and started firms accordingly.

In contrast, for many years Microsoft was known to prefer internal growth over acquisitions, usually in a fast-follower strategy, *i.e.*, basing development on the lessons learned through the economic experiment

78. *See, e.g.*, PETER ECKERSLEY, FRED VON LOHMANN & SETH SCHOEN, PACKET FORGERY BY ISPS: A REPORT ON THE COMCAST AFFAIR, ELECTRONIC FRONTIER FOUNDATION, Nov. 28, 2007, http://www.eff.org/files/eff_comcast_report2.pdf; *Comcast Throttles BitTorrent Traffic, Seeding Impossible*, TORRENT FREAK, Aug. 17, 2007, http://torrentfreak.com/comcast-throttles-bittorrent-traffic-seeding-impossible.

conducted by other firms. That did not deter entry in application software, but it motivated firms to start young companies with no planning for acquisition, and to expect the potential for imitation from the very firm with whom they had to partner in order to reach users. More recently, Microsoft has changed its stance about acquisitions, particularly in areas related to cloud computing, and that has raised a number of questions among VCs about funding firms in related areas.

The legal environment also shapes appropriability conditions. This, too, has received much attention from others, so for this discussion I will keep my observations brief. For example, the changing legal and regulatory conditions of the late 1990s adversely affected the basic costs and viability of a wide range of CLEC business plans. To say the least, the environment went from friendly to hostile in a few years, and, not surprisingly, entry of young start-ups declined as a result.[79]

As noted earlier, large firms and incumbent firms can be entrepreneurial, too. That is why some entrepreneurial actions by large firms, though otherwise puzzling, may have a silver lining. For example, Nokia continues to struggle to find new initiatives beyond hardware design, whether it involves buying Navteq or starting new music services that anger its carrier partners. In light of the relevance of entrepreneurship, we should salute them. As the provider of almost half the smart phones in the world[80] and close to 40% of all cell phones, Nokia has considerable clout. Yet it refuses to stand still. It continues to restructure, a sign of taking entrepreneurial risks in advance of new opportunities in new markets.

As another example, Cisco's attempt to get into video conferencing by purchasing Webex seems strategically incongruent, because it involves integrating a large software firm into one that specializes in equipment markets. Yet, the merger also introduces the company to a wide range of new opportunities and challenges related to developing tele-presence. Even if I am skeptical that Cisco will gain a return on its investment in Webex, I am interested to see what it makes of its entrepreneurial action and this departure from prior strategic action.

Once again, established firms with market power do not tend to benefit from entrepreneurial inventiveness that threatens their economic rents. That raises concerns that dominant firms will attempt to shape development costs, entry speeds, and appropriability conditions to serve

79. NUECHTERLEIN & WEISER, *supra* note 12, at 15–16, 69–114 (examining the tension between the competitive goals of the Telecommunications Act of 1996 and the unavoidability of economies of scale); *see also* Greenstein & Mazzeo, *Differentiated Entry into Competitive Telephony*, 54 J. INDUS. ECON. 323–350 (Sept. 2006).

80. Kevin O'Marah, *Feasting on a Content Economy: Nokia Bites Apple*, AMR RESEARCH, June 30, 2008, http://www.amrresearch.com/content/View.aspx?compURI=tcm:7-37691.

its own strategic priorities. For example, established firms would rather buy out an entrepreneurial inventor than see it become an independent company and potentially compete.

Once again, Wall Street analysts may stress different factors than those that benefit society at large or users. In part that is because the value of many start-ups may be privately held for some time, not evaluated by any stock market pricing, as with established firms. In addition, as with a multiplicity of economic experiments and the regularity of standards competition, there are general benefits to buyers from competition between start-ups and established firms that Wall Street does not necessarily value. Even when such start-ups have only small chances of success in the long term, such competition enlarges buyer choice, multiples opportunities for learning, and enhances urgency at the established organizations.

More broadly, just as with unfettered experimentation and competition between standards, entrepreneurial inventiveness yields benefits at the level of the market even if the benefits are small at the level of the firm. Such entrepreneurship will serve a purpose in the plans of innovative entrants. It will generate reactions from other competitors or imitators. Once again, this fosters incentives that ultimately lead to lower prices and more innovative products, and sooner. Users benefit from that, and policy can encourage it.

D. Absence of Unilateral Bargaining

Negotiation shows up in every firm's life. From some of the above examples, it is easy to see why: suppliers complain about growing costs and suggest alternatives, technologists suggest alternative methods for accomplishing a task, programmers complain about the poor quality of code and seek to push out release dates, stockholders demand higher profits, buyers complain about tight budgets and threaten to choose another option. Managers caught between such complaints must constantly negotiate with many participants. More to the point, managers inside the Internet value chain have an especially difficult task because the addition of technical interrelatedness adds one more layer of complexity to an already tough negotiating task.

In a network with a high degree of technical interrelatedness, there are general gains to all parties from bringing routines into business processes and activities, much like there are gains to adopting standards and platforms to coordinate activities. While there may be no better way to reduce complexity, adopting such routines may require negotiation between multiple parties.

Such negotiation offers no guarantee of success. Many outcomes are possible. Occasionally both parties want an agreement, but just as often

one party will desire it more than the other. Alternatively, one party may have an ability to generate a better deal than the other, and sees bargaining as an opportunity to generate a strategic advance or gain additional revenue. As a general rule, the structure of bargaining sometimes can work out to a Goldilocks equilibrium that is just right—not too hot and not too cold—but more often it does not. One firm gets too powerful or another prominent bargainer loses its way.

In the extreme, bargaining becomes one-sided, with one party asking for something while the other refuses to provide it or only agrees to it at a high cost. The simplest manifestation of this extreme situation arises when the more powerful party declares a "take-it-or-leave-it" offer, leaving other parties no choice.

Given such a range of possible outcomes, how can we tell what signifies a healthy market? In short, the absence of one-sided bargaining.

That said, it is not as simple as it sounds. The presence of one-sided bargaining by itself is not bad. That is, the absence of one-sided bargaining is a sign of health, while the presence of one-sided bargaining is a sign of potential illness, which might have adverse consequences that might spread. The key question is whether the parties who receive such take-it-or-leave-it offers have access to reasonable alternatives. This will take some explaining.

Let me illustrate the role of negotiations with a comparatively uncontroversial example. Intel has a series of agreements with numerous OEMs about putting the Intel Inside and Centrino brands on their products to signal to users that the laptop includes a Wi-Fi compatible motherboard and antennae designed by Intel. In addition, Intel often includes certain compensation for the marketing expenses of putting the Intel copyrighted jingle inside a commercial.

A breakdown of negotiations can arise from one-sided bargaining, but this example illustrates that breakdowns can occur for other reasons as well. A few years ago Dell refused to carry the Centrino branded systems, and, accordingly, did not receive the quid pro quo compensation. Both parties went on their merry way for many years. Dell continued to carry both Intel products, but after that incident began to more prominently distribute designs with AMD chips. At the same time Intel reached deals with every other major OEM, and succeeded in making Centrino a feature of the majority of notebooks in use.

What does this example illustrate? First, that Intel's market power had its limits with Dell. It eventually reached a point in its negotiation with Dell where Intel gave Dell a take-it-or-leave-it offer and, indeed, Dell chose to leave it (unlike virtually everyone else in the industry). Second, as long as Dell had plenty of other options, the losses to Dell or society at large were not too large. Indeed, there might have been gains,

since Dell's choices translated into more buyer options beyond the Centrino.

One-sided bargaining can have some serious consequences, however. Some years ago there was a proposal to let all Internet participants simply negotiate compensation between them, so that Google/Yahoo/Disney would negotiate with Comcast/Time-Warner/Verizon, and every other possible combination. Intel's example suggests the problem with such a proposal: imagine the uproar among users in the locations where such negotiations failed to come to resolution and no other close substitutes existed. It would be far worse than the brief uproar last year among Yankee fans who could not get local baseball telecasts due to a negotiation breakdown between Major Leagues Baseball, the Yankees, and a local cable provider.

Indeed, this did happen a few years ago when negotiations broke down between Cogent and Sprint.[81] However, the situation was easy to misunderstand and misinterpret. In this case, the absence of market power reduced the policy concerns affiliated with the breakdown of bargaining, albeit some policy concerns still remained. That requires explanation.

Specifically, Cogent and Sprint were exchanging traffic through a third party and, like other backbone firms, sought to connect directly in a peering arrangement that bypassed the third party.[82] That is not a trivial step. First, it required the building of appropriate lines and equipment, which cost money. Second, as with other peering, it required measuring traffic directly to verify that traffic was sufficiently symmetrical back and forth. The two firms negotiated an agreement for building the connection and the terms for breach—that is, what type of traffic experience would justify ending the peering.

After building this connection, Cogent stated that it was satisfied with the traffic flow, while Sprint stated it was not. There was no dispute about the symmetry of the traffic back and forth, but there was a disagreement about its level. Sprint argued that Cogent did not provide enough traffic to justify a peering relationship.[83]

After declaring Cogent could no longer peer with it, Sprint did not immediately de-peer. Rather, it unilaterally declared that the two companies were in a paying relationship, as Sprint would do with any small ISP. Sprint then began to send bills to Cogent. Cogent argued that it had met the conditions for peering, and that Sprint's claims were

81. Scott Woolley, *The Day the Web Went Dead*, FORBES.COM, Dec. 2, 2008, http://www.forbes.com/2008/12/01/cogent-sprint-regulation-tech-enter-cz_sw_1202cogent.html.

82. *Id.*

83. Woolley, *supra* note 81.

disingenuous attempts to make money. In essence, Cogent refused to pay, and both companies put considerable spin on events.[84]

This standoff went on for many months until Sprint's management decided to shut down its side of the peering.[85] That action had consequences for users on both networks who did not multi-home, i.e., did not use more than one backbone firm. One set of exclusive Sprint users could not reach another set of exclusive Cogent users.[86]

To make a long story short, users of both carriers were angry. Cogent publically blamed Sprint's decision to de-peer, and, for reasons not made public, after a few days Sprint's management gave in, reversing the de-peering.[87] Soon after the two firms came to a long-term agreement whose details were not disclosed publically.[88] In other words, as of this writing, this negotiating tactic hurt users, but it is unclear which firm won the negotiation. It is not clear how much money changed hands (or will change hands).

The Sprint-Cogent case suggests four lessons. First, any outcome depends on the circumstances surrounding the use of the tactic. No generality could hold for all circumstances. Second, user (dis)satisfaction plays a role in those negotiations, but it is not the only determinant. Third, any rule about interconnection will have tactical consequences for users. For example, a must-carry rule for interconnection would simply have narrowed the set of actions Sprint could take.

The fourth lesson is more subtle. Several news stories tried to make an inference about the managerial style of Cogent's CEO, Dave Schaffer, since this is not the first negotiation breakdown his firm has encountered. That focus misses the forest for the trees. The personality of a CEO is not the point. Only his entrepreneurial vision is, because at the heart of this example lies a potential competition policy issue.

Specifically, Cogent has a "entrepreneurial" distinct vision about how to attract customers and serve their needs. To execute that vision Cogent necessarily must interact or exchange traffic with the very firms with whom it competes. It is not hard to interpret negotiation breakdown initiated by an incumbent firm as a tactic to discourage an entrepreneurial vision and deter an economic experiment by an entrant. In general, competition policy issues always arise any time an existing firm can shape the costs of an entrepreneurial entrant. If further

84. *Id.*

85. *Id.*

86. Numerous computer scientists and networking experts have pointed out to me that both Sprint and Cogent could have adjusted their routing tables in advance to prevent users from being cutoff. Hence, there is a sense in which both parties played brinkmanship and bear responsibility for imposing costs on their users.

87. Woolley, *supra* note 81.

88. *Id.*

investigation reveals that Sprint possessed market power, their refusal to interconnect would be especially disturbing.

In other words, if Sprint were found to have market power, then its attempts to bargain over interconnection are potentially more problematic.[89] Its actions could be interpreted as an attempt to shape competition.

One tactical gain from market power, for example, is the ability to ignore customer complaints at the retail level while pursuing other tactical goals, say, at the wholesale level, where interconnection takes place. It appears that Sprint's capitulation to its user base is, however, evidence that Sprint's management does not have the ability to ignore its users for very long.

I raised this example for a reason. In short, one-sided negotiations and bargaining breakdowns are, by themselves, insufficient to conclude definitively there is a problem. In the presence of market power, however, it is much more likely a sign of lack of innovative health.

Now consider the lessons from negotiated arrangements that are a bit more one-sided, and happened in the presence of market power. These days Apple offers a standard contract to all application developers for the iPhone about how their services will be sold, requiring them to sign non-disclosure agreements as a condition for inclusion on the Apple Web page where applications are sold.

What dispute arose? Developers complained that the non-disclosure agreement was too tight, even for firms whose applications were ultimately rejected. This made headlines when Apple "clarified" its policies, announcing that even the non-disclosure letter was subject to the non-disclosure agreement.[90] Apple argued that anything done by a developer for Apple could not be shared with others, even if Apple refused it for the iPhone after review. After considerable uproar on blogs and developer list serves,[91] Apple relented on this provision, but it continued to argue that it had a right to protect its innovations through use of these agreements.[92]

Once again, Apple made a take-it-or-leave-it offer and imposed conditions on others. The negotiation breakdown was not necessarily symptomatic of a problem, however, as long as alternatives existed. In

89. Because events suggest that Sprint in fact lacked market power, I am not concluding that Sprint acted in an attempt to shape competition.

90. *See, e.g.*, Arnold Kim, *Apple Extends Non-Disclosure to App Store Rejection Letters*, MACRUMORS.COM, Sept. 23, 2008, http://www.macrumors.com/2008/09/23/apple-extends-non-disclosure-to-app-store-rejection-letters/.

91. *Id.* (receiving over 1,000 negative ratings and over 400 comments).

92. Gregg Keizer, *Apple Drops iPhone NDA Gag Order*, COMPUTERWORLD, Oct. 1, 2008, http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9116007.

this case, however, not many alternatives existed for reaching Apple iPhone users. If a developer wanted to reach the Apple user-base they had to distribute through Apple's outlets and accept the condition. No other avenue for reaching them existed. Apple controlled them all.

If other devices are substitutes for the Apple iPhone, then this situation is less worrisome. If developers wanted to reach users through alternative devices for mobile computing, they had to reach deals with those device providers. Initially few existed, but increasingly announcements are being made about new entry. As some gain market share, the situation may change.

These examples illustrate several general points. Breakdown happens for many reasons, and those should be considered distinct from the reasons shaping one-sided negotiations, which can look similar. One-sided negotiations, in contrast, involve one party with enough bargaining power to make a take-it-or-leave-it offer that others have no choice but to accept.

What factors mitigate the public policy issues in Apple's case? Generally speaking, suppliers in young markets get wider licenses from a court just because they produce for a nascent set of users for new devices and services. For many good reasons there is a general presumption that no firm introducing a new product has market power at the early stage of growth, as it is subject to competition from established brands with established distribution channels and large market share.

So why does Apple get any scrutiny at all? Questions arose here because Apple is unique. The attention is a testament to Apple's unusual recent success commercializing small devices, such as the iPod, and its rather unique place today as a firm that every developer expects to succeed with users, even with a new product, unlike just about every other firm.

Perhaps a more famous example of one-sided bargaining came from Microsoft in the mid 1990s. While I do not mean to single out Microsoft in the use of one-sided negotiations, several circumstances contribute to a disproportionate number of examples in the essay from Microsoft's conduct in the 1990s. First, due to the antitrust trial,[93] many of its internal memos became public, providing a unique and well-documented window on how such negotiations were conducted. Second, Bill Gates was remarkably adept at pressing his negotiating advantages when he had them. His behavior provides good illustrations of how one-sided negotiations can become. Third, and similar to the Apple example, many developers wanted access to the users of Windows. However, in this case, the alternatives were quite limited, and, so, courts had no issue

93. United States v. Microsoft Corp., 147 F.3d 935 (D.C. Cir. 1998).

concluding that Microsoft's attempts to make access more or less difficult for those developers had real consequences.[94] The user base for Windows was large, so the stakes were quite high for many developers if they faced even a small degree of problems.[95]

For instance, prior to the rollout of Windows 95, Microsoft began to dictate conditions to its application developers. It started mildly, with design specifications that application providers were required to comply with, such as specifications for pull-down menus, and other processes that had to follow the available APIs. Most firms acquiesced to these for lack of any alternative, and some grumbled about it at the time.

The more controversial dictates started showing up closer to the rollout of the system. Perhaps the most interesting and dramatic example of bargaining arose in the deal between Apple and Microsoft. The newly returned Steve Jobs took over Apple when it was in a dire financial position, and one avenue for a quick infusion of cash was to settle a patent dispute with Microsoft. Microsoft, in turn, was willing to settle the dispute quickly only as part of a comprehensive deal that included Apple making Internet Explorer the default browser for the Mac.[96]

Transcripts of Microsoft e-mail (made public later) showed that Gates and other Microsoft executives discussed how to hint to Jobs that it was possible for a delay in the release of Word for the Mac, a threat to gain movement from Jobs, since such a delay could hurt Apple's slumping sales. In retrospect, even a hint of this delay to Jobs was remarkable, since the development of Word was (actually) proceeding in a timely manner. After the fact, it is difficult to know what role such a threat played in addition to all the factors at work.[97] As it turned out, Jobs accepted the deal for Internet Explorer and his own customers booed him soundly at a convention when he initially announced it.[98] It surely was not the way for a newly returned CEO to curry favor with customers, but he was over a barrel at the time.

Still, for sheer unpleasantness, the negotiation between Compaq

---

94. United States v. Microsoft Corp., 87 F. Supp. 2d 30 (D.D.C. 2000).

95. *Compare* Daniel Rubinfeld, *Maintenance of a Monopoly:* U.S. v. Microsoft, *in* THE ANTITRUST REVOLUTION: ECONOMICS, COMPETITION, AND POLICY 476, 476–501 (John E. Kwoka, Jr. & Lawrence J. White eds., 2008), *with* Timothy F. Bresnahan, *The Economics of the Microsoft Case* (2009) (unpublished manuscript), *available at* http://www.stanford.edu/~tbres/Microsoft/The_Economics_of_The_Microsoft_Case.pdf; *with* WILLIAM H. PAGE & JOHN E. LOPATKA, THE MICROSOFT CASE: ANTITRUST, HIGH TECHNOLOGY, AND CONSUMER WELFARE (2007).

96. John Markoff, *Microsoft Comes to the Aid of a Struggling Apple*, N.Y. TIMES, Aug. 7, 1997, at A1, *available at* http://www.nytimes.com/1997/08/07/business/microsoft-comes-to-the-aid-of-a-struggling-apple.html.

97. *See* BANK, *supra* note 42, at 12.

98. *See, e.g.*, Peter Burrows et al., *Is This Apple's New Plan?,* BUS. WK., Aug. 25, 1997, http://www.businessweek.com/archives/1997/b3541160.arc.htm.

and Microsoft has no equivalent. Compaq had heard from many customers who wanted Netscape browsers, and featured it prominently. As a reward for listening, Microsoft publicly roughed up Compaq. In 1996, for example, an employee at Compaq removed the Internet Explorer icon from shipped versions of computers. The employees viewed this as part of a range of actions to keep the icons less confusing, orienting them toward business obligations and toward the applications users wanted. Microsoft believed Compaq had a business obligation to display Internet Explorer, and it sent a letter to Compaq threatening to cut off its operating system license in sixty days if a removed Internet Explorer icon was not put back on all new systems[99] and the dispute did not come to resolution.[100]

Compaq capitulated on the dispute quickly. At the time it left everyone in the industry with the strong impression that Microsoft chose to make an example of Compaq, demonstrating the drawbacks to being a business partner that did not play by Microsoft's rules.

Why did it leave that impression? Because of the way negotiations took place. That dispute could have (and should have) been settled with a few phone calls to the right senior executives, or, at most, arbitration.[101] That did not happen in part because Microsoft's executives urgently wanted to keep their browser available in competition with Netscape, and they did not want to give the appearance of ceding even a temporary disadvantage. As was frequently pointed out in public forums, this was but one of several alleged strong-arm tactics that most computer company executives refused to discuss in public for fear of retaliation from Microsoft.[102] Similarly, no senior executive at Microsoft ever apologized, nor disavowed the action, nor did the firm ever give back any of the strategic gains it reaped from the action, which left the impression that the negotiating method was not an accident.

As it happened here, in time the executive team at Microsoft concluded that it had not been sufficiently strict with its business

99. Letter from Don Hardwick, Group Manager, OEM Sales Div., Microsoft Corp., to Celeste Dunn, Vice President, Consumer Software Business Unit, Compaq (June 6, 1996), http://www.usdoj.gov/atr/cases/exhibits/650.pdf.

100. United States v. Microsoft Corp., 84 F. Supp. 2d 9, 59–60 (D.D.C. 1999).

101. Indeed, it was apparently settled through a few phone calls, but only after the threatening letter had been sent, which makes one wonder how much of the public discussion was simply making the best of it by putting lipstick on a pig. Thereafter, the Netscape and Internet Explorer icons appeared on both desktops for a short period, but Compaq renegotiated its contracts with others. *See* Declan McCullagh, *Compaq: It Was All a Big Mix-Up*, WIRED, Feb. 16, 1999, http://www.wired.com/politics/law/news/1999/02/17938; *see also Microsoft*, 84 F. Supp. 2d at 59–60.

102. Rajiv Chandrasekaran, *Jabs at Company Figure into Trial*, WASH. POST, Jan. 27, 1999, at E01, *available at* http://www.washingtonpost.com/wp-srv/business/longterm/microsoft/stories/1999/jabs012799.htm.

partners, which motivated taking further action, accordingly to Judge Jackson's recounting of events.[103] Thereafter, Microsoft inserted clauses into operating system licenses, which included restrictions on OEMs, including restrictions on the "out of the box" experience for users when they first fired-up their systems. Those same contractual restrictions later were used to prevent OEMs from adding help screens for users of the Netscape browser, among other issues that helped Microsoft's competitive aims—albeit, by driving up OEM service expenses.[104]

Why are one-sided negotiations a bad sign for the innovative environment? They reveal one big problem: using their negotiating leverage, managers at the firm doing the dictating can find leeway to justify actions that make their own lives better or easier, even when it comes at the expense of others. That can become a detriment to innovation, especially when one-sided negotiations begin to serve defensive purposes of the dominant firm to the detriment of others. It can restrict the conduct of economic experiments and hinder the realization of competitive benefits from unfettered standards competition.

In the above examples, Microsoft tried to reduce Netscape's ability to distribute its products and made it difficult for users to find alternatives. It appears we can recognize unhealthy negotiations for innovative entrants when established firms prevent distributors from installing help screens for their users through contracting clauses, and when distributors complain about restrictions that limit users' options to modify their products.

To be sure, it is difficult to assess whether one-sided negotiation contributes to a negative outcome in general. While such open questions cannot be resolved entirely in a short essay such as this, consider three key questions as a start for diagnosing any specific example: (1) Does a firm dictating conditions possess market power and employ it in its bargaining behavior? (2) Are non-innovative tactics being employed to shape innovative behavior by others? (3) Are users being restricted for reasons that have any relationship to a product's merits and functionality?[105]

---

103. *See Microsoft*, 84 F. Supp. 2d at 59–62.

104. After the introduction of these restrictions Hewlett Packard sent a letter to Microsoft with the strongly worded lines:

> We must have the ability to decide how our system is presented to our end users. If we had a choice of another supplier, based on your actions in this area, I assure you [that you] would not be our supplier of choice. I strongly urge you to have your executives review your decisions and to change this unacceptable policy.

Id. at 62.

105. *See, e.g.*, Shane Greenstein, Market Structure and Innovation: A Brief Synopsis of Recent Thinking for the Federal Trade Commission, Testimony for the Federal Trade

As illustration, consider the first screen restrictions (on help screens) that Microsoft employed on OEMs. The answer to the above questions would be yes, yes, and no, suggesting they are too one-sided and unhealthy. Not surprisingly, even while the appellate court reviewing Judge Jackson used dozens of pages to admonish his talking to reporters, its members could not bring themselves to alter his ruling about the use of first screen restrictions. That is, these were among the provisions the appellate court cited as violations of antitrust law.[106]

The epilogue to this episode is informative. As it turned out, the bright light of the court's inquiry turned into a partial antiseptic—albeit it was a slow acting one from the viewpoint of those wanting strong action taken against Microsoft.[107] Publicity about these actions had an effect on developers, who have increasingly moved to open source platforms. In 2006, several years after the antitrust trial, Microsoft took public action to counter developer defections.

Microsoft publicly declared that it had adopted a set of principles that bound the firm to remain consistent in its actions over time.[108] This action directly addressed one of the issues that perennially arose in the 1990s—accusations that Microsoft's employees altered APIs or other firm technologies in self-interested ways that discriminated between business partners.[109] This was thought to be a policy that application

---

Commission Hearings on *Competition and Intellectual Property Law in the Knowledge Based Economy,* Washington, D.C. (2002) (transcript available at http://www.ftc.gov/opp/intellect/shanemitchell.pdf); Bresnahan, Greenstein, and Henderson, *supra* note 42; Phil Weiser, *Regulating Interoperability: Lessons from AT&T, Microsoft and Beyond* (Working Paper), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1344828.

106. As further illustration of these questions, consider the negotiating breakdown between Sprint and Cogent. Focusing on Sprint's action, how would the questions come out? No, maybe, no. Indeed, Sprint capitulated to Cogent precisely because it lacked market power, which also is why it was not a situation of one-sided negotiations.

107. Timothy F. Bresnahan, *The Economics of the Microsoft Case* (Working Paper), *available at* http://www.stanford.edu/~tbres/Microsoft/The_Economics_of_The_Microsoft_Case.pdf.

108. Microsoft, Windows Principles: Empowering Choice, Opportunity, and Interoperability, http://www.microsoft.com/About/CorporateCitizenship/US/PromotingInnovation/WindowsPrinciples.mspx; *see also* Weiser, *supra* note 105, at 11.

109. One among the many provision seems particular aimed at these concerns:

> 1. APIs. Microsoft provides the developer community with a broad range of innovative operating system services, through documented application programming interfaces (APIs), for use in developing state-of-the-art applications. The U.S. antitrust ruling requires that Microsoft disclose all of the interfaces internal to Windows called by "middleware" within the operating system, such as the browser, the media player, and so forth. In this way, competitors in these categories will know that they can plug into Windows to get services in the same way that these built-in Windows features do. This has worked well, and Microsoft will continue to disclose these interfaces even after the U.S. antitrust ruling expires. In fact, we will go further, extending our API commitment to the benefit of all software developers. Going forward, Microsoft will ensure that all the interfaces within Windows called

developers would find encouraging, since it relieved concerns about the potential waste of time and effort out of negotiations. In 2008 Microsoft announced another set of principles for remaining consistent in its interoperability designs, and these reinforced the earlier points.[110]

Notice a key subtlety: Microsoft committed to consistency. *Consistent* policies from a firm are those that change slowly at most, allowing for the planning of others. They are changed without caprice, without an ad hoc approach, and without seemingly arbitrary timing, in other words, without actions that necessarily diminish the returns on others' innovative, long-term investments.

Consistency has great importance in interdependent value chains. Other firms will not make long-term investments if they fear not making a return on that investment due to changes by others, which are out of their control. Entrepreneurs will not take action if they fear conditions will change arbitrarily on them later, or systemically to their disadvantage. Firms will not undertake costly economic experiments if they cannot assure themselves that other firms won't interfere with the conditions that support learning from their market experience.

In adopting a commitment to consistency, Microsoft did not give up its rights to retain secrets (e.g., remain less than transparent) nor to give up its right to retain managerial discretion (e.g., exclude participation from outsiders) after engaging with business partners. Instead, Microsoft committed to not arbitrarily alter or apply what was decided unilaterally by management, inviting business partners (i.e., especially developers and OEMs) to inquire whether they receive treatment similar to another partner of Microsoft's (i.e., another developer's competitor).[111]

Will consistency lead to fewer take-it-or-leave-it offers? It depends on one's view. Many books have been written about the managerial preference of Bill Gates, and many of Microsoft's practices arose from his preferences.[112] His recent retirement suggests the firm would display less variance in its contracting activities in any event. In that case, the answer would seem to be yes.

Once again, this viewpoint differs from the standard approach on

---

> by any other Microsoft product, such as the Microsoft Office system or Windows Live, will be disclosed for use by the developer community generally. That means that anything that Microsoft products can do in terms of how they plug into Windows, competing products will be able to do as well.

Id.

110. Microsoft, Interoperability Principles: Open Connections, Standards Support, Data Portability, http://www.microsoft.com/interop/principles/default.mspx; *see also* Weiser, *supra* note 105, at 11–12 .

111. *Id.*

112. *See, e.g.,* BANK, *supra* note 42.

Wall Street. That view typically stresses the profitability of being a leading firm and the benefits to employees working there—up to a point. For example, IBM in the 1970s had a great market position in large enterprise computing. Working there paid better than anything else in computing, albeit every veteran of that era talks about how internal politics consumed the organization.

More recently, working at Microsoft's Windows or Office division or Intel's microprocessor division has had its benefits, since those divisions have been awash in billions. That enabled these companies to fund some rather ambitious internal projects, which was fun for many employees, although more fun for those managers who won the internal debates than those who lost.

The focus on Microsoft's profitability, or on IBM's in the prior era, is simply too narrow a frame for thinking about the role of negotiations in shaping industry-wide innovative activity. The innovative health of many participants requires a broader vision and analysis, looking beyond the consequences of actions benefiting the largest incumbent firm.

In conclusion, the absence of one-sided bargaining is a sign of health. Absence of one-sided negotiation indicates that no firm has largely shaped the actions of others. Such actions do not need to be unhealthy, and can arise for a variety of reasons. But it can be unhealthy when dominant firms face incentives to shape the behavior of innovative entrants and competitors in ways that benefit only the dominant firm. Such actions have the potential to limit innovative behavior. Particularly worrisome is a firm with market power imposing constraints which undermines economic experimentation, standards competition, and entrepreneurial entry.

III. HEALTHY INNOVATIVE COMPETITION FROM DOMINANT FIRMS

If this essay has any broad lessons, they are these two observations: First, it is myopic for policy to cede full discretion over innovation in an evolving value chain to any firm who happens to have market power today, whether it is Microsoft, Comcast, or whomever. Second, it also is myopic to cede full discretion over policy to slow moving regulation and court decisions, particularly when courts do not have reason to consider the range of policies to nurture long term investments by innovative actors in a complex and interrelated value chain.

What is an alternative to such myopic policy? A third way, one that offers clear and predictable policy guidelines, coupled with administrative processes for quick resolution of disputes. This essay has stressed that such a regulatory policy would consist of more than just four sentences. It would stress four signs of innovative health as part of transparent and

consistent guidelines that private actors could use to anticipate policy.

This section will illustrate the steps such reasoning might take, and how it could apply the four principles, but will not develop a full argument. That is, I briefly will consider behavioral analysis consistent with examining the health of an innovative market. For that purpose I will illustrate the issues raised during negotiations between Cogent and Sprint, and those raised by Comcast after it throttled Bit-Torrent traffic.

First, consider Cogent's negotiation breakdowns with other firms, particularly Sprint. The breakdowns generally were short, but had consequences for users nonetheless.

Let's start with the four principles. This breakdown deprived some users of full access to the Internet, violating any strict interpretation of the first of the FCC's principals, i.e., access to the lawful Internet. A strict interpretation of the first of the four principles might require firms to interconnect under all circumstances, removing de-peering as a negotiating tactic. That policy would seem to be motivated by a desire to protect the user experience.

That is problematic, in my view, because this behavioral rule got to the right answer for the wrong reasons, and, thus, sets a policy precedent with little value for others. As pointed out above, it misses several key insights about what those negotiations entailed, and what curtailing them would effect.

Rather, these events raise issues in the competition policy for interconnection between an entrepreneurial firms and an established firm. In my view all parties behaved in ways that did not facilitate a path towards a Coasian solution. Non-participants in decision making were hurt by the actions of the parties involved. Government intervention was, thus, merited.

The four principles fail to draw attention to many of the relevant competition policy issues in this case. For example, de-peering by an incumbent firm could be a tactic in discouraging a new entrant's entrepreneurial behavior, or in discouraging an economic experiment. The key question is: Would the guidelines be implemented differently if they were preserving economic experiments or preventing incumbent firms from discouraging entrepreneurial entry? In my view this example illustrates that the answer is certainly yes.

As a second example, consider Comcast's unilateral declaration to throttle P2P applications on its lines with resets. The FCC eventually intervened, arguing that Comcast could not single out a specific application for such action. Comcast has responded with new proposals for ways to manage its traffic.

As it played out, one striking feature about this event was the willingness of all parties to act without asking for anyone else's

permission. P2P users acted as if they could run any application, irrespective of its consequences for others, even when it degraded the quality of service for neighbors during peak-load time periods. Comcast acted as if it had full discretion to manage its data over its facilities without informing anyone, even its own customers.

Using the analysis above, I would interpret this behavior as symptomatic of the lack of fully specified contracts (for the issues under dispute) between parties who (it would seem) actually have a contract. In other words, the development of P2P applications put both parties in a situation unanticipated by their original contract, which required a renegotiation of its terms.

In my view all parties behaved in ways that did not facilitate building trust between them, and, thus, their behavior departed very far from anything that positively contributed to a path aimed towards a Coasian solution. Once again, government intervention was, thus, merited.

To be sure, the basic economics of incomplete contracting partially favors giving discretion to Comcast's management. Management could internalize the externality one user imposes on others—managing traffic for many users' general benefit. That is, P2P applications, like BitTorrent, can impose large negative externalities on other users, particularly in cable architectures during peak-load time periods. Such externalities can degrade the quality of service to the majority of users without some sort of limitation or restriction.

That does not imply, however, that Comcast has unfettered discretion to manage the situation. There is at least one additional incomplete contract to consider, that between Comcast and other providers of applications presently in the market. Arguably, there is a public policy issue regarding those innovative entrepreneurs who are not in the market at present, but might be in the near term. It would be quite difficult for Comcast and future entrants to reach a Coasian bargain—some of them do not even exist yet! In brief, Comcast's actions also have consequences for long-run innovative incentives by other application providers.

In that sense, Comcast's behavior had many less appealing aspects, such as its lack of transparency, and the lack of participation from others in decision making, as well as its virtually one-sided negotiating stance with all other application providers and lack of clear statements about its own actions,[113] until the FCC intervened.[114] Moreover, the firm's initially

113. ELECTRONIC FRONTIER FOUNDATION, PACKET FORGERY BY ISPS: A REPORT ON THE COMCAST AFFAIR, http://www.eff.org/wp/packet-forgery-isps-report-comcast-affair.

114. Formal Complaint of Free Press and Public Knowledge Against Comcast

inflexible public stance appeared to be aimed at shaping the willingness of others to experiment. Arguably, if it persisted, it would also shape proposals for new applications using new standards, and entrepreneurial initiatives whose businesses depended on Comcast's actions. It is not obviously healthy for innovation to give Comcast's management unchecked discretion to make take-it-or-leave-it offers to providers of any application its management believes harms users.

Here, once again, the FCC's principles fall short of being guidelines. They do not direct attention towards the salient issues in several new areas of application development or an analysis that takes into account the signs of innovative health. An entrepreneur's returns on investments will depend on their ability to transmit data over Comcast's lines. The key question is: does the lack of rules for Comcast's behavior encourage or discourage entrepreneurship in new applications? In my view this example illustrates that the absence of a clear limitation to Comcast's discretion reduces investment incentives for any entrepreneur who anticipates putting their application over Comcast's lines.

Both these examples illustrate one additional aspect in which the FCC principles fall short of being guidelines. Both cases raise general issues that are likely to arise again, perhaps with different participants, perhaps in different locations. In neither case do the FCC principles translate into clear, positive behavioral guidelines for the firms under scrutiny. While there are implications for what not to do, there is scant, positive guidance for what to do.

Imagine the discussion taking place within the management at Comcast and Cogent, or any other firm who wants to learn lessons from watching the actions of these firms. When making their tactical strategies, they surely must be asking: What sort of behavior will generate a positive/negative policy response? By what norms for consistent, transparent, and participatory decision making, if any, will this firm's actions be judged? They had little information from which to forecast policy.

Consistency and transparency are virtues for policy making. If only the FCC elaborated on their meaning in publically available guidelines, these firms could anticipate what potential issues their own actions might trigger. That is the benefit of guidelines that go beyond four sentences. Guidelines remove impediments to anticipating the reaction of policy makers to a firm's actions. That might not improve the quality of decision makers, but almost certainly it will reduce the likelihood of running afoul of well reasoned guidelines, clearly articulated in advance. That has to improve the quality of managerial action.

Corporation for Secretly Degrading Peer-to-Peer Applications, *Memorandum Opinion & Order*, 23 FCC Rcd. 13,028 (Aug. 20, 2008).

Barbara van Schewick
Assistant Professor of Law and (by courtesy) Electrical Engineering
Director, Center for Internet and Society
Stanford Law School

Opening Statement at the
Federal Communications Commission's Workshop on
Innovation, Investment and the Open Internet
in Cambridge, MA on January 13, 2010,
WC Docket No. 07-52, GN Docket No. 09-191

Good evening. My name is Barbara van Schewick. I'm an assistant professor at Stanford Law School and direct the Center for Internet and Society there. I also have a courtesy appointment at Stanford's electrical engineering department. I have a law degree and a PhD in computer science. For the past nine years, my research has focused on the relationship between Internet architecture, innovation and regulation. My book "Internet Architecture and Innovation" will be published by MIT Press this spring.

The Internet has created an enormous amount of application innovation. I'm here today to explain which factors made this possible.

I'll do this through three stories. Here is the first:

In the mid-nineties, a software engineer in Silicon Valley named Pierre Omidyar thought: "Wouldn't it be nice if we could buy and sell stuff over the Internet, using auctions to determine the price?" Most of his friends thought he was crazy: "Strangers will never buy from strangers online," they said. But Pierre Omidyar didn't care about that. He stayed home over Labor Day weekend 1995, wrote his software and put it online. Nine months later, so many people were using the platform that he decided to quit his day job and focus on it fulltime. The platform, renamed eBay, became a huge success, and today, more than 88 million people worldwide are using it to buy or sell things over the Internet.

Three aspects are important here:

Pierre Omidyar had an idea for an application. People thought it was crazy, but it didn't matter. In particular, it didn't matter whether network providers believed that "strangers will ever buy from strangers online." And that's because on the Internet, network providers don't have to do anything to enable new applications to run. This is a consequence of the Internet's architecture. The original Internet was based on a design principle called the end-to-end arguments. This design principle was first described by Jerome Saltzer, David Reed and David Clark who are here today. Following this design principle, the network was designed to be as general as possible in order to support a wide variety of applications with different needs. So when a new application comes along, the network doesn't have to be changed to allow the application to run. All the innovator has to do is write a program that runs on a computer attached to the Internet.

As a result, an innovator does not have to convince network providers that her application is useful, or will be commercially successful. The only person who needs to be convinced that this is a good idea is the innovator herself. This greatly increases the chance that innovative ideas will be realized.

Second: When the application has been written, the network does not need to be changed before the application can run on the network. If you want to use it, you install it on your computer. That's it. The only person who needs to be convinced that this application may be useful is the person who actually wants to use it. This greatly increases the chances that people can actually use the new application.

Third: In this architecture, it doesn't cost a lot to develop new applications. You need access to a computer, be able to program, and time to actually write the program. This greatly increases the number and type of people who can develop new applications. Like Pierre Omidyar, you don't have to be an employee of a firm or have outside funding to realize your idea for an application. Because the biggest investment is often the design and programming of the application itself, innovators can develop an application in their free time or as a side project. Under these conditions, an application does not have to produce a profit in the future to cover the costs of developing it. Instead, a wide range of benefits may be sufficient to cover the development costs.

Thus, the architecture allows innovators with a wide range of motivations and funding models to develop applications.

So, three aspects:

Innovators independently decide whether to realize innovative ideas. They do not need support, or "permission" from network providers in order to innovate.

Users independently decide which applications they want to use.

The low costs of application innovation enable a very large and diverse group of people to develop new applications.

That was the first story. Here is the second:

In 2002, two European entrepreneurs named Niklas Zennström and Janus Friis thought: "Wouldn't it be nice if we could use peer-to-peer software to make phone calls over the Internet?" At the time, most network engineers didn't think this was possible. They thought that Internet telephony would require special treatment from the network (something we call "Quality of Service"). Network providers weren't really interested in pursuing the technology because it was a huge threat to their business model. But Zennström and Friis didn't care about all this. They went ahead, developed their software, the software became Skype, and today, more than 500 million people worldwide are using Skype to place phone calls over the Internet.

Again, we have two entrepreneurs who had an innovative idea for an application. Network engineers didn't think it would work, but it didn't matter. Nothing new so far.

The application constituted a huge threat to network providers' business models, but it didn't matter. And for Zennström and Friis, it didn't matter because there was nothing network providers could do about it. And there was nothing network providers could do about this, because the Internet's architecture prevented them from interfering with the applications and content on their networks. As I said, the Internet was based on the end-to-end arguments. As a consequence of this design, the network couldn't distinguish between different applications and content (it was "application-blind"), and as a result, network providers couldn't control the applications and content on their network.

Today, that's different. Today, sophisticated technology is available that enables network providers to identify the applications and content on their network and control their execution.

Thus, the original Internet was application-blind, today's Internet is not. Does it matter?

Imagine you have this great idea for a video platform that will revolutionize the way people watch TV. Once they have used your application, they will never want to go back to cable again. Of course there are risks. The technology may not work. Users may not like your product. Your business model may be wrong. But in the application-blind network, you know that you will get a fair chance in the market place. You will be able to compete with other applications on the merits.

In today's network, cable providers may squash you. The network can turn against you any time and block your application or slow it down. There are many reasons why network providers may want to do so. Maybe your application competes with theirs; maybe they just want a share of your profits. Maybe they don't like your content, or your application is slowed down to manage bandwidth. Whatever the network provider's reason, if your application gets blocked, your project fails, and you won't be able to reap its benefits. And accounting for this possibility, you (or potential investors) may decide not to pursue your idea.

Third story:

When YouTube came on the market, it competed with Google Video. You Tube was better. It won.

In today's Internet, things might have been different. In an application-aware network, the network provider can ask applications to pay an access fee. There are many ways in which it could do so, and all of them would be bad for application innovation.

Let's focus on one possibility: When YouTube came on the market, network providers might have said: "Google, you are big. You have lots of money. Why don't you give us some of this money, and we will give Google Video better transport." Imagine Google pays. Suddenly, Google Video is so much better. Not because it's the better product, but because Google is rich, and Google was able to strike a deal. In such a world, network providers get to decide who is

successful, by deciding who gets a deal. Suddenly network providers, not users, get to pick winners and losers on the Internet.

Three stories, different factors. How do changes in these factors affect the amount and type of application innovation?

Some changes may affect the benefits and costs of innovation. An innovator decides to innovate if the benefits (broadly defined) are larger than the costs. Increase the costs or reduce the expected benefits (for example, through access charges or discrimination), and some innovations may not be justified any more.

Some changes may affect the size or diversity of the innovator pool. Others may let network providers, not users, choose which applications will be successful and how the network can be used. For example, access charges may reduce the profits of all affected application developers, but they may hit certain types of innovators (for example, those with no or little outside funding) particularly hard.

Why are these things important?

If there is uncertainty (e.g., about technology or user needs) or if user needs are very heterogeneous, a larger and more diverse group of innovators will produce more and better applications than a smaller, less diverse group of innovators, and that innovation will better meets user needs.

What's the intuition here?

If there is uncertainty, nobody really knows in advance which applications will work, or which applications will be successful. Under these circumstances, economic theory suggests that it is best to try out many different ideas, and see what happens. Some will succeed, some will fail, but trying is the only way to find out. And because different people will have different ideas and different views of the world, more and more diverse people will have more and more diverse ideas. Tim Berners-Lee looked at the Internet and saw a giant web of shared information; Pierre Omidyar saw an online marketplace, and Jeff Bezos saw an online bookstore. A larger and more diverse group of potential innovators will also realize more of the ideas that are known. For

example, start-ups may have an incentive to realize ideas that established firms wouldn't pursue. Users have an incentive to create applications that manufacturers won't produce.

By contrast, fewer innovators, or less diverse innovators, will try fewer things, leaving valuable ideas on the table.

But widespread experimentation is only part of the mechanism that produces innovation under uncertainty. The second is: Who gets to decide which applications become successful? Users or network providers? Does it make a difference?

I argue it does – because users and network providers will choose different applications. There are two reasons for this:

First, users and network providers use different criteria when choosing which applications will be successful. Users choose the applications that best meet their needs. That's easy. Network providers may use different criteria: "Does this application compete with my own application? Does it create a lot of bandwidth? Does my preferred vendor offer network management tools that happen to block this application?" Consider Skype. Many mobile providers in Europe do not allow their users to use Skype over the mobile Internet. If you look at user forums, you will see that users don't like this. They want to use Skype on their cell phone. But if users use Skype, they don't make as many traditional cell phone calls, and voice revenue shrinks. So network providers make a decision that's different from what users would choose.

And second, even in those cases where network providers would like to choose the applications that users want, they don't necessarily know what that is. That's the uncertainty I talked about earlier. In many cases, nobody knows whether an application will be successful until users actually try it. Network providers cannot replace this.

Beyond innovation, user choice is also important if we want the Internet to provide the maximum value for society – but that's another story.

Thus, if network providers pick winners and losers on the Internet, if they decide how users can use the network, users may end up with applications that they would not have chosen, and may be forced to use the Internet in a way that does not create the value it could.

In sum, there are a number of aspects that foster application innovation:

Innovators independently choose which applications they want to pursue; users independently choose which applications they want to use. The application-blindness of the network ensures that the network provider cannot interfere with these choices, that it cannot distort competition among applications or reduce application developers' profits through access charges. Finally, the low costs of innovation not only make many more applications worth pursuing, but also allow a large and diverse group of people to become innovators, which in turn increases the overall amount and quality of innovation.

But why do we care so much about application innovation? Why should policy makers care about it?

I have a longer answer to this question. It explains how application innovation contributes to economic growth and how it creates value for society in all areas of society. But my time is almost up. Therefore, let me just say this:

Did you ever try to explain to your partner's grandmother why she should get the Internet? I did. Although I'm a computer scientist, I didn't say: "Grandma, you have to get the Internet! It's so cool! It lets you send packets back and forth." No, I said: "If you get the Internet, you can call us and see your grandchildren on the screen. And if we have new pictures, you'll be able to see them immediately after we send them. And you can read about everything you can possibly imagine …"

Thus, the Internet does not create value through its existence alone. It creates value by enabling us to do the things we want to do, do things we never knew we wanted to do, or do things more efficiently. Applications are the tools that let us realize this value, in all areas of society. And by protecting the factors that have fostered application innovation in the past, we can make sure the Internet will be even more useful and valuable in the future.

Thank you for your attention.

RAND Journal of Economics
Vol. 37, No. 3, Autumn 2006
pp. 668–691

Competition in two-sided markets

Mark Armstrong*

Many markets involve two groups of agents who interact via "platforms," where one group's benefit from joining a platform depends on the size of the other group that joins the platform. I present three models of such markets: a monopoly platform; a model of competing platforms where agents join a single platform; and a model of "competitive bottlenecks" where one group joins all platforms. The determinants of equilibrium prices are (i) the magnitude of the cross-group externalities, (ii) whether fees are levied on a lump-sum or per-transaction basis, and (iii) whether agents join one platform or several platforms.

1. Introduction

■ There are many examples of markets in which two or more groups of agents interact via intermediaries or "platforms." Surplus is created—or destroyed in the case of negative externalities—when the groups interact. Of course, there are countless examples where firms compete to deal with two or more groups. Any firm is likely to do better if its products appeal to both men and women, for instance. However, in a set of interesting cases, cross-group externalities are present, and the benefit enjoyed by a member of one group depends upon how well the platform does in attracting custom from the other group. For instance, a heterosexual dating agency or nightclub can do well only if it succeeds in attracting business from both men and women. This article is about such markets. A brief list of other such markets includes: credit cards (for a given set of charges, a consumer is more likely to use a credit card that is accepted widely by retailers, while a retailer is more likely to accept a card that is carried by more consumers); television channels (where viewers typically prefer to watch a channel with fewer commercials, while an advertiser is prepared to pay more to place a commercial on a channel with more viewers); and shopping malls (where a consumer is more likely to visit a mall with a greater range of retailers, while a retailer is willing to pay more to locate in a mall with a greater number of consumers passing through). See Rochet and Tirole (2003) for further examples of two-sided markets.

As I shall argue in more detail, there are three main factors that determine the structure of prices offered to the two groups.

□ **Relative size of cross-group externalities.** If a member of group 1 exerts a large positive externality on each member of group 2, then group 1 will be targeted aggressively by platforms. In broad terms, and especially in competitive markets, it is group 1's benefit to the other group

* University College London; mark.armstrong@ucl.ac.uk.

The first version of this article was presented at the Econometric Society European Meeting held in Venice, August 2002. I am grateful to the Editor and two referees, to the audiences at many seminar presentations, to Simon Anderson, Carli Coetzee, Jacques Crémer, Jean-Charles Rochet, Jean Tirole, Xavier Vives, and especially to Julian Wright for discussions and corrections. The support of the Economic and Social Reaserch Council (UK) is gratefully acknowledged.

 Copyright © 2006, RAND.

that determines group 1's price, not how much group 1 benefits from the presence of group 2 (see Proposition 2 below). In a nightclub, if men gain more from interacting with women than vice versa, then we expect there to be a tendency for nightclubs to offer lower entry fees to women than to men.

Unless they act to tip the industry to monopoly, positive cross-group externalities act to intensify competition and reduce platform profit—see expression (13) below. To be able to compete effectively on one side of the market, a platform needs to perform well on the other side (and vice versa). This creates a downward pressure on the prices offered to both sides compared to the case where no externalities exist. This implies that platforms have an incentive to find ways to mitigate network effects. One method of doing this is discussed next.

□ **Fixed fees or per-transaction charges.** Platforms might charge for their services on a lump-sum basis, so that an agent's payment does not explicitly depend on how well the platform performs on the other side of the market. Alternatively, if feasible, the payment might be an explicit function of the platform's performance on the other side. One example of this latter practice occurs when a television channel or a newspaper makes its advertising charge an increasing function of the audience or readership it obtains. Similarly, a credit card network levies (most of) its charges on a per-transaction basis, and the bulk of a real estate agent's fees are levied only in the event of a sale. The crucial difference between the two charging bases is that cross-group externalities are weaker with per-transaction charges, since a fraction of the benefit of interacting with an extra agent on the other side is eroded by the extra payment incurred. If an agent pays a platform only in the event of a successful interaction, the agent does not need to worry about how well the platform does in its dealings with the other side. That is, to attract one side of the market, it is not so important that the platform first gets the other side "on board." Because externalities are lessened with per-transaction charging, it is plausible that platform profit is higher when this form of charging is used.[1] (See Propositions 3 and 5 for illustrations of this effect.) Finally, the distinction between the two forms of tariff only matters when there are competing platforms. When there is a monopoly platform (see Section 3), it makes no difference if tariffs are levied on a lump-sum or per-transaction basis.

□ **Single-homing or multi-homing.** When an agent chooses to use only one platform, it has become common to say the agent is "single-homing." When an agent uses several platforms, she is said to "multi-home." It makes a significant difference to outcomes whether groups single-home or multi-home. In broad terms, there are three cases to consider: (i) both groups single-home, (ii) one group single-homes while the other multi-homes, and (iii) both groups multi-home. If interacting with the other side is the primary reason for an agent to join a platform, then we might not expect case (iii) to be very common—if each member of group 2 joins all platforms, there is no need for any member of group 1 to join more than one platform—and so I do not analyze this configuration. (If all native French speakers also speak English, there is less incentive for a native English speaker to learn French.) Configuration (i) is discussed in Section 4. Although the analysis of that case provides useful insights about two-sided markets, it is hard to think of many actual markets that fit this configuration precisely.

By contrast, there are several important markets that resemble configuration (ii), and in Section 5 these are termed "competitive bottlenecks." Here, if it wishes to interact with an agent on the single-homing side, the multi-homing side has no choice but to deal with that agent's chosen platform. Thus, platforms have monopoly power over providing access to their single-homing customers for the multi-homing side. This monopoly power naturally leads to high prices being charged to the multi-homing side, and there will be too few agents on this side being served

[1] An exception to this occurs when the market tips to monopoly. Here, an incumbent's profits typically increase with the importance of network effects, since entrants find it hard to gain a toehold even when the incumbent sets high prices. This explains one conclusion of Caillaud and Jullien (2003), which is that equilibrium profit rises when platforms cannot use transaction charges.

© RAND 2006.

from a social point of view (Proposition 4).[2] By contrast, platforms do have to compete for the single-homing agents, and high profits generated from the multi-homing side are to a large extent passed on to the single-homing side in the form of low prices (or even zero prices).

2. Related literature

■ I discuss some of the related literature later as it becomes most relevant in the analysis (especially in Section 5). However, it is useful to discuss two pioneering articles up front.

Caillaud and Jullien (2003) discuss the case of competing matchmakers, such as dating agencies, real estate agents, and internet "business-to-business" websites. (See also van Raalte and Webers (1998).) There is potentially a rich set of contracting possibilities. For instance, a platform might have a subscription charge in combination with a charge in the event of a successful match. In addition, Caillaud and Jullien allow platforms to set negative subscription charges and to make their profit from taxing transactions on the platform. Caillaud and Jullien first examine the case where all agents must single-home. (I provide a parallel analysis in Section 4.) In this case, there is essentially perfect competition, and agents have no intrinsic preference for one platform over another except insofar as one platform has more agents from the other side or charges lower prices. Therefore, the efficient outcome is for all agents to use the same platform. Caillaud and Jullien's Proposition 1 shows that the only equilibria in this case involve one platform attracting all agents (as is efficient) and that platform making no profit. The equilibrium structure of prices involves negative subscription fees and maximal transaction charges, since this is the most profitable way to prevent entry. Caillaud and Jullien go on to analyze the more complicated case where agents can multi-home. They analyze several possibilities, but the cases most relevant for my article are what they term "mixed equilibria" (see their Propositions 8 and 11). These correspond to my competitive bottleneck situations, and they involve one side multi-homing and the other side single-homing. They find that the single-homing side is treated favorably (indeed, its price is necessarily no higher than its cost), while the multi-homing side has all its surplus extracted. I discuss the relationship between the two approaches in more detail in Section 5.

Another closely related article is Rochet and Tirole (2003). The flavor of their analysis can be understood in the context of the credit card market (although the analysis applies more widely). On one side of the market are consumers, on the other side is the set of retailers, and facilitating the interaction between these two groups are two competing credit card networks. For much of Rochet and Tirole's analysis, the credit card platforms levy charges purely on a per-transaction basis, and there are no lump-sum fees for either side. Suppose that one credit card offers a lower transaction fee to retailers than its rival. A retailer choosing between accepting just the cheaper card or accepting both cards faces a tradeoff. If it accepts just the cheaper card, then its consumers have a stark choice between paying by this card or not using a card at all. Alternatively, if the retailer accepts both cards, then (i) more consumers will choose to pay by some card but (ii) fewer consumers will use the retailer's preferred lower-cost card. If a credit card reduces its charge to retailers relative to its rival, this will "steer" some retailers that previously accepted both cards to accept only the lower-cost card. In a symmetric equilibrium, all retailers accept both credit cards (or neither), while consumers always use their preferred credit card. The share of the charges that are borne by the two sides depends on how closely consumers view the two cards as substitutes. If few consumers switch cards in response to a price cut on their side, then consumers should pay a large share of the total transaction charge; if consumers view the cards as close substitutes, then retailers will bear most of the charges in equilibrium. Rochet and Tirole also consider the case where there are fixed fees as well as per-transaction fees, under the assumption that consumers use a single card. This is essentially the same model as my competitive bottleneck model, and I discuss this part of their article in more detail in Section 5.

[2] This tendency toward high prices for the multi-homing side is tempered when the single-homing side benefits from having many agents from the other side on their platform. Then high prices to the multi-homing side will drive away that side and disadvantage the platform when it tries to attract the single-homing side.

© RAND 2006.

There are a number of modelling differences between my article and Rochet and Tirole (2003) that concern the specification of agents' utility, the structure of platforms' fees, and the structure of platforms' costs.[3] In both articles, agent j has gross utility from using platform i of the form

$$u_j^i = \alpha_j^i n^i + \zeta_j^i.$$

Here, n^i is the number of agents from the other side who are present on platform i, α_j^i is the benefit that agent j enjoys from interacting with each agent on the other side, and ζ_j^i is the fixed benefit the agent obtains from using that platform. Rochet and Tirole assume that ζ_j^i does not depend on i or j (and can be set equal to zero), but that α_j^i varies both with agent j and platform i. In Sections 3 and 4, by contrast, I assume that α_j^i does not depend on i or j but only on which side of the market the agent is on, while ζ_j^i depends on the agent and on the platform. (In Section 5, I suppose that the interaction term α for one side does vary across agents within a group.) The decision whether to make agents' heterogeneity to do with the interaction term α or the fixed benefit ζ has major implications for the structure of prices to the two sides in equilibrium. For instance, with a monopoly platform, the formulas for profit-maximizing prices look very different in the two articles. Moreover, when α_j^i depends on the platform i, an agent cares about which platform the transaction takes place on (if there is a choice): this effect plays a major role in Rochet and Tirole's analysis but is absent here.[4]

Turning to the structure of the platforms' fees, for the most part Rochet and Tirole assume that agents pay a per-transaction fee for each agent on the platform from the other side. If this fee is denoted γ^i, then agent j's net utility on platform i is $u_j^i = (\alpha_j^i - \gamma^i)n^i$ (when ζ is set equal to zero). This confirms the discussion in Section 1 that per-transaction charges act to reduce the size of network effects. In the monopoly platform case, an agent's incentive to join the platform does not depend on the platform's performance on the other side, and she will join if and only if $\alpha_j^i \geq \gamma^i$. The present article, especially in Section 4, assumes that platform charges are levied as a lump-sum fee, say p^i, in which case the agent's net utility is $u_j^i = \alpha n^i + \zeta_j^i - p^i$. The final modelling difference between the two articles is with the specification of costs: Rochet and Tirole assume mainly that a platform's costs are incurred on a per-transaction basis, so that if a platform has n_1 group-1 agents and n_2 group-2 agents, its total cost is cn_1n_2 for some per-transaction cost c. In the current paper, costs are often modelled as being incurred when agents join a platform, so that a platform's total cost is $f_1n_1 + f_2n_2$ for some per-agent costs f_1 and f_2. Which assumptions concerning tariffs and costs best reflect reality depends on the context. Rochet and Tirole's model is well suited to the credit card context, for instance, whereas the assumptions here are intended to apply to markets such as nightclubs, shopping malls, and newspapers.

3. Monopoly platform

■ This section presents the analysis for a monopoly platform. This framework does not apply to most of the examples of two-sided markets that come to mind, although there are a few applications. For instance, yellow pages directories are often a monopoly of the incumbent telephone company, shopping malls or nightclubs are sometimes far enough away from others that the monopoly paradigm might be appropriate, and sometimes there is only one newspaper or magazine in the relevant market.

Suppose there are two groups of agents, denoted 1 and 2. A member of one group cares about

[3] The assumptions in Caillaud and Jullien (2003) to do with utility and costs are closer to mine than to those of Rochet and Tirole. Caillaud and Jullien do not have any intrinsic product differentiation between the platforms. However, there is a benefit to join two platforms rather than one, since they assume that there is a better chance of a match between buyers and sellers when two platforms are involved.

[4] A recent article that encompasses these two approaches with a monopoly platform is Rochet and Tirole (2006), where simultaneous heterogeneity in both α and ζ is allowed. However, a full analysis of this case is technically challenging in the case of competing platforms.

© RAND 2006.

the number of the other group who use the platform. (For simplicity, I ignore the possibility that agents care also about the number from the same group who join the platform.) Suppose the utility of an agent is determined in the following way: if the platform attracts n_1 and n_2 members of the two groups, the utilities of a group-1 agent and a group-2 agent are respectively

$$u_1 = \alpha_1 n_2 - p_1; \quad u_2 = \alpha_2 n_1 - p_2, \tag{1}$$

where p_1 and p_2 are the platform's prices to the two groups. The parameter α_1 measures the benefit a group-1 agent enjoys from interacting with each group-2 agent, and α_2 measures the benefit a group-2 agent obtains from each group-1 agent. Expression (1) describes how utilities are determined as a function of the numbers of agents who participate. To close the demand model, I specify the numbers who participate as a function of the utilities: if the utilities offered to the two groups are u_1 and u_2, suppose the numbers of each group who join the platform are

$$n_1 = \phi_1(u_1); \quad n_2 = \phi_2(u_2)$$

for some increasing functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$.

Turning to the cost side, suppose the platform incurs a per-agent cost f_1 for serving group 1 and per-agent cost f_2 for group 2. Therefore, the firm's profit is $\pi = n_1(p_1 - f_1) + n_2(p_2 - f_2)$. If we consider the platform to be offering utilities $\{u_1, u_2\}$ rather than prices $\{p_1, p_2\}$, then the implicit price for group 1 is $p_1 = \alpha_1 n_2 - u_1$ (and similarly for group 2). Therefore, expressed in terms of utilities, the platform's profit is

$$\pi(u_1, u_2) = \phi_1(u_1)\left[\alpha_1\phi_2(u_2) - u_1 - f_1\right] + \phi_2(u_2)\left[\alpha_2\phi_1(u_1) - u_2 - f_2\right]. \tag{2}$$

Let the aggregate consumer surplus of group $i = 1, 2$ be $v_i(u_i)$, where $v_i(\cdot)$ satisfies the envelope condition $v_i'(u_i) \equiv \phi_i(u_i)$. Then welfare, as measured by the unweighted sum of profit and consumer surplus, is

$$w = \pi(u_1, u_2) + v_1(u_1) + v_2(u_2).$$

It is easily verified that the welfare-maximizing outcome has the utilities satisfying

$$u_1 = (\alpha_1 + \alpha_2)n_2 - f_1; \quad u_2 = (\alpha_1 + \alpha_2)n_1 - f_2.$$

From expression (1), the socially optimal prices satisfy

$$p_1 = f_1 - \alpha_2 n_2; \quad p_2 = f_2 - \alpha_1 n_1.$$

As one would expect, the optimal price for group 1, say, equals the cost of supplying service to a type-1 agent adjusted downward by the external benefit that an extra group-1 agent brings to the group-2 agents on the platform. (There are n_2 group-2 agents on the platform, and each one benefits by α_2 when an extra group-1 agent joins.) In particular, prices should be below cost if $\alpha_1, \alpha_2 > 0$.

From expression (2), the profit-maximizing prices satisfy

$$p_1 = f_1 - \alpha_2 n_2 + \frac{\phi_1(u_1)}{\phi_1'(u_1)}; \quad p_2 = f_2 - \alpha_1 n_1 + \frac{\phi_2(u_2)}{\phi_2'(u_2)}. \tag{3}$$

Thus, the profit-maximizing price for group 1, say, is equal to the cost of providing service (f_1), adjusted downward by the external benefit to group 2 ($\alpha_2 n_2$), and adjusted upward by a factor related to the elasticity of the group's participation. These profit-maximizing prices can be obtained in the familiar form of Lerner indices and elasticities, as recorded in the following result.

© RAND 2006.

Proposition 1. Write

$$\eta_1(p_1 \mid n_2) = \frac{p_1\phi_1'(\alpha_1 n_2 - p_1)}{\phi_1(\alpha_1 n_2 - p_1)}; \quad \eta_2(p_2 \mid n_1) = \frac{p_2\phi_2'(\alpha_2 n_1 - p_2)}{\phi_2(\alpha_2 n_1 - p_2)}$$

for a group's price elasticity of demand for a given level of participation by the other group. Then the profit-maximizing pair of prices satisfy

$$\frac{p_1 - (f_1 - \alpha_2 n_2)}{p_1} = \frac{1}{\eta_1(p_1 \mid n_2)}; \quad \frac{p_2 - (f_2 - \alpha_1 n_1)}{p_2} = \frac{1}{\eta_2(p_2 \mid n_1)}. \tag{4}$$

It is possible that the profit-maximizing outcome involves group 1, say, being offered a subsidized service, i.e., $p_1 < f_1$. From (4), this occurs if the group's elasticity of demand is high and/or the external benefit enjoyed by group 2 is large. Indeed, the subsidy might be so large that the price is negative (or zero, if negative prices are not feasible). This analysis applies, in a stylized way, to a market with a monopoly yellow pages directory. Such directories typically are distributed for free, and profits are made solely from charges to advertisers. The analysis might also apply to software markets in which one type of software is required to create files in a certain format and another type is required to read such files. Often, the reading software is supplied for free, while the writing software needs to be paid for. (For the analysis to apply accurately, though, there need to be two disjoint groups of agents: those who wish to read files and those who wish to create files. It does not readily apply when most people wish to perform both tasks.)

4. Two-sided single-homing

■ This model involves competing platforms but assumes for exogenous reasons that each agent chooses to join a single platform.

□ **Basic model.** The model extends the monopoly platform model in a natural way. There are two groups of agents, 1 and 2, and there are two platforms, A and B, which enable the two groups to interact. Groups 1 and 2 obtain the respective utilities $\{u_1^i, u_2^i\}$ if they join platform i. These utilities $\{u_1^i, u_2^i\}$ are determined in a similar manner to the monopoly model expressed in (1): if platform i attracts n_1^i and n_2^i members of the two groups, the utilities on this platform are

$$u_1^i = \alpha_1 n_2^i - p_1^i; \quad u_2^i = \alpha_2 n_1^i - p_2^i, \tag{5}$$

where $\{p_1^i, p_2^i\}$ are the respective prices charged by the platform to the two groups.

When group 1 is offered a choice of utilities u_1^A and u_1^B from the two platforms and group 2 is offered the choice u_2^A and u_2^B, suppose the number of each group who join platform i is given by the Hotelling specification

$$n_1^i = \frac{1}{2} + \frac{u_1^i - u_1^j}{2t_1}; \quad n_2^i = \frac{1}{2} + \frac{u_2^i - u_2^j}{2t_2}. \tag{6}$$

Here, agents in a group are assumed to be uniformly located along a unit interval with the two platforms located at the two endpoints, and $t_1, t_2 > 0$ are the product differentiation (or transport cost) parameters for the two groups that describe the competitiveness of the two sides of the market.

Putting (6) together with (5), and using the fact that $n_1^j = 1 - n_1^i$, gives the following implicit expressions for market shares:

$$n_1^i = \frac{1}{2} + \frac{\alpha_1(2n_2^i - 1) - (p_1^i - p_1^j)}{2t_1}; \quad n_2^i = \frac{1}{2} + \frac{\alpha_2(2n_1^i - 1) - (p_2^i - p_2^j)}{2t_2}. \tag{7}$$

Keeping its group-2 price fixed, expression (7) shows that an extra group-1 agent on a platform attracts a further α_2/t_2 group-2 agents to that platform.

© RAND 2006.

To focus on market-sharing equilibria, suppose the network externality parameters $\{\alpha_1, \alpha_2\}$ are small compared to the differentiation parameters $\{t_1, t_2\}$. (If network effects were large compared to brand preferences, then there could be equilibria only where one platform corners both sides of the market.) The necessary and sufficient condition for a market-sharing equilibrium to exist is

$$4t_1t_2 > (\alpha_1 + \alpha_2)^2, \tag{8}$$

and this inequality is assumed to hold throughout the following analysis.

Suppose platforms A and B offer the respective price pairs (p_1^A, p_2^A) and (p_1^B, p_2^B). Given these prices, solving the simultaneous equations (7) implies that market shares are

$$n_1^i = \frac{1}{2} + \frac{1}{2}\frac{\alpha_1(p_2^j - p_2^i) + t_2(p_1^j - p_1^i)}{t_1t_2 - \alpha_1\alpha_2}; \quad n_2^i = \frac{1}{2} + \frac{1}{2}\frac{\alpha_2(p_1^j - p_1^i) + t_1(p_2^j - p_2^i)}{t_1t_2 - \alpha_1\alpha_2}. \tag{9}$$

(Assumption (8) implies that the denominators $t_1t_2 - \alpha_1\alpha_2$ are positive.) As one would expect, if $\alpha_1, \alpha_2 > 0$, demand by the two groups is complementary, in the sense that a platform's market share for one group is decreasing in its price for the other group.

As with the monopoly model, suppose each platform has a per-agent cost f_1 for serving group 1 and f_2 for serving group 2. Therefore, platform i's profit is

$$(p_1^i - f_1)\left[\frac{1}{2} + \frac{1}{2}\frac{\alpha_1(p_2^j - p_2^i) + t_2(p_1^j - p_1^i)}{t_1t_2 - \alpha_1\alpha_2}\right] + (p_2^i - f_2)\left[\frac{1}{2} + \frac{1}{2}\frac{\alpha_2(p_1^j - p_1^i) + t_1(p_2^j - p_2^i)}{t_1t_2 - \alpha_1\alpha_2}\right].$$

This expression is quadratic in platform i's prices, and it is concave in these prices if and only if (8) holds. Therefore, platform i's best response to j's prices is characterized by the first-order conditions. Given (8), one can check that there are no asymmetric equilibria. For the case of a symmetric equilibrium where each platform offers the same price pair (p_1, p_2), the first-order conditions for equilibrium prices are

$$p_1 = f_1 + t_1 - \frac{\alpha_2}{t_2}(\alpha_1 + p_2 - f_2); \quad p_2 = f_2 + t_2 - \frac{\alpha_1}{t_1}(\alpha_2 + p_1 - f_1). \tag{10}$$

Expressions (10) can be interpreted in the following manner. First, note that in a Hotelling model without cross-group externalities, the equilibrium price for group 1 would be $p_1 = f_1 + t_1$. In this two-sided setting, the price is adjusted downward by the factor $(\alpha_2/t_2)(\alpha_1 + p_2 - f_2)$. This adjustment factor can be decomposed into two parts. The term $(\alpha_1 + p_2 - f_2)$ represents the external benefit to a platform when an additional group-2 agent is attracted. To see this, note first that the platform makes profit $(p_2 - f_2)$ from an extra group-2 agent. Second, α_1 measures the extra revenue the platform can extract from its group-1 agents (without losing market share) when it has an extra group-2 agent. (An extra group-2 agent means the utility of a group-1 agent on the platform increases by α_1, while the utility of the agent on the rival platform falls by α_1. Therefore, the relative utility for group-1 agents on the platform increases by $2\alpha_1$. Since in equilibrium, a platform has half the group-1 agents, the extra revenue it can extract from its group-1 agents is α_1.) Thus $(\alpha_1 + p_2 - f_2)$ indeed represents the external benefit to a platform of attracting a marginal group-2 agent. Finally, as shown in (7), a platform attracts α_2/t_2 extra group-2 agents when it has an extra group-1 agent. In sum, the adjustment factor $\frac{\alpha_2}{t_2}(\alpha_1 + p_2 - f_2)$ measures the external benefit to the platform from attracting an extra group-1 agent; in other words, it measures the opportunity cost of raising the group-1 price by enough to cause one group-1 agent to leave. This discussion is summarized by an annotated version of formula (10):

$$p_1 = \underbrace{f_1}_{\text{cost}} + \underbrace{t_1}_{\text{market power}} - \underbrace{(\alpha_2/t_2)}_{\text{extra group-2 agents}} \times \underbrace{(\alpha_1 + p_2 - f_2).}_{\text{profit from an extra group-2 agent}} \tag{11}$$

© RAND 2006.

Finally, solving the simultaneous equations in (10) implies that $p_1 = f_1 + t_1 - \alpha_2$ and $p_2 = f_2 + t_2 - \alpha_1$. This discussion is summarized in the next proposition.

Proposition 2. Suppose (8) holds. Then the model with two-sided single-homing has a unique equilibrium that is symmetric. Equilibrium prices for group 1 and group 2 are given respectively by

$$p_1 = f_1 + t_1 - \alpha_2; \quad p_2 = f_2 + t_2 - \alpha_1. \tag{12}$$

Thus, a platform will target one group more aggressively than the other if that group is (i) on the more competitive side of the market and/or (ii) causes larger benefits to the other group than vice versa.[5]

While the expressions (12) are certainly "simple," they are not intuitive, which is why the discussion focused on (10). The fact that, say, group 1's price does not depend on its own externality parameter α_1 is surely an artifact of the Hotelling specification for consumer demand. The assumption that the total size of each group is fixed, so when platforms set low prices they steal business only from the rival rather than expand the overall market, greatly simplifies the analysis. While the solution as verbally described in equation (11) seems likely to hold more widely, the neat formulas in (12) will not. One disadvantage of using a framework with fixed group sizes, however, is that there is no scope for meaningful welfare analysis because prices are simply transfers between agents: any (symmetric) pair of prices chosen by platforms will yield the same total surplus.

It is useful to compare the competitive formulas (12) with the monopoly formulas (4). It turns out there is an extra effect with competition. From (7), a platform's own-price elasticity of demand given fixed and equal market share for the other group is $\eta_1 = p_1/t_1$ and $\eta_2 = p_2/t_2$ for group 1 and group 2 respectively. Thus, expression (12) may be rewritten as

$$\frac{p_1 - (f_1 - 2\alpha_2 n_2)}{p_1} = \frac{1}{\eta_1}; \quad \frac{p_2 - (f_2 - 2\alpha_1 n_1)}{p_2} = \frac{1}{\eta_2}.$$

Comparing these expressions with the monopoly formulas (4) shows that a duopolist puts twice as much emphasis on the external benefit from one group when it sets its price to the other group. The reason for this difference is simple. When a monopoly platform sets a high price that induces an agent from one side to leave, that agent disappears from the market. When a duopoly platform sets a high price that induces an agent from, say, group 1 to leave, that agent does not disappear but instead joins the rival platform, and this makes it harder to attract group-2 agents.

From (12), in equilibrium each platform makes profit

$$\pi = \frac{t_1 + t_2 - \alpha_1 - \alpha_2}{2}. \tag{13}$$

Assumption (8) guarantees that this profit is positive. Positive cross-group externalities act to reduce profit compared to the case where $\alpha_1 = \alpha_2 = 0$, since platforms have an additional incentive to compete hard for market share. Next, I discuss an extension where platforms can choose more complicated tariffs that reduce, or even reverse, these externalities.

□ **Alternative tariffs.** *Uniform prices.* In some contexts, it is natural to investigate the effect of price discrimination on prices and profit. For instance, are equilibrium profits higher or lower when nightclubs can charge different entry fees for women and men? Does the ability to target one side of the market without sacrificing revenues on the other side raise or lower profit? These issues can be addressed using the framework just discussed.

[5] A price in (12) can be negative if that side of the market involves a low cost, is competitive, or causes a large external benefit to the other side. It is often unrealistic to suppose negative prices are feasible, in which case the analysis needs to be adapted to incorporate the nonnegativity constraint—see Armstrong and Wright (forthcoming).

© RAND 2006.

Suppose that, say, $f_1 = f_2 = f$. (It makes little sense to discuss price discrimination if the costs are significantly different for the two groups.) Suppose platforms cannot set different prices to the two groups, and platform i chooses the uniform price p^i. (Perhaps sex discrimination laws prevent differential pricing by nightclubs.) Platform i's profit is $(p^i - f)(n_1^i + n_2^i)$, and a platform cares only about the total number of agents it attracts. From (9), total demand for platform i is

$$n_1^i + n_2^i = 1 + \frac{1}{2}\frac{t_1 + t_2 + \alpha_1 + \alpha_2}{t_1 t_2 - \alpha_1 \alpha_2}(p^j - p^i).$$

Therefore, the equilibrium uniform price is

$$p = f + 2\frac{t_1 t_2 - \alpha_1 \alpha_2}{t_1 + t_2 + \alpha_1 + \alpha_2}. \tag{14}$$

One can show that this uniform price (14) lies between the discriminatory prices in (12). Therefore, the consequence of a ban on price discrimination is that one group is made better off (the group that has the higher price with discriminatory pricing) while the other group is made worse off.

Equilibrium profits increase with discrimination if $p - f < (1/2)\{(p_1 - f) + (p_2 - f)\}$. Using expressions (12) and (14), it follows that this is the case if and only if

$$(t_1 - t_2)^2 > (\alpha_1 - \alpha_2)^2. \tag{15}$$

Condition (15) requires that the differences between the groups are more to do with differences in competitive conditions (t) than with differences in external benefits (α). Thus, when differences are largely due to differences in α, the ability of platforms to engage in price discrimination is damaging to their profits. Since total welfare is constant in this particular model, it follows that when condition (15) holds, consumers in aggregate are worse off when platforms engage in price discrimination. If, as seems plausible, price discrimination in, say, nightclubs, is more to do with asymmetries in cross-group benefits than with competitive conditions for the two groups, the use of price discrimination acts to make consumers in aggregate better off.

Two-part tariffs. The analysis so far has assumed that agents are charged a fixed fee to join a platform. There are several other kinds of tariffs that could be envisaged. For instance, Rochet and Tirole (2003) focus on the case where platforms levy charges on a per-transaction basis, i.e., the total charge to one group is proportional to the platform's realized market share of the other group. Alternatively, platforms could commit to supply agents with fixed utilities instead of charging a fixed price. Implicitly, there is then a commitment to reduce the charge that group-1 agents pay if it turns out that the market share for group 2 is smaller than expected, assuming that measurement problems do not preclude this. A more general formulation that encompasses these various possibilities is for platforms to offer a "two-part tariff," in which agents pay a fixed fee p together with a marginal price, γ, for each agent on the other side who joins the platform. That is to say, platform i's tariffs to groups 1 and 2 are respectively

$$T_1^i = p_1^i + \gamma_1^i n_2^i; T_2^i = p_2^i + \gamma_2^i n_1^i. \tag{16}$$

Special cases of this family of tariffs include (i) $\gamma_1^i = \gamma_2^i = 0$, where platforms compete in fixed fees as in the basic model presented above, and (ii) $\gamma_1^i = \alpha_1$ and $\gamma_2^i = \alpha_2$, where agents pay exactly the benefit they enjoy from interacting with an additional member of the other group. Thus, in case (ii) a platform commits to deliver a constant utility to customers, irrespective of its success on the other side of the market.

In general, each platform now has four degrees of freedom in its tariff choice. The analysis is more complicated than required in the basic model, and the details are left to the Appendix. This analysis is summarized in the following result.

© RAND 2006.

Proposition 3. Suppose assumption (8) holds. When platforms compete using two-part tariffs, a continuum of symmetric equilibria exist.[6] Let $0 \leq \gamma_1 \leq 2\alpha_1$ and $0 \leq \gamma_2 \leq 2\alpha_2$ be marginal prices aimed at group 1 and group 2 respectively. An equilibrium exists where the two platforms offer the same pair of two-part tariffs to group 1 and group 2 of the form

$$T_1 = p_1 + \gamma_1 n_2; \quad T_2 = p_2 + \gamma_2 n_1,$$

where the fixed fees are given by

$$p_1 = f_1 + t_1 - \alpha_2 + \frac{1}{2}(\gamma_2 - \gamma_1); \quad p_2 = f_2 + t_2 - \alpha_1 + \frac{1}{2}(\gamma_1 - \gamma_2). \tag{17}$$

The expressions (17) generalize (12) above. The profit of each platform in such an equilibrium is modified from (13) to be

$$\pi = \frac{t_1 + t_2 - \alpha_1 - \alpha_2}{2} + \frac{\gamma_1 + \gamma_2}{4}.$$

This profit is increasing in the marginal prices γ_1 and γ_2. The reason that high marginal prices yield high profit is that they reduce, or even overturn, the cross-group network effects that make the market so competitive.

Thus, when platforms can use more ornate tariffs that depend on the platform's success on the other side of the market, a continuum of symmetric equilibria exist that are ranked by the profit they generate. In technical terms, the source of the multiple equilibria is related to the multiple equilibria that exist in a (deterministic) supply function framework—see Section 3 of Klemperer and Meyer (1989). The common issue in the two settings is that a firm, for a given choice of tariff by its rival, has a continuum of best responses.

The question arises of which of the equilibria in Proposition 3 is selected. One suggestion might be that platforms coordinate on an equilibrium that generates high profits, i.e., on a pair of tariffs with large γ_i.[7] An alternative viewpoint is that the pure subscription tariffs analyzed in the basic model presented above are robust: if its rival offers a pure subscription tariff ($\gamma_1 = \gamma_2 = 0$), a platform has no incentive to offer a more ornate tariff that depends on its performance on the other side. However, more generally, this analysis suggests that, while it is straightforward to analyze the case of pure subscription tariffs (as in the basic model presented above) or the case of pure transaction tariffs (as emphasized in Rochet and Tirole (2003)), blending the two families of tariffs presents major problems for the predictive power of the model as formulated.

5. Competitive bottlenecks

■ This section presents the third and final model, which is termed a model of "competitive bottlenecks." The model of Section 4 is modified by supposing that while group 1 continues to deal with a single platform (to single-home), group 2 wishes to deal with each platform (to multi-home). Implicit in this model is the idea that group 2 puts more weight on the network benefits of being in contact with the widest population of group-1 consumers than it does on the costs of dealing with more than one platform. The crucial difference between this model and that discussed in Section 4 is that here group 2 does not make an "either-or" decision to join a platform. Rather, keeping the market shares for group 1 constant, a group-2 agent makes a decision to join one platform independently from its decision to join the other.[8] In this sense, there is no competition between platforms to attract group-2 customers.

[6] There are also asymmetric equilibria.

[7] As formulated here, there is no barrier to the platforms setting high prices γ and generating high profits (and low utilities for agents). A more realistic model would have an outside option for agents, which would put a ceiling on prices.

[8] This is true if group-2 agents are atomistic. If not, then when a "large" group-2 agent that has already joined platform A decides to join platform B, this will draw some group-1 agents away from platform A and cause a negative externality on the surplus from the agent's platform A interaction. Here, though, this possibility is ignored.

© RAND 2006.

There are several examples of markets where this framework seems a stylized representation. Consider the case of competing mobile telecommunications networks. (See Section 3.1 of Armstrong (2002) and Wright (2002) for formal models of this industry.) Subscribers wish to join at most one mobile network (i.e., they single-home). People on fixed telephony networks wish to call mobile subscribers. For a specified charge, someone can call any given mobile network, and in this sense, the people who call mobile networks multi-home. A mobile subscriber will choose the network with the tariff that leaves her with the most surplus. A mobile network's tariff has two ingredients: charges for subscription and outbound calls that affect the subscriber's utility directly, and charges the network makes to others for delivering calls *to* the subscriber (so-called call termination charges). Unless the subscriber cares directly about the welfare of people who might call her, the latter charges affect the subscriber's utility only insofar as they affect the number of calls she receives. (High termination charges will typically act to reduce the number of calls made to mobile networks, and this is detrimental to a subscriber's utility if she benefits from receiving calls.)

The tariffs that mobile networks set in equilibrium have low charges for subscription and outbound calls and high charges for call termination. In particular, the models predict that high profits made from call termination are passed on to subscribers in the form of subsidized handsets or similar inducements. More precisely, the equilibrium call-termination charge is chosen to maximize the welfare of mobile subscribers and mobile networks combined, and the interests of those who call mobile networks are ignored. This feature—that the single-homing side is treated well and the multi-homing side's interests are ignored in equilibrium—is a characteristic of the models presented below. A competitive bottleneck is present: even if the market for subscribers is highly competitive so that mobile networks have low equilibrium profits overall, there is no competition for providing communication services to these subscribers.

Other examples of this competitive bottleneck framework include: most people might read a single newspaper (perhaps due to time constraints), but advertisers might place ads in all relevant newspapers; consumers might choose to visit a single shopping mall (perhaps because of transport costs), but the same retailer might choose to open a branch in several malls; consumers might visit a single supermarket, but suppliers might wish to place their products on the shelves of several supermarkets; or a travel agent might use just one computerized airline reservation system, while airlines are forced to deal with all such platforms in order to gain access to each travel agent's customers. After analyzing a fairly general model of competitive bottlenecks, I shall present specific applications to newspapers and supermarkets.

□ **A general framework.** Suppose there are two, possibly asymmetric, platforms that facilitate interaction between two groups of agents. Suppose that group-2 agents are heterogeneous: if there are n_1^i group-1 agents on platform i, the number of group-2 agents prepared to pay a fixed fee p_2^i to join this platform is denoted

$$n_2^i = \phi^i(n_1^i, p_2^i), \tag{18}$$

where the function ϕ^i is decreasing in p_2^i and increasing in n_1^i. A group-2 agent's decision to join one platform does not depend on whether she chooses to join the rival platform. Let $R^i(n_1^i, n_2^i)$ denote platform i's revenue from group 2 when it has n_1^i group-1 agents and sets its group-2 price such that n_2^i group-2 agents choose to join the platform. Formally, R^i is defined by the relation

$$R^i(n_1^i, \phi^i(n_1^i, p_2^i)) \equiv p_2^i \phi^i(n_1^i, p_2^i). \tag{19}$$

Similarly to expression (5), platform i's group-1 utility u_1^i is given by

$$u_1^i = U^i(n_2^i) - p_1^i$$

if the platform charges p_1^i to group 1 and n_2^i group-2 agents join the platform. Here U^i is the (possibly nonlinear) function that measures the benefit that a group-1 agent enjoys with greater group-2 participation on the platform. (The function U^i might be decreasing, for instance when

© RAND 2006.

newspaper readers find advertisements to be a nuisance.) If a group-1 agent's utility is u_1^i with platform i, suppose the platform will attract

$$n_1^i = \Phi^i(u_1^i, u_1^j) \tag{20}$$

group-1 agents, where Φ^i is increasing in the first argument and decreasing in the second. If platform i's total cost of serving the two sides is denoted $C^i(n_1^i, n_2^i)$, its profit is

$$\pi^i = n_1^i p_1^i + R^i(n_1^i, n_2^i) - C^i(n_1^i, n_2^i). \tag{21}$$

Next, the number of group-2 agents on each platform in equilibrium is derived as a function of the equilibrium market shares for group 1. Suppose that in equilibrium, platform i offers utility $\hat{u}_1^i$ to its group-1 agents and attracts a number $\hat{n}_1^i$ of such agents (given by the function Φ^i). Then the platform must be maximizing its profits given this group-1 utility $\hat{u}_1^i$. Consider varying p_1^i and n_2^i so that utility $\hat{u}_1^i = U^i(n_2^i) - p_1^i$ is constant. Writing $p_1^i = U^i(n_2^i) - \hat{u}_1^i$ in (21) means that profit is

$$\pi^i = \hat{n}_1^i \left[U^i(n_2^i) - \hat{u}_1^i\right] + R^i(\hat{n}_1^i, n_2^i) - C^i(\hat{n}_1^i, n_2^i).$$

Given $\hat{n}_1^i$, platform i will choose to serve a number $\hat{n}_2^i$ of group-2 agents, where $\hat{n}_2^i$ maximizes

$$\hat{n}_1^i U^i(\cdot) + R^i(\hat{n}_1^i, \cdot) - C^i(\hat{n}_1^i, \cdot). \tag{22}$$

The equilibrium price to group 2 is $\hat{p}_2^i$, where this satisfies

$$\hat{n}_2^i = \phi^i(\hat{n}_1^i, \hat{p}_2^i). \tag{23}$$

For a given $\hat{n}_1^i$, notice that expression (22) measures the total surplus of platform i and its group-1 agents as the number of group-2 agents is varied. Therefore, the number of group-2 agents is chosen to maximize the joint interests of the platform and its group-1 agents, and the interests of group 2 are ignored. In general, this implies that there is a market failure, and there is a suboptimal number of group-2 agents on each platform for a given distribution of group-1 agents.

In more detail, suppose that the gross group-2 surplus on platform i when that platform has n_1^i group-1 agents and n_2^i group-2 agents is denoted $V^i(n_1^i, n_2^i)$. When there are no externalities within the set of group 2-agents, this surplus function differentiates to give the inverse demand function, so that

$$\frac{\partial}{\partial n_2^i} V^i(n_1^i, n_2^i) \equiv \frac{R^i(n_1^i, n_2^i)}{n_2^i}. \tag{24}$$

(The right-hand side of this expression is just the price paid by group-2 agents.) In some contexts, there are intragroup externalities present, and a group-2 agent might be better off if there were fewer other group-2 agents on the same platform, in which case the formula (24) is not valid. The most obvious examples of this phenomenon are where shops or advertisers are competing among themselves to sell to consumers, and this situation is discussed below. When there are no intragroup externalities present in group 2 and (24) holds, there is an unambiguous market failure present, in that there are too few group-2 agents on each platform for given numbers of group 1. To see this, note that given its group-1 market share, total surplus on platform i is maximized by choosing n_2^i to maximize

$$\hat{n}_1^i U^i(\cdot) + V^i(\hat{n}_1^i, \cdot) - C^i(\hat{n}_1^i, \cdot). \tag{25}$$

Since $V^i(n_1^i, n_2^i) - R^i(n_1^i, n_2^i)$, group 2's net aggregate surplus, is increasing in n_2^i, the maximizer of (25) is greater than the maximizer of (22), and there are too few group-2 agents served in equilibrium.

This discussion is summarized in the next proposition.

© RAND 2006.

Proposition 4. In the competitive bottleneck model, in any equilibrium the number of group-2 agents on a platform is chosen to maximize the joint surplus of the platform and its group-1 agents, and the interests of group 2 are ignored. Unless there are externalities within the set of group-2 agents, there are too few group-2 agents on each platform given the distribution of group-1 agents on each platform.

As with the mobile telephony case just discussed, in this model it does not make sense to speak of the competitiveness of "the market." There are two markets: the market for single-homing agents which is, to a greater or lesser extent, competitive, and a market for multi-homing agents where each platform holds a local monopoly. The excessive prices faced by the multi-homing side do not necessarily result in excess profits for platforms, since platforms might be forced by competitive pressure to transfer their monopoly revenues to the single-homing agents. Rather, the market failure is a suboptimal balance of prices to the two sides of the market.

Without putting more structure on this general model, it is hard to predict how their joint surplus is shared between platforms and group-1 agents. The price charged to group 1 will depend in part on the strength of competition in the group-1 market for consumers (i.e., on the form of the functions Φ^i). I investigate this further in the two applications that follow.

□ **Informative advertising on media platforms.** Consider a situation where advertisers wish to make contact with potential customers by placing ads on media platforms such as newspapers or yellow pages directories. Suppose there are two such platforms, A and B, which are assumed to be symmetric. Ads are placed on the platforms by monopoly retailers ("group 2"). (See the discussion below for the case where advertisers or retailers compete with each other.) Assume readers ("group 1") use one or the other platform but not both. A reader will purchase some quantity of a given product if she sees an ad for the product and the product's price allows the consumer to enjoy nonnegative consumer surplus. The cost of producing and distributing an individual copy of a newspaper/directory is $c(n_2)$ when it contains n_2 ads. If there are no other costs, the cost function $C(n_1, n_2)$ used in the general framework presented above takes the multiplicative form

$$C(n_1, n_2) = n_1 c(n_2). \tag{26}$$

Advertisers are characterized by the parameter α_2: a type-α_2 advertiser has a product that generates profit α_2 from each reader who sees its ad. Because each reader joins only one platform, an advertiser has no means with which to contact any particular reader except to place an ad on that reader's chosen platform. Thus, a type-α_2 advertiser values an ad placed on a platform with readership n_1 at $\alpha_2 n_1$.[9] If the fixed charge for placing an ad on platform i is p_2^i, then a type-α_2 advertiser will place an ad on the platform if $\alpha_2 n_1^i \geq p_2^i$. Suppose an advertiser's valuation α_2 is unknown to a platform. Specifically, suppose for each advertiser the valuation α_2 is independently and identically drawn from a distribution function $F(\alpha_2)$. From the viewpoint of the platforms, each advertiser is *ex ante* identical, and so a given platform will charge the same price p_2 to all advertisers. Therefore, the function ϕ in (18) is given by

$$\phi(n_1^i, p_2^i) = 1 - F(p_2^i / n_1^i).$$

With an advertiser demand function of this form, the revenue function R in (19) is proportional to the readership:

$$R(n_1, n_2) = n_1 r(n_2),$$

where $r(\cdot)$ is a platform's advertising-revenue-per-reader function. (This function $r(\cdot)$ is defined by $r(1 - F(\gamma)) \equiv (1 - F(\gamma))\gamma$.) In this case, expression (22) is proportional to n_1^i, and the

[9] Thus, an advertiser's payoff is proportional to the platform's readership. If a seller has limited supplies of the product (or, more generally, if its cost of production is convex), the seller only obtains benefit from the ad reaching a certain number of potential consumers, and then the linearity assumption would not be plausible.

© RAND 2006.

equilibrium advertising volume $\hat{n}_2$ maximizes

$$U(n_2) + r(n_2) - c(n_2) \tag{27}$$

regardless of the platform's performance on the group-1 side. Thus, a media platform's decision on its advertising volume can be made independently of the size of its readership.

Suppose, as in (6), there is a Hotelling functional form for the consumer market share function Φ^i in (20) given by

$$n_1^i = \frac{1}{2} + \frac{u_1^i - u_1^j}{2t}, \tag{28}$$

where t is the parameter that measures the competitiveness of the market for readers. Therefore, the symmetric equilibrium (if it exists) involves the two platforms sharing readers equally: $\hat{n}_1^i = \hat{n}_1^j = 1/2$. Expression (23) then shows that the equilibrium charge to advertisers, $\hat{p}_2$, is given by

$$\hat{n}_2 = 1 - F(2\hat{p}_2), \tag{29}$$

where $\hat{n}_2$ is given by (27). Proposition 4 then implies that the equilibrium involves too little advertising. In this stark model, advertisers do not gain or lose when the market for readers becomes more competitive. For instance, if two newspapers merge in this model, advertisers will not be made worse off, since they will be forced to pay the same monopoly prices in either event.

This analysis was simplified by assuming that cost in (26) was proportional to readership. A richer model would include a per-advertiser platform cost in addition, since it is plausible that a platform incurs costs in dealing with each advertiser. In this case we would have, say, $C(n_1, n_2) = n_1 c(n_2) + f_2 n_2$. One could analyze the application to shopping malls using similar techniques. Here, a shopping mall's cost function would plausibly take the additive form

$$C(n_1, n_2) = f_1 n_1 + f_2 n_2.$$

Also, it is plausible to model retailers (or advertisers in the newspaper application) as having a fixed cost of joining a platform (the fixed cost associated with starting a shop or designing an ad) in addition to the platform's charge. However, the fundamental conclusion—that there will be too few retailers/advertisers in equilibrium—will continue to hold in these richer settings, as shown in Proposition 4.

Turn next to the outcome for readers. As in Section 4, it turns out that the equilibrium price depends on the way in which advertising charges are levied. There are two natural ways to charge advertisers: (i) advertising charges are explicitly levied on a per-reader basis, and (ii) advertisers are charged a lump-sum fee for placing an ad.[10] The reason this makes a difference to the competitiveness of the market for readers is that it affects the profitability of a platform's deviation in the reader price. With case (i), if a platform attracts more readers, the number of ads does not change.[11] With case (ii), by contrast, having more readers acts to attract more advertisers (keeping the lump-sum advertising charge constant), and this in turn acts to attract still more readers if readers like ads. (Note that the charging basis for advertising does not affect the equilibrium number of ads, given by $\hat{n}_2$, which maximizes (27) above.) I shall now discuss these two families of tariffs in more detail.

Per-reader advertising charges. Suppose that platform i offers advertising space for a charge γ^i per reader on the platform. Then a type-α_2 advertiser will choose to join the platform if and only

[10] Unlike in Section 4, here I do not consider the more general class of two-part tariffs, where advertisers pay a fixed fee plus a charge per reader. From the earlier analysis, I expect there would be a problem of multiple equilibria if I did so.

[11] Conceptually, competing with per-reader charges is the same as competing in terms of quantities of advertising space, and the equilibrium price for ads is determined to clear the market, given the realized market shares for readers.

© RAND 2006.

if $\alpha_2 > \gamma^i$, and the number of ads does not depend on the number of readers on the platform. In this case the analysis is simple. Both platforms will choose the number of ads $\hat{n}_2$ that maximizes (27), and this generates advertising revenue per reader equal to $r(\hat{n}_2)$. Since the two platforms offer the same number of ads, a reader cares only about the relative price when deciding which platform to use. Given the specification in (28), the profit of platform i if it charges readers p_1^i (while the rival platform charges readers p_1^j) is

$$\pi^i = \left(\frac{1}{2} - \frac{p_1^i - p_1^j}{2t}\right)\left(p_1^i + r(\hat{n}_2) - c(\hat{n}_2)\right).$$

Therefore, the symmetric equilibrium price for readers is given by

$$p_1 = c(\hat{n}_2) + t - r(\hat{n}_2), \tag{30}$$

and a platform's revenue from advertising, $r(\hat{n}_2)$, is passed on to consumers in the form of a low price p_1. Platform profits are equal to the product differentiation parameter t.

When competition for readers is intense (t is small) or the advertising revenue $r(\hat{n}_2)$ is large, the price to readers is below cost. It might even be that the price in (30) is negative. If, as seems plausible, there is a nonnegativity constraint on prices, the equilibrium will involve group 1 being allowed onto the platform for free. This could be a rationale for why yellow pages directories and some newspapers are supplied to readers for free, why a shopping mall might not wish to charge consumers for entry even if it were feasible to do so, or why a broadcaster might wish not to charge viewers even when this is technologically feasible.

Lump-sum advertising charges. To analyze this more complex case, I need to calculate a platform's extra readership when it undercuts its rival's reader price. When a platform undercuts its rival on the reader side, it will clearly attract more readers; in consequence, the platform attracts more advertisers (given that its lump-sum charge for advertising is unchanged), which thereby attracts further readers if readers like ads. This feedback loop is absent when advertising charges are levied per reader. To be more precise, the advertising volume on platform i as a function of readership n_1^i is

$$\tilde{n}_2(n_1^i) = 1 - F(\hat{p}_2/n_1^i),$$

where the equilibrium lump-sum charge to advertisers $\hat{p}_2$ is given in expression (29). Using this notation, from (28) it follows that platform i's readership, n_1^i, must satisfy the following implicit equation:

$$n_1^i = \frac{1}{2} + \frac{U(\tilde{n}_2(n_1^i)) - p_1^i - \left[U(\tilde{n}_2(1 - n_1^i)) - p_1^j\right]}{2t}. \tag{31}$$

If (31) has multiple solutions for n_1^i, then there are multiple demand configurations consistent with the reader prices, and some method of choosing among the possible configurations is needed. To sidestep this issue, suppose model parameters are such that expression (31) has a unique solution in n_1^i for all relevant prices p_1^i and p_1^j. In Section 4, assumption (8) ruled out this possibility in the single-homing framework. It seems hard to find the precise corresponding assumption needed for the competitive bottleneck model. However, it is clear what is needed for there to be a unique solution to (31): t should be "large" relative to U' and $\tilde{n}_2'$. (This implies that the right-hand side of (31) is relatively flat as a function of n_1^i.) If $U' \leq 0$, so that readers either do not care about ads or actively dislike them, then the right-hand side of (31) is decreasing in n_1^i and there will always be a unique solution to (31). Armstrong and Wright (forthcoming) explore the issue of multiple consistent demand configurations and the existence of equilibrium in a related model where the group-2 agents are homogeneous (i.e., when $\tilde{n}_2'$ is large).

© RAND 2006.

Implicit differentiation of (31) implies that

$$\left.\frac{\partial n_1^i}{\partial p_1^i}\right|_{p_1^i=p_1^j} = \frac{-1}{2t - 2\tilde{n}_2'(1/2)U'(\hat{n}_2)}. \tag{32}$$

In particular, when readers like ads ($U' > 0$), a platform's reader demand is more elastic than was the case with per-reader advertising charges (when $\partial n_1^i/\partial p_1^i = -1/(2t)$). When readers dislike ads ($U' < 0$), by contrast, their demand is less elastic. The platform's profit is

$$\pi^i = n_1^i \left[p_1^i - c(\tilde{n}_2(n_1^i)) + r(\tilde{n}_2(n_1^i))\right].$$

Using expression (32), it follows that the equilibrium consumer price p_1 is

$$p_1 = c(\hat{n}_2) + t - r(\hat{n}_2) - \frac{1}{2}\tilde{n}_2'\left(\frac{1}{2}\right)\left[2U'(\hat{n}_2) + r'(\hat{n}_2) - c'(\hat{n}_2)\right].$$

The first-order condition for the fact that $\hat{n}_2$ maximizes expression (27) then implies

$$p_1 = c(\hat{n}_2) + t - r(\hat{n}_2) - \frac{1}{2}\tilde{n}_2'\left(\frac{1}{2}\right)U'(\hat{n}_2). \tag{33}$$

Discussion. Comparing the reader price in expression (33) with that for the per-reader charging case in (30), we can see that when advertising charges are levied on a lump-sum basis, the equilibrium price for readers is lower or higher than when they are levied on a per-reader basis according to whether U' is positive or negative. Platform profits are correspondingly lower or higher with lump-sum charging according to whether U' is positive or negative. These results are akin to those presented in Section 4, where the use of tariffs that depend positively on the platform's success on the other side of the market was seen to relax competition and boost profits. In the knife-edge case where readers do not care about ads at all ($U' = 0$), there is no difference between the regimes of lump-sum and per-reader charges for advertising. The reason for this is that with lump-sum charging there is no extra incentive for a platform to undercut its rival in the market for readers. While it is true that when platform A gains readership at the expense of platform B, B will find that its advertising demand shrinks; this effect no longer gives A any advantage in the market for readers. This discussion is summarized in the final proposition.

Proposition 5. In the model of informative advertising on media platforms, if advertisers do not compete between themselves for consumers, there will be too few ads in equilibrium from a social point of view. If readers like (dislike) ads, when platforms charge advertisers on a lump-sum basis the equilibrium reader price and platform profit is lower (higher) than when advertising charges are levied on a per-reader basis.

The model presented here extends Gabszewicz, Laussel, and Sonnac (2001). That article proposes a model of the newspaper industry in which readers single-home and advertisers multi-home. Therefore, although it is not emphasized in the article, they find there is monopoly pricing to advertisers and hence underprovision of advertising. Monopoly revenues from supplying advertising space are passed on to readers in the form of a subsidized price. Sometimes these revenues are so large that the newspaper is supplied to readers for free.[12] The main difference is that in their article, readers are assumed to be indifferent to ads ($U' = 0$ in the current notation). This implies that the outcome does not depend on whether lump-sum or per-reader advertising charges are used. As is clear from the previous discussion, when platforms use lump-sum advertising charges (which is arguably the more plausible scenario), the analysis is rather complicated when readers do care about advertising intensity.

[12] Gabszewicz, Laussel, and Sonnac (2001) have an initial stage in which newspapers choose their political stance before competing for readers and dealing with advertisers. When advertising revenues are so high that newspapers are offered for free, they show that newspapers will choose the same centrist stance, whereas if newspapers compete in prices for readers, they will choose to differentiate their politics.

© RAND 2006.

Another related article, in the context of the television industry, is Anderson and Coate (2005). For the most part, they assume that viewers single-home and advertisers multi-home. It is arguable that the single-homing assumption is less natural in the television context than in the newspaper context, especially since the introduction of the remote control, and this is why I focused my discussion on newspapers and yellow pages. In contrast to Gabszewicz, Laussel, and Sonnac (2001), Anderson and Coate (2005) assume that viewers care (negatively) about advertising intensity. They analyze both the case where viewers are not charged for viewing and, more relevant for the current article, the case where viewers pay to view a television channel. In the latter case, they find there is too little advertising in equilibrium, for exactly the same reasons as outlined above. They assume that advertising charges are levied on a per-viewer basis, which avoids many of the complexities discussed above.

A third article that deserves mention is Kaiser and Wright (2006), who use data from German magazines to analyze which of the three models in this article (monopoly platform, two-sided single-homing, or competitive bottlenecks) fits that industry best.[13] Magazines are grouped into genres (e.g., home improvement or photography), and model parameters are estimated for each of these three models. Within a genre, they find that readers are overwhelmingly single-homing, with about 8% of readers buying more than one magazine within a genre. Surprisingly, they also find that advertisers tend to be single-homing and to place an ad in just one magazine at a time (about 17% of advertisers multi-home). Perhaps magazines cater to differentiated readers, and many advertisers believe their product is suited to only one readership. In any event, the authors estimate that the two-sided single-homing model analyzed in Section 4 best explains the data. (However, they find that the market for readers is more competitive than the market for advertisers.)

□ **Supermarkets.** A second application of the competitive bottleneck model is to supermarkets and similar kinds of retailers. A commonly held view about the supermarket sector is that, provided competition for consumers is vigorous, consumers are treated well by supermarkets but supermarkets deal too aggressively with their suppliers. The model when applied to this industry can generate these stylized features.

Suppose two supermarkets compete to attract consumers. Consumers ("group 1") care both about the prices they pay and the range of products on offer. They visit either one supermarket or the other (but not both) over the relevant time period. Suppose there is a continuum of monopoly products ("group 2"), each of which could be supplied to either or both supermarkets. For simplicity, suppose each consumer wishes to buy one unit of each product so long as the price of the product is less than their reservation value, α_1. (Thus, consumers view the various products as being equally valuable and as neither substitutes nor complements in their utility function.) Suppose that supermarkets incur a cost c for selling each unit of any product.[14] Supermarkets are assumed to set retail prices to their consumers and to make take-it-or-leave-it offers to buy from their suppliers. In particular, supermarkets hold all the bargaining power when dealing with their suppliers.

As in the case of advertising on media platforms, suppose that the unit cost of supply for any product is unknown to a supermarket and that for each product the unit cost α_2 is independently and identically drawn from a distribution function $F(\alpha_2)$. From the supermarkets' viewpoint, each supplier is *ex ante* identical, and so a given supermarket will make the same offer to all suppliers to buy at the per-unit price p_2. (Note that, unlike other models discussed, this is a payment from the platform to the group-2 agents. Note also that the context makes it natural to assume platforms make their payments to suppliers on a per-transaction rather than lump-sum basis.) The number of suppliers that agree to this level of compensation is $F(p_2)$. If a supermarket sets a retail price p_1 per unit to consumers, a consumer's utility from visiting the supermarket is the number of

[13] Also see Rysman (2004), who estimates the size of cross-group externalities in the market for yellow pages.

[14] With some extra complexity, one could extend this simple framework to allow for per-consumer costs (e.g., car parking and checkout facilities) and for per-product costs.

© RAND 2006.

products multiplied by the net surplus per product

$$u_1 = F(p_2)(\alpha_1 - p_1), \tag{34}$$

whereas the supermarket's profit per consumer is

$$\pi = F(p_2)(p_1 - c - p_2). \tag{35}$$

Regardless of its market share of consumers, a supermarket will choose p_1 and p_2 to maximize its profit per consumer, π, subject to delivering a required utility u_1. Expressions (34) and (35) then imply that

$$p_2 \text{ maximizes } F(p_2)(\alpha_1 - c - p_2). \tag{36}$$

As with all the competitive bottleneck models, in equilibrium the joint surplus of the platforms and the single-homing group is maximized (supermarkets and consumers in this case, as given in expression (36)), and the interests of the multi-homing side (the suppliers) are ignored. The low level of compensation in (36) will exclude some relatively high-cost suppliers whose presence in the supermarkets is nevertheless efficient. (A supplier should supply if $\alpha_2 < \alpha_1 - c$, whereas the equilibrium price p_2 in expression (36) is strictly lower than $\alpha_1 - c$ and supply is inefficiently restricted.) In other words, payments to suppliers are too low from a social point of view and there are too few products on the shelves.

How well consumers are treated depends on competitive conditions on their side. If they choose their supermarket according to the Hotelling specification in (28), one can show their equilibrium utility u_1 is given by

$$u_1 = F(p_2)(\alpha_1 - c - p_2) - t,$$

so that consumers keep the joint surplus $F(p_2)(\alpha_1 - c - p_2)$ except for the market power element t retained by the supermarkets. (An explicit expression for the equilibrium per-unit price to consumers is $p_1 = c + p_2 + t/F(p_2)$.)

As with the previous model of informative advertising, the treatment of suppliers is not affected by the strength of competition between supermarkets for consumers. In this model, if two supermarkets merge, consumers would be treated less favorably but suppliers would not be affected. Their compensation in (36) anyway is the payment that would be chosen if there were a single monopoly supermarket.

In this model, supermarkets act to reverse the bargaining asymmetry that consumers might otherwise have in their dealings with individual monopoly suppliers.[15] For instance, suppose without the institution of supermarkets the suppliers sell directly to consumers. (Again, though, there are two "shopping centers" in the same locations as the supermarkets, and suppliers sell in both of these areas. For consistency with the supermarket analysis, suppose also that each supplier incurs the selling cost c per unit in addition to its supply cost α_2.) In this case, those suppliers with cost $c + \alpha_2$ below the consumer reservation price α_1 would choose to supply at the monopoly price $p_1 = \alpha_1$. (Because there is a continuum of suppliers, each supplier sets its price without regard for the effect its high price has on the number of consumers who visit the local shopping area.) This move to direct supply clearly makes consumers worse off, although it does improve efficiency because the competitive bottleneck is overcome and the range of products supplied is efficient.[16]

[15] Dobson and Waterson (1997) analyze a model where supermarkets bargain with suppliers over supply prices, and where a merger between two supermarkets improves the bargaining position of supermarkets and so drives down the equilibrium supply prices. In extreme cases they show that a merger between supermarkets might lead to lower retail prices for consumers.

[16] The change from a situation where suppliers sell their products directly to consumers as opposed to selling to a supermarket corresponds to the distinction made in Smith and Hay (2005) between "streets" and "supermarkets." They

© RAND 2006.

The same logic suggests that suppliers are likely to be better off when they act independently than when they integrate to form a supermarket. Independent suppliers do not take account of the negative effect on other suppliers in the same shopping area when they set high prices. When suppliers merge to form a supermarket, they internalize the effect that each supplier's pricing decision has on the overall attractiveness of the shopping center for consumers, and this makes competition between shopping centers more intense. Lack of coordination between independent suppliers acts as a kind of commitment to price high in a shopping area, and this boosts profit. (See Beggs (1994) for further analysis of this point.)

□ **Intraplatform competition.** An interesting issue concerns the extent of competition between retailers within platforms. For instance, a television channel might charge more for a car ad if it promised not to show a rival manufacturer's ad in the same slot. A shopping mall might charge a higher rent to a retailer with the promise that it will not let a competing retailer into the same mall. Implementing competition within a platform will often mean that retailers' prices and profits are lower than they would be with monopoly retailers.[17] Thus we may expect that if the platform allowed retailing competition, it would make less money from the retailing side of the market but more money from the consumer side (if it charged consumers for entry). A plausible hypothesis is: platforms will allow competition within the platform if consumers can be charged for entry, but if for exogenous reasons consumers have free entry, then platforms will restrict competition to drive up the revenues obtained from retailers.

This topic deserves a separate article to itself. Here I merely describe a stylized example to show the plausibility of the hypothesis. Suppose there are two platforms that can serve any number of consumers and retailers costlessly ($C(n_1, n_2) \equiv 0$ in the general model presented above). If consumers receive utility u_1^i from platform i, suppose the market share of platform i is given by expression (28). Suppose there is a single, homogeneous product supplied by a group of identical retailers. If the retail price for this product is P, each consumer demands a quantity $q(P)$ of the product. Let $v(P)$ be the consumer surplus associated with this demand function. Each retailer has marginal cost C for supplying a unit of the product. Suppose there is no fixed cost associated with a retailer locating in a given platform (other than the platform charge for entry).

A platform must decide whether to have retailing competition or not (i.e., whether to have more than one retailer on the platform). Suppose that a retailer chooses its price on a platform to reflect competitive conditions on that platform. (That is, retailers can price discriminate from one platform to the other.) If there is competition on the platform, the price of the product on the platform is equal to marginal cost C. If there is no competition on the platform, the product's price will be above cost, and the price-cost markup will depend on the extent of competition between platforms.

First, suppose platforms can charge consumers for access to the platform. In this case, it is optimal for platform i to choose to have competition on the platform. To see this, suppose that the rival platform offers consumers a utility u_1^j. (It does not matter whether this utility is achieved by means of competition within the rival platform or not.) If the product has price P^i on platform i and consumers pay p_1^i to gain access to the platform, consumers obtain utility $u_1^i = v(P^i) - p_1^i$ from that platform. Therefore, given (28), the joint profit of platform i and the retailer(s) on this platform is

$$\left(\frac{1}{2} + \frac{v(P^i) - p_1^i - u_1^j}{2t}\right)\left(q(P^i)(P^i - C) + p_1^i\right).$$

propose a slightly different model and find more generally that the comparison of the number of products supplied in the two situations is ambiguous.

[17] Stahl (1982) analyzes a model where a retailer's profit might increase if a retailer supplying an imperfect substitute locates in the same shopping center. The greater variety of products attracts more consumers to the shopping center, which might outweigh the more intense competition within the shopping center. (See Schutz and Stahl (1996) for a related search model.) By contrast, in the model presented here, competition drives profit to zero.

© RAND 2006.

Rewriting this in terms of $u_1^i = v(P^i) - p_1^i$, this joint profit is

$$\left(\frac{1}{2} + \frac{u_1^i - u_1^j}{2t}\right)\left(q(P^i)(P^i - C) + v(P^i) - u_1^i\right).$$

Clearly, for any (u_1^i, u_1^j) this joint profit is maximized by choosing $P^i = C$, i.e., by implementing competition, and moreover this joint profit is then entirely appropriated by the platform. Therefore, when consumers can be charged for access to the platform, the competitive option is the most profitable way for a platform to generate any given level of consumer utility. This is a dominant strategy and does not depend on whether the rival platform chooses to have competition. The platform obtains all of its revenue from the consumer side. It is straightforward to show that the equilibrium charge for access by consumers to the platforms is $p_1 = t$ (and the price for the product on the platforms is $P = C$).

Second, suppose that for some exogenous reason platforms cannot charge consumers for access and so must make their profit from the retailer side. (For instance, television channels historically could not charge viewers for access and so had to fund their service entirely from the advertising side.) The only way a platform can set a positive price to a retailer in this stark framework is if the retailer is a local monopoly, so platforms will restrict competition in order to obtain any revenue at all. The monopoly retailer on platform i will make profit

$$\left(\frac{1}{2} + \frac{v(P^i) - v(P^j)}{2t}\right) q(P^i)(P^i - C)$$

if it sets the price P^i and the monopoly on the rival platform sets the price P^j. The platform can then extract this profit by means of its charge to the retailer. Suppose the monopoly retailers on the two platforms are separately owned. One can then show that the equilibrium product price P is close to the monopoly price (i.e., the price that maximizes $q(P)(P - C)$) when t is large and is close to cost C when t is small. One can also show that consumers in this framework are always worse off when platforms can charge them for access compared to when they have free access, despite the fact that they pay a lower price for the product once they are on the platform. (See Corollary 1 in Armstrong and Vickers (2001).)

An article that relates to this discussion is Dukes and Gal-Or (2003). They have a model where competing producers place ads on competing media platforms. They assume that the platforms are broadcasters who do not charge viewers/listeners for access to the platform. Consistent with the above discussion, they find that a media platform (usually) sells advertising space exclusively to one firm from the producer oligopoly.

□ **Discussion.** Section 2 discussed two important precursors to this article. Let us now rejoin this discussion to point out the relationships between the three competitive bottleneck models in the articles. First, consider Proposition 11 in Caillaud and Jullien (2003). In the notation of Section 4 of the current article, those authors present a model in which there is no intrinsic product differentiation between the two platforms, where group-1 agents have a benefit α_1 from interacting with each group-2 agent, and where group-2 agents each have a benefit α_2 from interacting with each group-1 agent. (Thus there is no variation in α_2 as there was in this section.) Caillaud and Jullien show that there is a symmetric equilibrium of the following form (provided $\alpha_2/2 > f_2$). Group 1 single-homes and divides equally between the two platforms, while group-2 agents join both platforms. The price to group 1 equals their cost, $p_1 = f_1$, while the price to group 2 fully extracts their surplus, so $p_2 = \alpha_2/2$. This forms an equilibrium because a platform has no incentive to undercut its rival on either side of the market. If the platform sets a price $p_2 < \alpha_2/2$, this has no effect on group 2's choice and will not boost demand from that side. If the platform sets a price $p_1 < f_1$, this will attract all group-1 agents but will not affect demand by group 2, so this deviation will reduce the platform's profit given that the price is below cost.[18] Thus positive profits

[18] Here one important issue is not discussed. If the deviating platform simultaneously reduces p_1 and increases p_2, there are multiple consistent demand configurations, and for the stated prices to form an equilibrium, a particular choice

© RAND 2006.

can be sustained in equilibrium even when two identical platforms compete. A feature of Caillaud and Jullien's model is that demands are discontinuous in prices: a small price reduction to group 1 means that the platform attracts the entire set of group-1 agents, and this feature implies that it can never be optimal to set a price $p_1 > f_1$. (Due to the finite cross-elasticities in the present model, there is no reason to rule out above-cost pricing to the single-homing side.) However, a small price reduction to the multi-homing side has no effect on demand, and this provides the source of profits in this industry.

Second, consider Section 6 of Rochet and Tirole (2003). Up to that section, they consider either a monopolist charging two-part tariffs or platform competition in pure transaction prices. Using the notation for two-part tariffs in Section 4 of the current article, Section 6 of Rochet and Tirole supposes that platforms can be taken to compete in the "per-transaction" prices P_1^i and P_2^i, defined by

$$P_1^i = \gamma_1^i + \frac{p_1^i - f_1}{n_2^i}; \quad P_2^i = \gamma_2^i + \frac{p_2^i - f_2}{n_1^i}.$$

The interpretation of this game is that platforms commit to per-transaction prices and settle accounts with their customers once market shares and transactions have been recorded. It is important to note, though, that assuming platforms commit to per-transaction prices is not equivalent to assuming they offer two-part tariffs. Specifically, it is true that for a given pair of two-part tariffs offered by platform j, platform i's payoff only depends on its own tariffs via the summary prices P_1^i and P_2^i above. However, for the reasons outlined in Section 4, platform i's particular choice of two-part tariffs (among those tariffs with the same per-transaction prices P_1^i and P_2^i) does matter for platform j, since it affects j's incentive to deviate. Namely, an aggressive move by platform j has more impact on j's market share under competition in two-part tariffs, since its effect on platform i's customers is not dampened by a commitment to per-transaction prices. Equilibria therefore depend on the modelling choice of strategic variable. The assumption that platforms compete using two-part tariffs is perhaps more descriptive of existing two-sided markets.

There are at least three limitations to the present analysis of competitive bottlenecks. First, in the applications I made the simplifying assumption that the population of group-1 agents was constant. Thus, the fact that this group tends to be treated favorably in equilibrium has no effect on the number of such agents who choose to participate. If instead there were a market expansion effect, this would make group 2 better off, because they have more group-1 agents with whom to interact. In principle, it is conceivable that this effect could imply that the number of group-2 agents served is not too small from a social point of view. However, this turns out not to be possible. (See Armstrong (2002) for this analysis in the telecommunications context.)

Second, I made the convenient assumption that no group-1 agents multi-homed. A richer model would allow for some agents to multi-home (for instance, some people read two newspapers, some people might go to one supermarket for some products and another supermarket for other products, and so on). Platforms then no longer have a monopoly over providing access by group 2 to these multi-homing group-1 agents. So far, little progress has been made in extending the analysis to these mixed situations, and this is a fruitful topic for future research. (See Section 7.1 of Anderson and Coate (2005) for a first step in this direction.)

Third and finally, I did not consider a platform's incentive to require an otherwise multi-homing agent to deal with it exclusively. It is plausible in the context of the competitive bottleneck model that a platform might try to sign up group-2 agents exclusively, in order to give it an advantage in the market for group-1 agents. Of course, if platforms succeed in forcing group-2 agents to decide to deal with one platform or the other, then platforms will find themselves in the two-sided single-homing situation analyzed in Section 4. Because network effects are so strong in that situation, it is plausible that platforms find their equilibrium profits decrease when they

for the demand configuration needs to be made. See Armstrong and Wright (forthcoming) for further discussion of this issue.

© RAND 2006.

force group 2 to deal exclusively. (See Armstrong and Wright (forthcoming) for further analysis of exclusive contracts within this framework.)

Appendix

■ In this Appendix I supply the proof of Proposition 3. Suppose that platform i's tariffs take the form in (16) above. If a group-1 agent joins platform i, she obtains utility

$$u_1^i = (\alpha_1 - \gamma_1^i)n_2^i - p_1^i, \tag{A1}$$

and so the number of group-1 and group-2 agents who join platform i is

$$n_1^i = \frac{1}{2} + \frac{(\alpha_1 - \gamma_1^i)n_2^i - (\alpha_1 - \gamma_1^j)(1 - n_2^i) - (p_1^i - p_1^j)}{2t_1} \tag{A2}$$

$$n_2^i = \frac{1}{2} + \frac{(\alpha_2 - \gamma_2^i)n_1^i - (\alpha_2 - \gamma_2^j)(1 - n_1^i) - (p_2^i - p_2^j)}{2t_2}. \tag{A3}$$

By solving this pair of equations in n_1^i and n_2^i, one obtains the following explicit formulas for n_1^i and n_2^i in terms of the eight tariff parameters:

$$n_1^i = \frac{1}{2} + \frac{1}{2}\frac{\left(2\alpha_1 - \gamma_1^i - \gamma_1^j\right)\left(2p_2^j - 2p_2^i + \gamma_2^j - \gamma_2^i\right) + t_2(4p_1^j - 4p_1^i + 2\gamma_1^j - 2\gamma_1^i)}{4t_1t_2 - (2\alpha_1 - \gamma_1^j - \gamma_1^i)(2\alpha_2 - \gamma_2^j - \gamma_2^i)}$$

$$n_2^i = \frac{1}{2} + \frac{1}{2}\frac{\left(2\alpha_2 - \gamma_2^i - \gamma_2^j\right)\left(2p_1^j - 2p_1^i + \gamma_1^j - \gamma_1^i\right) + t_1(4p_2^j - 4p_2^i + 2\gamma_2^j - 2\gamma_2^i)}{4t_1t_2 - (2\alpha_1 - \gamma_1^j - \gamma_1^i)(2\alpha_2 - \gamma_2^j - \gamma_2^i)}.$$

Equilibrium prices are determined by the sensitivities of market shares to changes in the various prices. To determine symmetric equilibria, I calculate the derivative of market shares with respect to changes in prices, evaluated when the two platforms set the same quadruple of prices $(p_1, p_2, \gamma_1, \gamma_2)$:

$$\frac{\partial n_1^i}{\partial p_1^i} = -\frac{t_2}{2\Delta}; \quad \frac{\partial n_1^i}{\partial \gamma_1^i} = -\frac{t_2}{4\Delta} \tag{A4}$$

$$\frac{\partial n_2^i}{\partial p_2^i} = -\frac{t_1}{2\Delta}; \quad \frac{\partial n_2^i}{\partial \gamma_2^i} = -\frac{t_1}{4\Delta} \tag{A5}$$

$$\frac{\partial n_1^i}{\partial p_2^i} = -\frac{\alpha_1 - \gamma_1}{2\Delta}; \quad \frac{\partial n_1^i}{\partial \gamma_2^i} = -\frac{\alpha_1 - \gamma_1}{4\Delta} \tag{A6}$$

$$\frac{\partial n_2^i}{\partial p_1^i} = -\frac{\alpha_2 - \gamma_2}{2\Delta}; \quad \frac{\partial n_2^i}{\partial \gamma_1^i} = -\frac{\alpha_2 - \gamma_2}{4\Delta}, \tag{A7}$$

where $\Delta = t_1t_2 - (\alpha_1 - \gamma_1)(\alpha_2 - \gamma_2)$. Notice that in each case a small change in γ has exactly half the effect of a small change in the corresponding p. The reason is that with equal market shares, the effect on the total charge an agent must pay with a change in γ is half that with a change in p.

Platform i's profit is

$$\pi^i = (\gamma_1^i n_2^i + p_1^i - f_1)n_1^i + (\gamma_2^i n_1^i + p_2^i - f_2)n_2^i. \tag{A8}$$

Notice that expressions (A4)–(A7) imply that at any symmetric set of prices, we have

$$\frac{\partial \pi^i}{\partial p_1^i} = 2\frac{\partial \pi^i}{\partial \gamma_1^i}; \quad \frac{\partial \pi^i}{\partial p_2^i} = 2\frac{\partial \pi^i}{\partial \gamma_2^i}. \tag{A9}$$

We will see that this feature of the market will generate a multiplicity of symmetric equilibria.

First, I show that platform i's objective function is concave given its rival's choice $(p_1^j, p_2^j, \gamma_1^j, \gamma_2^j)$, as long as $0 \le \gamma_1^j \le 2\alpha_1$, and $0 \le \gamma_2^j \le 2\alpha_2$ are nonnegative and the maintained assumption (8) holds. I need to show this so that I can characterize equilibria in terms of the first-order conditions. Usually, verifying that a function of four variables is concave is a tedious matter. However, in this context, I can easily reduce the number of i's strategic variables to two, which greatly simplifies the calculation. Given the rival prices $(p_1^j, p_2^j, \gamma_1^j, \gamma_2^j)$, it turns out that i's profits are a function

© RAND 2006.

only of the utilities u_1^i and u_2^i it offers its consumers. When it offers this pair of utilities, it will attract a certain number n_1^i and n_2^i of each group—I will shortly derive this relationship explicitly—and by combining expressions (A1) and (A8), its total profit can be written in terms of the utilities as

$$\pi^i = (\alpha_1 n_2^i - u_1^i - f_1)n_1^i + (\alpha_2 n_1^i - u_2^i - f_2)n_2^i. \tag{A10}$$

That is to say, any choice of (p_1^i, γ_1^i) that leaves u_1^i unchanged in (A1) generates the same profits for the platform. One implication of this is that a platform has a continuum of best responses to its rival's choice of tariffs. If I show that platform i's profits are concave in (u_1^i, u_2^i), then I have done what is required.

To do this, I need to derive platform i's market shares as a function of its offered utilities and the rival's tariffs. Similarly to expressions (A2)–(A3), we have

$$n_1^i = \frac{1}{2} + \frac{u_1^i - ((\alpha_1 - \gamma_1^j)(1 - n_2^i) - p_1^j)}{2t_1}$$
$$n_2^i = \frac{1}{2} + \frac{u_2^i - ((\alpha_2 - \gamma_2^j)(1 - n_1^i) - p_2^j)}{2t_2}.$$

Solving this pair of equations gives the following explicit expressions for market shares:

$$n_1^i = \frac{1}{2} + \frac{(\alpha_1 - \gamma_1^j)(u_2^i - (\frac{\alpha_2 - \gamma_2^j}{2} - p_2^j)) + 2t_2(u_1^i - (\frac{\alpha_1 - \gamma_1^j}{2} - p_1^j))}{4t_1t_2 - (\alpha_1 - \gamma_1^j)(\alpha_2 - \gamma_2^j)}$$
$$n_2^i = \frac{1}{2} + \frac{(\alpha_2 - \gamma_2^j)(u_1^i - (\frac{\alpha_1 - \gamma_1^j}{2} - p_1^j)) + 2t_1(u_2^i - (\frac{\alpha_2 - \gamma_2^j}{2} - p_2^j))}{4t_1t_2 - (\alpha_1 - \gamma_1^j)(\alpha_2 - \gamma_2^j)}.$$

Notice that these market share functions are linear in utilities, and therefore profit in (A10) is quadratic in utilities. Profit is concave in utilities if (i) $\partial^2\pi^i/\partial(u_1^i)^2 < 0$ and (ii) the determinant of the matrix of second derivatives of π is positive. Tedious calculations show (i) holds if

$$4t_1t_2 > (\alpha_1 + \alpha_2)(\alpha_2 - \gamma_2^j) + (\alpha_1 - \gamma_1^j)(\alpha_2 - \gamma_2^j),$$

which is true when the maintained assumption (8) holds and $0 \leq \gamma_i^j \leq 2\alpha_i$. Also, (ii) holds if

$$16t_1t_2 > \left(2\alpha_1 + 2\alpha_2 - \gamma_1^j - \gamma_2^j\right)^2,$$

which holds under the same assumptions. I deduce that a platform's choice problem is concave.

Next, I characterize the symmetric equilibria. Suppose the two platforms choose the same pair of per-user charges (γ_1, γ_2). From (A8) and using (A4) and (A7), the first-order condition $\partial\pi^i/\partial p_1^i = 0$ evaluated at the symmetric fixed charges $p_1^j = p_1^i$ and $p_2^j = p_2^i$ yields

$$p_1 = f_1 + t_1 - \frac{\gamma_1 + \gamma_2}{2} - \frac{\alpha_2 - \gamma_2}{t_2}\left(\alpha_1 + p_2 - f_2 + \frac{\gamma_2 - \gamma_1}{2}\right),$$

and similarly

$$p_2 = f_2 + t_2 - \frac{\gamma_1 + \gamma_2}{2} - \frac{\alpha_1 - \gamma_1}{t_1}\left(\alpha_2 + p_1 - f_1 + \frac{\gamma_1 - \gamma_2}{2}\right).$$

Solving this pair of simultaneous equations in p_1 and p_2 implies that equilibrium fixed fees given the per-user charges (γ_1, γ_2) satisfy expressions (17). Secondly, when the fixed fees on each platform are given by (17), one can show using (A8) and (A9) that a platform's optimal choice of per-user fees is (γ_1, γ_2). Therefore, since the first-order approach is valid in this problem, the result is established. *Q.E.D.*

References

Anderson, S. and Coate, S. "Market Provision of Broadcasting: A Welfare Analysis." *Review of Economic Studies*, Vol. 72 (2005), pp. 947–972.

Armstrong, M. "The Theory of Access Pricing and Interconnection." In M. Cave, S. Majumdar, and I. Vogelsang, eds., *Handbook of Telecommunications Economics: Vol. I*. Amsterdam: North-Holland, 2002.

© RAND 2006.

——— AND VICKERS, J. "Competitive Price Discrimination." *RAND Journal of Economics*, Vol. 32 (2001), pp. 579–605.

——— AND WRIGHT, J. "Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts." *Economic Theory*, forthcoming.

BEGGS, A. "Mergers and Malls." *Journal of Industrial Economics*, Vol. 42 (1994), pp. 419–428.

CAILLAUD, B. AND JULLIEN, B. "Chicken and Egg: Competition Among Intermediation Service Providers." *RAND Journal of Economics*, Vol. 34 (2003), pp. 309–328.

DOBSON, P. AND WATERSON, M. "Countervailing Power and Consumer Prices." *Economic Journal*, Vol. 107 (1997), pp. 418–430.

DUKES, A. AND GAL-OR, E. "Negotiations and Exclusivity Contracts for Advertising." *Marketing Science*, Vol. 22 (2003), pp. 222–245.

GABSZEWICZ, J., LAUSSEL, D., AND SONNAC, N. "Press Advertising and the Ascent of the 'Pensé Unique.'" *European Economic Review*, Vol. 45 (2001), pp. 641–651.

KAISER, U. AND WRIGHT, J. "Price Structure in Two-Sided Markets: Evidence from the Magazine Industry." *International Journal of Industrial Organization*, Vol. 24 (2006), pp. 1–28.

KLEMPERER, P. AND MEYER, M. "Supply Function Equilibria in Oligopoly Under Uncertainty." *Econometrica*, Vol. 57 (1989), pp. 1243–1277.

ROCHET, J.-C. AND TIROLE, J. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association*, Vol. 1 (2003), pp. 990–1029.

——— AND ———. "Two-Sided Markets: A Progress Report." *RAND Journal of Economics*, Vol. 37 (2006), pp. 645–667.

RYSMAN, M. "Competition Between Networks: A Study of the Market for Yellow Pages." *Review of Economic Studies*, Vol. 71 (2004), pp. 483–512.

SCHULZ, N. AND STAHL, K. "Do Consumers Search for the Highest Price? Oligopoly Equilibrium and Monopoly Optimum in Differentiated-Products Markets." *RAND Journal of Economics*, Vol. 27 (1996), pp. 542–562.

SMITH, H. AND HAY, D. "Streets, Malls and Supermarkets." *Journal of Economics and Management Strategy*, Vol. 14 (2005), pp. 29–59.

STAHL, K. "Location and Spatial Pricing Theory with Nonconvex Transportation Cost Schedules." *Bell Journal of Economics*, Vol. 13 (1982), pp. 575–582.

VAN RAALTE, C. AND WEBERS, H. "Spatial Competition with Intermediated Matching." *Journal of Economic Behavior and Organization*, Vol. 34 (1998), pp. 477–488.

WRIGHT, J. "Access Pricing Under Competition: An Application to Cellular Networks." *Journal of Industrial Economics*, Vol. 50 (2002), pp. 289–316.

© RAND 2006.

A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue

JOHN MUSACCHIO *

Technology and Information Management, University of California, Santa Cruz

GALINA SCHWARTZ

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

JEAN WALRAND

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

Abstract

We address whether local ISPs should be allowed to charge content providers, who derive advertising revenue, for the right to access end-users. We compare two-sided pricing where such charges are allowed to one-sided pricing where they are prohibited. By deriving provider equilibrium actions (prices and investments), we determine which regime is welfare-superior as a function of a few key parameters. We find that two-sided pricing is more favorable when the ratio between parameters characterizing advertising rates and end-user price sensitivity is either low or high.

1 Introduction

Today, an Internet service provider (ISP) charges both end-users who subscribe to that ISP for their last-mile Internet access as well as content providers that are directly connected to the ISP. However, an ISP generally does not charge content providers that are not directly attached to it for delivering content to end-users. One of the focal questions in the network neutrality policy debate is whether these current charging practices should continue and be mandated by law, or if ISPs ought to be allowed to charge all content providers that deliver content to the ISP's end-users. Indeed the current network neutrality debate began when the CEO of AT&T suggested that such charges be allowed (see Whitacre, 2005).

To address this question, we develop a two-sided market model of the interaction of ISPs, end-users, and content providers. The model is closely related to the existing two-sided markets literature as we detail later in this section. In our model, the ISPs play the "platform" role that intermediates the two sides: content providers and end-users. We model a "neutral"

* Contact author: John Musacchio, Technology and Information Management Program, Jack Baskin School of Engineering, University of California Santa Cruz, 1156 High Street, MS: SOE 3. E-mail: johnm@soe.ucsc.edu
This work was supported by the National Science Foundation under Grants ITR/ANIR 0331659 and NeTS-FIND 0627161 and 0626397. We sincerely thank the anonymous reviewers and the editors, Julian Wright and in particular Marius Schwartz, whose thoughtful and detailed suggestions for improving this paper are greatly appreciated.

network as a regime in which ISPs are allowed to charge only content providers that buy their Internet access from them. We assume that with such a charging structure, the ISPs compete on price to attract content providers to buy their access from them, driving these prices to the ISP's costs. By normalizing the price content providers pay ISPs to be net of ISP connection cost, we can model a neutral network as having a zero price for content providers. This could be viewed as a market with one-sided pricing where ISPs only charge end-users. Conversely, two-sided pricing (corresponding to a "non-neutral" network) allows all ISPs to charge all content providers, thus permitting the ISPs to extract revenues from both content providers and end-users. Our contribution is the development of a model that relates the level of end-user usage to the investment decisions of providers of both types under both network regimes. We use our model to compare the welfare of both regimes.

The question we address in this work is part of the larger debate on network neutrality, which includes diverse issues such as whether service differentiation should be allowed, or whether charges for content constitute an impingement of freedom of speech (see Odlyzko, 2008; and Chong, 2007). In 2006 there was a considerable divergence of opinions on the subject of net-neutrality. Indeed, the issue was intensely debated by law and policy makers, and the imposition of restrictive network regulations on Internet service providers (ISPs) in order to achieve network neutrality seemed likely. Recently, the situation has begun to change. In June 2007, the Federal Trade Commission (FTC) issued a report, forcefully stating the lack of FTC support for network neutrality regulatory restraints, and warning of "potentially adverse and unintended effects of regulation" (FTC, 2007, p.159). Similarly, on September 7, 2007 the Department of Justice issued comments "cautioning against premature regulation of the Internet," (DOJ, 2007). Thus by the fall of 2007, the imminent threat of new regulation has diminished somewhat, and a consensus favoring the current (or unregulated) network regime seemed to have emerged. Still, the debate about network neutrality is far from over. Many prominent members of Congress support new network neutrality legislation, including two leading candidates for president in the 2008 election. We do not attempt to address all of the questions in the network neutrality debate. We only study the issue of whether ISPs ought to be allowed to charge content providers for accessing the end-users.

Our model is based on the ideas of two-sided markets, and there is a large literature on the subject. For a survey of two-sided markets, see for example, Rochet and Tirole (2006) and Armstrong (2006). The two-sided market literature studies markets in which a platform provider needs to attract two types of participants, and the presence of more of one type makes the platform more valuable to the other type. Rochet and Tirole (2006) define the market as two-sided, when the volume of realized transactions depends not solely on the aggregate price level charged to the two parties, but also on how this aggregate is divided between them. Using the two-sided market parlance, the ISPs of our model provide the platform, while end-users are one type of participant and content providers are the other type. As will become clear when we describe the details of our model, the end-users "single-home" or connect to one ISP. In a two-sided network, the content providers are forced to "multi-home" or pay multiple ISPs for delivering their content (see for example Armstrong, 2006 on "competitive bottlenecks"). In contrast, the content providers in a network with one-sided pricing "single-home" or pay just one ISP for connectivity. However, a content provider that pays one ISP enjoys the benefits of having connectivity to all the ISPs, because all the ISPs are interconnected. This is in contrast to most two-sided market models where the participants of one platform do not benefit from the presence of

participants of another platform, that is, Microsoft Xbox users do not benefit from more game makers writing Nintendo Wii games. This is an important structural difference.

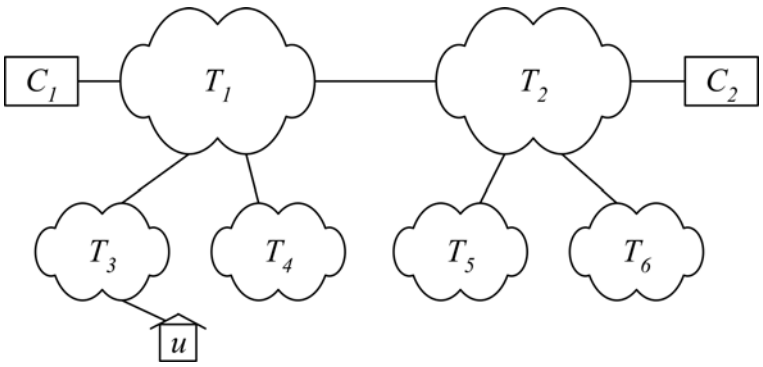

Figure 1: ISPs $\{T_1, \dots, T_6\}$, content providers $\{C_1, C_2\}$ and end-user u.

Other researchers have used the ideas of two-sided markets to study network neutrality. Hermalin and Katz (2007) model network neutrality as a restriction on the product space, and consider whether ISPs should be allowed to offer more than one grade of service. Hogendorn (2007) studies two-sided markets where intermediaries sit between "conduits" and content providers. In his context, net-neutrality means content has open access to conduits where an "open access" regime affords open access to the intermediaries. Weiser (2007) discusses policy issues related to two-sided markets.

The novelty of our model over other work in the two-sided market literature is our explicit modeling of platform investment choices. In the existing literature, the platform incurs the cost of serving the users, which usually is assumed linear in the number of users, but does not make an investment choice.

To introduce the notation and illustrate the arguments for and against the imposition of regulations that prohibit the ISPs from engaging in two-sided pricing, consider the network shown in Figure 1. The figure shows end-user u, ISPs $\{T_1, \dots, T_6\}$, and content providers $\{C_1, C_2\}$. In this network, T_1 and T_2 are transit ISPs (that is, the ISPs who operate portions of the Internet backbone), and $T_3 - T_6$ are local ISPs (i.e, residential ISPs who provide last-mile access for end-users). In the figure, the content providers are attached to a transit ISP whereas a typical end-user u is attached to a local ISP.

In our model, in a market with one-sided pricing, end-users and content providers pay only for their direct access. The transit ISPs charge the local ISPs for carrying their traffic. The transit ISPs typically enter in peering agreements under which they agree to carry each other's traffic, usually free of charge. The transit ISPs charge the content providers for their attachment. We model one-sided pricing ("neutral" network), as a case where the local ISPs are prohibited from charging content providers not buying access directly from them. Thus, with one-sided pricing, content provider C_1 pays ISP T_1 for its access to the Internet but does not pay any of the other ISPs. In contrast, with two-sided pricing ("non-neutral" network) ISP T_3 is able to charge content provider C_1 for carrying its content to end-user u, even though C_1 is not directly attached to T_3.

So far, we have distinguished between two classes of access providers: transit ISPs and local ISPs. The local ISPs provide the last-mile link to end-users (that is, residential consumers). The transit ISPs have no residential customers and serve only content providers, though there are notable examples of companies operating both transit and local networks. The core economic distinction between these ISP classes is that local ISPs are thought to

have more market power. This market power is a result of substantial scale economies in providing mass-market (residential) broadband access. These scale economies limit (today at least) the number of local ISPs in a specific residential area. In particular, today's end-user choice of local ISP is limited to one to three at most. To reflect this market power, in our model each local ISP has a monopoly over its user base.

In reverse, "first-mile" access for major content providers is supplied by transit ISPs, who are believed to be considerably more competitive, because individual content providers (i) generate a lot of traffic and therefore can economically support dedicated links from several suppliers, and (ii) can pick to locate themselves at the points where the network access is the cheapest (that is, to attach directly to the backbone).

Our main purpose of distinguishing transit and local ISPs was to help explain why under today's institutions, local ISPs do not charge content providers. First, content providers can buy access at their end (competitively) from transit ISPs; second, typically, transit ISPs do not pay for access to local ISPs (enjoy settlement-free "peering"); and third, local ISPs do not receive any other payments from content providers. Thus under current practices, the local ISPs do not charge content providers, except perhaps in the case where a company operates both transit and local networks. Even in such a case, one could argue that the ISP only charges competitively for general access rather than for the right to reach the ISP's users in particular. This is because, for instance, a content provider always has the option to switch to a non-integrated competitive transit provider. Thus, in our further analysis we abandon the distinction between local ISPs and transit ISPs, and focus on local ISPs only, which from now on we simply call ISPs.

The arguments for and against net-neutrality having to do with local ISPs charging content providers can be summarized as follows. See ACLU (2006), Farber (2007), Felten (2006), Lessig[1], Owen and Rosston (2003), Sidak (2006), and Yoo (2006) for more elaboration on these points.

Against Neutrality: This line of reasoning is usually expressed by ISPs. Say that C_1 is a source of video streams that require a large bandwidth. ISP T_3 may argue that to accommodate the traffic originating from C_1, he must make substantial investments, which he may not be able to recover from the end-users who buy access directly from him. However, the content providers make additional advertising revenues when end-users consume new high-bandwidth services, which justify their investments. If the ISPs cannot get a share of these additional revenues, they will not invest as much to increase the network capacity. The situation causes poor network quality, which reduces end-user demand, which in turn leads to further reduction of incentives to invest for both provider types.

For Neutrality: This line of reasoning is usually expressed by content providers. If every ISP can charge the content provider C_i for the right to access its end-users, the ability of ISPs to charge content providers would increase dramatically. ISPs would then charge content providers more. In turn, this would reduce the investment incentives for content providers thus lowering content quality.

To sum up, both lines of reasoning argue that their preferred regime would generate higher overall welfare. Clearly, both sides cannot be right; a more detailed analysis is required to clarify the trade-off.

This paper explores how provider investments and revenues differ with network regime. We assume that the number of ISPs and content providers is fixed. That is, we do not

[1] "The Lessig Blog," http://www.lessig.org/blog/archives/003391.shtml

consider the longer term impact of regulations of the ISPs on provider entry incentives (that is, we assume the fixed structure of the network industry).

In section 2, we propose an economic model that relates the investments and prices to revenue for content providers and ISPs. We analyze two-sided pricing (non-neutral) in section 3.1 and we analyze one-sided pricing (neutral) in section 3.2. Section 4 is devoted to

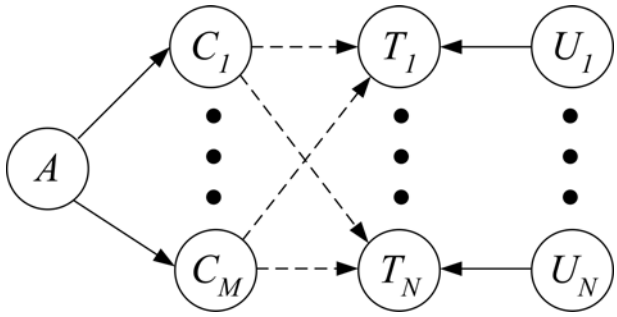

Figure 2: The direction of payments in the model. The dotted lines indicate payments made only with two-sided pricing ("non-neutral").

a comparison of the two regimes of the network. In section 5, we summarize our findings. The technical details of the analysis can be found in our working paper, see Musacchio *et al* (2009).

2 Model

Figure 2 illustrates our setting. In the model, there are M content providers and N ISPs. Each ISP T_n is attached to end-users U_n $(n = 1,2, \dots, N)$ and charges them p_n per click. The ISP T_n has a monopoly over its end-user base U_n. Thus, the end-users are divided between the ISPs, with each ISP having $1/N$ of the entire market. This assumption reflects the market power of local ISPs. Each ISP T_n charges each content provider C_m an amount equal to q_n per click. Content provider C_m invests c_m and ISP T_n invests t_n.

Recall from the Introduction that we measure q net of content providers' access payment (for network attachment), which is set at marginal cost due to the ISPs' competition. Accordingly, we measure the content provider per-user charges to advertisers (which we denote as a) net of the content provider's access payment.

We characterize usage of end-users U_n by the number of "clicks" B_n they make. Since Internet advertising is most often priced per click, clicks are a natural metric for expressing advertising revenue. It is a less natural metric for expressing ISP revenue from end-users, because ISPs do not charge users per click but rather base their charges on bits. However, it is convenient to use only one metric and argue that one could approximate one metric from knowledge of the other using an appropriate scaling factor. The rate B_n of clicks of end-users U_n depends on the price p_n but also on the quality of the network, which in-turn is determined by provider investments, as follows

$$B_n = \left\{\frac{1}{N^{1-w}}(c_1^v + \cdots + c_M^v)\left[(1-\rho)t_n^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w)\right]\right\} e^{-p_n/\theta} \tag{1}$$

where $\rho \in (0,1)$, $\theta > 0$, and $v, w \geq 0$ with $v + w < 1$. For a given network quality (the expression in the curly brackets) the rate of clicks exponentially decreases with price p_n.

The term $c_1^v + \cdots + c_M^v$ is the value of the content providers' investments as seen by a typical end-user. This expression is concave in the investments of the individual providers, and the interpretation is that each content provider adds value to the network. Also note that the structure of the expression is such that end-users value a network in which content is produced by numerous content providers higher than a network in which the content is provided by a single provider with the same cumulative investment. Our end-users' preference for content variety is similar to that of the classical monopolistic competition model by Dixit and Stiglitz (1977). The term in square brackets reflects the value of the ISP's investments for end-users. Clearly, users U_n value the investment made by their ISP, but they may also value the investments made by other ISPs. For instance, a user of one ISP might derive more value by having a better connection with users of another ISP. In our model, the parameter ρ captures this spill-over effect. When $\rho = 1$, end-users U_n value investments of all ISPs equally while when $\rho = 0$, they value only the investment of their ISP. When $\rho \in (0,1)$ end-users U_n value the investment of his ISP T_n more than investments of other ISPs $k \neq n$. The term ρ reflects the direct network effect of ISPs on end-users (not between them and content providers). This effect captures a typical network externality (see Thijssen, 2004 for a discussion of investment spill-over effects). The factor $1/N^{1-w}$ is a convenient normalization. It reflects the division of the end-user pool among N providers and it is justified as follows. Suppose there were no spill-over and each ISP were to invest t/N. The total rate of clicks should be independent of N. In our model, the total click rate is proportional to $(1/N^{1-w})(N(t/N)^w)$, which is indeed independent of N.

At this point, it is worth discussing why we have included certain features in the model. We chose to include N ISPs in the model, rather than just one, because there is a potential of free-riding in the two-sided pricing (non-neutral) case, which gives multiple ISPs the power to charge each content provider. This is because an ISP can increase his price to content providers and enjoy the additional revenue this increase causes, while the downside of inducing the content provider to invest less has to be borne by all of the ISPs. The magnitude of this effect will likely increase with N. We have also allowed for an arbitrary number of content providers, and included the spill-over term ρ, to study whether these features have a strong effect.

Returning to the model specification, the rate R_{mn} of clicks from end-users U_n to C_m is given by

$$R_{mn} = \frac{c_m^v}{c_1^v + \cdots + c_M^v} B_n. \tag{2}$$

Thus, the total rate of clicks for content provider C_m is given by

$$D_m = \sum_n R_{mn}. \tag{3}$$

For further intuition behind the expressions (1) and (3) for click rates, consider a symmetric case, with all providers investing equally ($t_n = t/N$ and $c_m = c/M$) and $\rho = 1$. Then, (1) and (3) reduce to

$$B_n(\text{symmetric case}) = \left\{\frac{1}{N} M^{1-v} c^v t^w\right\} e^{-p/\theta},\ D_m(\text{ symmetric case}) = \left\{\frac{1}{M^v} c^v t^w\right\} e^{-p/\theta}$$

respectively. Thus for both provider types, the click rates exponentially decrease with end-user price, increase with provider investments c and t, and exhibit decreasing return to investments. The only difference in the content provider and ISP click rate expressions are the effects of M and N. Recall that we included a normalization factor in (1) so that the same total investment split evenly across all ISPs should result in the same total click rate, regardless of N. On the other hand, we assumed that a user values variety in content, so that the same total investment split amongst multiple content providers is more valued than if it were split amongst fewer content providers. That is why B_n increases in M for fixed c, but D_m does not increase with N.

Leaving the symmetric case behind and returning to the general model, we assume that content providers charge a fixed amount a per click to the advertisers. Each content provider's objective is to maximize its profit which is equal to revenues from end-user clicks net of investment costs. Thus

$$\Pi_{Cm} = \sum_{n=1}^{N} (a - q_n) R_{mn} - \beta c_m \tag{4}$$

where the term $\beta > 1$ is the outside option (alternative use of funds c_m).

ISP T_n profit is

$$\Pi_{Tn} = (p_n + q_n) B_n - \alpha t_n \tag{5}$$

where $\alpha > 1$ is the outside option of the ISP. We assume providers of each type are identical and we will focus on finding symmetric equilibria for both one- and two-sided pricing.

3 The analysis of one- and two-sided pricing

To compare one-sided and two-sided pricing (neutral and non-neutral networks), we make the following assumptions.

- One-sided pricing (neutral network): In stage 1, each T_n simultaneously chooses (t_n, p_n). The price q_n charged to content providers is constrained to be 0. In stage 2, each C_m chooses c_m.
- Two-sided pricing (non-neutral network): In stage 1, each T_n simultaneously chooses (t_n, p_n, q_n). In stage 2, each C_m chooses c_m.

As we discussed in the Introduction, in one-sided pricing, the price q_n charged to content providers is zero by the following argument. Since content providers need only pay the ISP with which they are directly connected, the resulting price competition forces prices down to ISP costs. Thus the prices $\{q_n\}$ are zero after one normalizes them to be net of ISP cost. In contrast, in a network with two-sided pricing (non-neutral), all ISPs have the ability to charge each content provider. This permits ISPs to exercise their market power on content providers and allows them to extract a non-zero price.

In both cases, we assume that content providers observe ISP investments, and can subsequently adjust their investments based on the ISPs' choices. We justify this assumption by the difference in time and scale of the required initial investments. The investments of ISP tend to be longer-term investments in infrastructure, such as deploying networks of fibre-optic cable. Conversely, the investments of content-providers tend to be shorter-term and more ongoing in nature, such as development of content, making ongoing improvements to a search algorithm, or adding/replacing servers in a server farm.

3.1 Two-sided pricing

In a network with two-sided pricing (non-neutral network), each ISP chooses (t_n, p_n, q_n) and each content provider chooses c_m. To analyze this situation, we study how C_m chooses the optimal c_m for a given set of (t_n, p_n, q_n). We then substitute that value of c_m in the expression for Π_T and we optimize for (t_n, p_n, q_n).

The best choice for c_m given (t, p, q) maximizes

$$\begin{aligned}\Pi_{Cm} &= aD_m - \sum_n q_n R_{mn} - \beta c_m \\ &= N^{w-1} c_m^v \left[\sum_n (a - q_n)\left((1-\rho)t_n^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w)\right) e^{-p_n/\theta}\right] - \beta c_m.\end{aligned} \tag{6}$$

As a result of the cancelation between the denominator of R_{mn} (see (2)) and the expression for the click rate B_n (see (1)), the dependency of content provider m's revenue on the investments of other content providers disappears. Thus the expression for profit Π_{Cm} is independent of other content provider investments c_j, $j \neq m$. Therefore, each content provider's optimization is unaffected by the simultaneously made (but correctly anticipated in equilibrium) investment decisions of the other content providers. Assuming that the term in square brackets is positive, we find that

$$\beta c_m^{1-v} = vN^{w-1}\left[\sum_k (a - q_k)((1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w))e^{-p_k/\theta}\right] =: \beta c^{1-v}. \tag{7}$$

For that value of c_m, we find that

$$\Pi_{Tn} = MN^{w-1}(q_n + p_n)F_n e^{-p_n/\theta}\left(\frac{v}{N\beta}\right)^{v/(1-v)}\left[\sum_k (a - q_k)e^{-p_k/\theta}F_k\right]^{v/(1-v)} - \alpha t_n \quad (8)$$

where

$$F_n = (1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w) = \phi t_n^w + \frac{\rho}{N}\sum_{k\neq n} t_k^w \quad (9)$$

with

$$\phi := 1 - \rho + \frac{\rho}{N} < 1, \;\; if \;\; N > 1. \quad (10)$$

The ISP T_n chooses investment and prices (t_n, p_n, q_n) that maximize his profit given by equation (8). The simultaneous decisions of each of the ISPs affect each other, therefore in order to find a Nash equilibrium we need to identify a point where the best response functions intersect. Writing that the three corresponding partial derivatives of (8) are equal to zero, and then finding the symmetric intersection point of the best response functions, we find the following solution (see our working paper, Musacchio at al. (2008)):

$$p_n = p = \theta - a; \quad (11)$$

$$q_n = q = a - \theta\frac{v}{N(1-v)+v}; \quad (12)$$

$$t_n = t \;\; \text{with } (Nt)^{1-v-w} = x^{1-v}y^v e^{-(\theta-a)/\theta}; \quad (13)$$

$$c_m = c \;\; \text{with } c^{1-v-w} = x^w y^{1-w} e^{-(\theta-a)/\theta}; \quad (14)$$

$$\Pi_{Cm}^{1-v-w} = \Pi_C^{1-v-w} := \left(\frac{\theta v(1-v)}{N(1-v)+v}\right)^{1-v-w} x^w y^v e^{-(\theta-a)/\theta}; \quad (15)$$

$$\Pi_{Tn}^{1-v-w} = \Pi_T^{1-v-w} := \left(\frac{M\theta(N(1-v) - wN\phi(1-v) - vw)}{N(N(1-v)+v)}\right)^{1-v-w} x^w y^v e^{-(\theta-a)/\theta}; \quad (16)$$

$$\frac{\Pi_C}{c} = \frac{\beta(1-v)}{v}; \quad (17)$$

$$\frac{\Pi_T}{t} = \frac{\alpha}{w}\left[\frac{N(1-v)}{N\phi(1-v)+v} - w\right]; \quad (18)$$

$$B^{1-v-w} = M^{1-v-w}x^w y^v e^{-(\theta-a)/\theta} \quad (19)$$

where $B := \sum_n B_n = \sum_m D_m$ is the total click rate and

$$x := \frac{M\theta w}{\alpha}\frac{N\phi(1-v)+v}{N(1-v)+v} \;\; \text{and} \;\; y := \frac{\theta}{\beta}\frac{v^2}{N(1-v)+v}. \quad (20)$$

3.2 One-sided pricing

A network with one-sided pricing (neutral network) is similar to one with two-sided pricing (non-neutral), except that $q_n = 0$ as we argued in section 2 for $n = 1, \dots, N$. The best choice of c given $\{q_n = 0, p_n, t_n\}$ is such that

$$\beta c_m^{1-v} = vN^{-1}\left[\sum_k a((1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w))e^{-p_k/\theta}\right] =: \beta c^{1-v}.$$

For that value of c_m, we find that

$$\Pi_{Tn} = MN^{-1}p_n F_n e^{-p_n/\theta}(\frac{v}{\beta})^{v/(1-v)}\left[\sum_k ae^{-p_k/\theta}F_k\right]^{v/(1-v)} - \alpha t_n \tag{21}$$

where

$$F_n = \phi t_n^w + \frac{\rho}{N}\sum_{k\neq n} t_k^w.$$

The ISP T_n chooses investment and price (t_n, p_n) that maximize the above expression. We find a symmetric Nash equilibrium by writing that the two corresponding partial derivatives of (21) with respect to a single ISP actions are zero, and that the other ISPs make the same actions, and solving all of the resulting equations. This analysis leads to the following solutions (see our working paper, Musacchio at al. (2008)):

$$p_n = p_0 := \frac{\theta N(1-v)}{N(1-v)+v}; \tag{22}$$

$$q_m = 0; \tag{23}$$

$$t_n = t_0 \text{ where } (Nt_0)^{1-v-w} = x^{1-v}y_0^v e^{-p_0/\theta}; \tag{24}$$

$$c_m = c_0 \text{ where } c_0^{1-v-w} = x^w y_0^{1-w} e^{-p_0/\theta}; \tag{25}$$

$$\Pi_{Cm}^{1-v-w} = \Pi_{C0}^{1-v-w} := (a(1-v))^{1-v-w}x^w y_0^v e^{-p_0/\theta}; \tag{26}$$

$$\Pi_{Tn}^{1-v-w} = \Pi_{T0}^{1-v-w} := \left(\frac{M\theta(N(1-v) - wN\phi(1-v) - wv)}{N(N(1-v)+v)}\right)^{1-v-w} x^w y_0^v e^{-p_0/\theta}; \tag{27}$$

$$\frac{\Pi_{C0}}{c_0} = \frac{\beta(1-v)}{v}; \tag{28}$$

$$\frac{\Pi_{T0}}{t_0} = \frac{\alpha}{w}\left[\frac{N(1-v)}{N\phi(1-v)+v} - w\right]; \tag{29}$$

$$B_0^{1-v-w} = M^{1-v-w}x^w y_0^v e^{-p_0/\theta} \tag{30}$$

where B_0 is the total click rate, x is given in (20), and

$$y_0 := \frac{av}{\beta}. \tag{31}$$

4 Comparison

In this section, we compare the Nash equilibria of the two regimes. In section 4.1, we derive expressions for the welfare of end-users, and the ratio of social welfare with one- vs. two-sided pricing. In section 4.2, we demonstrate that the return on investments is the same in both regimes. In section 4.3, we compare the profits and social welfare of the two regimes for a range of parameters.

4.1 User welfare and social welfare

Before proceeding we define the following notation.

$$\pi := \frac{v}{N(1-v)+v} \quad \text{and} \quad \delta := \frac{a}{\theta}. \tag{32}$$

In order to compute end-user welfare, we use the total click rate to proxy the aggregate user demand. This enables us to calculate consumer surplus and use it to measure end-user welfare. We compute the consumer surplus by taking the integral of the demand function from the equilibrium price to infinity. This integral is taken with the investment levels of content providers and ISPs fixed. We find

$$W_U(\text{two}-\text{sided}) = M\theta x^{w/(1-v-w)} y^{v/(1-v-w)} e^{-\frac{\theta-a}{\theta(1-v-w)}}.$$

The expression for the one-sided case is the same, but with y exchanged for y_0 and $\theta - a$ in the exponent exchanged with p_0. The ratio of the social welfare with one- vs. two-sided pricing has the form

$$\frac{W_U(\text{one}-\text{sided}) + N\Pi_T(\text{one}-\text{sided}) + M\Pi_C(\text{one}-\text{sided})}{W_U(\text{two}-\text{sided}) + N\Pi_T(\text{two}-\text{sided}) + M\Pi_C(\text{two}-\text{sided})}$$

$$= \frac{1+\delta(1-v)+(\pi/v)(N(1-v)-wN\phi(1-v)-wv)}{1+\pi(1-v)+(\pi/v)(N(1-v)-wN\phi(1-v)-wv)} [(\delta/\pi)^v e^{\pi-\delta}]^{1/(1-w-v)}.$$

4.2 Return on investment

We use the definition $\delta = a/\theta$ and equations (11), (12), (22), and (23) to relate prices in the one- and two-sided pricing regimes. This permits us to formulate the following proposition.

Proposition 1: The *price* p *end-users pay in the two-sided pricing case is given by*

$$p = \theta(1-\delta) = \theta - a.$$

Also, we note that

$$p + q = p_0 = \theta(1-\pi).$$

Moreover,

$$\frac{\Pi_C}{c} = \frac{\Pi_{C0}}{c_0} \quad and \quad \frac{\Pi_T}{t} = \frac{\Pi_{T0}}{t_0}.$$

Thus, from Proposition 1, the total revenue per click of the ISPs is the same in both regimes and so are the rate of return on investments of the content and ISPs.

Despite the fact that the rate of return on investments are the same in both regimes, as we will see in the next subsection, the size of those investments and resulting profits may be quite different across regimes.

4.3 Comparative statics

Dividing the expressions for one-sided pricing by the corresponding expressions for the two-sided pricing case, we define ratios such as

$$r(\Pi_C) := \left(\frac{\Pi_C(\text{ one} - \text{sided })}{\Pi_C(\text{two} - \text{sided })}\right)^{1-v-w}$$

where $\Pi_C(\ two - sided\ )$ is the profit per content provider in the two-sided case as expressed in (15) and $\Pi_C(\ one - sided\ )$ is the profit per content provider with one-sided pricing (26). We define $r(c), r(t),$ and $r(\Pi_T)$ similarly. We find

$$r(\Pi_T) = r(t) = r(B) = \left(\frac{\delta}{\pi}\right)^{v} e^{\pi-\delta}; \tag{33}$$

$$r(\Pi_C) = r(c) = \left(\frac{\delta}{\pi}\right)^{1-w} e^{\pi-\delta}. \tag{34}$$

Figure 3 shows the ratios of profts with one- vs. two-sided pricing for both content providers and ISPs. Figure 4 shows the same ratios, but for different values of v and w. The figures show that for small or large values of a/θ, the ratio of advertising revenue per click to the constant characterizing price sensitivity of end-users, two-sided pricing is preferable to both content providers and ISPs. (Here we say "preferable" in that the profits are larger, though we have seen that the rate of return on investments are the same.) For mid range values of a/θ, one-sided pricing is preferable to both, though the values of a/θ where the transition between one-sided being preferable to two-sided are not exactly the same for content providers and ISPs. Furthermore, as N, the number of ISPs increases, the range of a/θ values for which one-sided pricing is superior increases, while also the degree by which it is superior (in terms of revenues to content providers and ISPs) increases.

These results can be explained by the following reasoning. When a/θ is large, the content providers' revenues from advertising are relatively high, and the ISPs' revenue from end-users are relatively low. Because of this, the ISPs incentives to invest are suboptimal (too low relative to the socially optimal ones), unless they can extract some of the content providers' advertising revenue by charging the content providers. Thus in the one-sided pricing case, the ISPs under invest, making the rewards for them as well as content providers less than it could have been with two-sided pricing.

It is important to note that when a/θ is larger than 1, the price p charged end-users becomes negative in the two-sided case as can be seen from (11). If end-users were actually paid to click, intuition suggests that they would click an unbounded amount and therefore our exponential model of demand (1) would not be valid in this region of a/θ. However, one could interpret price p to be net any variable costs ν to end users – similar to how we

define q.[2] With this interpretation, the price p could be negative while the actual prices users see positive, so long as $|p| < \nu$. We therefore show numerical results in our plots for a/θ as large as 1.2.

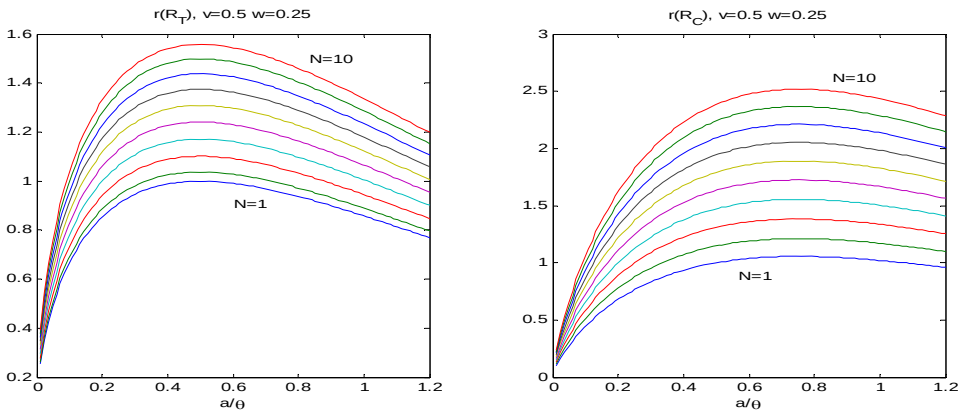

Figure 3: The ratios of profits ($v = 0.5, w = 0.25$) for different values of N.

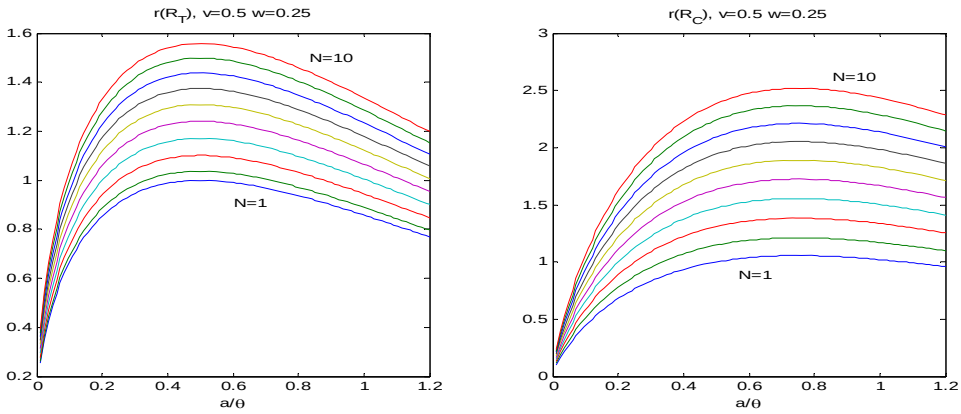

Figure 4: The ratios of profits ($v = 0.25, w = 0.5$) for different values of N.

When a/θ is very small, the content providers' advertising revenue is relatively low, and the ISP's end-user revenue is relatively high. In order to get the content providers to invest adequately, the ISPs need to pay the content providers. That is why for small enough a/θ the price q is negative (see Figure 5), representing a per click payment from the ISPs to the content providers.

Interestingly, even in the two-sided pricing case, our content providers obviously get some share of the surplus generated jointly by them and the ISPs. This is in contrast to the multi-homing case of Armstrong (2006) (which is roughly analogous to our two-sided pricing case), where the surplus is fully extracted from group-2 agents (content providers in our case) – the entire surplus is shared between the platform and the end-users. We do not have this extreme result, because in our model content providers invest *after* ISPs announce

[2] If we suppose such a normalization, it may be more natural to express end-user demand in terms of the price p+ν actually seen by users. However, supposing demand was exponential in p+ν rather than exponential in p would only change expression (1) for click rate by a constant factor. This constant factor disappears in any expression comparing the relative revenues or welfares of the two regimes.

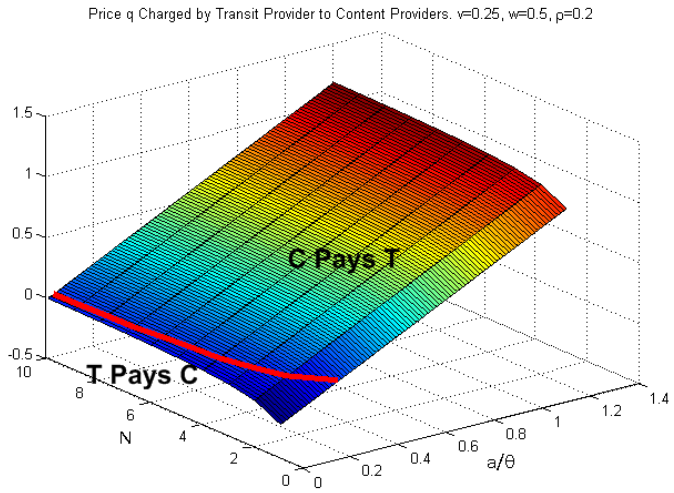

Figure 5: The price q (that the ISPs charge the content providers).

prices. Thus, ISP commitment to the declared prices permits content providers to retain a positive fraction of the surplus.

Finally, when a/θ is in the intermediate range, in between the two extremes, both content providers and ISPs have adequate incentive to invest. However, another effect comes into play – ISP free riding becomes an important factor when N is large. As N increases in the two-sided pricing case there are more ISPs that levy a charge against each content provider. As the price ISPs charge content providers increases, it becomes less attractive for content providers to invest. Thus, an ISP choosing the price to charge content providers is balancing the positive effect of earning more revenue per click from content providers versus the negative effect of having fewer clicks because the content providers have reduced their investment. But each ISP sees the entire gain of raising its price, but the loss is borne by all N ISPs. Consequently, the ISPs overcharge the content providers in Nash equilibrium, and the degree of this overcharging increases with N. This is analogous to the tragedy of the commons where people overexploit a public resource. Another perhaps more direct analogy is the "castles on the Rhine effect" where each castle owner is incentivized to increase transit tolls on passing traffic excessively by ignoring the fact that the resulting reduction in traffic harms not only him, but also other castle owners. When all castles do the same, the traffic on the Rhine decreases (Kay, 1990). The extent of this negative externality, and hence the degree of over-charging, increases with N.

Figure 6 shows a three dimensional plot of the ratio of social welfare for one- vs. two-sided pricing. The plot shows how the ratio changes for different N and a/θ. The second panel of Figure 6 is a simplified version of the first panel. It depicts only the boundaries in the parameter space where one-sided pricing is preferable to two-sided and vice versa.

It is also worthwhile to note that spill-over parameter ρ and the number of content providers M do not appear in the expression for the ratio of content provider revenue between the two regimes (34) nor do they appear in the ratio of revenues for ISPs (33). This is in spite of the fact that ρ and M do appear in the expressions for both the one-sided and two-sided pricing equilibria. This suggests that the spill-over effect and number of content providers have little or no effect on the comparative welfare of the two regimes.

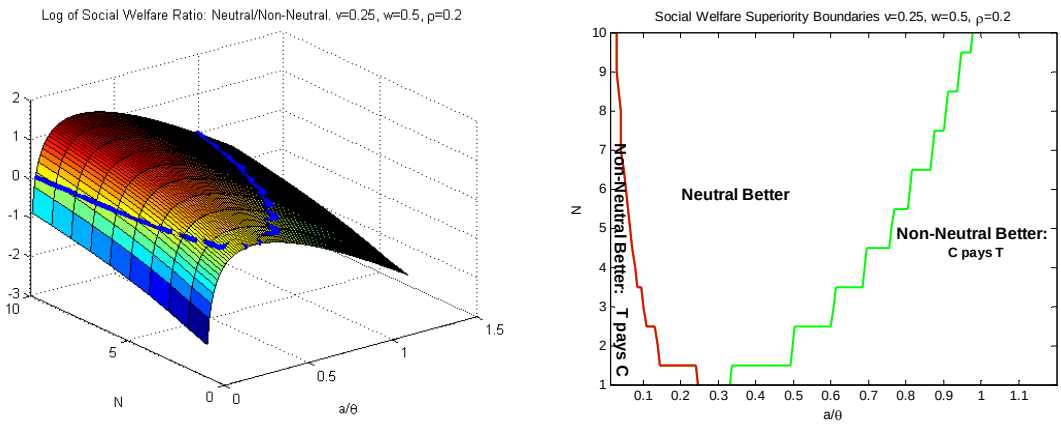

Figure 6: Left: The log of social welfare ratio (one-sided to two-sided). Right: Regions of social welfare superiority (one-sided vs. two-sided).

5 Conclusions

Our analysis seeks to address whether content providers should pay ISP(s) only for the right to access the network at their end, or whether other ISPs (that is, the ISPs other than the one through which a specific content provider is attached to the network) may also charge for the "right" to reach end-users. We call the former "one-sided pricing", and the latter "two-sided pricing". We suggest that the imposition of regulation that precludes the ISPs from two-sided pricing corresponds to a "neutral" network, while the practice of two-sided pricing corresponds to a "non-neutral" network.

We study how each pricing regime affects investment incentives of transit and content providers. We show that parameters such as advertising rate, end-user price sensitivity, and the number of ISPs influence whether one- or two-sided pricing achieves a higher social welfare. From our results, when the ratio of advertising rates to the constant characterizing price sensitivity is an extreme value, either large or small, two-sided pricing is preferable. If the ratio of advertising rates to the constant characterizing price sensitivity is not extreme, then an effect like the "castles on the Rhine effect" becomes more important. ISPs in a two-sided pricing regime have the potential to over charge content providers, and this effect becomes stronger as the number of ISPs increases.

In our comparison of one- and two-sided pricing we assume that the ISPs choose their strategy first and that the content providers follow. We justify this assumption by the difference in the time and scale of the required initial investments. While we feel this is a valid assumption, it is worth noting that if we had modelled the game such that the content providers and ISPs reverse their order of play, the game would not have a meaningful equilibrium. (In such a reversed model with two-sided pricing, the ISPs would charge an unbounded price to content providers after they had committed to their content investments. This is because at this point the content providers would have no action left to take in the game. This of course would lead to content providers investing zero in the first stage of the game.)

We make other key modeling choices, such as that the ISPs have local monopolies over their users. We believe that if we had studied a model where each ISP is a duopolist, which better models the degree of choice most end-users have today, our results would have been qualitatively similar. However, there are a number of competing effects such a model may

introduce. First, such a scenario would reduce the market power of ISPs over end-users, thus reducing the revenue the ISPs could extract from them. Since two-sided pricing provides ISPs with another source of revenue to justify their investments, this effect would tend to increase the parameter region for which two-sided pricing is social welfare superior. Second, a duopolist competing on product quality invests more than a monopolist, so this would tend to increase efficiency of one-sided pricing. Third, if the model were changed from having N to $2N$ ISPs, then the free riding or "castles on the Rhine" effect would grow, tending to reduce the welfare of the two-sided pricing case. The net effect of all these individual effects would of course depend on the detailed specifications of such a model.

Another assumption we make is that all ISPs have identical payoff functions. This would not be the case if the ISPs served different population sizes or different population densities for instance. Our assumption makes the analysis much easier, but we feel that if this assumption were relaxed, the results would be qualitatively similar though closed form solutions might not exist. We feel the same way about our assumption that content providers have identical payoff functions.

Our two-sided pricing model assumes that in-bound traffic to a local ISP could be identified as originating at a particular content provider, in order for the ISP to levy the appropriate charge to the content provider. This assumption would not strictly hold if content providers had some way of reaching end-users of an ISP without paying this ISP for end-user traffic. For instance, if there were a second ISP that enjoyed settlement free peering with the first ISP, the content provider could route its traffic through the second ISP and thus avoid the first ISP charge for end-user access. This strategy may be facilitated by the fact that the end-users of both ISPs send traffic to each other, and perhaps the traffic from the content provider could be masked in some way to look like traffic originating from the second ISP's end-users. However, the communication protocols of the Internet require that packets be labeled with the origin (IP) address. It seems unlikely today that a large content provider could have the origin addresses of its traffic falsified in a way that would both prevent ISPs from being able to charge the content provider while still enabling end-users to send traffic in the reverse direction back to the content provider. However, it is certainly possible that technology would be developed to enable such a strategy in the future, especially if there were an economic incentive for developing it.

There are several assumptions we make that might be viewed as limitations of our model. First, we fix the number of network providers, independent of the network regime. This assumption is realistic in the short-run, but in the long-run the number of providers entering the network industry is likely to differ with regime. Second, we do not consider heterogeneity in the providers nor in the end-users. Third, we assume full commitment to the declared prices, that is the ISPs cannot later adjust the prices that they declared initially. We also have not modeled the price content providers charge advertisers as a decision variable, but we have modeled the price ISPs charge end-users. Though we feel that the present model correctly captures focal features of the underlying network environment, in future work we plan to extend our model to permit endogenous industry structure, that is, we will study how network regime affects the number of content providers and ISPs.

6 References

ACLU (2006) "Internet Freedom and Innovation at Risk: Why Congress Must Restore Strong Net Neutrality Protection," ACLU Press Release, September 9 2006. http://www.aclu.org/freespeech/internet/26829res20060922.html

Armstrong, M. (2006) "Competition in Two Sided Markets," *RAND Journal of Economics*, 37: 668-691.

Chong, R. (2007) "The 31 Flavors of the Net Neutrality Debate: Beware The Trojan Horse," *Advanced Communications Law and Policy Institute, Scholarship Series*, New York Law School, December 2007.

Dixit, A. and J. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67: 297-308.

Department of Justice (2007) "Comments on Network Neutrality in Federal Communications Commission Proceeding," Press Release, September 7, 2007. http://www.usdoj.gov/opa/pr/2007/September/07_at_682.html

Farber, D. and M. Katz (2007) "Hold Off On Net Neutrality," *Washington Post*, Jan. 22.

Felten, E. (2006) "The Nuts and Bolts of Network Neutrality," Center for Information Technology Policy, Princeton, July 6, 2006. http://itpolicy.princeton.edu/pub/neutrality.pdf

Federal Trade Commission (2007) "Broadband Connectivity Competition Policy," Report, June 2007.

Hermalin, B. and M. Katz (2007) "The Economics of Product Line Restrictions With an Application to the Network Neutrality Controversy," *Information Economics and Policy*, 19: 215-48.

Hogendorn, C. (2007) "Broadband Internet: Net Neutrality Versus Open Access," *International Economics and Economic Policy*, 4: 185-208.

Kay, J. A. (1990) "Tax Policy: A Survey," *Economic Journal*, 100: 18-75.

Musacchio, J., G. Schwartz and J. Walrand (2009) "A Two-Sided Market Analysis of Provider Investment Incentives With an Application to the Net-Neutrality Issue," Technical Report UCSC-SOE-09-08, School of Engineering, University of California, Santa Cruz, March 2009. http://www.soe.ucsc.edu/ johnm/Publications/neutrality.pdf

Odlyzko, A. (2008) "Network neutrality, Search Neutrality, and the Never-Ending Conflict Between Efficiency and Fairness in Markets," *Review of Network Economics*, 8: 40-60.

Owen, B. and G. Rosston (2003) "Local Broadband Access: *Primum Non Nocere or Primum Processi?* A Property Rights Approach," Stanford Law and Economics Olin Working Paper No. 263.

Rochet, J.-C. and J. Tirole (2006) "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 37: 645-667.

Sidak, G. (2006) "A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet," *Journal of Competition Law & Economics*, 2: 349-474.

Thijssen, J. (2004) *Investment Under Uncertainty, Coalition Spillovers and Market Evolution in a Game Theoretic Perspective*. Series: Theory and Decision Library C, 35. Springer: New York.

Weiser, P. (2007) "Report from the Center for the New West Putting Network Neutrality in Perspective," Center for the New West Discussion Paper, January 11, 2007. http://www.centerfornewwest.org/pdf/TelecomSummary.pdf

Whitacre, E. (2005) Interview by R. O. Crockett "At SBC, It's All About 'Scale and Scope'," *Business Week*, November 7.

Yoo, C. S. (2006) "Network Neutrality and the Economics of Congestion," *Georgetown Law Journal*, 94: 1847-1908.

A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue

JOHN MUSACCHIO *

Technology and Information Management, University of California, Santa Cruz

GALINA SCHWARTZ

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

JEAN WALRAND

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

Abstract

We address whether local ISPs should be allowed to charge content providers, who derive advertising revenue, for the right to access end-users. We compare two-sided pricing where such charges are allowed to one-sided pricing where they are prohibited. By deriving provider equilibrium actions (prices and investments), we determine which regime is welfare-superior as a function of a few key parameters. We find that two-sided pricing is more favorable when the ratio between parameters characterizing advertising rates and end-user price sensitivity is either low or high.

1 Introduction

Today, an Internet service provider (ISP) charges both end-users who subscribe to that ISP for their last-mile Internet access as well as content providers that are directly connected to the ISP. However, an ISP generally does not charge content providers that are not directly attached to it for delivering content to end-users. One of the focal questions in the network neutrality policy debate is whether these current charging practices should continue and be mandated by law, or if ISPs ought to be allowed to charge all content providers that deliver content to the ISP's end-users. Indeed the current network neutrality debate began when the CEO of AT&T suggested that such charges be allowed (see Whitacre, 2005).

To address this question, we develop a two-sided market model of the interaction of ISPs, end-users, and content providers. The model is closely related to the existing two-sided markets literature as we detail later in this section. In our model, the ISPs play the "platform" role that intermediates the two sides: content providers and end-users. We model a "neutral"

* Contact author: John Musacchio, Technology and Information Management Program, Jack Baskin School of Engineering, University of California Santa Cruz, 1156 High Street, MS: SOE 3. E-mail: johnm@soe.ucsc.edu
This work was supported by the National Science Foundation under Grants ITR/ANIR 0331659 and NeTS-FIND 0627161 and 0626397. We sincerely thank the anonymous reviewers and the editors, Julian Wright and in particular Marius Schwartz, whose thoughtful and detailed suggestions for improving this paper are greatly appreciated.

network as a regime in which ISPs are allowed to charge only content providers that buy their Internet access from them. We assume that with such a charging structure, the ISPs compete on price to attract content providers to buy their access from them, driving these prices to the ISP's costs. By normalizing the price content providers pay ISPs to be net of ISP connection cost, we can model a neutral network as having a zero price for content providers. This could be viewed as a market with one-sided pricing where ISPs only charge end-users. Conversely, two-sided pricing (corresponding to a "non-neutral" network) allows all ISPs to charge all content providers, thus permitting the ISPs to extract revenues from both content providers and end-users. Our contribution is the development of a model that relates the level of end-user usage to the investment decisions of providers of both types under both network regimes. We use our model to compare the welfare of both regimes.

The question we address in this work is part of the larger debate on network neutrality, which includes diverse issues such as whether service differentiation should be allowed, or whether charges for content constitute an impingement of freedom of speech (see Odlyzko, 2008; and Chong, 2007). In 2006 there was a considerable divergence of opinions on the subject of net-neutrality. Indeed, the issue was intensely debated by law and policy makers, and the imposition of restrictive network regulations on Internet service providers (ISPs) in order to achieve network neutrality seemed likely. Recently, the situation has begun to change. In June 2007, the Federal Trade Commission (FTC) issued a report, forcefully stating the lack of FTC support for network neutrality regulatory restraints, and warning of "potentially adverse and unintended effects of regulation" (FTC, 2007, p.159). Similarly, on September 7, 2007 the Department of Justice issued comments "cautioning against premature regulation of the Internet," (DOJ, 2007). Thus by the fall of 2007, the imminent threat of new regulation has diminished somewhat, and a consensus favoring the current (or unregulated) network regime seemed to have emerged. Still, the debate about network neutrality is far from over. Many prominent members of Congress support new network neutrality legislation, including two leading candidates for president in the 2008 election. We do not attempt to address all of the questions in the network neutrality debate. We only study the issue of whether ISPs ought to be allowed to charge content providers for accessing the end-users.

Our model is based on the ideas of two-sided markets, and there is a large literature on the subject. For a survey of two-sided markets, see for example, Rochet and Tirole (2006) and Armstrong (2006). The two-sided market literature studies markets in which a platform provider needs to attract two types of participants, and the presence of more of one type makes the platform more valuable to the other type. Rochet and Tirole (2006) define the market as two-sided, when the volume of realized transactions depends not solely on the aggregate price level charged to the two parties, but also on how this aggregate is divided between them. Using the two-sided market parlance, the ISPs of our model provide the platform, while end-users are one type of participant and content providers are the other type. As will become clear when we describe the details of our model, the end-users "single-home" or connect to one ISP. In a two-sided network, the content providers are forced to "multi-home" or pay multiple ISPs for delivering their content (see for example Armstrong, 2006 on "competitive bottlenecks"). In contrast, the content providers in a network with one-sided pricing "single-home" or pay just one ISP for connectivity. However, a content provider that pays one ISP enjoys the benefits of having connectivity to all the ISPs, because all the ISPs are interconnected. This is in contrast to most two-sided market models where the participants of one platform do not benefit from the presence of

participants of another platform, that is, Microsoft Xbox users do not benefit from more game makers writing Nintendo Wii games. This is an important structural difference.

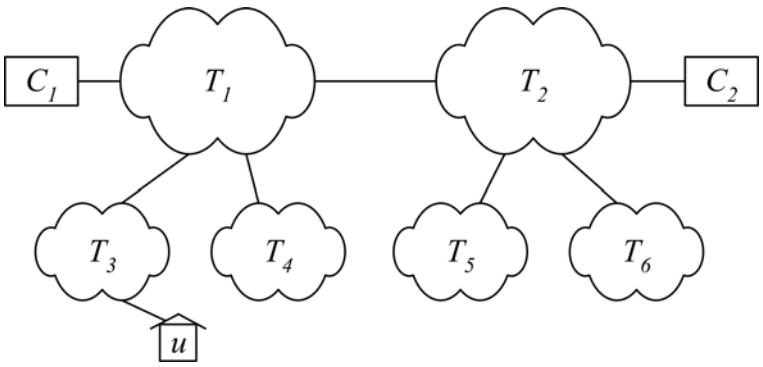

Figure 1: ISPs $\{T_1, \dots, T_6\}$, content providers $\{C_1, C_2\}$ and end-user u.

Other researchers have used the ideas of two-sided markets to study network neutrality. Hermalin and Katz (2007) model network neutrality as a restriction on the product space, and consider whether ISPs should be allowed to offer more than one grade of service. Hogendorn (2007) studies two-sided markets where intermediaries sit between "conduits" and content providers. In his context, net-neutrality means content has open access to conduits where an "open access" regime affords open access to the intermediaries. Weiser (2007) discusses policy issues related to two-sided markets.

The novelty of our model over other work in the two-sided market literature is our explicit modeling of platform investment choices. In the existing literature, the platform incurs the cost of serving the users, which usually is assumed linear in the number of users, but does not make an investment choice.

To introduce the notation and illustrate the arguments for and against the imposition of regulations that prohibit the ISPs from engaging in two-sided pricing, consider the network shown in Figure 1. The figure shows end-user u, ISPs $\{T_1, \dots, T_6\}$, and content providers $\{C_1, C_2\}$. In this network, T_1 and T_2 are transit ISPs (that is, the ISPs who operate portions of the Internet backbone), and $T_3 - T_6$ are local ISPs (i.e, residential ISPs who provide last-mile access for end-users). In the figure, the content providers are attached to a transit ISP whereas a typical end-user u is attached to a local ISP.

In our model, in a market with one-sided pricing, end-users and content providers pay only for their direct access. The transit ISPs charge the local ISPs for carrying their traffic. The transit ISPs typically enter in peering agreements under which they agree to carry each other's traffic, usually free of charge. The transit ISPs charge the content providers for their attachment. We model one-sided pricing ("neutral" network), as a case where the local ISPs are prohibited from charging content providers not buying access directly from them. Thus, with one-sided pricing, content provider C_1 pays ISP T_1 for its access to the Internet but does not pay any of the other ISPs. In contrast, with two-sided pricing ("non-neutral" network) ISP T_3 is able to charge content provider C_1 for carrying its content to end-user u, even though C_1 is not directly attached to T_3.

So far, we have distinguished between two classes of access providers: transit ISPs and local ISPs. The local ISPs provide the last-mile link to end-users (that is, residential consumers). The transit ISPs have no residential customers and serve only content providers, though there are notable examples of companies operating both transit and local networks. The core economic distinction between these ISP classes is that local ISPs are thought to

have more market power. This market power is a result of substantial scale economies in providing mass-market (residential) broadband access. These scale economies limit (today at least) the number of local ISPs in a specific residential area. In particular, today's end-user choice of local ISP is limited to one to three at most. To reflect this market power, in our model each local ISP has a monopoly over its user base.

In reverse, "first-mile" access for major content providers is supplied by transit ISPs, who are believed to be considerably more competitive, because individual content providers (i) generate a lot of traffic and therefore can economically support dedicated links from several suppliers, and (ii) can pick to locate themselves at the points where the network access is the cheapest (that is, to attach directly to the backbone).

Our main purpose of distinguishing transit and local ISPs was to help explain why under today's institutions, local ISPs do not charge content providers. First, content providers can buy access at their end (competitively) from transit ISPs; second, typically, transit ISPs do not pay for access to local ISPs (enjoy settlement-free "peering"); and third, local ISPs do not receive any other payments from content providers. Thus under current practices, the local ISPs do not charge content providers, except perhaps in the case where a company operates both transit and local networks. Even in such a case, one could argue that the ISP only charges competitively for general access rather than for the right to reach the ISP's users in particular. This is because, for instance, a content provider always has the option to switch to a non-integrated competitive transit provider. Thus, in our further analysis we abandon the distinction between local ISPs and transit ISPs, and focus on local ISPs only, which from now on we simply call ISPs.

The arguments for and against net-neutrality having to do with local ISPs charging content providers can be summarized as follows. See ACLU (2006), Farber (2007), Felten (2006), Lessig[1], Owen and Rosston (2003), Sidak (2006), and Yoo (2006) for more elaboration on these points.

Against Neutrality: This line of reasoning is usually expressed by ISPs. Say that C_1 is a source of video streams that require a large bandwidth. ISP T_3 may argue that to accommodate the traffic originating from C_1, he must make substantial investments, which he may not be able to recover from the end-users who buy access directly from him. However, the content providers make additional advertising revenues when end-users consume new high-bandwidth services, which justify their investments. If the ISPs cannot get a share of these additional revenues, they will not invest as much to increase the network capacity. The situation causes poor network quality, which reduces end-user demand, which in turn leads to further reduction of incentives to invest for both provider types.

For Neutrality: This line of reasoning is usually expressed by content providers. If every ISP can charge the content provider C_i for the right to access its end-users, the ability of ISPs to charge content providers would increase dramatically. ISPs would then charge content providers more. In turn, this would reduce the investment incentives for content providers thus lowering content quality.

To sum up, both lines of reasoning argue that their preferred regime would generate higher overall welfare. Clearly, both sides cannot be right; a more detailed analysis is required to clarify the trade-off.

This paper explores how provider investments and revenues differ with network regime. We assume that the number of ISPs and content providers is fixed. That is, we do not

[1] "The Lessig Blog," http://www.lessig.org/blog/archives/003391.shtml

consider the longer term impact of regulations of the ISPs on provider entry incentives (that is, we assume the fixed structure of the network industry).

In section 2, we propose an economic model that relates the investments and prices to revenue for content providers and ISPs. We analyze two-sided pricing (non-neutral) in section 3.1 and we analyze one-sided pricing (neutral) in section 3.2. Section 4 is devoted to

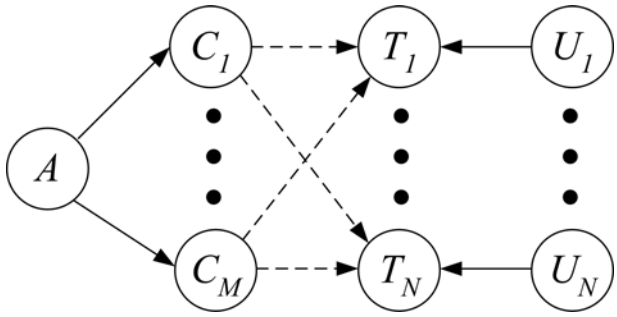

Figure 2: The direction of payments in the model. The dotted lines indicate payments made only with two-sided pricing ("non-neutral").

a comparison of the two regimes of the network. In section 5, we summarize our findings. The technical details of the analysis can be found in our working paper, see Musacchio *et al* (2009).

2 Model

Figure 2 illustrates our setting. In the model, there are M content providers and N ISPs. Each ISP T_n is attached to end-users U_n $(n = 1,2, \dots, N)$ and charges them p_n per click. The ISP T_n has a monopoly over its end-user base U_n. Thus, the end-users are divided between the ISPs, with each ISP having $1/N$ of the entire market. This assumption reflects the market power of local ISPs. Each ISP T_n charges each content provider C_m an amount equal to q_n per click. Content provider C_m invests c_m and ISP T_n invests t_n.

Recall from the Introduction that we measure q net of content providers' access payment (for network attachment), which is set at marginal cost due to the ISPs' competition. Accordingly, we measure the content provider per-user charges to advertisers (which we denote as a) net of the content provider's access payment.

We characterize usage of end-users U_n by the number of "clicks" B_n they make. Since Internet advertising is most often priced per click, clicks are a natural metric for expressing advertising revenue. It is a less natural metric for expressing ISP revenue from end-users, because ISPs do not charge users per click but rather base their charges on bits. However, it is convenient to use only one metric and argue that one could approximate one metric from knowledge of the other using an appropriate scaling factor. The rate B_n of clicks of end-users U_n depends on the price p_n but also on the quality of the network, which in-turn is determined by provider investments, as follows

$$B_n = \left\{\frac{1}{N^{1-w}}(c_1^v + \cdots + c_M^v)\left[(1-\rho)t_n^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w)\right]\right\} e^{-p_n/\theta} \tag{1}$$

where $\rho \in (0,1)$, $\theta > 0$, and $v, w \geq 0$ with $v + w < 1$. For a given network quality (the expression in the curly brackets) the rate of clicks exponentially decreases with price p_n.

The term $c_1^v + \cdots + c_M^v$ is the value of the content providers' investments as seen by a typical end-user. This expression is concave in the investments of the individual providers, and the interpretation is that each content provider adds value to the network. Also note that the structure of the expression is such that end-users value a network in which content is produced by numerous content providers higher than a network in which the content is provided by a single provider with the same cumulative investment. Our end-users' preference for content variety is similar to that of the classical monopolistic competition model by Dixit and Stiglitz (1977). The term in square brackets reflects the value of the ISP's investments for end-users. Clearly, users U_n value the investment made by their ISP, but they may also value the investments made by other ISPs. For instance, a user of one ISP might derive more value by having a better connection with users of another ISP. In our model, the parameter ρ captures this spill-over effect. When $\rho = 1$, end-users U_n value investments of all ISPs equally while when $\rho = 0$, they value only the investment of their ISP. When $\rho \in (0,1)$ end-users U_n value the investment of his ISP T_n more than investments of other ISPs $k \neq n$. The term ρ reflects the direct network effect of ISPs on end-users (not between them and content providers). This effect captures a typical network externality (see Thijssen, 2004 for a discussion of investment spill-over effects). The factor $1/N^{1-w}$ is a convenient normalization. It reflects the division of the end-user pool among N providers and it is justified as follows. Suppose there were no spill-over and each ISP were to invest t/N. The total rate of clicks should be independent of N. In our model, the total click rate is proportional to $(1/N^{1-w})(N(t/N)^w)$, which is indeed independent of N.

At this point, it is worth discussing why we have included certain features in the model. We chose to include N ISPs in the model, rather than just one, because there is a potential of free-riding in the two-sided pricing (non-neutral) case, which gives multiple ISPs the power to charge each content provider. This is because an ISP can increase his price to content providers and enjoy the additional revenue this increase causes, while the downside of inducing the content provider to invest less has to be borne by all of the ISPs. The magnitude of this effect will likely increase with N. We have also allowed for an arbitrary number of content providers, and included the spill-over term ρ, to study whether these features have a strong effect.

Returning to the model specification, the rate R_{mn} of clicks from end-users U_n to C_m is given by

$$R_{mn} = \frac{c_m^v}{c_1^v + \cdots + c_M^v} B_n. \tag{2}$$

Thus, the total rate of clicks for content provider C_m is given by

$$D_m = \sum_n R_{mn}. \tag{3}$$

For further intuition behind the expressions (1) and (3) for click rates, consider a symmetric case, with all providers investing equally ($t_n = t/N$ and $c_m = c/M$) and $\rho = 1$. Then, (1) and (3) reduce to

$$B_n(\text{symmetric case}) = \left\{\frac{1}{N} M^{1-v} c^v t^w\right\} e^{-p/\theta},\ D_m(\text{ symmetric case}) = \left\{\frac{1}{M^v} c^v t^w\right\} e^{-p/\theta}$$

respectively. Thus for both provider types, the click rates exponentially decrease with end-user price, increase with provider investments c and t, and exhibit decreasing return to investments. The only difference in the content provider and ISP click rate expressions are the effects of M and N. Recall that we included a normalization factor in (1) so that the same total investment split evenly across all ISPs should result in the same total click rate, regardless of N. On the other hand, we assumed that a user values variety in content, so that the same total investment split amongst multiple content providers is more valued than if it were split amongst fewer content providers. That is why B_n increases in M for fixed c, but D_m does not increase with N.

Leaving the symmetric case behind and returning to the general model, we assume that content providers charge a fixed amount a per click to the advertisers. Each content provider's objective is to maximize its profit which is equal to revenues from end-user clicks net of investment costs. Thus

$$\Pi_{Cm} = \sum_{n=1}^{N} (a - q_n) R_{mn} - \beta c_m \tag{4}$$

where the term $\beta > 1$ is the outside option (alternative use of funds c_m).

ISP T_n profit is

$$\Pi_{Tn} = (p_n + q_n) B_n - \alpha t_n \tag{5}$$

where $\alpha > 1$ is the outside option of the ISP. We assume providers of each type are identical and we will focus on finding symmetric equilibria for both one- and two-sided pricing.

3 The analysis of one- and two-sided pricing

To compare one-sided and two-sided pricing (neutral and non-neutral networks), we make the following assumptions.

- One-sided pricing (neutral network): In stage 1, each T_n simultaneously chooses (t_n, p_n). The price q_n charged to content providers is constrained to be 0. In stage 2, each C_m chooses c_m.
- Two-sided pricing (non-neutral network): In stage 1, each T_n simultaneously chooses (t_n, p_n, q_n). In stage 2, each C_m chooses c_m.

As we discussed in the Introduction, in one-sided pricing, the price q_n charged to content providers is zero by the following argument. Since content providers need only pay the ISP with which they are directly connected, the resulting price competition forces prices down to ISP costs. Thus the prices $\{q_n\}$ are zero after one normalizes them to be net of ISP cost. In contrast, in a network with two-sided pricing (non-neutral), all ISPs have the ability to charge each content provider. This permits ISPs to exercise their market power on content providers and allows them to extract a non-zero price.

In both cases, we assume that content providers observe ISP investments, and can subsequently adjust their investments based on the ISPs' choices. We justify this assumption by the difference in time and scale of the required initial investments. The investments of ISP tend to be longer-term investments in infrastructure, such as deploying networks of fibre-optic cable. Conversely, the investments of content-providers tend to be shorter-term and more ongoing in nature, such as development of content, making ongoing improvements to a search algorithm, or adding/replacing servers in a server farm.

3.1 Two-sided pricing

In a network with two-sided pricing (non-neutral network), each ISP chooses (t_n, p_n, q_n) and each content provider chooses c_m. To analyze this situation, we study how C_m chooses the optimal c_m for a given set of (t_n, p_n, q_n). We then substitute that value of c_m in the expression for Π_T and we optimize for (t_n, p_n, q_n).

The best choice for c_m given (t, p, q) maximizes

$$\begin{aligned}\Pi_{Cm} &= aD_m - \sum_n q_n R_{mn} - \beta c_m \\ &= N^{w-1} c_m^v \left[\sum_n (a - q_n)\left((1-\rho)t_n^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w)\right) e^{-p_n/\theta}\right] - \beta c_m.\end{aligned} \tag{6}$$

As a result of the cancelation between the denominator of R_{mn} (see (2)) and the expression for the click rate B_n (see (1)), the dependency of content provider m's revenue on the investments of other content providers disappears. Thus the expression for profit Π_{Cm} is independent of other content provider investments c_j, $j \neq m$. Therefore, each content provider's optimization is unaffected by the simultaneously made (but correctly anticipated in equilibrium) investment decisions of the other content providers. Assuming that the term in square brackets is positive, we find that

$$\beta c_m^{1-v} = vN^{w-1}\left[\sum_k (a - q_k)((1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w))e^{-p_k/\theta}\right] =: \beta c^{1-v}. \tag{7}$$

For that value of c_m, we find that

$$\Pi_{Tn} = MN^{w-1}(q_n + p_n)F_n e^{-p_n/\theta}\left(\frac{v}{N\beta}\right)^{v/(1-v)}\left[\sum_k (a - q_k)e^{-p_k/\theta}F_k\right]^{v/(1-v)} - \alpha t_n \quad (8)$$

where

$$F_n = (1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w) = \phi t_n^w + \frac{\rho}{N}\sum_{k\neq n} t_k^w \quad (9)$$

with

$$\phi := 1 - \rho + \frac{\rho}{N} < 1, \;\; if \;\; N > 1. \quad (10)$$

The ISP T_n chooses investment and prices (t_n, p_n, q_n) that maximize his profit given by equation (8). The simultaneous decisions of each of the ISPs affect each other, therefore in order to find a Nash equilibrium we need to identify a point where the best response functions intersect. Writing that the three corresponding partial derivatives of (8) are equal to zero, and then finding the symmetric intersection point of the best response functions, we find the following solution (see our working paper, Musacchio at al. (2008)):

$$p_n = p = \theta - a; \quad (11)$$

$$q_n = q = a - \theta\frac{v}{N(1-v)+v}; \quad (12)$$

$$t_n = t \;\; \text{with } (Nt)^{1-v-w} = x^{1-v}y^v e^{-(\theta-a)/\theta}; \quad (13)$$

$$c_m = c \;\; \text{with } c^{1-v-w} = x^w y^{1-w} e^{-(\theta-a)/\theta}; \quad (14)$$

$$\Pi_{Cm}^{1-v-w} = \Pi_C^{1-v-w} := \left(\frac{\theta v(1-v)}{N(1-v)+v}\right)^{1-v-w} x^w y^v e^{-(\theta-a)/\theta}; \quad (15)$$

$$\Pi_{Tn}^{1-v-w} = \Pi_T^{1-v-w} := \left(\frac{M\theta(N(1-v) - wN\phi(1-v) - vw)}{N(N(1-v)+v)}\right)^{1-v-w} x^w y^v e^{-(\theta-a)/\theta}; \quad (16)$$

$$\frac{\Pi_C}{c} = \frac{\beta(1-v)}{v}; \quad (17)$$

$$\frac{\Pi_T}{t} = \frac{\alpha}{w}\left[\frac{N(1-v)}{N\phi(1-v)+v} - w\right]; \quad (18)$$

$$B^{1-v-w} = M^{1-v-w}x^w y^v e^{-(\theta-a)/\theta} \quad (19)$$

where $B := \sum_n B_n = \sum_m D_m$ is the total click rate and

$$x := \frac{M\theta w}{\alpha}\frac{N\phi(1-v)+v}{N(1-v)+v} \;\; \text{and} \;\; y := \frac{\theta}{\beta}\frac{v^2}{N(1-v)+v}. \quad (20)$$

3.2 One-sided pricing

A network with one-sided pricing (neutral network) is similar to one with two-sided pricing (non-neutral), except that $q_n = 0$ as we argued in section 2 for $n = 1, \dots, N$. The best choice of c given $\{q_n = 0, p_n, t_n\}$ is such that

$$\beta c_m^{1-v} = vN^{-1}\left[\sum_k a((1-\rho)t_k^w + \frac{\rho}{N}(t_1^w + \cdots + t_N^w))e^{-p_k/\theta}\right] =: \beta c^{1-v}.$$

For that value of c_m, we find that

$$\Pi_{Tn} = MN^{-1}p_n F_n e^{-p_n/\theta}(\frac{v}{\beta})^{v/(1-v)}\left[\sum_k ae^{-p_k/\theta}F_k\right]^{v/(1-v)} - \alpha t_n \tag{21}$$

where

$$F_n = \phi t_n^w + \frac{\rho}{N}\sum_{k\neq n} t_k^w.$$

The ISP T_n chooses investment and price (t_n, p_n) that maximize the above expression. We find a symmetric Nash equilibrium by writing that the two corresponding partial derivatives of (21) with respect to a single ISP actions are zero, and that the other ISPs make the same actions, and solving all of the resulting equations. This analysis leads to the following solutions (see our working paper, Musacchio at al. (2008)):

$$p_n = p_0 := \frac{\theta N(1-v)}{N(1-v)+v}; \tag{22}$$

$$q_m = 0; \tag{23}$$

$$t_n = t_0 \text{ where } (Nt_0)^{1-v-w} = x^{1-v}y_0^v e^{-p_0/\theta}; \tag{24}$$

$$c_m = c_0 \text{ where } c_0^{1-v-w} = x^w y_0^{1-w} e^{-p_0/\theta}; \tag{25}$$

$$\Pi_{Cm}^{1-v-w} = \Pi_{C0}^{1-v-w} := (a(1-v))^{1-v-w}x^w y_0^v e^{-p_0/\theta}; \tag{26}$$

$$\Pi_{Tn}^{1-v-w} = \Pi_{T0}^{1-v-w} := \left(\frac{M\theta(N(1-v) - wN\phi(1-v) - wv)}{N(N(1-v)+v)}\right)^{1-v-w} x^w y_0^v e^{-p_0/\theta}; \tag{27}$$

$$\frac{\Pi_{C0}}{c_0} = \frac{\beta(1-v)}{v}; \tag{28}$$

$$\frac{\Pi_{T0}}{t_0} = \frac{\alpha}{w}\left[\frac{N(1-v)}{N\phi(1-v)+v} - w\right]; \tag{29}$$

$$B_0^{1-v-w} = M^{1-v-w}x^w y_0^v e^{-p_0/\theta} \tag{30}$$

where B_0 is the total click rate, x is given in (20), and

$$y_0 := \frac{av}{\beta}. \tag{31}$$

4 Comparison

In this section, we compare the Nash equilibria of the two regimes. In section 4.1, we derive expressions for the welfare of end-users, and the ratio of social welfare with one- vs. two-sided pricing. In section 4.2, we demonstrate that the return on investments is the same in both regimes. In section 4.3, we compare the profits and social welfare of the two regimes for a range of parameters.

4.1 User welfare and social welfare

Before proceeding we define the following notation.

$$\pi := \frac{v}{N(1-v)+v} \quad \text{and} \quad \delta := \frac{a}{\theta}. \tag{32}$$

In order to compute end-user welfare, we use the total click rate to proxy the aggregate user demand. This enables us to calculate consumer surplus and use it to measure end-user welfare. We compute the consumer surplus by taking the integral of the demand function from the equilibrium price to infinity. This integral is taken with the investment levels of content providers and ISPs fixed. We find

$$W_U(\text{two}-\text{sided}) = M\theta x^{w/(1-v-w)} y^{v/(1-v-w)} e^{-\frac{\theta-a}{\theta(1-v-w)}}.$$

The expression for the one-sided case is the same, but with y exchanged for y_0 and $\theta - a$ in the exponent exchanged with p_0. The ratio of the social welfare with one- vs. two-sided pricing has the form

$$\frac{W_U(\text{one}-\text{sided}) + N\Pi_T(\text{one}-\text{sided}) + M\Pi_C(\text{one}-\text{sided})}{W_U(\text{two}-\text{sided}) + N\Pi_T(\text{two}-\text{sided}) + M\Pi_C(\text{two}-\text{sided})}$$

$$= \frac{1+\delta(1-v)+(\pi/v)(N(1-v)-wN\phi(1-v)-wv)}{1+\pi(1-v)+(\pi/v)(N(1-v)-wN\phi(1-v)-wv)} [(\delta/\pi)^v e^{\pi-\delta}]^{1/(1-w-v)}.$$

4.2 Return on investment

We use the definition $\delta = a/\theta$ and equations (11), (12), (22), and (23) to relate prices in the one- and two-sided pricing regimes. This permits us to formulate the following proposition.

Proposition 1: The *price* p *end-users pay in the two-sided pricing case is given by*

$$p = \theta(1-\delta) = \theta - a.$$

Also, we note that

$$p + q = p_0 = \theta(1-\pi).$$

Moreover,

$$\frac{\Pi_C}{c} = \frac{\Pi_{C0}}{c_0} \quad and \quad \frac{\Pi_T}{t} = \frac{\Pi_{T0}}{t_0}.$$

Thus, from Proposition 1, the total revenue per click of the ISPs is the same in both regimes and so are the rate of return on investments of the content and ISPs.

Despite the fact that the rate of return on investments are the same in both regimes, as we will see in the next subsection, the size of those investments and resulting profits may be quite different across regimes.

4.3 Comparative statics

Dividing the expressions for one-sided pricing by the corresponding expressions for the two-sided pricing case, we define ratios such as

$$r(\Pi_C) := \left(\frac{\Pi_C(\text{ one} - \text{sided })}{\Pi_C(\text{two} - \text{sided })}\right)^{1-v-w}$$

where $\Pi_C(\ two - sided\ )$ is the profit per content provider in the two-sided case as expressed in (15) and $\Pi_C(\ one - sided\ )$ is the profit per content provider with one-sided pricing (26). We define $r(c), r(t),$ and $r(\Pi_T)$ similarly. We find

$$r(\Pi_T) = r(t) = r(B) = \left(\frac{\delta}{\pi}\right)^{v} e^{\pi-\delta}; \tag{33}$$

$$r(\Pi_C) = r(c) = \left(\frac{\delta}{\pi}\right)^{1-w} e^{\pi-\delta}. \tag{34}$$

Figure 3 shows the ratios of profts with one- vs. two-sided pricing for both content providers and ISPs. Figure 4 shows the same ratios, but for different values of v and w. The figures show that for small or large values of a/θ, the ratio of advertising revenue per click to the constant characterizing price sensitivity of end-users, two-sided pricing is preferable to both content providers and ISPs. (Here we say "preferable" in that the profits are larger, though we have seen that the rate of return on investments are the same.) For mid range values of a/θ, one-sided pricing is preferable to both, though the values of a/θ where the transition between one-sided being preferable to two-sided are not exactly the same for content providers and ISPs. Furthermore, as N, the number of ISPs increases, the range of a/θ values for which one-sided pricing is superior increases, while also the degree by which it is superior (in terms of revenues to content providers and ISPs) increases.

These results can be explained by the following reasoning. When a/θ is large, the content providers' revenues from advertising are relatively high, and the ISPs' revenue from end-users are relatively low. Because of this, the ISPs incentives to invest are suboptimal (too low relative to the socially optimal ones), unless they can extract some of the content providers' advertising revenue by charging the content providers. Thus in the one-sided pricing case, the ISPs under invest, making the rewards for them as well as content providers less than it could have been with two-sided pricing.

It is important to note that when a/θ is larger than 1, the price p charged end-users becomes negative in the two-sided case as can be seen from (11). If end-users were actually paid to click, intuition suggests that they would click an unbounded amount and therefore our exponential model of demand (1) would not be valid in this region of a/θ. However, one could interpret price p to be net any variable costs ν to end users – similar to how we

define q.[2] With this interpretation, the price p could be negative while the actual prices users see positive, so long as $|p| < \nu$. We therefore show numerical results in our plots for a/θ as large as 1.2.

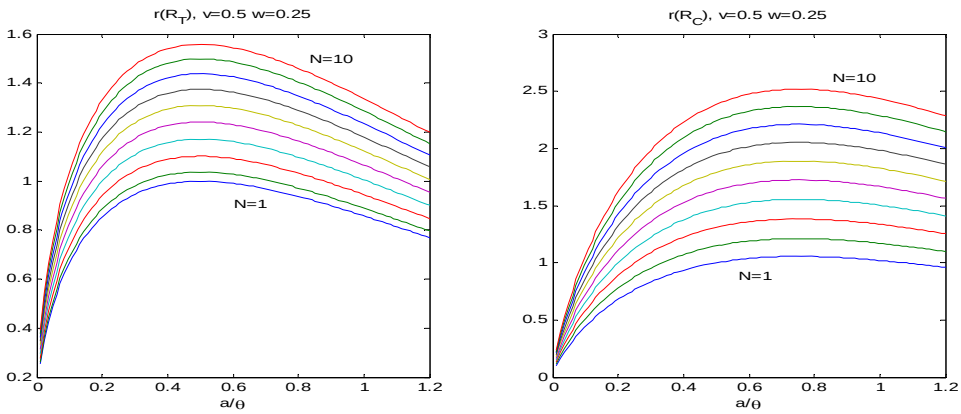

Figure 3: The ratios of profits ($v = 0.5, w = 0.25$) for different values of N.

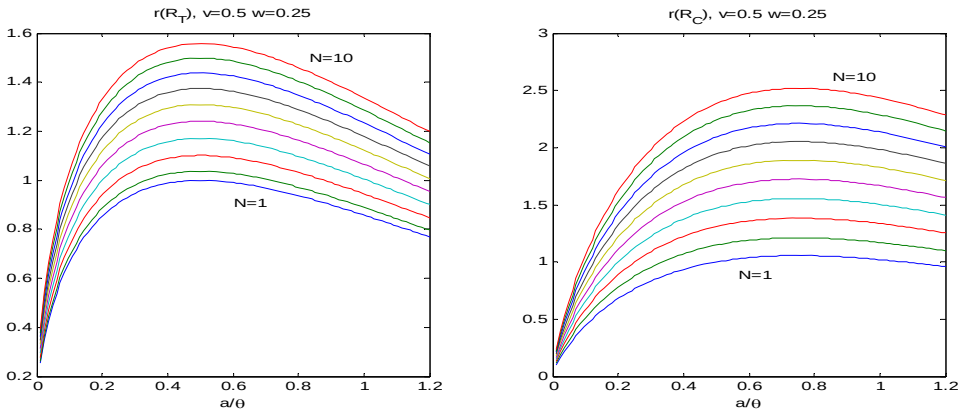

Figure 4: The ratios of profits ($v = 0.25, w = 0.5$) for different values of N.

When a/θ is very small, the content providers' advertising revenue is relatively low, and the ISP's end-user revenue is relatively high. In order to get the content providers to invest adequately, the ISPs need to pay the content providers. That is why for small enough a/θ the price q is negative (see Figure 5), representing a per click payment from the ISPs to the content providers.

Interestingly, even in the two-sided pricing case, our content providers obviously get some share of the surplus generated jointly by them and the ISPs. This is in contrast to the multi-homing case of Armstrong (2006) (which is roughly analogous to our two-sided pricing case), where the surplus is fully extracted from group-2 agents (content providers in our case) – the entire surplus is shared between the platform and the end-users. We do not have this extreme result, because in our model content providers invest *after* ISPs announce

[2] If we suppose such a normalization, it may be more natural to express end-user demand in terms of the price p+ν actually seen by users. However, supposing demand was exponential in p+ν rather than exponential in p would only change expression (1) for click rate by a constant factor. This constant factor disappears in any expression comparing the relative revenues or welfares of the two regimes.

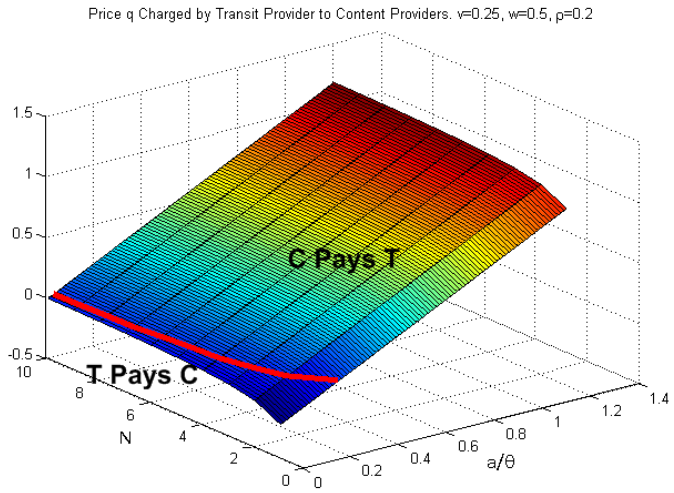

Figure 5: The price q (that the ISPs charge the content providers).

prices. Thus, ISP commitment to the declared prices permits content providers to retain a positive fraction of the surplus.

Finally, when a/θ is in the intermediate range, in between the two extremes, both content providers and ISPs have adequate incentive to invest. However, another effect comes into play – ISP free riding becomes an important factor when N is large. As N increases in the two-sided pricing case there are more ISPs that levy a charge against each content provider. As the price ISPs charge content providers increases, it becomes less attractive for content providers to invest. Thus, an ISP choosing the price to charge content providers is balancing the positive effect of earning more revenue per click from content providers versus the negative effect of having fewer clicks because the content providers have reduced their investment. But each ISP sees the entire gain of raising its price, but the loss is borne by all N ISPs. Consequently, the ISPs overcharge the content providers in Nash equilibrium, and the degree of this overcharging increases with N. This is analogous to the tragedy of the commons where people overexploit a public resource. Another perhaps more direct analogy is the "castles on the Rhine effect" where each castle owner is incentivized to increase transit tolls on passing traffic excessively by ignoring the fact that the resulting reduction in traffic harms not only him, but also other castle owners. When all castles do the same, the traffic on the Rhine decreases (Kay, 1990). The extent of this negative externality, and hence the degree of over-charging, increases with N.

Figure 6 shows a three dimensional plot of the ratio of social welfare for one- vs. two-sided pricing. The plot shows how the ratio changes for different N and a/θ. The second panel of Figure 6 is a simplified version of the first panel. It depicts only the boundaries in the parameter space where one-sided pricing is preferable to two-sided and vice versa.

It is also worthwhile to note that spill-over parameter ρ and the number of content providers M do not appear in the expression for the ratio of content provider revenue between the two regimes (34) nor do they appear in the ratio of revenues for ISPs (33). This is in spite of the fact that ρ and M do appear in the expressions for both the one-sided and two-sided pricing equilibria. This suggests that the spill-over effect and number of content providers have little or no effect on the comparative welfare of the two regimes.

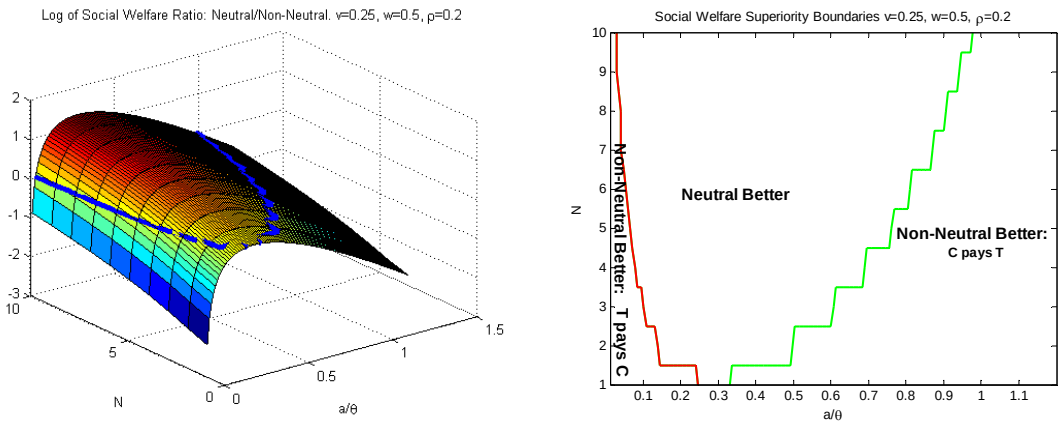

Figure 6: Left: The log of social welfare ratio (one-sided to two-sided). Right: Regions of social welfare superiority (one-sided vs. two-sided).

5 Conclusions

Our analysis seeks to address whether content providers should pay ISP(s) only for the right to access the network at their end, or whether other ISPs (that is, the ISPs other than the one through which a specific content provider is attached to the network) may also charge for the "right" to reach end-users. We call the former "one-sided pricing", and the latter "two-sided pricing". We suggest that the imposition of regulation that precludes the ISPs from two-sided pricing corresponds to a "neutral" network, while the practice of two-sided pricing corresponds to a "non-neutral" network.

We study how each pricing regime affects investment incentives of transit and content providers. We show that parameters such as advertising rate, end-user price sensitivity, and the number of ISPs influence whether one- or two-sided pricing achieves a higher social welfare. From our results, when the ratio of advertising rates to the constant characterizing price sensitivity is an extreme value, either large or small, two-sided pricing is preferable. If the ratio of advertising rates to the constant characterizing price sensitivity is not extreme, then an effect like the "castles on the Rhine effect" becomes more important. ISPs in a two-sided pricing regime have the potential to over charge content providers, and this effect becomes stronger as the number of ISPs increases.

In our comparison of one- and two-sided pricing we assume that the ISPs choose their strategy first and that the content providers follow. We justify this assumption by the difference in the time and scale of the required initial investments. While we feel this is a valid assumption, it is worth noting that if we had modelled the game such that the content providers and ISPs reverse their order of play, the game would not have a meaningful equilibrium. (In such a reversed model with two-sided pricing, the ISPs would charge an unbounded price to content providers after they had committed to their content investments. This is because at this point the content providers would have no action left to take in the game. This of course would lead to content providers investing zero in the first stage of the game.)

We make other key modeling choices, such as that the ISPs have local monopolies over their users. We believe that if we had studied a model where each ISP is a duopolist, which better models the degree of choice most end-users have today, our results would have been qualitatively similar. However, there are a number of competing effects such a model may

introduce. First, such a scenario would reduce the market power of ISPs over end-users, thus reducing the revenue the ISPs could extract from them. Since two-sided pricing provides ISPs with another source of revenue to justify their investments, this effect would tend to increase the parameter region for which two-sided pricing is social welfare superior. Second, a duopolist competing on product quality invests more than a monopolist, so this would tend to increase efficiency of one-sided pricing. Third, if the model were changed from having N to $2N$ ISPs, then the free riding or "castles on the Rhine" effect would grow, tending to reduce the welfare of the two-sided pricing case. The net effect of all these individual effects would of course depend on the detailed specifications of such a model.

Another assumption we make is that all ISPs have identical payoff functions. This would not be the case if the ISPs served different population sizes or different population densities for instance. Our assumption makes the analysis much easier, but we feel that if this assumption were relaxed, the results would be qualitatively similar though closed form solutions might not exist. We feel the same way about our assumption that content providers have identical payoff functions.

Our two-sided pricing model assumes that in-bound traffic to a local ISP could be identified as originating at a particular content provider, in order for the ISP to levy the appropriate charge to the content provider. This assumption would not strictly hold if content providers had some way of reaching end-users of an ISP without paying this ISP for end-user traffic. For instance, if there were a second ISP that enjoyed settlement free peering with the first ISP, the content provider could route its traffic through the second ISP and thus avoid the first ISP charge for end-user access. This strategy may be facilitated by the fact that the end-users of both ISPs send traffic to each other, and perhaps the traffic from the content provider could be masked in some way to look like traffic originating from the second ISP's end-users. However, the communication protocols of the Internet require that packets be labeled with the origin (IP) address. It seems unlikely today that a large content provider could have the origin addresses of its traffic falsified in a way that would both prevent ISPs from being able to charge the content provider while still enabling end-users to send traffic in the reverse direction back to the content provider. However, it is certainly possible that technology would be developed to enable such a strategy in the future, especially if there were an economic incentive for developing it.

There are several assumptions we make that might be viewed as limitations of our model. First, we fix the number of network providers, independent of the network regime. This assumption is realistic in the short-run, but in the long-run the number of providers entering the network industry is likely to differ with regime. Second, we do not consider heterogeneity in the providers nor in the end-users. Third, we assume full commitment to the declared prices, that is the ISPs cannot later adjust the prices that they declared initially. We also have not modeled the price content providers charge advertisers as a decision variable, but we have modeled the price ISPs charge end-users. Though we feel that the present model correctly captures focal features of the underlying network environment, in future work we plan to extend our model to permit endogenous industry structure, that is, we will study how network regime affects the number of content providers and ISPs.

6 References

ACLU (2006) "Internet Freedom and Innovation at Risk: Why Congress Must Restore Strong Net Neutrality Protection," ACLU Press Release, September 9 2006. http://www.aclu.org/freespeech/internet/26829res20060922.html

Armstrong, M. (2006) "Competition in Two Sided Markets," *RAND Journal of Economics*, 37: 668-691.

Chong, R. (2007) "The 31 Flavors of the Net Neutrality Debate: Beware The Trojan Horse," *Advanced Communications Law and Policy Institute, Scholarship Series*, New York Law School, December 2007.

Dixit, A. and J. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67: 297-308.

Department of Justice (2007) "Comments on Network Neutrality in Federal Communications Commission Proceeding," Press Release, September 7, 2007. http://www.usdoj.gov/opa/pr/2007/September/07_at_682.html

Farber, D. and M. Katz (2007) "Hold Off On Net Neutrality," *Washington Post*, Jan. 22.

Felten, E. (2006) "The Nuts and Bolts of Network Neutrality," Center for Information Technology Policy, Princeton, July 6, 2006. http://itpolicy.princeton.edu/pub/neutrality.pdf

Federal Trade Commission (2007) "Broadband Connectivity Competition Policy," Report, June 2007.

Hermalin, B. and M. Katz (2007) "The Economics of Product Line Restrictions With an Application to the Network Neutrality Controversy," *Information Economics and Policy*, 19: 215-48.

Hogendorn, C. (2007) "Broadband Internet: Net Neutrality Versus Open Access," *International Economics and Economic Policy*, 4: 185-208.

Kay, J. A. (1990) "Tax Policy: A Survey," *Economic Journal*, 100: 18-75.

Musacchio, J., G. Schwartz and J. Walrand (2009) "A Two-Sided Market Analysis of Provider Investment Incentives With an Application to the Net-Neutrality Issue," Technical Report UCSC-SOE-09-08, School of Engineering, University of California, Santa Cruz, March 2009. http://www.soe.ucsc.edu/ johnm/Publications/neutrality.pdf

Odlyzko, A. (2008) "Network neutrality, Search Neutrality, and the Never-Ending Conflict Between Efficiency and Fairness in Markets," *Review of Network Economics*, 8: 40-60.

Owen, B. and G. Rosston (2003) "Local Broadband Access: *Primum Non Nocere or Primum Processi?* A Property Rights Approach," Stanford Law and Economics Olin Working Paper No. 263.

Rochet, J.-C. and J. Tirole (2006) "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 37: 645-667.

Sidak, G. (2006) "A Consumer-Welfare Approach to Network Neutrality Regulation of the Internet," *Journal of Competition Law & Economics*, 2: 349-474.

Thijssen, J. (2004) *Investment Under Uncertainty, Coalition Spillovers and Market Evolution in a Game Theoretic Perspective*. Series: Theory and Decision Library C, 35. Springer: New York.

Weiser, P. (2007) "Report from the Center for the New West Putting Network Neutrality in Perspective," Center for the New West Discussion Paper, January 11, 2007. http://www.centerfornewwest.org/pdf/TelecomSummary.pdf

Whitacre, E. (2005) Interview by R. O. Crockett "At SBC, It's All About 'Scale and Scope'," *Business Week*, November 7.

Yoo, C. S. (2006) "Network Neutrality and the Economics of Congestion," *Georgetown Law Journal*, 94: 1847-1908.

Platform Competition in Two-Sided Markets

Jean-Charles Rochet* Jean Tirole†

December 13, 2002

Abstract

Many if not most markets with network externalities are two-sided. To succeed, platforms in industries such as software, portals and media, payment systems and the Internet, must "get both sides of the market on board ". Accordingly, platforms devote much attention to their business model, that is to how they court each side while making money overall. The paper builds a model of platform competition with two-sided markets. It unveils the determinants of price allocation and end-user surplus for different governance structures (profit-maximizing platforms and not-for-profit joint undertakings), and compares the outcomes with those under an integrated monopolist and a Ramsey planner.

1 Introduction

Buyers of videogame consoles want games to play on; game developers pick platforms that are or will be popular among gamers. Cardholders value credit or debit card only to the extent that these are accepted by the merchants they patronize; affiliated merchants benefit from a widespread diffusion of cards among consumers. More generally, many if not most markets with network externalities are characterized by the presence of two distinct sides whose ultimate benefit stems from interacting through a common platform. Platform owners or sponsors in these industries must address the celebrated "chicken-and-egg problem" and be careful to "get both sides on board". Despite much theoretical

*IDEI and GREMAQ (CNRS UMR 5604), Toulouse

†IDEI and GREMAQ (CNRS UMR 5604), Toulouse, CERAS (CNRS URA 2036), and MIT

progress made in the last two decades on the economics of network externalities and widespread strategy discussions of the chicken-and-egg problem, two-sided markets have received scant attention. The purpose of this paper is to start filling this gap.

The recognition that many markets are multi-sided leads to new and interesting, positive and normative questions. Under multi-sidedness, platforms must choose a price structure and not only a price level for their service. For example, videogame platforms such as Sony, Sega and Nintendo make money on game developers through per-unit royalties on games and fixed fees for development kits and treat the gamers side as a loss leader. Interestingly, operating system platforms for the PC and handheld devices have adopted the opposite *business model* and aim at making money on consumers. The choice of a business model seems to be key to the success of a platform and receives much corporate attention. Table 1 provides a few illustrations[1] of the two-sided markets and shows that platforms often treat one side as a profit center and the other as a loss leader, or, at best, as financially neutral. A number of these illustrations are discussed in "mini-case studies" in section 7. And Table 2 lists a few important segments of the new economy that will be searching for a proper business model in the next few years. Such conventional wisdom about business models found in the trade press and summarized in Table 1 is of course subject to criticism. To reason in terms of profit centers, costs are often "intuitively," but

[1]There are of course other illustrations, for example *scientic journals*, that must match readers and authors. Interestingly, the *Bell Journal of Economics* for a number of years after it was launched was sent for free to anyone who requested it. There is currently much discussion of how the business model for scientific journals will evolve with electronic publishing. The list of *social gatherings* examples of cross-subsidization could be extended to include dating or marital agencies which may charge only one side of the market.

A couple of explanations regarding markets that will not be discussed in section 7: *Social gatherings*: celebrities often do not pay or are paid to come to social happenings as they attract other participants (who may then be charged an hefty fee); similarly, in some conferences, star speakers are paid while others pay. *Real estate*: In many countries buyers are not charged for visiting real estate properties and thus marginal visits are heavily subsidized. To be certain, the sale price reflects the real estate agency fee, but this does not imply that the arrangement is neutral (see section 8). *Shopping malls*: shoppers are subsidized. They don't pay for parking; in France they can also buy gasoline at a substantial discount. *Discount coupon books*: These are given away to consumers. Intermediaries charge merchants for the service. *Browsers*: The picture given in Table 1 is a bit simplistic. In particular, Netscape initially made about one third of its revenue on the client side before giving the software away. But Netscape always viewed the software running on top of the operating system on the web servers as a major source of profit.

arbitrarily allocated to either side of the market. Yet, the conventional wisdom points at some more fundamental logic related to prices and surpluses on both sides of the market. A major objective of our paper is to unveil this logic and the determinants of the choice of a business model.

ILLUSTRATIONS OF EXISTING BUSINESS MODELS

Product	loss leader/break-even segment/ subsidized segment	profit-making segment/ subsidizing segment
SOFTWARE		
Videogames	consumers (consoles)	software developers
Streaming media	consumers	servers
Browsers	users	web servers
Operating systems (Windows; Palm, Pocket PC)	application developers (development tools, support, functionality,...)	clients
Text processing	reader/viewer	writer
PORTALS AND MEDIA		
Portals	"eyeballs"	advertizers
Newspapers	readers	advertizers
(Charge-free) TV networks	viewers	advertizers
PAYMENT SYSTEMS		
Credit and differed debit cards (Visa, MasterCard, Amex,...)	cardholders	merchants
On-line debit cards	merchants	cardholders

(continued...)

(continued...)

OTHERS...		
Social gatherings	celebrities in social happenings	other participants
Shopping malls	consumers (free parking, cheap gas,...)	shops
Discount coupon books (Want Advertizer)	consumers	merchants
(Legacy) Internet	websites	dial-up consumers
Real estate	buyers	sellers

Table 1: existing business models

LOOKING FORWARD:

Platform	Two sides	Instruments of cost allocation or cross-subsidization
B2B	buyers / sellers	design of auctions, information flows,...
Internet backbone services	consumers / websites	termination (settlement) charges
Pools and standards	relevant sides	level of royalties, inclusiveness of pools,...
Software as a service (.Net vs Java,...)	consumers / application developers	development tools and other efforts to create an applications development environment, backward compatibility, pricing,...

Table 2: prospective applications

From both positive and normative viewpoints, two-sided markets differ from the textbook treatment of multiproduct oligopoly or monopoly. The interaction between the two sides gives rise to strong complementarities, but the corresponding externalities are not internalized by end users, unlike in the multiproduct literature (the same consumer buys the razor and the razor blade). In this sense, our theory is a cross between network economics, which emphasizes such externalities, and the literature on (monopoly or competitive) multiproduct pricing, which stresses cross-elasticities. For example, socially optimal "Ramsey" prices are not driven solely by superelasticity formulae but also reflect each side's contribution to the other side's surplus.

Some new questions raised by two-sided markets are more specific to the existence of competition between platforms. In a number of markets, a fraction of end users on one or the two sides connect to several platforms. Using the Internet terminology, we will say that they *multihome*. For example, many merchants accept both American Express and Visa; furthermore, some consumers have both Amex and Visa cards in their pockets. Many consumers have the Internet Explorer and the Netscape browsers installed on their PC, and a number of websites are configured optimally for both browsers. Readers may subscribe to multiple newspapers, B2B exchange members may buy or sell their wares on several exchanges, and real estate sellers and buyers may use the services of multiple real estate agencies. Competitive prices on one market then depend on the extent of multihoming on the other side of the market. For example, when Visa reduces the (transaction-proportional) charge paid by the merchants,[2] merchants become more tempted to turn down the more costly Amex card as long as a large fraction of Amex customers also owns a Visa card. More generally, multihoming on one side intensifies price competition on the other side as platforms use low prices in an attempt to "steer" end users on the latter side toward an exclusive relationship.[3]

[2]The mechanism through which this reduction operates is indirect and is described in section 7.

[3]The occurence of steering is easiest to visualize in those illustrations in which platforms charge per-end-user-transaction fees: The seller of a house or a $B2B$ supplier may only list the house or the wares on the cheapest platform.

The paper studies how the price allocation between the two sides of the market is affected by a) platform governance (for-profit vs not-for-profit), b) end users' cost of multihoming, c) platform differentiation, d) platforms' ability to use volume-based pricing, e) the presence of same-side externalities, and f) platform compatibility. It also investigates how privately optimal pricing structures compare with socially optimal ones.

The paper is organized as follows. Section 2 describes the simplest version of the model, in which end-users incur no fixed cost and platform pricing is linear on both sides of the market, and analyzes the (profit maximizer and Ramsey planner) monopoly benchmarks. Section 3 derives equilibrium behavior when two (for-profit or not-for-profit) platforms compete. Section 4 obtains some comparative statics in order to help predict the choice of business model. Section 5 compares the price structures in the case of linear demands. Section 6 generalizes the model and results in order to allow for fixed user costs and nonlinear platform pricing. Section 7 summarizes the main results and provides seven "mini case studies" to illustrate how our theory may shed light on existing and future business models. Last, Section 8 concludes with some general considerations about two-sided markets.

As we discussed, our work puts network economics and multiproduct pricing together. From the early work of Rohlfs (1974) to the recent theoretical advances and applications to antitrust through the pioneering work of Katz-Shapiro (1985, 1986) and Farrell-Saloner (1985, 1986), a large body of literature has developed on network industries. To make progress, however, this literature has ignored multisidedness and the price allocation question. In contrast, the competitive multiproduct pricing literature (e.g., Baumol et al 1982, Wilson 1993) has carefully described the interdependency of pricing decisions but it has

In industries in which platforms do not charge per-end-user-transaction fees, steering is more subtle as it operates through effort substitution. For example, a software platform offering better software development kits, support, and application programming interfaces not only encourages the development of applications optimized to this platform, but is also likely to induce application developers to devote less attention to rival platforms. A portal or TV network's cut in advertising rates induces advertisers to advertise more on their medium and to substitute away from other media. A shopping mall's cut in rental prices or improved layout may induce a shop to increase its size or appeal and lead the latter to neglect or abandon its outlets in rival shopping malls, and so forth.

not considered the affiliation externalities that lie at the core of the network economics literature. In contrast with the buyer of a razor, who internalizes the impact of his purchase on the demand and surplus attached to razor blades, our end-users do not internalize the impact of their purchase on the other side of the market.

Our paper is most closely related to the recent theoretical literature on chicken-and-egg problems.[4] This literature however assumes either that there is a monopoly platform (Baye-Morgan 2001, Rochet-Tirole 2002, Schmalensee 2002) or that platforms are fully interconnected (Laffont et al 2001) and so end-users enjoy the same level of connectivity regardless of the platform they select. Parker and Van Alstyne (2000) study monopoly pricing in a situation in which the demand for one good depends (linearly) on its price and on the quantity of the other good sold. They characterize the price structure as a function of the network externality coefficients. They then look at the incentive of a producer of a good to enter a (complementary or substitute) market with another incumbent producer. With complements, entry losses may be profitable because entry puts pressure on price and boosts the profit of the core business. Caillaud and Jullien (2001) study competition among intermediaries. In their model, platforms act as matchmakers and can use sophisticated pricing (registration fees, and possibly transaction fees provided the intermediaries observe transactions). Indeed, one of their contributions is to show that dominant firms are better off charging transactions rather than registrations when deterring entry. They also show that competition is more intense when platforms cannot deter multihoming. Their contribution is complementary to ours. For example, it assumes homogeneous populations on either side, and thus abstracts from the elasticity-related issues studied in our paper. Last, in a model related to that of Caillaud-Jullien, Jullien (2001) shows that an entrant represents a much stronger competitive threat on an incumbent platform when third-degree price discrimination is feasible. The ability to "divide and conquer" forces profit down, so much so that the incumbent may prefer platform compatibility.

[4]The policy implications of two-sidedness are discussed in Evans (2002). The reader will find further illustrations of two-sided markets and an interesting analysis thereof in Armstrong (2002).

2 Monopoly platform benchmark

The two-sided markets described heretofore differ is some respects, and we therefore should not aim at capturing all specificities of all industries. Our strategy will be to include a number of key ingredients common to our illustrations in a basic model, and then to generalize our analysis in order to extend its relevance to various two-sided markets. For the moment, we assume that end users incur no fixed usage cost and that platform pricing is linear. This basic model is a good representation of the credit card market; the reader may want to keep this in mind, although it will be clear that the insights have much broader generality.

Economic value is created by "interactions" or "transactions" between pairs of end users, buyers (superscript B) and sellers (superscript S). Buyers are heterogenous in that their gross surpluses b^B associated with a transaction differ. Similarly, sellers' gross surplus b^S from a transaction differ. Such transactions are mediated by a platform. The platform's marginal cost of a transaction is denoted by $c \geq 0$.

As an illustration, consider the case of payment cards. The buyer wants to purchase a bundle of goods or services from the merchant at a certain price p. In our vocabulary, a "transaction" takes place if and only if the buyer pays by card instead of using another payment instrument (say, cash). Benefits b^B and b^S correspond to differences in utility of buyers and sellers when they pay by card rather than cash. Under the No Surcharge Rule (very often imposed by payment card networks)[5] the merchant is not able to charge different retail prices for card and cash payments. Therefore the distributions of b^B and b^S are independent of the prices chosen by platforms and merchants, and can be taken as exogenous.

In the absence of fixed usage costs and fixed fees, the buyers' (sellers') demand depends only on the price p^B (respectively, p^S) charged by the monopoly platform. There are

[5]Even in the countries where the No Surcharge Rule is not imposed, as in the UK, it turns out that merchants seldom charge different prices for card and cash payments. We discuss in Section 8 the possible reasons for this fact, and more generally for the non-neutrality of the price structure in two-sided markets.

network externalities in that the surplus of a buyer with gross per transaction surplus b^B, $(b^B - p^B)N^S$ depends on the number of sellers N^S, but the buyers' "quasi"-demand:[6]

$$N^B = \Pr(b^B \geq p^B) = D^B(p^B)$$

is independent of the number of sellers. Similarly, let

$$N^S = \Pr(b^S \geq p^S) = D^S(p^S)$$

denote the sellers' quasi-demand for platform services. Consider a (buyer, seller) pair. Without loss of generality we can assume that each such pair corresponds to one potential transaction.

In contrast with search models à la Baye-Morgan (2001) or Caillaud-Jullien (2001), we take as given the matching process between buyers and sellers, and focus on the proportion of such matches that effectively results in a "transaction".[7] Assuming for simplicity the independence between b^B and b^S, the proportion (or volume) of transactions is equal to the product $D^B(p^B)D^S(p^S)$.[8]

We consider in turn the case of a private monopoly, and that of a public monopoly maximizing social welfare subject to budget balance.

2.1 Private monopoly

A private monopoly chooses selects prices so as to maximize total profit:

$$\pi = (p^B + p^S - c)D^B(p^B)D^S(p^S).$$

Assuming that D^B and D^S are log concave, it is easy to see that π is also log concave

[6]The word "quasi"-demand is used to reflect the fact that, in a two-sided market, actual demand depends on the decisions of both types of users (buyers and sellers in our terminology). In our specification, this demand is simply the product of the quasi-demands of buyers and sellers.

[7]In the payment card example, a "transaction" between a cardholder and a merchant means that the payment is by card rather than by cash.

[8]This multiplicative formula was first used by Schmalensee (2002). Most of our results can be extended to the more general case where b^B and b^S are not independent, in which case the transaction volume Q has a more general expression $Q(p^B, p^S) = \Pr(b^B \geq p^B, b^S \geq p^S)$.

(jointly in (p^B, p^S)). Its maximum is characterized by the first-order conditions:

$$\begin{aligned}\frac{\partial(\log \pi)}{\partial p^B} &= \frac{1}{p^B + p^S - c} + \frac{(D^B)'}{D^B} &= 0,\\ \frac{\partial(\log \pi)}{\partial p^S} &= \frac{1}{p^B + p^S - c} + \frac{(D^S)'}{D^S} &= 0.\end{aligned}$$

In particular:

$$(D^B)' D^S = D^B (D^S)'.$$

This condition characterizes the values of p^B and p^S that maximize volume for a given total price p: The volume impact of a small (absolute) variation of prices has to be the same on both sides. If we introduce the elasticities of quasi-demands:

$$\eta^B = -\frac{p^B (D^B)'}{D^B} \quad \text{and} \quad \eta^S = -\frac{p^S (D^S)'}{D^S},$$

the private monopoly prices can be characterized by a two-sided formula that is reminiscent of Lerner's formula:

$$p^B + p^S - c = \frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}. \tag{1}$$

In fact, the total price $p = p^B + p^S$ chosen by the private monopoly is given by the classical Lerner formula:

$$\frac{p - c}{p} = \frac{1}{\eta}, \quad \text{or} \quad p = \frac{\eta}{\eta - 1} c, \tag{2}$$

where $\eta = \eta^B + \eta^S$, the total volume elasticity, is assumed to exceed 1. What is new in formula (1) is the way in which this total price is allocated between the two sides of the market:

$$p^B = \frac{\eta^B}{\eta} p = \frac{\eta^B}{\eta - 1} c, \tag{3}$$

and

$$p^S = \frac{\eta^S}{\eta} p = \frac{\eta^S}{\eta - 1} c. \tag{4}$$

Proposition 1 : *(i) A monopoly platform's total price, $p = p^B + p^S$, is given by the standard Lerner formula for elasticity equal to the sum of the two elasticities, $\eta = \eta^B + \eta^S$:*

$$\frac{p-c}{p} = \frac{1}{\eta}. \tag{2}$$

(ii) The price structure is given by the ratio of elasticities (and not inverse elasticities):

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}. \tag{5}$$

2.2 Ramsey pricing

We consider now the case of a Ramsey monopolist maximizing welfare subject to budget balance, and derive the Ramsey formulae in our context[9]. The net surpluses on each side for an average transaction are given by standard formulae:

$$V^k(p^k) = \int_{p^k}^{+\infty} D^k(t)dt$$

for $k \in \{B, S\}$.

Under budget balance, social welfare is highest when the sum of both sides' net surpluses:

$$W = V^S(p^S)D^B(p^B) + V^B(p^B)D^S(p^S),$$

is maximized subject to the constraint:

$$p^B + p^S = c.$$

The first-order, "cost allocation" condition is:

$$\frac{\partial W}{\partial p^B} = \frac{\partial W}{\partial p^S}.$$

This gives:

$$V^S(D^B)' - D^B D^S = -D^S D^B + V^B(D^S)'.$$

After simplification, we obtain a characterization of Ramsey prices:

[9] A similar formula is derived in Laffont et al. (2001) in a model in which network externalities are reaped through platform interconnection.

Proposition 2 *: Ramsey prices embody the average surpluses created on the other side of the market and are characterized by two conditions:*

$$p^B + p^S = c \quad \textit{(budget balance)}, \tag{6}$$

and

$$\frac{p^B}{\eta^B}\left[\frac{V^B}{D^B}\right] = \frac{p^S}{\eta^S}\left[\frac{V^S}{D^S}\right] \ \textit{(cost allocation)}. \tag{7}$$

Condition (7) characterizes the price structure that maximizes social surplus for a given total price p. Returning to the formula yielding the private monopolist's price structure,

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}, \tag{1'}$$

the additional terms in formula (7) (the bracketed terms) reflect the average surpluses per transaction for buyers and sellers. [Later, when we compare price structures across governance forms, we will compare prices for a given price level. That is, we will say that two governance forms generate the same price structure if they give rise to the same prices for a given price level target $p = p^B + p^S$. Of course different governance forms generate different price levels.]

3 Competing platforms

3.1 Modeling

We now assume that two platforms compete for the markets (we will also look at the case in which both platforms are jointly owned, in order to compare the outcome under platform competition with those obtained in section 2 in the private monopoly and Ramsey cases). *End-users' benefits.* As earlier, buyers and sellers are heterogenous: Their benefits from transacting vary across the two populations and are private information. These benefits are denoted b_i^B for the buyer (when the transaction takes place on platform i) and b^S for

the seller, and are drawn from continuous distributions.[10] The proportional fees charged by platform i are p_i^B for buyers and p_i^S for sellers. A buyer with gross surplus b_i^B from transacting on platform i is willing to use that platform provided that $b_i^B \geq p_i^B$. However, the buyer prefers to transact on platform j if $b_j^B - p_j^B > b_i^B - p_i^B$. Similarly, a seller with type b^S is willing to trade on platform i provided that $b^S \geq p_i^S$, and prefers to trade on platform j if $p_j^S < p_i^S$.

Notice that a transaction can occur only if the two sides have at least one platform in common; that is, there exists at least one platform on which both are willing to trade. If both "multihome" (are affiliated with both platforms), the choice of platform is a priori indeterminate. In accordance with our illustrations, we assume that, whenever a seller is affiliated with the two platforms, the buyer chooses the one on which the transaction takes place.[11]

Transaction volumes. The buyers' behavior generates "quasi-demand functions":

$$D_i^B = D_i^B(p_i^B) = \Pr(b_i^B - p_i^B > 0), \tag{8}$$

and

$$d_i^B(p_1^B, p_2^B) = \Pr\left[b_i^B - p_i^B > \max(0, b_j^B - p_j^B)\right]. \tag{9}$$

D_i^B is the proportion of buyers who are willing to use platform i when the seller is affiliated only with platform i. Similarly, d_i^B is the proportion of buyers who are willing to trade on platform i when the seller multihomes. By construction, these functions satisfy the following properties:

$$d_i^B \leq D_i^B \leq d_1^B + d_2^B. \tag{10}$$

[10] For simplicity, we assume that the seller's gross surplus does not depend on the platform where the transaction takes place. Furthermore, when performing the welfare analysis, we equate these benefits with the social values of the service brought about by the platforms. However, sellers may exert externalities on each other. For example, a seller's acceptation of a payment card may affect rival sellers. The welfare analysis (but not the positive one) must be amended correspondingly. For more on this, see Rochet-Tirole (2002).

[11] This assumption is satisfied by most of our illustrations: a cardholder selects the card when the merchant accepts multiple cards, the reader or viewer selects the newspaper, portal or TV network, the videogame user selects the platform if the game is written for several consoles, etc. Notice that this assumption introduces a slight asymmetry between the two sides of the market.

We assume that the distribution of (b_1^B, b_2^B) is symmetric, which implies that demand functions are also symmetric: $D_1^B(p^B) = D_2^B(p^B) \equiv \hat{D}^B(p^B)$ and $d_1^B(p_1^B, p_2^B) \equiv d_2^B(p_2^B, p_1^B)$. When prices are equal $p_1^B = p_2^B = p^B$, we will use the simplified notation:

$$d^B(p^B) \equiv d_i^B(p^B, p^B).$$

We focus for the moment on symmetric prices: $p_1^B = p_2^B = p^B$ and $p_1^S = p_2^S = p^S$. A seller of type b^S affiliates with both platforms when $b^S \geq p^S$ and none otherwise. The transaction volumes on each platform are thus equal to

$$Q = d^B(p^B) D^S(p^S). \tag{11}$$

The sellers' net surplus is, as earlier,

$$V^S(p^S) = \int_{p^S}^{+\infty} D^S(t)dt,$$

while the buyers' net surplus is

$$\begin{aligned} V^B(p_1^B, p_2^B) &= \int_{p_1^B}^{+\infty} d_1^B(t_1, p_2^B)dt_1 + \int_{p_2^B}^{+\infty} D_2^B(t_2)dt_2 \\ &= \int_{p_2^B}^{+\infty} d_2^B(p_1^B, t_2)dt_2 + \int_{p_1^B}^{+\infty} D_1^B(t_1)dt_1. \end{aligned}$$

Joint ownership benchmarks. The private monopoly and Ramsey benchmarks studied in Section 2 correspond to the situation in which both platforms are under joint ownership and charge identical prices. For instance,

$$D^B(p^B) = 2d^B(p^B)$$

where

$$d^B(p^B) = d_1^B(p^B, p^B) = d_2^B(p^B, p^B).$$

Governance. We assume that the two platforms are controlled by competing entities, either profit-maximizing firms (Section 3.3) or not-for-profit associations (Section 3.4). Important examples of such associations can be found in the payment card industry (Visa and MasterCard). In such associations, prices for buyers and sellers are determined by competition (both intra and inter platforms) on downstream markets (issuing banks on the buyers' side, and acquirers on the sellers' side).

3.2 Transaction volumes for asymmetric prices

In order to analyze competition, we need to determine transaction volumes on each platform for arbitrary prices, thus extending formula (11) to nonsymmetric prices. Suppose that platform 1 is cheaper for sellers: $p_1^S < p_2^S$. A seller of type b^S has three possibilities:[12] no trade, affiliation with platform 1 only, affiliation with both platforms. The first possibility is optimal whenever $b^S \leq p_1^S$. The choice between the other two possibilities involves a trade-off between a lower volume (when affiliated with platform 1 only) and an obligation to trade on the most expensive platform (when affiliated with both platforms). The corresponding expected net surpluses of a seller of type b^S are respectively $(b^S - p_1^S)D_1^B(p_1^B)$ and $(b^S - p_1^S)d_1^B(p_1^B, p_2^B) + (b^S - p_2^S)d_2^B(p_1^B, p_2^B)$. The seller chooses to multihome when b^S is large enough, more precisely when

$$b^S > \hat{b}_{12} \equiv \frac{p_2^S d_2^B - p_1^S(D_1^B - d_1^B)}{d_2^B - (D_1^B - d_1^B)}. \tag{12}$$

We can now summarize sellers' optimal decisions:

- sellers with low types ($b^S \leq p_1^S$) do not trade,
- sellers with high types ($b^S \geq \hat{b}_{12}$) trade on both platforms,
- sellers with intermediate types ($p_1^S < b^S < \hat{b}_{12}$) only trade on the less expensive platform (here, platform 1).

By undercutting the rival platform, each platform thus induces some sellers (those with intermediate types) to stop multihoming, a strategy known as "steering". The formulae for $p_1^S > p_2^S$ are obtained by permutation of indices. When p_1^S and p_2^S converge to the same price p^S, $\hat{b}_{12}$ and $\hat{b}_{21}$ both converge also to p^S, which establishes continuity of the formulae giving $\hat{b}_{12}$ and $\hat{b}_{21}$.

Let us denote by σ_i $(i = 1, 2)$ the following indices:

$$\sigma_i = \frac{d_1^B + d_2^B - D_j^B}{d_i^B} \qquad i, j = 1, 2;\ i \neq j.$$

[12] Affiliation with platform 2 only is clearly dominated.

Given property (3), σ_i belongs to the interval $[0,1]$. It measures the "loyalty" of consumers of platform i, i.e. the proportion of them who stop trading when platform i ceases to be available. We call σ_i the "singlehoming" index of platform i. It is equal to 0 when buyer demand faced by the seller is independent of whether the seller is affiliated with platform i ($d_1^B + d_2^B = D_j^B$). It is equal to 1 when all platform i buyers are lost when the seller stops affiliating with that platform ($D_j^B = d_j^B$). For a symmetric price configuration (with $D_1^B = D_2^B = \hat{D}^B$), we have

$$\sigma_1 = \sigma_2 = \sigma = 2 - \frac{\hat{D}^B}{d^B}.$$

Starting from a symmetric price structure, suppose platform 1 decreases p_1^S by a small amount ε. This increases demand for platform 1 in two ways: The platform attracts new merchants ($p_1^S - \varepsilon \leq b_1^S < p_1^S$) and "steers" former multihoming merchants ($p_1^S < b_1^S < \hat{b}_{12}$). Given that $\frac{\partial \hat{b}_{12}}{\partial p_1^S} = 1 - \frac{1}{\sigma_2}$, the effectiveness of steering depends on σ_2: it is nil when $\sigma_2 = 1$ and infinite when $\sigma_2 = 0$.

We are now in a position to determine the volume of transactions on each platform as a function of prices p_i^B and p_i^S. We restrict ourselves to the case $p_1^S \leq p_2^S$ (the case $p_2^S < p_1^S$ is obtained by symmetry). Let us denote by D^S the sellers' "quasi-demand function":

$$D^S(p^S) = \Pr(b^S > p^S).$$

From the affiliation decisions derived above, a proportion $D^S(\hat{b}_{12})$ of sellers multihome, while a proportion $D^S(p_1^S) - D^S(\hat{b}_{12})$ are affiliated only with platform 1. Assuming that the probability of a meeting between a buyer and a seller is independent of their types, the total expected volumes of transactions on the platforms are:

$$Q_1 = d_1^B(p_1^B, p_2^B) D^S(\hat{b}_{12}) + D_1^B(p_1^B) \left\{ D^S(p_1^S) - D^S(\hat{b}_{12}) \right\}, \tag{13}$$

for platform 1, and:

$$Q_2 = d_2^B(p_1^B, p_2^B) D^S(\hat{b}_{12}), \tag{14}$$

for platform 2, where $\hat{b}_{12}$ is given by formula (12). As already noticed, these formulae are continuous across the "diagonal" $p_1^S = p_2^S$:

$$\lim_{\substack{p_1^S \to p^S \\ p_2^S \to p^S}} Q_i = d_i^B(p_1^B, p_2^B) D^S(p^S).$$

3.3 Competition between proprietary platforms

Proprietary platforms choose prices so as to maximize profit. Consider for example platform 1's profit:

$$\pi_1 = (p_1^B + p_1^S - c) Q_1. \tag{15}$$

As in the case of a monopolist, this maximization can be decomposed into the choice of a price level, $p_1 = p_1^B + p_1^S$, and that of a price structure given a price level. The first-order conditions are:

$$Q_1 + (p_1^B + p_1^S - c)\frac{\partial Q_1}{\partial p_1^B} = Q_1 + (p_1^B + p_1^S - c)\frac{\partial Q_1}{\partial p_1^S} = 0,$$

or

$$\frac{\partial Q_1}{\partial p_1^S} = \frac{\partial Q_1}{\partial p_1^B} = -\frac{Q_1}{p_1^B + p_1^S - c}. \tag{16}$$

The following analysis is complex, as it must handle a potential lack of smoothness of the objective function. It can be skipped in a first reading. The end result (Proposition 3) is remarkably simple, though.

Recall the expressions of volumes on both systems, when, say, $p_1^S \leq p_2^S$:

$$Q_1 = d_1^B(p_1^B, p_2^B) D^S(\hat{b}_{12}) + D_1^B(p_1^B)\{D^S(p_1^S) - D^S(\hat{b}_{12})\}, \tag{13}$$

$$Q_2 = d_2^B(p_1^B, p_2^B) D^S(\hat{b}_{12}), \tag{14}$$

where

$$\hat{b}_{12} = \frac{p_2^S d_2^B - p_1^S(D_1^B - d_1^B)}{d_2^B - (D_1^B - d_1^B)}. \tag{12}$$

We focus on symmetric equilibria ($p_i^S \equiv p^S, p_i^B \equiv p^B$), for which volumes have simpler expressions:

$$Q_i = d_i^B(p^B, p^B)D^S(p^S).$$

While

$$\frac{\partial Q_1}{\partial p_1^B} = \frac{\partial d_1^B}{\partial p_1^B}(p^B, p^B)D^S(p^S), \tag{17}$$

the first derivative in formula (17) is not necessarily well defined since volumes have a different expression according to whether $p_1^S \leq p_2^S$ or $p_1^S > p_2^S$:

$$Q_1 = d^B(p^B)D^S(\hat{b}_{12}) + \hat{D}^B(p^B)\{D^S(p_1^S) - D^S(\hat{b}_{12})\}$$

when $p_1^S < p_2^S$, and

$$Q_1 = d^B(p^B)D^S(\hat{b}_{21})$$

when $p_1^S > p_2^S$. Interestingly, Q_1 turns out to be differentiable[13] even at $p_1^S = p_2^S$. Indeed, at symmetric prices:

$$\frac{\partial Q_1}{\partial p_1^S} = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B}. \tag{18}$$

[13]The left- and right-derivatives of Q_1 with respect to p_1^S at $p_1^S = p_2^S = p^S$ (implying $\hat{b}_{12} = \hat{b}_{21} = p^S$) are:

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_L = (D^S)' \frac{\partial \hat{b}_{12}}{\partial p_1^S}[d^B - \hat{D}^B] + (D^S)'\hat{D}^B,$$

and

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_R = (D^S)' \frac{\partial \hat{b}_{21}}{\partial p_1^S} d^B.$$

Moreover

$$\frac{\partial \hat{b}_{12}}{\partial p_1^S} = -\frac{\hat{D}^B - d^B}{2d^B - \hat{D}^B}, \quad \text{and} \quad \frac{\partial \hat{b}_{21}}{\partial p_1^S} = \frac{d^B}{2d^B - \hat{D}^B}.$$

And so

$$\begin{aligned}
\left(\frac{\partial Q_1}{\partial p_1^S}\right)_L &= (D^S)' \left[\frac{(\hat{D}^B - d^B)^2}{2d^B - \hat{D}^B} + \hat{D}^B\right] = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B} \\
\left(\frac{\partial Q_1}{\partial p_1^S}\right)_R &= (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B}.
\end{aligned}$$

Thus Q_1 is differentiable with respect to p_1^S.

Using (16), (17) and (18) we obtain a simple form for the first-order condition for a symmetric equilibrium:

$$\frac{\partial d_i^B}{\partial p_i^B} D^S = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B},$$

or:

$$\left(\frac{2d^B - \hat{D}^B}{d^B}\right)\left(-\frac{\partial d_i^B/\partial p_i^B}{d^B}\right) = -\frac{(D^S)'}{D^S}.$$

The first term on the left-hand side of this latter formula is the singlehoming index σ defined earlier, which measures the proportion of "unique customers". The second term is the ratio of the own-brand elasticity of demand for buyers

$$\eta_o^B = -\frac{p^B \partial d_i^B/\partial p_i^B}{d^B}$$

over the buyers' price p^B. Finally, the last term is the ratio of the elasticity of sellers' demand over sellers' price. Thus we can state:

Proposition 3 *: A symmetric equilibrium of the competition between proprietary platforms is characterized by:*

$$p^B + p^S - c = \frac{p^B}{\eta_o^B} = \frac{p^S}{(\eta^S/\sigma)}.$$

The formulae are thus the same as in the monopoly platform case, except that a) on the buyer side, the demand elasticity η^B is replaced by the (higher) own-brand elasticity η_o^B, and b) on the seller side, the demand elasticity η^S is replaced by the equivalent of an own-brand elasticity η^S/σ. When all buyers singlehome ($\sigma = 1$), the own-brand elasticity and the demand elasticity coincide. But as multihoming becomes more widespread (σ decreases), the possibility of steering increases the own-brand elasticity η^S/σ.

3.4 Competition between associations

When platforms are run by not-for-profit cooperatives owned by members (operators on the buyer and seller sides), prices paid by the end users are set by the members and not by the platforms. Platforms however have an important say in the price structure,

especially if competition among members is intense on both sides of the market. In our model, an association's only strategic decision is the choice of access charges between members. Neglecting platform costs, the zero-profit condition implies that these access "charges" exactly offset each other as one side receives the charge paid by the other side. For example in the payment card industry the access charge is called the interchange fee and is paid by acquirers (the sellers' banks) to issuers (the buyers' banks).[14] This section studies the access charge chosen by competing associations and compares the corresponding prices for final users (buyers and sellers) with those resulting from competition between profit-maximizing systems. While the section is currently most relevant to the payment card industry, its potential applicability is much broader. For example, reflecting recent concerns about unequal access to B2B exchanges, some have suggested that these exchanges be run as non-profit associations. Furthermore, and as will be observed in section 7.2, networks of interconnected networks (e.g. communication networks) are economically similar to non-profit platforms.

The members compete on two downstream markets, the buyer and the seller downstream markets. Given access charge a_i on platform i, the net marginal costs for a member of platform i of serving a buyer and a seller, respectively, are $c^B - a_i$ and $c^S + a_i$, where c^B and c^S represent the gross marginal costs incurred by the members on each side of the market. We make the simplifying assumption that intraplatform competition results in constant equilibrium margins charged by members on downstream markets: m^B on the buyers' side and m^S on the sellers' side. Equilibrium prices are thus given by:

$$p_i^B = c^B - a_i + m^B, \quad p_i^S = c^S + a_i + m^S.$$

This assumption is for example satisfied if (a) members belong to a single association and are differentiated in a direction orthogonal to that of platform differentiation;[15] and (b) members on a given platform are little differentiated. Intense intraplatform competition

[14] The determination of access charges within associations has so far only been studied in the context of the payment card industry and under the assumption of a monopoly platform (Rochet-Tirole (2002), Schmalensee (2002)).

[15] Mathematically, in a generalized Hotelling framework, the "transportation cost" for an end-user when

then results in Hotelling competition between members taking as given (as a first-order approximation) the number of end users on the platform (which is basically determined by the platforms' access charges given that the members' markups are small).[16]

Under this simplifying assumption, the profits of all members of an association are proportional to the volume of transactions on the association's platform. The interests of all members are thus completely aligned. Regardless of its exact structure the association selects the access charge so as to maximize its volume. Furthermore the total price on each system is constant:

$$p_i^B + p_i^S = c + m, \tag{19}$$

where $m = m^B + m^S$ is the total margin on downstream markets and $c = c^B + c^S$.

Last, in order to be able to compare the association with the cases of a monopolist and of competing proprietary platforms, we must assume that the quasi-demand functions are the same. That is, the members are only selling the varieties of each platform that the proprietary platforms were selling. Because we kept quasi-demand functions quite general, there is no difficulty in assuming this is indeed the case.

The outcome of the competition between the two associations is characterized by two price vectors (p_i^B, p_i^S), $i = 1, 2$, such that: for all i, (p_i^B, p_i^S) maximizes the volume Q_i on system i subject to (19), taking as given the price vector (p_j^B, p_j^S) on the other system.

The first-order conditions for a symmetric equilibrium are given by

$$p^B + p^S = c + m, \tag{20}$$

selecting a (platform, member) pair is the sum of the transportation cost to the platform and that to the member.

[16] If members have dual membership instead (eg. they are both affiliated with Visa and MasterCard, or they provide support or write applications for two cooperatively designed operating systems or videogame platforms), then requirement (b) is unnecessary in that margins are constant even if member differentiation is not small relative to platform differentiation: See Hausman et al. (2003). But one must then inquire into the associations' governance structure. Our treatment carries over as long as governance leads each association to maximize its volume.

(condition on total price) and the equivalent of condition (5):

$$\frac{\partial Q_i}{\partial p_i^B} = \frac{\partial Q_i}{\partial p_i^S}, \tag{21}$$

(same impact on volume of a marginal price increase on each side of the market).

The analysis of the price structure is therefore identical to that for proprietary platforms. The price *level* is lower for associations with healthy competition among their members but may exceed the proprietary platforms price level if double marginalization is strong.

Proposition 4 : *A symmetric equilibrium of the competition between associations is characterized by*

$$p^B + p^S = c + m$$

and

$$\frac{p^B}{\sigma \eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

Comparing now Proposition 2 and 4, we see that even when downstream markets are perfectly competitive (the margin m converges to zero) and so the price level is socially optimal, competition between not-for-profit associations need not generate an efficient outcome. Indeed, the condition for an efficient price structure (given in Proposition 2) is:

$$\frac{p^B}{\eta^B}\left[\frac{V^B}{D^B}\right] = \frac{p^S}{\eta^S}\left[\frac{V^S}{d^S}\right], \tag{7}$$

while the condition characterizing competition between associations is different:

$$\frac{p^B}{\sigma \eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

This is natural, as (a) the associations do not internalize the end-users' surpluses, and (b) the associations aim at steering sellers (which is reflected by the presence of σ) and stealing buyers (as indicated by the presence of η_o^B) away from the rival association, while market share considerations play no role in a Ramsey program. It is therefore perhaps remarkable that the two conditions coincide in the special case of linear demands, which we explore in detail in Section 5.

4 Determinants of business model

Let θ be a parameter that affects the volume of transactions on the platforms. In this section, we consider the impact of a small variation in θ on user prices p^B and p^S, depending on industry structure (monopoly or duopoly) and on the platforms' governance structure (for-profits or associations). We concentrate on three important determinants of industry conduct and performance:

Marquee buyers: In the first application, θ represents a (small) uniform shift in sellers' surpluses, due to the presence of marquee buyers on the other side of the market. As a result, the sellers' demand function becomes:

$$D^S(p^S, \theta) = D^S(p^S - \theta).$$

Installed bases/captive buyers: In the second application, θ represents the (small) mass of buyers who are loyal to their platform, independently of prices. Such buyers, say, are tied by long-term contracts. As a result, the buyers' demand functions become:

$$d_i^B(p_1^B, p_2^B, \theta) = d_i^B(p_1^B, p_2^B) + \theta, \;\; D^B(p^B, \theta) = D^B(p^B) + \theta, \; \hat{D}^B(p^B, \theta) = \hat{D}^B(p^B) + \theta.$$

Multihoming: In the third application, θ represents an exogenous increase in the single-homing index of buyers. Assume for example that d^B does not depend on θ, while D^B decreases in θ. Then $\sigma(p^B, \theta) = 2 - \frac{D^B(p^B,\theta)}{d^B(p^B)}$ is increasing in θ, while η_o^B does not depend on θ.[17]

Proposition 5 analyses the impact of small variations of θ on the prices p^B and p^S.

Proposition 5 *: (i) In the case of a monopoly platform (for-profit or association) and with log concave demand functions, the seller price increases when there are marquee buyers and decreases when there are captive buyers. The buyer price moves in the opposite direction.*

[17] This is for example the case in the Hotelling specification presented in Section 5, when the marginal transportation cost of buyers increases only for distances in the noncompetitive hinterland of the rival platform, so that d_i^B is unaffected while D^B decreases.

(ii) The same result holds under competition between associations, except that the comparative statics with respect to captive buyers requires a regularity condition.
(iii) In the case of competing associations, an increase in the multihoming index of buyers (keeping demand elasticities constant) leads to an increase in the buyer price and a decrease in the seller price.

Intuitively, marquee buyers make the platform more attractive for the sellers. The platform then raises its price p^S to sellers, which reduces the de facto marginal cost, $c - p^S$, of provision of the service to the buyers. The buyer price therefore falls. The intuition is similar in the case of captive buyers. Captive buyers allow the platform to raise the price p^B to buyers, thus reducing the de facto marginal cost $c - p^B$ of serving sellers. A regularity condition however is required here in the case of platform competition, which creates a countervailing steering effect: Each platform's buyer membership is then "more unique" to the platform, and so it is more costly for a seller to forgo the platform. Last, an increase in multihoming makes steering more attractive and puts a downward pressure on the seller price.

5 Linear demands

We illustrate the results obtained so far in a variant of the Hotelling model, where a buyer's preferences for platforms are represented by his location x on a line. Buyers are uniformly distributed on a line of length $(\Delta + 2\delta)$. Platform 1 and 2 are symmetrically located at a distance $\Delta/2$ of the origin of the line ($x = -\frac{\Delta}{2}$ for platform 1 and $x = \Delta/2$ for platform 2). The number Δ parametrizes the degree of substitutability between platforms. Buyers have also access to outside options, represented conventionally by two other symmetric platforms (denoted $1'$ and $2'$), located further away from the origin ($x = -\frac{\Delta}{2} - \delta$ and $x = \frac{\Delta}{2} + \delta$) and charging the same, exogenous, price p_0. The number δ will serve us as a measure of the weight of "unique consumers". When using a platform located at distance d, buyers incur a quadratic transportation cost $\frac{1}{2}d^2$, (the transportation cost parameter

is normalized to 1 without loss of generality).

Proposition 6 (proved in the Appendix) exhibits three main implications of the linear case. First, the price *structure* is the same regardless of whether the industry is served by a private monopoly, competing proprietary platforms or competing platforms. Second, if demand is linear on the seller side as well, then this common price structure is Ramsey optimal. Taken together, these results show that without detailed information about the demand structure, one should not expect clear comparisons of price structures across governance mechanisms. Nor are policy interventions to alter the price structure (as opposed to the price level) likely to be solidly grounded. Third, Proposition 6 provides sufficient conditions for the second-order conditions to be satisfied in the linear demand case.

Proposition 6 *: Suppose that the buyers' quasi-demand is described by an Hotelling model, with uniform distribution and outside options with distance Δ between the two platforms and distance δ between each platform and its nearest outside option, and that the market is not covered (not all potential buyers buy).*

(i) • The buyer singlehoming index is equal to:

$$\sigma = \Delta/(\Delta + \delta),$$

and decreases when the platforms become more substitutable.

• The platforms' ability to steer (discourage through undercutting sellers from multihoming) decreases with the buyer singlehoming index.

• On the buyer side, total elasticity is equal to own-brand elasticity times the singlehoming index:

$$\eta^B = \eta_o^B \sigma.$$

(ii) The price structure is the same under a monopoly platform, competing proprietary platforms and competing associations. It satisfies

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}.$$

(iii) If furthermore seller demand is linear, then the price structure in the three environments is Ramsey optimal.

(iv) The price vectors given in formulae (31) and (32) satisfy the second-order conditions for an equilibrium if and only if $\frac{\delta}{\Delta}$ *is smaller than* $\frac{1+\sqrt{5}}{2}$.

6 Generalization to fixed user fees and usage costs

In many of the examples presented in the introduction, fixed costs, either fixed fees charged by the platforms or fixed usage costs, play an important role. In order to demonstrate the robustness of our results to the introduction of fixed costs, we now adapt our model accordingly. To simplify the analysis, we assume that buyers singlehome (for example, consumers read a single newspaper or connect to a single portal). Second, we focus on the symmetric equilibrium. There is a sizeable literature on tipping in the presence of user fixed costs and we have little to add to this literature. Last, we first look at the case in which there is no direct exchange of money between the two sides of the market, as is the case for advertising in newspaper, TV and portals; we will later show how to extend the analysis to cases, such as videogames, exhibiting direct monetary transactions between end-users.

Platforms incur fixed costs C^B and C^S per buyer and seller, as well as marginal cost c per transaction between them (presumably $c = 0$ for advertising). Let platform i charge fixed fees A_i^B and A_i^S and variable charges $a_i^B N_i^S$ and $a_i^S N_i^B$ to buyers and sellers, where N_i^B and N_i^S are the numbers of buyers (eyeballs) and sellers (advertisers) connected to platform i. A buyer with (possibly negative) average benefit b_i^B of receiving an ad and with fixed usage cost γ_i^B (also possibly negative) has net utility

$$U_i^B = (b_i^B - a_i^B)N_i^S - A_i^B - \gamma_i^B.$$

Similarly, a seller with average benefit b^S of reaching a consumer and with fixed cost γ^S of designing an ad for this newspaper has net utility:

$$U_i^S = (b^S - a_i^S)N_i^B - A_i^S - \gamma^S.$$

The buyers are heterogenous over parameters (b_i^B, γ_i^B) and sellers are heterogenous over parameters (b^S, γ^S).

The strategic choices for the platforms are the per "transaction" (eyeball viewing an ad) markups:

$$p_i^B \equiv a_i^B + \frac{(A_i^B - C^B)}{N_i^S} \text{ and } p_i^S \equiv a_i^S + \frac{(A_i^S - C^S)}{N_i^B}.$$

Assuming that readers buy a single newspaper, the number of copies sold by newspaper i is given by

$$N_i^B = \Pr(U_i^B > \max(0, U_j^B)),$$

which is equal to some function d_i^B of prices (p_1^B, p_2^B) and numbers of ads (N_1^S, N_2^S) of the two newspapers

$$N_i^B = d_i^B(p_1^B, N_1^S; p_2^B, N_2^S) \equiv \Pr\left((b_i^B - p_i^B)N_i^S - C^B - \gamma_i^B \geq \max\left[0, (b_j^B - p_j^B)N_j^S - C^B - \gamma_j^B\right]\right). \tag{23}$$

N_i^S is itself a function of p_i^S and N_i^B:

$$N_i^S = D^S(p_i^S, N_i^B) = \Pr\left((b^S - p_i^S)N_i^B > \gamma^S\right). \tag{24}$$

These formulas are valid provided fixed costs for buyers are high enough so that no buyer buys the two newspapers (no multihoming for buyers). Substituting (24) into (23), and solving for (N_1^B, N_2^B), one obtains demand functions for the buyers:

$$N_i^B = n_i^B(p_1^B, p_1^S; p_2^B, p_2^S).$$

Let us define the own -and cross- elasticities for buyer demand:

$$\eta_o^B \equiv -\frac{\partial n_i^B}{\partial p_i^B}\frac{p_i^B}{n_i^B} \text{ and } \eta_S^B \equiv -\frac{\partial n_i^B}{\partial p_i^S}\frac{p_i^S}{n_i^B}.$$

On the seller side, we define the own-price elasticity and the network elasticity:

$$\eta^S \equiv -\frac{\partial D^S}{\partial p^S}\frac{p^S}{D^S} \text{ and } \eta_N^S \equiv \frac{\partial D^S}{\partial N^B}\frac{N^B}{D^S}.$$

With this notation, the formulae for transaction volumes and platform profit look remarkably similar to the ones obtained earlier. Platform i maximizes:

$$\pi_i = (p_i^B + p_i^S - c)N_i^B N_i^S.$$

Simple computations yield:

Proposition 7 *: A symmetric equilibrium is characterized by prices (p^B, p^S) satisfying:*

$$p^B + p^S - c = \frac{p^B}{\eta_o^B(1+\eta_N^S)} = \frac{p^S}{\eta^S + \eta_S^B(1+\eta_N^S)}.$$

While we simplified the model by assuming singlehoming ($\sigma = 1$), the presence of fixed costs implies that network externalities impact not only end-user surpluses, but also demands. For example, on the buyer side, the own price elasticity η_o^B is multiplied by a factor greater than 1 to account for the fact that when a platform reduces its buyer price, more buyers connect to the platform, inducing more sellers to connect and further increasing buyer demand. And similarly on the seller side.

In some more structured applications, the formulae in Proposition 7 simplify. For example, in the advertising example, it is reasonable to assume that sellers incur *no fixed usage cost* ($\gamma^S \equiv 0$), since the advertising campaign has already been prepared for other media. In this case formula (24) shows that D^S does not depend on N^B, so that $\eta_N^S = 0$, and

$$p^B + p^S - c = \frac{p^B}{\eta_o^B} = \frac{p^S}{\eta^S + \eta_S^B}.$$

Last, let us turn to the (videogame or operating system) case in which the transaction between the seller and the buyer involves a price charged by the seller to the buyer.

Additional complications arise because of this monetary transaction between buyers and sellers. The equilibrium price of this transaction is then determined by competitive forces in the market for videogames or software applications and depends on the pricing policies of platforms. To illustrate how to extend the model to reflect this, we assume

that sellers have market power and no marginal cost and that buyers differ only in the fixed cost of learning how to install and use an operating system or a console (and in the identity of their preferred applications). They receive gross surplus $\underline{v}$ for a fraction α of the applications (where the corresponding applications are drawn in i.i.d. fashion among consumers) and $\bar{v} > \underline{v}$ for a fraction $(1-\alpha)$. When α is large (so that $(1-\alpha)\bar{v} < \underline{v}$), it is efficient for the platforms to induce developers to charge the low price $p = \underline{v}$, so that buyers buy all games and receive a net marginal surplus $b^B = (1-\alpha)(\bar{v} - \underline{v})$. Then we can assume w.l.o.g. that $a^S = 0$, so that $b^S = \underline{v}$. Using the same notation as above, the net utilities of a typical buyer and a typical seller are

$$\begin{aligned} U_i^B &= b^B N_i^S - A_i^B - \gamma_i^B, \\ U_i^S &= b^S N_i^B - A_i^S - \gamma^S. \end{aligned}$$

Denoting again by p_i^B and p_i^S the per transaction mark-ups:

$$p_i^B = \frac{A_i^B - C^B}{N_i^S} \quad \text{and} \quad p_i^S = \frac{A_i^S - C^S}{N_i^B},$$

and d_i^B, D^S the associated demand functions, we obtain the same formulae as in Proposition 7.

7 Summary and mini case studies

Let us now summarize the paper's key insights. The main contribution has been to derive simple formulae governing the price structure in two-sided markets, and this for a wide array of governance structures (private monopoly, Ramsey planner, competition between for-profit or non-profit platforms). But we also obtained more specific insights. On the public policy side:

1) The Ramsey price structure does not correspond to a "fair cost allocation". Rather, like private business models, it aims at getting both sides on board.

2) The main conceptual difference between private and Ramsey price structures is that the latter takes into account the average net surplus created on the other side of the

market when attracting an end user on one side. Yet, private business models do not exhibit any obvious price structure bias (indeed, in the special case of linear demands, all private price structures are Ramsey optimal price structures).

On the *business model* front, we obtained:

3) Monopoly and competitive platforms design their price structure so as to get both sides on board.

4) An increase in multihoming on the buyer side facilitates steering on the seller side and results in a price structure more favorable to sellers.

5) The presence of marquee buyers (buyers generating a high surplus on the seller side) raises the seller price and (in the absence of price discrimination on the buyer side) lowers the buyer price.

6) Captive buyers tilt the price structure to the benefit of sellers.

We now develop seven "mini case studies" meant to emphasize the attention paid by platforms to the pricing structure. A rigorous validation of testable implications 3) through 6) lies beyond the scope of this paper, and we hope that future research will perform the econometric studies needed to confirm or infirm these hypotheses in specific industries. We only offer some casual empiricism; this preliminary evidence seems quite encouraging for the theory.

7.1 Credit and debit cards

The payment industry offers a nice illustration of implications 3) through 6). Historically, the business model for credit and differed debit cards has been to attract cardholders and induce them to use their cards. Visa and MasterCard are not-for-profit associations owned by over 6,000 bank (and nonbank) members. The associations centrally set interchange fees to be paid by acquirers (the merchants' banks) to issuers (the cardholders' banks). These interchange fees are proportional to transaction volume. A higher interchange fee is, via the competition among issuers, partly or fully passed through to consumers in the form of lower card fees and higher card benefits, which encourages card ownership and usage;

and, via the competition among acquirers, partly or fully passed through to merchants, who pay a higher merchant discount (the percentage of the sale price that the merchant must pay the acquirer), which discourages merchant acceptance. The associations' choice of interchange fees have typically favored cardholders over merchants who kept accepting the card despite the high level of the merchant discounts (implication 3).[18]

American Express, a for-profit closed system, works on the same business model, with an even higher degree of cross-subsidization. Traditionally, it has charged a much higher merchant discount.[19] It could afford to do so because the Amex clientele -in particular the corporate card clientele- was perceived as very attractive by merchants (hypothesis 5). The gap between Amex's and the associations' merchant discounts has narrowed in the 1990s as more and more Amex customers got also a Visa card or MasterCard. Such "multihoming" by a fraction of cardholders made it less costly for merchants to turn down Amex cards (implication 4).

The on-line debit card market in the US has adopted an entirely different business model. Rather than courting consumers, it has wooed merchants through a low interchange fee. One key difference with credit and differed debit cards is that consumers indeed do not need to be courted (they already have in their pocket an ATM card, that they can use as an on-line debit card; so in a sense they are "captive"), while merchants, to perform on-line debit, must install costly pinpads (which most of them have not yet done).[20] This emphasizes the relevance of hypotheses 3) and 6).

7.2 Internet

In the Internet, the instrument of cross-subsidization is the termination or settlement charge (or lack thereof) between backbones. The termination charge for off-net traffic

[18] Looking forward, it is likely that merchant card acceptance will become more elastic with the (ongoing) advent of on-line debit and the (future) introduction of Webplatforms.

[19] And thus implicitly a much higher "interchange fee". For Amex, the interchange fee is only implicit, since the company is vertically integrated and performs the three roles of issuer, system and acquirer.

[20] The on-line offerings were first made by regional ATM networks. A number of these networks have now been consolidated and converted into a for-profit platform (Concord ESF).

between two backbones is, as its name indicates, the price paid by the originating backbone to the terminating backbone for terminating the traffic handed over. It can be shown[21] that it is optimal to charge the "off-net cost" to end users for marginal incoming and outgoing traffic. That is, backbones should charge as if the traffic were off net, even though a fraction of the traffic is actually on net. The charge for incoming (outgoing) traffic decreases (increases) one-for-one with the termination charge. This implies that a high termination charge is indirectly borne by end users, like websites, whose volume of outgoing traffic far exceeds the volume of incoming traffic, and benefits end users, such as dial-up customers, who mostly receive traffic (downloads).

An internet in which backbones exchange traffic at a uniform volume-proportional termination charge is similar to the case of a single not-for-profit platform. This analogy can be best depicted by envisioning backbones exchanging traffic at public peering points.[22] An "association" running these public peering points and keeping track of bilateral termination charges would be similar to a credit-card association recording traffic between acquirers and issuers, with the termination charge the counterpart of the interchange fee. A network of interconnected networks therefore resembles a single not-for profit platform.

The Internet is still mostly run by pre-deregulation legacy arrangements, according to which the backbones charge nothing to each other for terminating traffic. This business model is currently being reexamined and it is quite possible that, as is the case for regular telephony, positive termination charges will be levied in the future. The legacy arrangements may well have made sense in an epoch in which the posting of content on the Web had to be encouraged. A key question now is whether a change in industry conditions motivates a move toward paying settlements.

[21] See Laffont et al. (2001) and Jeon et al. (2001) for derivations of this result in different environments.
[22] Even though, in practice, they mainly exchange their traffic at bilateral peering points.

7.3 Portals and media

The business model of (non-pay) TV and to a large extent newspapers has been to use viewers and readers as a loss leader, who attract advertizers. This business model has been adopted with a vengeance by Internet portals, which have supplied cheap or free Internet access as well as free content (share quotes, news, e-mail, etc....) to consumers. The profit center has been advertising revenue, including both fixed charges for banner placement and proportional referral fees.[23]

Interestingly the portal industry is considering whether to stick to this business model or move to for-fee content. For example, Yahoo! is now starting to charge fees for services such as real-time share-quote services or auction services. A number of content sites have appeared that charge substantial fees for on-line content.[24]

7.4 Videogames

Our last four case studies are drawn from the software industry. The videogame market is a typical two-sided one. A platform cannot sell the console without games to play on and cannot attract game developers without the prospect of an installed base of consumers. In its thirty years of existence, the video game industry has had four leading platforms, Atari, Nintendo and Sega, and finally Sony. The business model that has emerged uses consoles as the loss leader and draws platform profit from applications development. To be certain, history has repeatedly shown that technically impressive platforms (e.g., Mattel in 1981, Panasonic in 1993, and Sega in 1985 and after 1995) fail when few quality games are written for them. But attracting game developers is only a necessary condition. In fact, the business model currently employed by Nintendo, Sega and Sony is to charge software developers a fixed fee together with a per-unit royalty on the games they produce.[25]

[23] See Elfenbein-Lerner (2001) for a thorough analysis of contracts in recent Internet Portal Alliances.

[24] See, e.g., the Economist (April 14, 2001, p65) for more details.

[25] Initially, Nintendo placed a chip in its console. The console would not work unless an authenticating chip was present in the game cartridge. Encryption techniques allow platform manufacturers to meter game sales.

Microsoft released in the fall of 2001 the Xbox in competition with Sony's dominant PlayStation 2. Interestingly, Microsoft manufactures the Xbox console and uses it as a loss leader. While courting the developers[26] by using the familiar X86 chip and Windows platform and by not charging for the Xbox Prototype kit, Microsoft has stated that it intends to draw revenue from royalties.

Although the industry's business model involves drawing revenue from developers, platforms can only go so far in taxing the latter. A key factor in Sony's PlayStation's victory over the Sega Saturn and Nintendo 64 was that Sony offered a development platform and software application that was much cheaper (about $10,000 par seat) and much easier to use (as it was PC based) than it two rivals'.[27]

7.5 Streaming-media technology

Streaming-media platforms incorporate encoding, compression, scripting and delivery technologies to allow the delivery of streaming content, facilitate content creation and permit interactivity; for example, it is central to conferencing and Webcast. The current competition is among the RealNetworks, Microsoft and Apple platforms.

The streaming-media industry is still in its infancy and it is probably too early to point at "the" business model. The current business mostly, but not exclusively, subsidizes the client side. RealNetworks and Apple offer two clients, a basic, free client and a better, non-free one. RealNetworks, the current leader charges significant amounts on the server side for RealServer Plus and its upgrades. Apple in contrast is currently free on the server side, but has the disadvantage on running only on Macs. Microsoft's Windows Media is free (bundled with the operating systems).

[26] In September 2000, 157 developers were working on Xbox games. The Xbox is launched with 26 games. Interestingly, Electronic Arts (the maker of Fifa, SimCity and James Bond) was able to impose special conditions on Microsoft.

[27] See Cringely (2001) for more detail. Sony sold its console below cost and made the money back on game royalties.

7.6 Operating systems

Both sides in the Microsoft browser case agreed that a key competitive advantage of Windows over competing operating systems is its large installed base of applications. Windows' business model is basically the opposite of that of videogame platforms. Windows makes money on users and as a first approximation does not make or loses money on applications developers.[28] It fixes the Applications Programming Interfaces 3 or 4 years in advance (a costly strategy) and invests heavily in developer support. This strategy proved profitable in its contest with Apple and IBM's OS/2. Apple has no integrated developers system tools allowing developers to test their programs; the latter had to buy an IBM workstation and a compiler. IBM viewed developer kits as a profit center.[29] While other factors undoubtedly played a role in the competition among the three platforms, observers usually agree that Microsoft's choice of business model helped Windows establish dominance.

7.7 Text processing

A key issue confronting purchasers of text processing software is whether they will be able to "communicate" with people who don't make the same choice. Commercial software vendors have in this respect converged on the following business model: They offer a downgraded version of the paying software as "freeware". This free version allows "nonusers" to open, view and print, but not edit documents prepared with the paying software, and copy information from those documents to other applications. Examples of

[28] We are unaware of "hard data" on this and just report the industry's conventional wisdom. Nor do we have any hard data for handheld computer operating systems. Handheld computers operating systems, dominated by Palm's platform (75% market share in the US) and Microsoft's Pocket PC software, have adopted a business model that is similar to Windows for PC operating systems. Palm and Microsoft apparently charge about 10% of the hardware's wholesale price ($5 to $15) to hardware manufacturers. Both provide standard user interfaces and central support and development tools for developers of third-party software. For more detail, see http://www5.zdnet.com/zdnn/stories/news/0,4586,2714210,00.html?chkpt-zdhpnews01.

[29] Software developer kits were sold at about $600.

such free viewers are Word Viewer, PDF Viewer, and Scientific Viewer.[30]

8 Final thoughts about two-sided markets

Our premise is that many (probably most) markets with network externalities are two- (or multiple-) sided markets. A market with network externalities is a two-sided market if platforms can effectively cross-subsidize between different categories of end users that are parties to a transaction. That is, the volume of transactions on and the profit of a platform depend not only on the total price charged to the parties to the transaction, but also on its decomposition. There are two reasons why platforms may be unable to perform such cross-subsidization:

a) *Both sides of the market coordinate their purchases* . A debit card platform negotiating with a government for the handling of inter-agency financial transactions, an Internet operator offering an Intranet solution to a company, or a streaming-media platform offering streaming audio and video to a firm primarily for internal use all deal with a single party. A subsidization of the client side by the server side for example does not affect the total price of the software service and, ceteris paribus, does not affect the demand for the platform.[31]

b) *Pass-through and neutrality.* Even when end users on the two sides of the market act independently, monetary transfers between them may undo the redistributive impact and prevent any cross-subsidization. The value-added tax is an epitome of the possibility of neutrality. First-year economic students are taught that it really does not matter whether

[30]For Scientific Word, a mathematics software program adding a user interface and various other functions on to LaTeX.

[31]Mobile and fixed telephone services, for which most users are both callers and receivers, cannot be treated as one-sided markets. A high termination charge raises the marginal cost of calls and lowers the marginal cost of call receptions. In other words, the termination charge is an instrument of cross-subsidization similar to the interchange fee in credit card markets. Telephone users are on both sides of the market for *di®erent* communications only. For a given communication, end users are on a single side and (unless they are engaged in a repeated relationship) their consumption behaviors depend on their own price (calling price for the caller, receiving price for the receiver). As a consequence, the choice of termination charge is not neutral. See Jeon et al (2001) for more detail.

the seller or the buyer pays the tax. In the end, prices adjust so that any tax paid by the seller is passed through to the consumer. If such neutrality holds, then the markets discussed above should be treated as one-sided markets, that is markets in which only the total per transaction price charged by the platform matters and not its decomposition between end users.[32]

Yet thinking of the markets discussed in this paper as one-sided markets just runs counter to evidence. First, the platform owners in all these industries devote much attention to their price-allocation business model: Is it more important to woo one side or the other? The quest for "getting both sides on board" makes no sense in a world in which only the total price for the end user interaction, and not its decomposition, matters. And the trade press would not contain so many descriptions of "chicken- and-egg" problems. Second, the end users themselves are also very sensitive to the allocation of cost between them, indicating that some actual redistribution is taking place. Merchants vocally object to increases in interchange fees, and website operators will do so if settlement charges are introduced in the Internet. End users would not react in this way if charges were passed through. There are three broad reasons why neutrality does not hold in practice:

a) Transaction costs. "Transaction costs" refer to a broad range of frictions that make it costly for one side of the market to pass through a redistribution of charges to the other side. Often, these transaction costs are associated with small stakes for individual transactions (which can become substantial when applied to a large number of transactions). The cost of thinking about including the pass-through, writing it into a contract, advertizing it to the other side and enforcing the covenant may then be prohibitive. For example, contractual relationships between a supplier, a buyer and their banks may not specify on which payment system the settlement of the transaction will occur.

A second type of transaction cost has to do with the absence of a low-cost billing

[32] "Neutrality" refers to the pass-through property and a priori bears no connotation with respect to the well-being of end users and platforms and to social welfare. While neutrality reduces the number of instruments at the disposal of a given platform, it is not clear whether it helps or hurts the platforms in their rivalry. Similarly, neutrality a priori may be good or bad for end users and social welfare.

system. Suppose that an academic downloads a PDF file of another academic's paper. The micropayments that would be required for pass-through would probably require a costly third-party billing system to be developed cooperatively by Internet backbones and service providers.[33]

A third transaction cost is the impossibility of monitoring and recording the actual transaction or interaction. In the portal and media example, neutrality would imply that when the platform (portal, TV network, newspaper) raises the price of advertizing, this price increase translates into a smaller amount of money given by the advertizer to the consumer for "listening to the ad". But "listening" is not easily measurable (except for the monitoring of clicks in the Internet, and even then it is impossible to measure whether the consumer pays genuine attention[34]). In practice therefore, the viewer/reader receives no compensation from the advertizer and neutrality does not obtain.

b) Volume-insensitive costs. Neutrality also fails when at least one side of the market incurs costs that a) are influenced by the platform and b) are not proportional to the number of transactions on the platform. For example, while software developers incur some costs, such as the per-game royalties paid by game developers, that are proportional to sales, many costs are insensitive and affected by the platform: The fixed development cost is influenced by platform through software design, and so is the fixed charge for the development kit. On the user side, getting familiar with the platform's user interface may also involve some fixed costs.[35] End user transaction-insensitive prices and non-price attributes of a platform affect the number of end users or applications, but not directly the terms of the transactions between the end users.

[33]See Laffont et al (2001) for a demonstration that termination charges are neutral in the Internet in the absence of the frictions considered in this section.

[34]Such ways of charging consumers have been considered. For example, a start-up called CyberGold devised a way to pay viewers of ads on the web provided they peruse the Web ad to its last page. Advertisers were concerned about both moral hazard (clicking through ads without being really interested) and adverse selection (clickers would not be the high-demand consumers): See B. Ziegler's "Are Advertisers Ready to Pay Viewers", *Wall Street Journal*, 11/14/1996.

[35]Similarly, end users seem to be averse to being "nickelled and dimed" by Internet portals (perhaps because they have a hard time thinking through the total amounts at stake) and flat fees are still quite popular in that industry.

The portal-and-media and real estate example offers another illustration of this phenomenon. The advertizing cost of locating a potential buyer ought to be treated by the seller as a sunk cost when choosing the price to offer to the potentially interested buyer.

c) Platform-determined constraints on pass-through. The platform may also take steps that limit the extent of pass-through. A case in point is the no-discrimination-rule adopted by credit card associations (Visa, MasterCard) and for-profits (Amex).[36] Merchants do not pass the merchant discount through only to cardholders. So there is only a partial passthrough between the two sides.

These reasons, which have been embodied in our model, explain why markets with network externalities are predominantly two-sided markets.

[36]In the US, the associations' no-discrimination-rule takes a weaker form. Namely, merchants are not allowed to impose surcharges on card payments; but they can offer discounts for cash purchases! That very few do is an interesting fact, that is probably related to the transaction costs category. In Rochet-Tirole (2002), we abstract from such transaction costs and show that the level of the interchange fee is irrelevant if the no-discrimination rule is lifted.

References

[1] Armstrong, M. (2002) "Competition in Two-Sided Markets," presentation at ESEM meeting, Venice, August.

[2] Baumol, W.J., Panzar J. and R. Willig (1982) *Contestable Markets and the Theory of Industry Structure*, NY: Harcourt Brace, Jovanovich.

[3] Baye M., and J. Morgan (2001) "Information Gatekeepers on the Internet and the Competitiveness of Homogenous Product Markets," *American Economic Review*, 91: 454-474.

[4] Caillaud, B. and B. Jullien (2001) "Chicken and Egg: Competing Matchmakers," forthcoming, *Rand Journal of Economics*.

[5] Cowell, A. (2001) "Real Networks Skins Apple, Microsoft," http://www.networkcomputing.com/1204/1204f1b1.html (February 19).

[6] Cringely, R. (2001) "The Plumber is Back: How Nintendo and Microsoft Are about to Open a Case of Whoop-Ass on Sony's Video Game Empire," http://www.pbs.org/cringely/pulpit/pulpit 20010322.html (March 22).

[7] Elfenbein, D., and J. Lerner (2001) "Links and Hyperlinks: An Empirical Analysis of Internet Portal Alliances, 1995-1999," mimeo, Harvard University.

[8] Evans, D. (2002) "The Antitrust Economics of Two-Sided Markets," mimeo, AEI-Brookings Joint Center.

[9] Farrell, J. and G. Saloner (1985) "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, 16: 70-83.

[10] —— (1986) "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *American Economic Review*, 76: 940-955.

[11] Fudenberg, D., and J. Tirole (2000) "Pricing Under the Threat of Entry by a Sole Supplier of a Network Good," *Journal of Industrial Economics*, vol. 48(4):373-390.

[12] Hausman, J., G.K. Leonard, and J. Tirole (2003) "On Nonexclusive Membership in Competing Joint Ventures", *Rand Journal of Economics*, Spring.

[13] Jeon, D.S., Laffont, J.J., and J. Tirole (2001) "On the Receiver Pays Principle," mimeo.

[14] Jullien, B. (2001)"Competing with Network Externalities and Price Competition," mimeo, IDEI, Toulouse.

[15] Katz, M., and C. Shapiro (1985) "Network Externalities, Competition, and Compatibility," *American Economic Review*, 75: 424-440.

[16] —— (1986) "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, 94: 822-841.

[17] Laffont, J.J., Marcus, S., P. Rey, and J. Tirole (2001) "Internet Interconnection and the Off-Net-Cost Pricing Principle," forthcoming *Rand Journal of Economics*.

[18] Parker, G. and M. Van Alstyne (2000) "Information Complements, Substitutes, and Strategic Product Design," mimeo, Tulane University and University of Michigan.

[19] Rochet, J.C., and J. Tirole (2002) "Cooperation among Competitors: The Economics of Payment Card Associations," *Rand Journal of Economics*, 33(4), 1-22.

[20] Rysman, M. (2000) "Competition between Networks: A Study of the Market for Yellow Pages," IPS DP 104, Boston University.

[21] Schmalensee, R., (2002) "Payment Systems and Interchange Fees", *Journal of Industrial Economics*, 50: 103-122.

[22] Stahl, D. (1988) "Bertrand Competition for Inputs and Walrasian Outcomes," *American Economic Review*, 78: 189-201.

[23] Wilson, R. (1993) *Nonlinear Pricing*, Oxford University Press.

[24] Yanelle, M.O. (1997) "Banking Competition and Market Efficiency," *The Review of Economic Studies*, 64(219): 215-40.

APPENDIX

Proof of Proposition 5

(i) Monopoly. Total volume is given by

$$Q = D^B(p^B, \theta) D^S(p^S, \theta).$$

We assume that D^B and D^S are (strictly) log concave with respect to prices, so that the first-order conditions are sufficient for the maximization of volume under a constant margin (case of an association) and for the maximization of profit (case of a for-profit monopoly).

(i1) Monopoly association. The buyer price p^B induced by an association is characterized by

$$\varphi(p^B, \theta) = \lambda^B(p^B, \theta) - \lambda^S(c + m - p^B, \theta) = 0,$$

where $\lambda^B(p^B, \theta) = -(D^B(p^B, \theta))'/(D^B(p^B, \theta))$ and $\lambda^S(p^S, \theta) = -(D^S(p^S, \theta))'/(D^S(p^S, \theta))$ denote the "sensitivities" of demands and $c + m$ is the (fixed) total price. We can apply the implicit function theorem to φ, given that $\frac{\partial \varphi}{\partial p^B} > 0$. This is because the strict log concavity of demands implies that sensitivities are increasing. Thus p^B is differentiable in θ and $\frac{dp^B}{d\theta}$ has the same sign as $\left(-\frac{\partial \varphi}{\partial \theta}\right)$.

We just have to compute $\frac{\partial \varphi}{\partial \theta}$ in our two examples:

Marquee buyers:

$$\varphi(p^B, \theta) = \lambda^B(p^B) - \lambda^S(c + m - \theta - p^B)$$

$$\frac{\partial \varphi}{\partial \theta} = (\lambda^S)' > 0 \quad \text{(since } D^S \text{ is log concave).}$$

Thus $\frac{dp^B}{d\theta}$ is negative.

Captive buyers:

$$\varphi(p^B, \theta) = \frac{-(D^B)'}{D^B + \theta} - \lambda^S(c + m - p^B)$$

$$\frac{\partial \varphi}{\partial \theta} = \frac{(D^B)'}{(D^B + \theta)^2} < 0.$$

And so $\frac{dp^B}{d\theta}$ is positive.

(i2) For-profit monopoly. The maximum of the (log) profit is characterized by two conditions:

$$\begin{cases} \lambda^B(p^B, \theta) - \frac{1}{p^B + p^S - c} & = 0 \\ \lambda^S(p^S, \theta) - \frac{1}{p^B + p^S - c} & = 0. \end{cases}$$

Denoting by $\varphi(p, \theta)$the (vector) function on the left-hand side, we can apply the implicit function theorem (this time in $\mathbb{R}^2$) given that the Jacobian $\frac{D\varphi}{Dp}$ is nonsingular (by the strict concavity of $\log \pi$, the determinant of $\frac{D\varphi}{Dp}$ is positive). Thus we obtain:

$$\frac{dp}{d\theta} = -\left(\frac{D\varphi}{Dp}\right)^{-1} \frac{\partial \varphi}{\partial \theta},$$

where

$$\left(\frac{D\varphi}{Dp}\right)^{-1} = \frac{1}{\det \frac{D\varphi}{Dp}} \begin{bmatrix} \frac{\partial \lambda^S}{\partial p^S} + \frac{1}{(p-c)^2} & -\frac{1}{(p-c)^2} \\ -\frac{1}{(p-c)^2} & \frac{\partial \lambda^B}{\partial p^B} + \frac{1}{(p-c)^2} \end{bmatrix}$$

and

$$\frac{\partial \varphi}{\partial \theta} = \begin{pmatrix} \frac{\partial \lambda^B}{\partial \theta} \\ \frac{\partial \lambda^S}{\partial \theta} \end{pmatrix}.$$

Marquee buyers: $\frac{\partial \lambda^B}{\partial \theta} = 0$, $\frac{\partial \lambda^S}{\partial \theta} < 0$

$$\begin{pmatrix} \frac{dp^B}{d\theta} \\ \frac{dp^S}{d\theta} \end{pmatrix} = \frac{-\frac{\partial \lambda^S}{\partial \theta}}{\det \frac{D\varphi}{Dp}} \begin{bmatrix} -\frac{1}{(p-c)^2} \\ \frac{\partial \lambda^B}{\partial p^B} + \frac{1}{(p-c)^2}. \end{bmatrix}.$$

Thus $\frac{dp^B}{d\theta} < 0$ and $\frac{dp^S}{d\theta} > 0$.

Captive buyers: $\frac{\partial \lambda^B}{\partial \theta} < 0$, $\frac{\partial \lambda^S}{\partial \theta} = 0$, and so:

$$\begin{pmatrix} \frac{dp^B}{d\theta} \\ \frac{dp^S}{d\theta} \end{pmatrix} = \frac{-\frac{\partial \lambda^B}{\partial \theta}}{\det \frac{D\varphi}{Dp}} \begin{bmatrix} \frac{\partial \lambda^S}{\partial p^S} + \frac{1}{(p-c)^2} \\ -\frac{1}{(p-c)^2}. \end{bmatrix}$$

Thus $\frac{dp^B}{d\theta} > 0$ and $\frac{dp^S}{d\theta} < 0$.

(ii) Competing associations. In the case of associations, the equilibrium buyer price is characterized by:

$$\lambda_0^B(p^B, \theta)\sigma(p^B, \theta) - \lambda^S(c + m - p^B, \theta) = 0,$$

where

$$\lambda_0^B(p^B, \theta) = \frac{-\frac{\partial d_1^B}{\partial p_1^B}}{d_1^B}(p^B, p^B, \theta)$$

is the "own-price sensitivity" of buyer demand and

$$\sigma(p^B, \theta) = 2 - \frac{\hat{D}^B(p^B, \theta)}{d^B(p^B, \theta)}.$$

In order to determine the monotonicity properties of p^B with respect to θ, we apply the implicit function theorem to the left-hand side of the above equation:

$$\psi(p^B, \theta) = \lambda_0^B(p^B, \theta)\sigma(p^B, \theta) - \lambda^S(c + m - p^B, \theta).$$

However we need additional assumptions to ensure that $\frac{\partial \psi}{\partial p^B} > 0$, so that $p^B(\theta)$ is indeed (locally) unique and differentiable, for two reasons:

• Possible nonexistence of equilibrium, due to the fact that the volume on system i is not necessarily quasiconcave with respect to (p_i^B, p_i^S). The proof of Proposition 6 will observe that the candidate for equilibrium (i.e. the solution of $\psi = 0$) may sometimes be destabilized by "double deviations" of the form $(p^B + \varepsilon, p^S - \varepsilon)$.

• The possible presence of strategic complementarities that may generate a multiplicity of equilibria.

We will assume away these difficulties and postulate that $\frac{\partial\psi}{\partial p^B} > 0$ (regularity condition).[37] In this case, $p^B(\theta)$ is (locally) unique, differentiable, and $\frac{dp^B}{d\theta}$ has the same sign as $-\frac{\partial\psi}{\partial\theta}$. We then just have to determine the sign of $\frac{\partial\psi}{\partial\theta}$:
Marquee buyers:

$$\begin{aligned}\psi(p^B,\theta) &= \lambda_0^B(p^B)\sigma(p^B) - \lambda^S(c+m-\theta-p^B)\\ \frac{\partial\psi}{\partial\theta} &= (\lambda^S)' > 0.\end{aligned}$$

Captive buyers:

$$\begin{aligned}\psi(p^B,\theta) &= \left(\frac{-\frac{\partial d_1^B}{\partial p_1^B}}{d_1^B+\theta}\right)\left(2-\frac{D^B+\theta}{d^B+\theta}\right) - \lambda^S(c+m-p^B)\\ \frac{\partial\psi}{\partial\theta} &= -\frac{\lambda_0^B\sigma}{d^B+\theta} - \frac{d^B-D^B}{(d^B+\theta)^2}\lambda_0^B = \frac{-\lambda_0^B}{d^B+\theta}\left[\sigma+\frac{d^B-D^B}{d^B+\theta}\right] < 0.\end{aligned}$$

An increase in the number of captive buyers has two opposite effects. First, and as in the monopoly case, the captive customers reduce the elasticity of buyer demand, calling for a higher buyer price. Second, captive customers make steering more attractive, which pushes toward a higher seller price. The first effect dominates the second.
(iii) Increase in singlehoming. Again, we focus on competing associations. The buyer price at equilibrium is determined by:

$$\psi(p^B,\theta) = \lambda_0^B(p^B)\sigma^B(p^B,\theta) - \lambda^S(c+m-p^B).$$

By the same reasoning as above, $\frac{\partial\sigma^B}{\partial\theta} > 0$ implies $\frac{\partial\psi}{\partial\theta} > 0$ and $\frac{dp^B}{d\theta} > 0$. ■

Proof of Proposition 6

(a) Price structure. Letting $T = p_0 + \frac{\delta(\Delta+\delta)}{2}$, the quasi-demands are given by:[38]

$$d_1^B(p_1^B,p_2^B) = \frac{p_2^B-p_1^B}{\Delta} + \frac{T-p_1^B}{\delta}, \tag{25}$$

and

$$D_1^B(p_1^B) = (T-p_1^B)\left(\frac{1}{\delta}+\frac{1}{\Delta+\delta}\right). \tag{26}$$

[37] This regularity condition is satisfied when $\frac{\partial\lambda_0^B}{\partial p^B}$ and $\frac{\partial\sigma}{\partial p^B}$ are positive.

[38] The expressions of quasi-demands are easily deduced from the locations of marginal buyers:
• x_1 is indifferent between 1 and 1′: $p_1^B + \frac{1}{2}\left(x_1+\frac{\Delta}{2}\right)^2 = p_0 + \frac{1}{2}\left(x_1+\frac{\Delta}{2}+\delta\right)^2$, which gives: $x_1 = \frac{p_1^B-p_0}{\delta} - \frac{\Delta+\delta}{2}$;
• x_2 is indifferent between 1 and 2: $p_1^B + \frac{1}{2}\left(x_2+\frac{\Delta}{2}\right)^2 = p_2^B + \frac{1}{2}\left(x_2-\frac{\Delta}{2}\right)^2$, which gives: $x_2 = \frac{p_2^B-p_1^B}{\Delta}$;
• x_3 is indifferent between 1 and 2′: $p_1^B + \frac{1}{2}\left(x_3+\frac{\Delta}{2}\right)^2 = p_0 + \frac{1}{2}\left(x_3-\frac{\Delta}{2}-\delta\right)^2$, which gives: $x_3 = \frac{p_0-p_1^B}{\Delta+\delta} + \delta/2$.

The expressions of d_2^B and D_2^B are obtained by symmetry. Due to the linearity of these expressions, several simplifications appear. For example, the singlehoming index is price independent:

$$\sigma = 2 - \frac{D_1^B(p^B)}{d_1^B(p^B, p^B)} = \frac{\Delta}{\Delta + \delta}.$$

Similarly the expression of the marginal seller (who is indifferent between multihoming and singlehoming with the cheapest platform), does not depend on buyers' prices. For example, when $p_1^S \leq p_2^S$, formula (12) gives:

$$\hat{b}_{12} = p_2^S \frac{d_2^B}{d_1^B + d_2^B - D_1^B} + p_1^S \frac{d_1^B - D_1^B}{d_1^B + d_2^B - D_1^B}.$$

Hence (for $p_1^B = p_2^B = p^B$):

$$\hat{b}_{12} = p_2^S + \frac{\delta}{\Delta}(p_2^S - p_1^S).$$

and so $\hat{b}_{12}$ does not depend on p^B. Furthermore, steering is particularly powerful (in that undercutting induces many sellers to stop multihoming) when most consumers multihome, that is when σ is low.

Another simplification that appears when buyers' quasi-demand is linear is that the ratio of total elasticity to own-brand elasticity is equal to the single homing index σ:

$$\frac{\eta^B}{\eta_o^B} = \frac{\frac{\partial d_1^B}{\partial p_1^B} + \frac{\partial d_1^B}{\partial p_2^B}}{\frac{\partial d_1^B}{\partial p_1^B}} = \frac{\frac{1}{\delta}}{\frac{1}{\Delta} + \frac{1}{\delta}} = \frac{\Delta}{\Delta + \delta} = \sigma.$$

This property implies that the price structure under platform competition (between for-profits or between associations) is the same as under a monopoly platform:

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}.$$

Consider for example a decrease in Δ. As the platforms become more substitutable, buyer multihoming increases (σ falls); this induces platforms to steer, resulting in low prices on the seller side. However, competition also becomes more intense on the buyer side, resulting in lower buyer prices (p^B falls) and thereby in a higher opportunity cost $(c - p^B)$ of servicing sellers. For linear demand on the buyer side, these two effects offset.

Last, let us compare the common price structure with that of the Ramsey optimum. A useful property of linear demands is that the revenue (price times quantity) is equal to twice the product of the net surplus and the elasticity of demand. This property implies that if seller's quasi-demand is linear as well, (7) is equivalent to (5), and so the common price structure is Ramsey optimal.

(b) Second-order conditions. In the Hotelling model:

$$\begin{aligned}
\pi_i &= (p_i^B + p_i^S - c)Q_i; \\
\frac{\partial \pi_i}{\partial p_i^B} &= Q_i + (p_i^B + p_i^S - c)\frac{\partial Q_i}{\partial p_i^B}; \\
\frac{\partial \pi_i}{\partial p_i^S} &= Q_i + (p_i^B + p_i^S - c)\frac{\partial Q_i}{\partial p_i^S}; \\
\frac{\partial^2 \pi_i}{(\partial p_i^B)^2} &= 2\frac{\partial Q_i}{\partial p_i^B}; \quad \frac{\partial^2 \pi_i}{(\partial p_i^S)^2} = 2\frac{\partial Q_i}{\partial p_i^S}; \\
\frac{\partial^2 \pi_i}{\partial p_i^B \partial p_i^S} &= \frac{\partial Q_i}{\partial p_i^B} + \frac{\partial Q_i}{\partial p_i^S} + (p_i^B + p_i^S - c)\frac{\partial^2 Q_i}{\partial p_i^B \partial p_i^S}.
\end{aligned}$$

At a symmetric equilibrium of the game between competing proprietary platforms, we have

$$\frac{\partial Q_i}{\partial p_i^B} = \frac{\partial Q_i}{\partial p_i^S} = -(p-c)\frac{(\Delta+\delta)^2}{\Delta^2\delta} \equiv \alpha < 0.$$

Therefore the second-order condition is satisfied whenever the Hessian determinant of π_i is nonnegative:

$$\begin{aligned}
H &= \frac{\partial^2 \pi_i}{(\partial p_i^B)^2} \cdot \frac{\partial^2 \pi_i}{(\partial p_i^S)^2} - \left(\frac{\partial^2 \pi_i}{\partial p_i^B \partial p_i^S}\right)^2 \\
H &= 4\alpha^2 - (2\alpha+\beta)^2 = -\beta(4\alpha+\beta),
\end{aligned}$$

where

$$\beta \equiv (p_i^B + p_i^S - c)\frac{\partial^2 Q_i}{\partial p_i^B \partial p_i^S}$$

has a different expression in the two regions:

$$\begin{aligned}
\beta_1 &= (p-c)\frac{(\Delta^2 + \delta\Delta - \delta^2)}{\Delta^2\delta} \text{ when } p_1^S < p_2^S, \text{ and} \\
\beta_2 &= (p-c)\frac{(\Delta+\delta)^2}{\Delta^2\delta} \text{ when } p_1^S > p_2^S.
\end{aligned}$$

The second-order condition is always satisfied in the second region, since $\beta_2 = -\alpha > 0$ so that $H = 3\alpha^2 > 0$. In the first region, it is easy to see that $\beta_1 + 4\alpha$ is always negative. Thus the second-order condition is satisfied if and only if $\beta_1 \geq 0$, which is equivalent to

$$\delta^2 - \delta\Delta - \Delta^2 \leq 0$$

or

$$\frac{\delta}{\Delta} \leq \frac{1+\sqrt{5}}{2}.$$

When this condition is not satisfied, there is no symmetric equilibrium in pure strategies. The only candidate equilibrium (p^B, p^S) can be destabilized by a "double-deviation", where one of the platforms (say platform 1) increases p_1^B by ε and simultaneously decreases

p_1^S by the same amount. The first order increase in profit is zero (as guaranteed by the first-order conditions) but the second-order increase is positive:

$$\Delta\pi_1 \sim \left[\frac{\partial^2\pi_1}{(\partial p_1^B)^2} + \frac{\partial^2\pi_1}{(\partial p_1^S)^2} - 2\frac{\partial^2\pi_1}{\partial p_1^B \partial p_1^S}\right]\varepsilon^2 = -2\beta_1\varepsilon^2 > 0.$$

Finally, equilibrium prices can be obtained explicitly if we assume that the sellers' quasi-demand is also linear:

$$D^S(p^S) = A - p^S. \tag{27}$$

The volume on platform 1 when $p_1^S \leq p_2^S$ is:

$$\begin{aligned} Q_1 &= \left(\frac{p_2^B - p_1^B}{\Delta} + \frac{T - p_1^B}{\delta}\right)\left[A - p_2^S - \frac{\delta}{\Delta}(p_2^S - p_1^S)\right] \\ &+ \left(\frac{1}{\delta} + \frac{1}{\Delta+\delta}\right)(T - p_1^B)(p_2^B - p_1^S)\left(1 + \frac{\delta}{\Delta}\right). \end{aligned} \tag{28}$$

When $p_1^S > p_2^S$, the expression is simpler:

$$Q_1 = \left[A - p_1^S - \frac{\delta}{\Delta}(p_1^S - p_2^S)\right]\left[\frac{p_2^B - p_1^B}{\Delta} + \frac{T - p_1^B}{\delta}\right]. \tag{29}$$

In Proposition 2 we have shown that a symmetric equilibrium between competing associations must satisfy condition (22):

$$\frac{p^B}{\sigma\eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

Using formulae (25), (26) and (27) and after simplifications, this condition becomes:

$$p^B - p^S = T - A. \tag{30}$$

Recall that this condition is necessarily satisfied in a symmetric equilibrium between competing platforms, independently of their governance structure. However the value of the total price is different:

$$p^B + p^S = c + m$$

for associations, and

$$p^B + p^S - c = \frac{\Delta}{\Delta+\delta}(T - p^B) = \frac{\Delta}{\Delta+\delta}(A - p^S),$$

for proprietary platforms. The resulting equilibrium prices are:

$$p_A^B = \frac{1}{2}(c + m + T - A), \quad p_A^S = \frac{1}{2}(c + m - T + A), \tag{31}$$

for associations, and

$$p_P^B = \frac{c - A + T\left(1 + \frac{\Delta}{\Delta+\delta}\right)}{2 + \frac{\Delta}{\Delta+\delta}}, \quad p_P^S = \frac{c - T + A\left(1 + \frac{\Delta}{\Delta+\delta}\right)}{2 + \frac{\Delta}{\Delta+\delta}} \tag{32}$$

for proprietary systems.

PLATFORM COMPETITION IN TWO-SIDED MARKETS

Jean-Charles Rochet
Université de Toulouse,
Institut D'Economie Industrielle

Jean Tirole
Institut D'Economie Industrielle,
Centre d'Etudes et de Recherches en
Analyse Socio-Economique

Abstract
Many if not most markets with network externalities are two-sided. To succeed, platforms in industries such as software, portals and media, payment systems and the Internet, must "get both sides of the market on board." Accordingly, platforms devote much attention to their business model, that is, to how they court each side while making money overall. This paper builds a model of platform competition with two-sided markets. It unveils the determinants of price allocation and end-user surplus for different governance structures (profit-maximizing platforms and not-for-profit joint undertakings), and compares the outcomes with those under an integrated monopolist and a Ramsey planner. (JEL: L5, L82, L86, L96)

1. Introduction

Buyers of video game consoles want games to play on; game developers pick platforms that are or will be popular among gamers. Cardholders value credit or debit cards only to the extent that these are accepted by the merchants they patronize; affiliated merchants benefit from a widespread diffusion of cards among consumers. More generally, many if not most markets with network externalities are characterized by the presence of two distinct sides whose ultimate benefit stems from interacting through a common platform. Platform owners or sponsors in these industries must address the celebrated "chicken-and-egg problem" and be careful to "get both sides on board." Despite much theoretical progress made in the last two decades on the economics of network externalities and widespread strategy discussions of the chicken-and-egg problem, two-sided markets have received scant attention. The purpose of this paper is to start filling this gap.

The recognition that many markets are multisided leads to new and interesting positive and normative questions. Under multisidedness, platforms must choose a price structure and not only a price level for their service. For example, video game platforms such as Sony, Sega and Nintendo make money on game

E-mail addresses: Rochet: rochet@cict.fr; Tirole: tirole@cict.fr

© 2003 by the European Economic Association

developers through per-unit royalties on games and fixed fees for development kits and treat the gamers side as a loss leader. Interestingly, operating system platforms for the PC and handheld devices have adopted the opposite business model and aim at making money on consumers. The choice of a business model seems to be key to the success of a platform and receives much corporate attention. Table 1 provides a few illustrations[1] of the two-sided markets and shows that platforms often treat one side as a profit center and the other as a loss leader, or, at best, as financially neutral. A number of these illustrations are discussed in "mini case studies" in Section 7. And Table 2 lists a few important segments of the new economy that will be searching for a proper business model in the next few years. Such conventional wisdom about business models found in the trade press and summarized in Table 1 is of course subject to criticism. To reason in terms of profit centers, costs are often "intuitively," but arbitrarily allocated to either side of the market. Yet, the conventional wisdom points at some more fundamental logic related to prices and surpluses on both sides of the market. A major objective of our paper is to unveil this logic and the determinants of the choice of a business model.

From both positive and normative viewpoints, two-sided markets differ from the textbook treatment of multiproduct oligopoly or monopoly. The interaction between the two sides gives rise to strong complementarities, but the corresponding externalities are not internalized by end users, unlike in the multiproduct literature (the same consumer buys the razor and the razor blade). In this sense, our theory is a cross between network economics, which emphasizes such externalities, and the literature on (monopoly or competitive) multiproduct pricing, which stresses cross-elasticities. For example, socially optimal "Ramsey" prices are not driven solely by superelasticity formulae but also reflect each side's contribution to the other side's surplus.

Some new questions raised by two-sided markets are more specific to the existence of competition between platforms. In a number of markets, a fraction of end users on one or the two sides connect to several platforms. Using the

1. There are of course other illustrations, for example scientific journals, that must match readers and authors. Interestingly, the *Bell Journal of Economics* for a number of years after it was launched was sent for free to anyone who requested it. There is currently much discussion of how the business model for scientific journals will evolve with electronic publishing. The list of social gatherings examples of cross-subsidization could be extended to include dating or marital agencies which may charge only one side of the market.

A couple of explanations regarding markets that will not be discussed in Section 7: Social gatherings: celebrities often do not pay or are paid to come to social happenings as they attract other participants (who may then be charged a hefty fee); similarly, in some conferences, star speakers are paid while others pay. *Real estate*: In many countries buyers are not charged for visiting real estate properties and thus marginal visits are heavily subsidized. To be certain, the sale price reflects the real estate agency fee, but this does not imply that the arrangement is neutral (see Section 8). *Shopping malls*: shoppers are subsidized. They don't pay for parking; in France they can also buy gasoline at a substantial discount. *Discount coupon books*: These are given away to consumers. Intermediaries charge merchants for the service. *Browsers*: The picture given in Table 1 is a bit simplistic. In particular, Netscape initially made about one third of its revenue on the client side before giving the software away. But Netscape always viewed the software running on top of the operating system on the web servers as a major source of profit.

TABLE 1. ILLUSTRATIONS OF EXISTING BUSINESS MODELS

Product	Loss leader/break-even segment/subsidized segment	Profit-making segment/ subsidizing segment
	Software	
Video games	consumers (consoles)	software developers
Streaming media	consumers	servers
Browsers	users	Web servers
Operating systems (Windows; Palm, Pocket PC)	application developers (development tools, support, functionality, . . .)	clients
Text processing	reader/viewer	writer
	Portals and media	
Portals	"eyeballs"	advertisers
Newspapers	readers	advertisers
(Charge-free) TV networks	viewers	advertisers
	Payment systems	
Credit and differed debit cards (Visa, MasterCard, Amex, . . .)	cardholders	merchants
Online debit cards	merchants	cardholders
	Others	
Social gatherings	celebrities in social happenings	other participants
Shopping malls	consumers (free parking, cheap gas)	shops
Discount coupon books (Want Advertiser)	consumers	merchants
(Legacy) Internet	Web sites	dial-up consumers
Real estate	buyers	sellers

Internet terminology, we will say that they "multihome." For example, many merchants accept both American Express and Visa; furthermore, some consumers have both Amex and Visa cards in their pockets. Many consumers have the Internet Explorer and the Netscape browsers installed on their PC, and a number of Web sites are configured optimally for both browsers. Readers may subscribe to multiple newspapers, B2B exchange members may buy or sell their wares on

TABLE 2. PROSPECTIVE APPLICATIONS

Platform	Two sides	Instruments of cost allocation or cross-subsidization
B2B	buyers/sellers	design of auctions, information flows
Internet backbone services	consumers/websites	termination (settlement) charges
Pools and standards	relevant sides	level of royalties, inclusiveness of pools
Software as a service (.Net vs. Java)	consumers/application developers	development tools and other efforts to create an applications development environment, backward compatibility, pricing

several exchanges, and real estate sellers and buyers may use the services of multiple real estate agencies. Competitive prices on one market then depend on the extent of multihoming on the other side of the market. For example, when Visa reduces the (transaction-proportional) charge paid by the merchants,[2] merchants become more tempted to turn down the more costly Amex card as long as a large fraction of Amex customers also owns a Visa card. More generally, multihoming on one side intensifies price competition on the other side as platforms use low prices in an attempt to "steer" end users on the latter side toward an exclusive relationship.[3]

This paper studies how the price allocation between the two sides of the market is affected by a) platform governance (for-profit vs. not-for-profit), b) end users' cost of multihoming, c) platform differentiation, d) platforms' ability to use volume-based pricing, e) the presence of same-side externalities, and f) platform compatibility. It also investigates how privately optimal pricing structures compare with socially optimal ones.

The paper is organized as follows. Section 2 describes the simplest version of the model, in which end users incur no fixed cost and platform pricing is linear on both sides of the market, and analyzes the (profit maximizer and Ramsey planner) monopoly benchmarks. Section 3 derives equilibrium behavior when two (for-profit or not-for-profit) platforms compete. Section 4 obtains some comparative statics in order to help predict the choice of business model. Section 5 compares the price structures in the case of linear demands. Section 6 generalizes the model and results in order to allow for fixed user costs and nonlinear platform pricing. Section 7 summarizes the main results and provides seven "mini case studies" to illustrate how our theory may shed light on existing and future business models. Last, Section 8 concludes with some general considerations about two-sided markets.

As we discussed, our work puts network economics and multiproduct pricing together. From the early work of Rohlfs (1974) to the recent theoretical advances and applications to antitrust through the pioneering work of Katz and Shapiro (1985, 1986) as well as Farrell and Saloner (1985, 1986), a large body of literature has developed on network industries. To make progress, however, this literature has ignored multisidedness and the price allocation question. In contrast, the competitive multiproduct pricing literature (e.g., Baumol, Panzar,

2. The mechanism through which this reduction operates is indirect and is described in section 7.

3. The occurrence of steering is easiest to visualize in those illustrations in which platforms charge per-end-user-transaction fees: The seller of a house or a B2B supplier may only list the house or the wares on the cheapest platform.

In industries in which platforms do not charge per-end-user-transaction fees, steering is more subtle as it operates through effort substitution. For example, a software platform offering better software development kits, support, and application programming interfaces not only encourages the development of applications optimized to this platform, but is also likely to induce application developers to devote less attention to rival platforms. A portal or TV network's cut in advertising rates induces advertisers to advertise more on their medium and to substitute away from other media. A shopping mall's cut in rental prices or improved layout may induce a shop to increase its size or appeal and lead the latter to neglect or abandon its outlets in rival shopping malls, and so forth.

and Willig 1982, Wilson 1993) has carefully described the interdependency of pricing decisions but it has not considered the affiliation externalities that lie at the core of the network economics literature. In contrast with the buyer of a razor, who internalizes the impact of his purchase on the demand and surplus attached to razor blades, our end-users do not internalize the impact of their purchase on the other side of the market.

Our paper is most closely related to the recent theoretical literature on chicken-and-egg problems.[4] This literature however assumes either that there is a monopoly platform (Baye and Morgan 2001, Rochet and Tirole 2002, Schmalensee 2002) or that platforms are fully interconnected (Laffont, Marcus, Rey, Tirole 2003) and so end-users enjoy the same level of connectivity regardless of the platform they select. Parker and Van Alstyne (2000) study monopoly pricing in a situation in which the demand for one good depends (linearly) on its price and on the quantity of the other good sold. They characterize the price structure as a function of the network externality coefficients. They then look at the incentive of a producer of a good to enter a (complementary or substitute) market with another incumbent producer. With complements, entry losses may be profitable because entry puts pressure on price and boosts the profit of the core business. Caillaud and Jullien (2003) study competition among intermediaries. In their model, platforms act as matchmakers and can use sophisticated pricing (registration fees, and possibly transaction fees provided the intermediaries observe transactions). Indeed, one of their contributions is to show that dominant firms are better off charging transactions rather than registrations when deterring entry. They also show that competition is more intense when platforms cannot deter multihoming. Their contribution is complementary to ours. For example, it assumes homogeneous populations on either side, and thus abstracts from the elasticity-related issues studied in our paper. Last, in a model related to that of Caillaud and Jullien (2003), Jullien (2001) shows that an entrant represents a much stronger competitive threat on an incumbent platform when third-degree price discrimination is feasible. The ability to "divide and conquer" forces profit down, so much so that the incumbent may prefer platform compatibility.

2. Monopoly Platform Benchmark

The two-sided markets described heretofore differ in some respects, and we therefore should not aim at capturing all specificities of all industries. Our strategy will be to include a number of key ingredients common to our illustrations in a basic model, and then to generalize our analysis in order to

4. The policy implications of two-sidedness are discussed in Evans (2003). The reader will find further illustrations of two-sided markets and an interesting analysis thereof in Armstrong (2002). Rysman (2000) is the first empirical paper to estimate network effects in a two-sided context, namely the market for Yellow Pages. It is also related to the earlier literature on competition between intermediaries: Stahl (1988) and Yanelle (1997).

extend its relevance to various two-sided markets. For the moment, we assume that end users incur no fixed usage cost and that platform pricing is linear. This basic model is a good representation of the credit card market; the reader may want to keep this in mind, although it will be clear that the insights have much broader generality.

Economic value is created by "interactions" or "transactions" between pairs of end users, buyers (superscript B) and sellers (superscript S). Buyers are heterogenous in that their gross surpluses b^B associated with a transaction differ. Similarly, sellers' gross surplus b^S from a transaction differ. Such transactions are mediated by a platform. The platform's marginal cost of a transaction is denoted by $c \geq 0$.

As an illustration, consider the case of payment cards. The buyer wants to purchase a bundle of goods or services from the merchant at a certain price p. In our vocabulary, a "transaction" takes place if and only if the buyer pays by card instead of using another payment instrument (say, cash). Benefits b^B and b^S correspond to differences in utility of buyers and sellers when they pay by card rather than cash. Under the No Surcharge Rule (very often imposed by payment card networks)[5] the merchant is not able to charge different retail prices for card and cash payments. Therefore the distributions of b^B and b^S are independent of the prices chosen by platforms and merchants, and can be taken as exogenous.

In the absence of fixed usage costs and fixed fees, the buyers' (sellers') demand depends only on the price p^B (respectively, p^S) charged by the monopoly platform. There are network externalities in that the surplus of a buyer with gross per transaction surplus b^B, $(b^B - p^B)N^S$, depends on the number of sellers N^S, but the buyers' "quasi-demand function":[6]

$$N^B = \Pr(b^B \geq p^B) = D^B(p^B)$$

is independent of the number of sellers. Similarly, let

$$N^S = \Pr(b^S \geq p^S) = D^S(p^S)$$

denote the sellers' quasi-demand for platform services. Consider a (buyer, seller) pair. Without loss of generality we can assume that each such pair corresponds to one potential transaction.

In contrast with search models à la Baye and Morgan (2001) or Caillaud and Jullien (2003), we take as given the matching process between buyers and sellers, and focus on the proportion of such matches that effectively results in a

5. Even in the countries where the No Surcharge Rule is not imposed, as in the UK, it turns out that merchants seldom charge different prices for card and cash payments. We discuss in Section 8 the possible reasons for this fact, and more generally for the non-neutrality of the price structure in two-sided markets.

6. The word "quasi-demand function" is used to reflect the fact that, in a two-sided market, actual demand depends on the decisions of both types of users (buyers and sellers in our terminology). In our specification, this demand is simply the product of the quasi-demands of buyers and sellers.

"transaction."[7] Assuming for simplicity the independence between b^B and b^S, the proportion (or volume) of transactions is equal to the product $D^B(p^B) \times D^S(p^S)$.[8]

We consider in turn the case of a private monopoly, and that of a public monopoly maximizing social welfare subject to budget balance.

2.1 Private Monopoly

A private monopoly chooses prices so as to maximize total profit:

$$\pi = (p^B + p^S - c)\, D^B(p^B)\, D^S(p^S).$$

Assuming that D^B and D^S are log concave, it is easy to see that π is also log concave (jointly in (p^B, p^S)). Its maximum is characterized by the first-order conditions:

$$\frac{\partial(\log \pi)}{\partial p^B} = \frac{1}{p^B + p^S - c} + \frac{(D^B)'}{D^B} = 0,$$

$$\frac{\partial(\log \pi)}{\partial p^S} = \frac{1}{p^B + p^S - c} + \frac{(D^S)'}{D^S} = 0.$$

In particular:

$$(D^B)' D^S = D^B (D^S)'.$$

This condition characterizes the values of p^B and p^S that maximize volume for a given total price p: The volume impact of a small (absolute) variation of prices has to be the same on both sides. If we introduce the elasticities of quasi-demands:

$$\eta^B = -\frac{p^B (D^B)'}{D^B} \qquad \text{and} \qquad \eta^S = -\frac{p^S (D^S)'}{D^S},$$

the private monopoly prices can be characterized by a two-sided formula that is reminiscent of Lerner's formula:

$$p^B + p^S - c = \frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}. \tag{1}$$

In fact, the total price $p = p^B + p^S$ chosen by the private monopoly is given by the classical Lerner formula:

7. In the payment card example, a "transaction" between a cardholder and a merchant means that the payment is by card rather than by cash.

8. This multiplicative formula was first used by Schmalensee (2002). Most of our results can be extended to the more general case where b^B and b^S are not independent, in which case the transaction volume Q has a more general expression $Q(p^B, p^S) = \Pr(b^B \geq p^B, b^S \geq p^S)$.

$$\frac{p-c}{p} = \frac{1}{\eta}, \qquad \text{or} \qquad p = \frac{\eta}{\eta - 1}c, \tag{2}$$

where $\eta = \eta^B + \eta^S$, the total volume elasticity, is assumed to exceed 1. What is new in formula (1) is the way in which this total price is allocated between the two sides of the market:

$$p^B = \frac{\eta^B}{\eta}p = \frac{\eta^B}{\eta - 1}c, \tag{3}$$

and

$$p^S = \frac{\eta^S}{\eta}p = \frac{\eta^S}{\eta - 1}c. \tag{4}$$

PROPOSITION 1. (*i*) *A monopoly platform's total price*, $p = p^B + p^S$, *is given by the standard Lerner formula for elasticity equal to the sum of the two elasticities*, $\eta = \eta^B + \eta^S$:

$$\frac{p-c}{p} = \frac{1}{\eta}. \tag{2}$$

(*ii*) *The price structure is given by the ratio of elasticities* (*and not inverse elasticities*):

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}. \tag{5}$$

2.2 Ramsey Pricing

We consider now the case of a Ramsey monopolist maximizing welfare subject to budget balance, and derive the Ramsey formulae in our context.[9] The net surpluses on each side for an average transaction are given by standard formulae:

$$V^k(p^k) = \int_{p^k}^{+\infty} D^k(t)dt$$

for $k \in \{B, S\}$.

Under budget balance, social welfare is highest when the sum of both sides' net surpluses:

$$W = V^S(p^S)D^B(p^B) + V^B(p^B)D^S(p^S),$$

9. A similar formula is derived in Laffont et al. (2003) in a model in which network externalities are reaped through platform interconnection.

is maximized subject to the constraint:

$$p^B + p^S = c.$$

The first-order, "cost allocation" condition is:

$$\frac{\partial W}{\partial p^B} = \frac{\partial W}{\partial p^S}.$$

This gives:

$$V^S(D^B)' - D^B D^S = -D^S D^B + V^B(D^S)'.$$

After simplification, we obtain a characterization of Ramsey prices:

PROPOSITION 2. *Ramsey prices embody the average surpluses created on the other side of the market and are characterized by two conditions:*

$$p^B + p^S = c \text{ (budget balance)}. \tag{6}$$

and

$$\frac{p^B}{\eta^B}\left[\frac{V^B}{D^B}\right] = \frac{p^S}{\eta^S}\left[\frac{V^S}{D^S}\right] \text{ (cost allocation)}. \tag{7}$$

Condition (7) characterizes the price structure that maximizes social surplus for a given total price p. Returning to the formula yielding the private monopolist's price structure,

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}, \tag{1'}$$

the additional terms in Formula (7) (the bracketed terms) reflect the average surpluses per transaction for buyers and sellers. (Later, when we compare price structures across governance forms, we will compare prices for a given price level. That is, we will say that two governance forms generate the same price structure if they give rise to the same prices for a given price level target $p = p^B + p^S$. Of course different governance forms generate different price levels.)

3. Competing Platforms

3.1 Modeling

We now assume that two platforms compete for the markets (we will also look at the case in which both platforms are jointly owned, in order to compare the outcome under platform competition with those obtained in Section 2 in the private monopoly and Ramsey cases).

3.1.1 End users' benefits. As earlier, buyers and sellers are heterogenous: Their benefits from transacting vary across the two populations and are private information. These benefits are denoted b_i^B for the buyer (when the transaction takes place on platform i) and b^S for the seller, and are drawn from continuous distributions.[10] The proportional fees charged by platform i are p_i^B for buyers and p_i^S for sellers. A buyer with gross surplus b_i^B from transacting on platform i is willing to use that platform provided that $b_i^B \geq p_i^B$. However, the buyer prefers to transact on platform j if $b_j^B - p_j^B > b_i^B - p_i^B$. Similarly, a seller with type b^S is willing to trade on platform i provided that $b^S \geq p_i^S$, and prefers to trade on platform j if $p_j^S < p_i^S$.

Notice that a transaction can occur only if the two sides have at least one platform in common; that is, there exists at least one platform on which both are willing to trade. If both multihome (are affiliated with both platforms), the choice of platform is a priori indeterminate. In accordance with our illustrations, we assume that, whenever a seller is affiliated with the two platforms, the buyer chooses the one on which the transaction takes place.[11]

3.1.2 Transaction Volumes. The buyers' behavior generates "quasi-demand functions":

$$D_i^B = D_i^B(p_i^B) = \Pr(b_i^B - p_i^B > 0), \tag{8}$$

and

$$d_i^B(p_1^B, p_2^B) = \Pr[b_i^B - p_i^B > \max(0, b_j^B - p_j^B)]. \tag{9}$$

D_i^B is the proportion of buyers who are willing to use platform i when the seller is affiliated only with platform i. Similarly, d_i^B is the proportion of buyers who are willing to trade on platform i when the seller multihomes. By construction, these functions satisfy the following properties:

$$d_i^B \leq D_i^B \leq d_1^B + d_2^B. \tag{10}$$

We assume that the distribution of (b_1^B, b_2^B) is symmetric, which implies that demand functions are also symmetric: $D_1^B(p^B) = D_2^B(p^B) \equiv \hat{D}^B(p^B)$ and $d_1^B(p_1^B, p_2^B) \equiv d_2^B(p_2^B, p_1^B)$. When prices are equal $p_1^B = p_2^B = p^B$, we will use the simplified notation:

10. For simplicity, we assume that the seller's gross surplus does not depend on the platform where the transaction takes place. Furthermore, when performing the welfare analysis, we equate these benefits with the social values of the service brought about by the platforms. However, sellers may exert externalities on each other. For example, a seller's acceptation of a payment card may affect rival sellers. The welfare analysis (but not the positive one) must be amended correspondingly. For more on this, see Rochet and Tirole (2002).

11. This assumption is satisfied by most of our illustrations: a cardholder selects the card when the merchant accepts multiple cards, the reader or viewer selects the newspaper, portal or TV network, the videogame user selects the platform if the game is written for several consoles, etc. Notice that this assumption introduces a slight asymmetry between the two sides of the market.

$$d^B(p^B) \equiv d_i^B(p^B, p^B).$$

We focus for the moment on symmetric prices: $p_1^B = p_2^B = p^B$ and $p_1^S = p_2^S = p^S$. A seller of type b^S affiliates with both platforms when $b^S \geq p^S$ and none otherwise. The transaction volumes on each platform are thus equal to

$$Q = d^B(p^B) D^S(p^S). \tag{11}$$

The sellers' net surplus is, as earlier,

$$V^S(p^S) = \int_{p^S}^{+\infty} D^S(t)dt,$$

while the buyers' net surplus is

$$V^B(p_1^B, p_2^B) = \int_{p_1^B}^{+\infty} d_1^B(t_1, p_2^B)dt_1 + \int_{p_2^B}^{+\infty} D_2^B(t_2)dt_2$$

$$= \int_{p_2^B}^{+\infty} d_2^B(p_1^B, t_2)dt_2 + \int_{p_1^B}^{+\infty} D_1^B(t_1)dt_1.$$

3.1.3 Joint Ownership Benchmarks: The private monopoly and Ramsey benchmarks studied in Section 2 correspond to the situation in which both platforms are under joint ownership and charge identical prices. For instance,

$$D^B(p^B) = 2d^B(p^B)$$

where

$$d^B(p^B) = d_1^B(p^B, p^B) = d_2^B(p^B, p^B).$$

3.1.4 Governance: We assume that the two platforms are controlled by competing entities, either profit-maximizing firms (Section 3.3) or not-for-profit associations (Section 3.4). Important examples of such associations can be found in the payment card industry (Visa and MasterCard). In such associations, prices for buyers and sellers are determined by competition (both intra and inter platforms) on downstream markets (issuing banks on the buyers' side, and acquire on the sellers' side).

3.2 Transaction Volumes for Asymmetric Prices

In order to analyze competition, we need to determine transaction volumes on each platform for arbitrary prices, thus extending formula (11) to nonsymmetric prices. Suppose that Platform 1 is cheaper for sellers: $p_1^S < p_2^S$. A seller of type

b^S has three possibilities:[12] no trade, affiliation with Platform 1 only, affiliation with both platforms. The first possibility is optimal whenever $b^S \leq p_1^S$. The choice between the other two possibilities involves a trade-off between a lower volume (when affiliated with Platform 1 only) and an obligation to trade on the most expensive platform (when affiliated with both platforms). The corresponding expected net surpluses of a seller of type b^S are respectively $(b^S - p_1^S)D_1^B(p_1^B)$ and $(b^S - p_1^S)d_1^B(p_1^B, p_2^B) + (b^S - p_2^S)d_2^B(p_1^B, p_2^B)$. The seller chooses to multihome when b^S is large enough, more precisely when

$$b^S > \hat{b}_{12} \equiv \frac{p_2^S d_2^B - p_1^S(D_1^B - d_1^B)}{d_2^B - (D_1^B - d_1^B)}. \tag{12}$$

We can now summarize sellers' optimal decisions:

- sellers with low types ($b^S \leq p_1^S$) do not trade,
- sellers with high types ($b^S \geq \hat{b}_{12}$) trade on both platforms,
- sellers with intermediate types ($p_1^S < b^S < \hat{b}_{12}$) trade only on the less expensive platform (here, Platform 1).

By undercutting the rival platform, each platform thus induces some sellers (those with intermediate types) to stop multihoming, a strategy known as "steering." The formulae for $p_1^S > p_2^S$ are obtained by permutation of indices. When p_1^S and p_2^S converge to the same price p^S, $\hat{b}_{12}$ and $\hat{b}_{21}$ both converge also to p^S, which establishes continuity of the formulae giving $\hat{b}_{12}$ and $\hat{b}_{21}$.

Let us denote by $\sigma_i (i = 1, 2)$ the following indices:

$$\sigma_i = \frac{d_1^B + d_2^B - D_j^B}{d_i^B} \qquad i, j = 1, 2;\ i \neq j.$$

Given property (10), σ_i belongs to the interval [0, 1]. It measures the "loyalty" of consumers of platform i, i.e. the proportion of them who stop trading when platform i ceases to be available. We call σ_i the "singlehoming" index of Platform i. It is equal to 0 when buyer demand faced by the seller is independent of whether the seller is affiliated with Platform i ($d_1^B + d_2^B = D_j^B$). It is equal to 1 when all Platform i buyers are lost when the seller stops affiliating with that platform ($D_j^B = d_j^B$). For a symmetric price configuration (with $D_1^B = D_2^B = \hat{D}^B$), we have

$$\sigma_1 = \sigma_2 = \sigma = 2 - \frac{\hat{D}^B}{d^B}.$$

Starting from a symmetric price structure, suppose platform 1 decreases p_1^S by a small amount ε. This increases demand for Platform 1 in two ways: The platform attracts new merchants ($p_1^S - \varepsilon \leq b_1^S < p_1^S$) and "steers" former multihoming merchants ($p_1^S < b_1^S < \hat{b}_{12}$). Given that $(\partial \hat{b}_{12}/\partial p_1^S) = 1 - (1/\sigma_2)$, the effectiveness of steering depends on σ_2: it is nil when $\sigma_2 = 1$ and infinite when $\sigma_2 = 0$.

12. Affiliation with Platform 2 only is clearly dominated.

We are now in a position to determine the volume of transactions on each platform as a function of prices p_i^B and p_i^S. We restrict ourselves to the case $p_1^S \leq p_2^S$ (the case $p_2^S < p_1^S$ is obtained by symmetry). Let us denote by D^S the sellers' "quasi-demand function":

$$D^S(p^S) = \Pr(b^S > p^S).$$

From the affiliation decisions derived previously, a proportion $D^S(\hat{b}_{12})$ of sellers multihome, while a proportion $D^S(p_1^S) - D^S(\hat{b}_{12})$ are affiliated only with Platform 1. Assuming that the probability of a meeting between a buyer and a seller is independent of their types, the total expected volumes of transactions on the platforms are:

$$Q_1 = d_1^B(p_1^B, p_2^B)\, D^S(\hat{b}_{12}) + D_1^B(p_1^B)\{D^S(p_1^S) - D^S(\hat{b}_{12})\}, \tag{13}$$

for Platform 1, and:

$$Q_2 = d_2^B(p_1^B, p_2^B)\, D^S(\hat{b}_{12}), \tag{14}$$

for Platform 2, where $\hat{b}_{12}$ is given by formula (12). As already noticed, these formulae are continuous across the "diagonal" $p_1^S = p_2^S$:

$$\lim_{\substack{p_1^S \to p^S \\ p_2^S \to p^S}} Q_i = d_i^B(p_1^B, p_2^B)\, D^S(p^S).$$

3.3 Competition between Proprietary Platforms

Proprietary platforms choose prices so as to maximize profit. Consider, for example, Platform 1's profit:

$$\pi_1 = (p_1^B + p_1^S - c)Q_1. \tag{15}$$

As in the case of a monopolist, this maximization can be decomposed into the choice of a price level, $p_1 = p_1^B + p_1^S$, and that of a price structure given a price level. The first-order conditions are:

$$Q_1 + (p_1^B + p_1^S - c)\,\frac{\partial Q_1}{\partial p_1^B} = Q_1 + (p_1^B + p_1^S - c)\,\frac{\partial Q_1}{\partial p_1^S} = 0,$$

or

$$\frac{\partial Q_1}{\partial p_1^S} = \frac{\partial Q_1}{\partial p_1^B} = -\frac{Q_1}{p_1^B + p_1^S - c}. \tag{16}$$

The following analysis is complex, as it must handle a potential lack of smoothness of the objective function. It can be skipped in a first reading. The end result (Proposition 3) is remarkably simple, though.

Recall the expressions of volumes on both systems, when, say, $p_1^S \leq p_2^S$:

$$Q_1 = d_1^B(p_1^B, p_2^B)\, D^S(\hat{b}_{12}) + D_1^B(p_1^B)\{D^S(p_1^S) - D^S(\hat{b}_{12})\}, \tag{13}$$

$$Q_2 = d_2^B(p_1^B, p_2^B)\, D^S(\hat{b}_{12}), \tag{14}$$

where

$$\hat{b}_{12} = \frac{p_2^S d_2^B - p_1^S(D_1^B - d_1^B)}{d_2^B - (D_1^B - d_1^B)}. \tag{12}$$

We focus on symmetric equilibria ($p_i^S \equiv p^S$, $p_i^B \equiv p^B$), for which volumes have simpler expressions:

$$Q_i = d_i^B(p^B, p^B)\, D^S(p^S).$$

While

$$\frac{\partial Q_1}{\partial p_1^B} = \frac{\partial d_1^B}{\partial p_1^B}(p^B, p^B)\, D^S(p^S), \tag{17}$$

the first derivative in formula (17) is not necessarily well defined since volumes have a different expression according to whether $p_1^S < p_2^S$ or $p_1^S > p_2^S$:

$$Q_1 = d^B(p^B)\, D^S(\hat{b}_{12}) + \hat{D}^B(p^B)\{D^S(p_1^S) - D^S(\hat{b}_{12})\}$$

when $p_1^S < p_2^S$, and

$$Q_1 = d^B(p^B)\, D^S(\hat{b}_{21})$$

when $p_1^S > p_2^S$. Interestingly, Q_1 turns out to be differentiable[13] even at $p_1^S = p_2^S$. Indeed, at symmetric prices:

13. The left- and right-derivatives of Q_1 with respect to p_1^S at $p_1^S = p_2^S = p^S$ (implying $\hat{b}_{12} = \hat{b}_{21} = p^S$) are:

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_L = (D^S)' \frac{\partial \hat{b}_{12}}{\partial p_1^S}[d^B - \hat{D}^B] + (D^S)'\hat{D}^B,$$

and

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_R = (D^S)' \frac{\partial \hat{b}_{21}}{\partial p_1^S} d^B.$$

Moreover

$$\frac{\partial \hat{b}_{12}}{\partial p_1^S} = -\frac{\hat{D}^B - d^B}{2d^B - \hat{D}^B}, \qquad \text{and} \qquad \frac{\partial \hat{b}_{21}}{\partial p_1^S} = \frac{d^B}{2d^B - \hat{D}^B}.$$

And so

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_L = (D^S)' \left[\frac{(\hat{D}^B - d^B)^2}{2d^B - \hat{D}^B} + \hat{D}^B\right] = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B}$$

$$\left(\frac{\partial Q_1}{\partial p_1^S}\right)_R = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B}.$$

Thus Q_1 is differentiable with respect to p_1^S.

$$\frac{\partial Q_1}{\partial p_1^S} = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B}. \tag{18}$$

Using (16), (17), and (18) we obtain a simple form for the first-order condition for a symmetric equilibrium:

$$\frac{\partial d_i^B}{\partial p_i^B} D^S = (D^S)' \frac{(d^B)^2}{2d^B - \hat{D}^B},$$

or:

$$\left(\frac{2d^B - \hat{D}^B}{d^B}\right)\left(-\frac{\partial d_i^B/\partial p_i^B}{d^B}\right) = -\frac{(D^S)'}{D^S}.$$

The first term on the left-hand side of this latter formula is the singlehoming index σ defined earlier, which measures the proportion of "unique customers." The second term is the ratio of the own-brand elasticity of demand for buyers

$$\eta_o^B = -\frac{p^B \partial d_i^B/\partial p_i^B}{d^B}$$

over the buyers' price p^B. Finally, the last term is the ratio of the elasticity of sellers' demand over sellers' price. Thus we can state:

PROPOSITION 3. *A symmetric equilibrium of the competition between proprietary platforms is characterized by:*

$$p^B + p^S - c = \frac{p^B}{\eta_o^B} = \frac{p^S}{(\eta^S/\sigma)}.$$

The formulae are thus the same as in the monopoly platform case, except that a) on the buyer side, the demand elasticity η^B is replaced by the (higher) own-brand elasticity η_o^B, and b) on the seller side, the demand elasticity η^S is replaced by the equivalent of an own-brand elasticity η^S/σ. When all buyers singlehome ($\sigma = 1$), the own-brand elasticity and the demand elasticity coincide. But as multihoming becomes more widespread (σ decreases), the possibility of steering increases the own-brand elasticity η^S/σ.

3.4 Competition between Associations

When platforms are run by not-for-profit cooperatives owned by members (operators on the buyer and seller sides), prices paid by the end users are set by the members and not by the platforms. Platforms however have an important say in the price structure, especially if competition among members is intense on both sides of the market. In our model, an association's only strategic decision

is the choice of access charges between members. Neglecting platform costs, the zero-profit condition implies that these access "charges" exactly offset each other as one side receives the charge paid by the other side. For example in the payment card industry the access charge is called the interchange fee and is paid by acquirers (the sellers' banks) to issuers (the buyers' banks).[14] This section studies the access charge chosen by competing associations and compares the corresponding prices for final users (buyers and sellers) with those resulting from competition between profit-maximizing systems. While the section is currently most relevant to the payment card industry, its potential applicability is much broader. For example, reflecting recent concerns about unequal access to B2B exchanges, some have suggested that these exchanges be run as nonprofit associations. Furthermore, and as will be observed in Section 7.2, networks of interconnected networks (e.g., communication networks) are economically similar to nonprofit platforms.

The members compete on two downstream markets, the buyer and the seller downstream markets. Given access charge a_i on platform i, the net marginal costs for a member of platform i of serving a buyer and a seller, respectively, are $c^B - a_i$ and $c^S + a_i$, where c^B and c^S represent the gross marginal costs incurred by the members on each side of the market. We make the simplifying assumption that intraplatform competition results in constant equilibrium margins charged by members on downstream markets: m^B on the buyers' side and m^S on the sellers' side. Equilibrium prices are thus given by:

$$p_i^B = c^B - a_i + m^B, \qquad p_i^S = c^S + a_i + m^S.$$

This assumption is for example satisfied if (a) members belong to a single association and are differentiated in a direction orthogonal to that of platform differentiation;[15] and (b) members on a given platform are little differentiated. Intense intraplatform competition then results in Hotelling competition between members taking as given (as a first-order approximation) the number of end users on the platform (which is basically determined by the platforms' access charges given that the members' markups are small).[16]

Under this simplifying assumption, the profits of all members of an association are proportional to the volume of transactions on the association's

14. The determination of access charges within associations has so far only been studied in the context of the payment card industry and under the assumption of a monopoly platform [Rochet and Tirole (2002), Schmalensee (2002)].

15. Mathematically, in a generalized Hotelling framework, the "transportation cost" for an end-user when selecting a (platform, member) pair is the sum of the transportation cost to the platform and that to the member.

16. If members have dual membership instead (e.g., they are both affiliated with Visa and MasterCard, or they provide support or write applications for two cooperatively designed operating systems or videogame platforms), then requirement (b) is unnecessary in that margins are constant even if member differentiation is not small relative to platform differentiation: See Hausman, Leonard, and Tirole (2003). But one must then inquire into the associations' governance structure. Our treatment carries over as long as governance leads each association to maximize its volume.

platform. The interests of all members are thus completely aligned. Regardless of its exact structure the association selects the access charge so as to maximize its volume. Furthermore the total price on each system is constant:

$$p_i^B + p_i^S = c + m, \tag{19}$$

where $m = m^B + m^S$ is the total margin on downstream markets and $c = c^B + c^S$.

Last, in order to be able to compare the association with the cases of a monopolist and of competing proprietary platforms, we must assume that the quasi-demand functions are the same. That is, the members are only selling the varieties of each platform that the proprietary platforms were selling. Because we kept quasi-demand functions quite general, there is no difficulty in assuming this is indeed the case.

The outcome of the competition between the two associations is characterized by two price vectors (p_i^B, p_i^S), $i = 1, 2$, such that: for all i, (p_i^B, p_i^S) maximizes the volume Q_i on system i subject to (19), taking as given the price vector (p_j^B, p_j^S) on the other system.

The first-order conditions for a symmetric equilibrium are given by

$$p^B + p^S = c + m, \tag{20}$$

(condition on total price) and the equivalent of condition (5):

$$\frac{\partial Q_i}{\partial p_i^B} = \frac{\partial Q_i}{\partial p_i^S}, \tag{21}$$

(same impact on volume of a marginal price increase on each side of the market).

The analysis of the price structure is therefore identical to that for proprietary platforms. The price level is lower for associations with healthy competition among their members but may exceed the proprietary platforms price level if double marginalization is strong.

PROPOSITION 4. *A symmetric equilibrium of the competition between associations is characterized by*

$$p^B + p^S = c + m$$

and

$$\frac{p^B}{\sigma \eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

Comparing now Propositions 2 and 4, we see that even when downstream markets are perfectly competitive (the margin m converges to zero) and so the price level is socially optimal, competition between not-for-profit associations need not generate an efficient outcome. Indeed, the condition for an efficient price structure (given in Proposition 2) is:

$$\frac{p^B}{\eta^B}\left[\frac{V^B}{D^B}\right] = \frac{p^S}{\eta^S}\left[\frac{V^S}{d^S}\right], \tag{7}$$

while the condition characterizing competition between associations is different:

$$\frac{p^B}{\sigma\eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

This is natural, as (*a*) the associations do not internalize the end-users' surpluses, and (*b*) the associations aim at steering sellers (which is reflected by the presence of σ) and stealing buyers (as indicated by the presence of η_o^B) away from the rival association, while market share considerations play no role in a Ramsey program. It is therefore perhaps remarkable that the two conditions coincide in the special case of linear demands, which we explore in detail in Section 5.

4. Determinants of Business Model

Let θ be a parameter that affects the volume of transactions on the platforms. In this section, we consider the impact of a small variation in θ on user prices p^B and p^S, depending on industry structure (monopoly or duopoly) and on the platforms' governance structure (for-profits or associations). We concentrate on three important determinants of industry conduct and performance:

Marquee buyers: In the first application, θ represents a (small) uniform shift in sellers' surpluses, due to the presence of marquee buyers on the other side of the market. As a result, the sellers' demand function becomes:

$$D^S(p^S, \theta) = D^S(p^S - \theta).$$

Installed Bases/Captive Buyers: In the second application, θ represents the (small) mass of buyers who are loyal to their platform, independently of prices. Such buyers, say, are tied by long-term contracts. As a result, the buyers' demand functions become:

$$d_i^B(p_1^B, p_2^B, \theta) = d_i^B(p_1^B, p_2^B) + \theta,$$
$$D^B(p^B, \theta) = D^B(p^B) + \theta,$$
$$\hat{D}^B(p^B, \theta) = \hat{D}^B(p^B) + \theta.$$

Multihoming: In the third application, θ represents an exogenous increase in the singlehoming index of buyers. Assume for example that d^B does not depend on

θ, while D^B decreases in θ. Then $\sigma(p^B, \theta) = 2 - [D^B(p^B, \theta)/d^B(p^B)]$ is increasing in θ, while η_o^B does not depend on θ.[17]

Proposition 5 analyzes the impact of small variations of θ on the prices p^B and p^S.

PROPOSITION 5.

(1) In the case of a monopoly platform (for-profit or association) and with log concave demand functions, the seller price increases when there are marquee buyers and decreases when there are captive buyers. The buyer price moves in the opposite direction.
(2) The same result holds under competition between associations, except that the comparative statics with respect to captive buyers requires a regularity condition.
(3) In the case of competing associations, an increase in the multihoming index of buyers (keeping demand elasticities constant) leads to an increase in the buyer price and a decrease in the seller price.

Intuitively, marquee buyers make the platform more attractive for the sellers. The platform then raises its price p^S to sellers, which reduces the de facto marginal cost, $c - p^S$, of provision of the service to the buyers. The buyer price therefore falls. The intuition is similar in the case of captive buyers. Captive buyers allow the platform to raise the price p^B to buyers, thus reducing the de facto marginal cost $c - p^B$ of serving sellers. A regularity condition however is required here in the case of platform competition, which creates a countervailing steering effect: Each platform's buyer membership is then "more unique" to the platform, and so it is more costly for a seller to forgo the platform. Last, an increase in multihoming makes steering more attractive and puts a downward pressure on the seller price.

5. Linear Demands

We illustrate the results obtained so far in a variant of the Hotelling model, where a buyer's preferences for platforms are represented by his location x on a line. Buyers are uniformly distributed on a line of length $(\Delta + 2\delta)$. Platforms 1 and 2 are symmetrically located at a distance $\Delta/2$ of the origin of the line ($x = -(\Delta/2)$ for Platform 1 and $x = \Delta/2$ for Platform 2). The number Δ parametrizes the degree of substitutability between platforms. Buyers have also access to

17. This is, for example, the case in the Hotelling specification presented in Section 5, when the marginal transportation cost of buyers increases only for distances in the noncompetitive hinterland of the rival platform, so that d_i^B is unaffected while D^B decreases.

outside options, represented conventionally by two other symmetric platforms (denoted $1'$ and $2'$), located further away from the origin ($x = -(\Delta/2) - \delta$ and $x = (\Delta/2) + \delta$) and charging the same, exogenous, price p_0. The number δ will serve us as a measure of the weight of "unique consumers." When using a platform located at distance d, buyers incur a quadratic transportation cost $\frac{1}{2}d^2$, (the transportation cost parameter is normalized to 1 without loss of generality).

Proposition 6 (proved in the Appendix) exhibits three main implications of the linear case. First, the price structure is the same regardless of whether the industry is served by a private monopoly, competing proprietary platforms or competing associations. Second, if demand is linear on the seller side as well, then this common price structure is Ramsey optimal. Taken together, these results show that without detailed information about the demand structure, one should not expect clear comparisons of price structures across governance mechanisms. Nor are policy interventions to alter the price structure (as opposed to the price level) likely to be solidly grounded. Third, Proposition 6 provides sufficient conditions for the second-order conditions to be satisfied in the linear demand case.

PROPOSITION 6. *Suppose that the buyers' quasi-demand is described by an Hotelling model, with uniform distribution and outside options with distance Δ between the two platforms and distance δ between each platform and its nearest outside option, and that the market is not covered (not all potential buyers buy).*

(1) • The buyer singlehoming index is equal to:

$$\sigma = \Delta/(\Delta + \delta),$$

and decreases when the platforms become more substitutable.

- *The platforms' ability to steer (undercut in order to discourage sellers from multihoming) decreases with the buyer singlehoming index.*
- *On the buyer side, total elasticity is equal to own-brand elasticity times the singlehoming index:*

$$\eta^B = \eta_o^B \sigma.$$

(2) The price structure is the same under a monopoly platform, competing proprietary platforms and competing associations. It satisfies

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}.$$

(3) If furthermore seller demand is linear, then the price structure in the three environments is Ramsey optimal.

(4) The price vectors given in formulae (31) and (32) satisfy the second-order conditions for an equilibrium if and only if δ/Δ is smaller than $(1 + \sqrt{5})/2$.

6. Generalization to Fixed User Fees and Usage Costs

In many of the examples presented in the Introduction, fixed costs, either fixed fees charged by the platforms or fixed usage costs, play an important role. In order to demonstrate the robustness of our results to the introduction of fixed costs, we now adapt our model accordingly. To simplify the analysis, we assume that buyers singlehome (for example, consumers read a single newspaper or connect to a single portal). Second, we focus on the symmetric equilibrium. There is a sizeable literature on tipping in the presence of user fixed costs and we have little to add to this literature. Last, we first look at the case in which there is no direct exchange of money between the two sides of the market, as is the case for advertising in newspapers, TVs, and portals; we will later show how to extend the analysis to cases, such as videogames, exhibiting direct monetary transactions between end-users.

Platforms incur fixed costs C^B and C^S per buyer and seller, as well as marginal cost c per transaction between them (presumably $c = 0$ for advertising). Let platform i charge fixed fees A_i^B and A_i^S and variable charges $a_i^B N_i^S$ and $a_i^S N_i^B$ to buyers and sellers, where N_i^B and N_i^S are the numbers of buyers (eyeballs) and sellers (advertisers) connected to platform i. A buyer with (possibly negative) average benefit b_i^B of receiving an ad and with fixed usage cost γ_i^B (also possibly negative) has net utility

$$U_i^B = (b_i^B - a_i^B) N_i^S - A_i^B - \gamma_i^B.$$

Similarly, a seller with average benefit b^S of reaching a consumer and with fixed cost γ^S of designing an ad for this newspaper has net utility:

$$U_i^S = (b^S - a_i^S) N_i^B - A_i^S - \gamma^S.$$

The buyers are heterogenous over parameters (b_i^B, γ_i^B) and sellers are heterogenous over parameters (b^S, γ^S).

The strategic choices for the platforms are the per "transaction" (eyeball viewing an ad) markups:

$$p_i^B \equiv a_i^B + \frac{(A_i^B - C^B)}{N_i^S} \qquad \text{and} \qquad p_i^S \equiv a_i^S + \frac{(A_i^S - C^S)}{N_i^B}.$$

Assuming that readers buy a single newspaper, the number of copies sold by newspaper i is given by

$$N_i^B = \Pr(U_i^B > \max(0, U_j^B)),$$

which is equal to some function d_i^B of prices (p_1^B, p_2^B) and numbers of ads (N_1^S, N_2^S) of the two newspapers

$$\begin{aligned} N_i^B &= d_i^B(p_1^B, N_1^S; p_2^B, N_2^S) \equiv \Pr((b_i^B - p_i^B)N_i^S - C^B - \gamma_i^B \\ &\geq \max[0, (b_j^B - p_j^B)N_j^S - C^B - \gamma_j^B]). \end{aligned} \tag{23}$$

N_i^S is itself a function of p_i^S and N_i^B:

$$N_i^S = D^S(p_i^S, N_i^B) = \Pr((b^S - p_i^S)N_i^B > \gamma^S). \tag{24}$$

These formulas are valid provided fixed costs for buyers are high enough so that no buyer buys the two newspapers (no multihoming for buyers). Substituting (24) into (23), and solving for (N_1^B, N_2^B), one obtains demand functions for the buyers:

$$N_i^B = n_i^B(p_1^B, p_1^S; p_2^B, p_2^S).$$

Let us define the own- and cross-elasticities for buyer demand:

$$\eta_o^B \equiv -\frac{\partial n_i^B}{\partial p_i^B}\frac{p_i^B}{n_i^B} \qquad \text{and} \qquad \eta_S^B \equiv -\frac{\partial n_i^B}{\partial p_i^S}\frac{p_i^S}{n_i^B}.$$

On the seller side, we define the own-price elasticity and the network elasticity:

$$\eta^S \equiv -\frac{\partial D^S}{\partial p^S}\frac{p^S}{D^S} \qquad \text{and} \qquad \eta_N^S \equiv \frac{\partial D^S}{\partial N^B}\frac{N^B}{D^S}.$$

With this notation, the formulae for transaction volumes and platform profit look remarkably similar to the ones obtained earlier. Platform i maximizes:

$$\pi_i = (p_i^B + p_i^S - c)N_i^B N_i^S.$$

Simple computations yield:

PROPOSITION 7. *A symmetric equilibrium is characterized by prices* (p^B, p^S) *satisfying:*

$$p^B + p^S - c = \frac{p^B}{\eta_o^B(1 + \eta_N^S)} = \frac{p^S}{\eta^S + \eta_S^B(1 + \eta_N^S)}.$$

While we simplified the model by assuming singlehoming ($\sigma = 1$), the presence of fixed costs implies that network externalities impact not only end-user surpluses, but also demands. For example, on the buyer side, the own price elasticity η_o^B is multiplied by a factor greater than 1 to account for the fact that when a platform reduces its buyer price, more buyers connect to the platform, inducing more sellers to connect and further increasing buyer demand. And similarly on the seller side.

In more structured applications, the formulae in Proposition 7 may simplify. For example, in the advertising example, it is reasonable to assume that sellers

incur no fixed usage cost ($\gamma^S \equiv 0$), since the advertising campaign has already been prepared for other media. In this case formula (24) shows that D^S does not depend on N^B, so that $\eta_N^S = 0$, and

$$p^B + p^S - c = \frac{p^B}{\eta_o^B} = \frac{p^S}{\eta^S + \eta_S^B}.$$

Last, let us turn to the (videogame or operating system) case in which the transaction between the seller and buyer involves a price charged by the seller to the buyer.

Additional complications arise because of this monetary transaction between buyers and sellers. The equilibrium price of this transaction is then determined by competitive forces in the market for videogames or software applications and depends on the pricing policies of platforms. To illustrate how to extend the model to reflect this, we assume that sellers have market power and no marginal cost and that buyers differ only in the fixed cost of learning how to install and use an operating system or a console (and in the identity of their preferred applications). They receive gross surplus $\underline{v}$ for a fraction α of the applications (where the corresponding applications are drawn in i.i.d. fashion among consumers) and $\bar{v} > \underline{v}$ for a fraction $(1 - \alpha)$. When α is large (so that $(1 - \alpha)\bar{v} < \underline{v}$), it is efficient for the platforms to induce developers to charge the low price $p = \underline{v}$, so that buyers buy all games and receive a net marginal surplus $b^B = (1 - \alpha)(\bar{v} - \underline{v})$. Then we can assume w.l.o.g. that $a^S = 0$, so that $b^S = \underline{v}$. Using the same notation as above, the net utilities of a typical buyer and a typical seller are

$$U_i^B = b^B N_i^S - A_i^B - \gamma_i^B,$$

$$U_i^S = b^S N_i^B - A_i^S - \gamma^S.$$

Denoting again by p_i^B and p_i^S the per transaction mark-ups:

$$p_i^B = \frac{A_i^B - C^B}{N_i^S} \qquad \text{and} \qquad p_i^S = \frac{A_i^S - C^S}{N_i^B},$$

and d_i^B, D^S the associated demand functions, we obtain the same formulae as in Proposition 7.

7. Summary and Mini Case Studies

Let us now summarize the paper's key insights. The main contribution has been to derive simple formulae governing the price structure in two-sided markets, and this for a wide array of governance structures (private monopoly, Ramsey planner, competition between for-profit or nonprofit platforms). But we also obtained more specific insights. On the public policy side:

1) The Ramsey price structure does not correspond to a "fair cost allocation." Rather, like private business models, it aims at getting both sides on board.
2) The main conceptual difference between private and Ramsey price structures is that the latter takes into account the average net surplus created on the other side of the market when attracting an end user on one side. Yet, private business models do not exhibit any obvious price structure bias (indeed, in the special case of linear demands, all private price structures are Ramsey optimal price structures).

On the business model front, we obtained:

3) Monopoly and competitive platforms design their price structure so as to get both sides on board.
4) An increase in multihoming on the buyer side facilitates steering on the seller side and results in a price structure more favorable to sellers.
5) The presence of marquee buyers (buyers generating a high surplus on the seller side) raises the seller price and (in the absence of price discrimination on the buyer side) lowers the buyer price.
6) Captive buyers tilt the price structure to the benefit of sellers.

We now develop seven "mini case studies" meant to emphasize the attention paid by platforms to the pricing structure. A rigorous validation of testable implications 3) through 6) lies beyond the scope of this paper, and we hope that future research will perform the econometric studies needed to confirm or infirm these hypotheses in specific industries. We only offer some casual empiricism; this preliminary evidence seems quite encouraging of the theory.

7.1 Credit and Debit Cards

The payment industry offers a nice illustration of implications 3) through 6). Historically, the business model for credit and differed debit cards has been to attract cardholders and induce them to use their cards. Visa and MasterCard are not-for-profit associations owned by over 6,000 bank (and nonbank) members. The associations centrally set interchange fees to be paid by acquirers (the merchants' banks) to issuers (the cardholders' banks). These interchange fees are proportional to transaction volume. A higher interchange fee is, via the competition among issuers, partly or fully passed through to consumers in the form of lower card fees and higher card benefits, which encourages card ownership and usage; and via competition among acquirers, partly or fully passed through to merchants, who pay a higher merchant discount (the percentage of the sale price that the merchant must pay the acquirer), which discourages merchant acceptance. The associations' choice of interchange fees have typi-

cally favored cardholders over merchants who kept accepting the card despite the high level of the merchant discounts (implication 3).[18]

American Express, a for-profit closed system, works on the same business model, with an even higher degree of cross-subsidization. Traditionally, it has charged a much higher merchant discount.[19] It could afford to do so because the Amex clientele—in particular the corporate card clientele—was perceived as very attractive by merchants (implication 5). The gap between Amex's and the associations' merchant discounts has narrowed in the 1990s as more and more Amex customers got also a Visa card or MasterCard. Such "multihoming" by a fraction of cardholders made it less costly for merchants to turn down Amex cards (implication 4).

The online debit card market in the United States has adopted an entirely different business model. Rather than courting consumers, it has wooed merchants through a low interchange fee. One key difference with credit and differed debit cards is that consumers indeed do not need to be courted (they already have in their pocket an ATM card, that they can use as an online debit card; so in a sense they are "captive"), while merchants, to perform online debit, must install costly pinpads (which most of them have not yet done).[20] This emphasizes the relevance of implications 3 and 6).

7.2 Internet

In the Internet, the instrument of cross-subsidization is the termination or settlement charge (or lack thereof) between backbones. The termination charge for off-Net traffic between two backbones is, as its name indicates, the price paid by the originating backbone to the terminating backbone for terminating the traffic handed over. It can be shown[21] that it is optimal to charge the "off-Net cost" to end users for marginal incoming and outgoing traffic. That is, backbones should charge as if the traffic were off-Net, even though a fraction of the traffic is actually on-Net. The charge for incoming (outgoing) traffic decreases (increases) one-for-one with the termination charge. This implies that a high termination charge is indirectly borne by end users, like websites, whose volume of outgoing traffic far exceeds the volume of incoming traffic, and benefits end users, such as dial-up customers, who mostly receive traffic (downloads).

18. Looking forward, it is likely that merchant card acceptance will become more elastic with the (ongoing) advent of online debit and the (future) introduction of Web platforms.

19. And thus implicitly a much higher "interchange fee." For Amex, the interchange fee is only implicit, since the company is vertically integrated and performs the three roles of issuer, system and acquirer.

20. The online offerings were first made by regional ATM networks. A number of these networks have now been consolidated and converted into a for-profit platform (Concord ESF).

21. See Laffont et al. (2003) and Jeon, Laffont, and Tirole (2003) for derivations of this result in different environments.

An Internet in which backbones exchange traffic at a uniform volume-proportional termination charge is similar to the case of a single not-for-profit platform. This analogy can be best depicted by envisioning backbones exchanging traffic at public peering points.[22] An "association" running these public peering points and keeping track of bilateral termination charges would be similar to a credit-card association recording traffic between acquirers and issuers, with the termination charge the counterpart of the interchange fee. A network of interconnected networks therefore resembles a single not-for profit platform.

The Internet is still mostly run by pre-deregulation legacy arrangements, according to which the backbones charge nothing to each other for terminating traffic. This business model is currently being reexamined and it is quite possible that, as is the case for regular telephony, positive termination charges will be levied in the future. The legacy arrangements may well have made sense in an epoch in which the posting of content on the Web had to be encouraged. A key question now is whether a change in industry conditions motivates a move toward paying settlements.

7.3 Portals and Media

The business model of (nonpay) TV, and to a large extent, newspapers has been to use viewers and readers as a loss leader, who attract advertizers. This business model has been adopted with a vengeance by Internet portals, which have supplied cheap or free Internet access as well as free content (share quotes, news, e-mail, etc.) to consumers. The profit center has been advertising revenue, including both fixed charges for banner placement and proportional referral fees.[23]

Interestingly the portal industry is considering whether to stick to this business model or move to for-fee content. For example, Yahoo! is now starting to charge fees for services such as real-time share-quote services or auction services. A number of content sites have appeared that charge substantial fees for on-line content.[24]

7.4 Video Games

Our last four case studies are drawn from the software industry. The video game market is a typical two-sided one. A platform cannot sell the console without games to play on and cannot attract game developers without the prospect of an installed base of consumers. In its thirty years of existence, the video game

22. Even though, in practice, they mainly exchange their traffic at bilateral peering points.

23. See Elfenbein and Lerner (2001) for a thorough analysis of contracts in recent Internet Portal Alliances.

24. See, e.g., the *Economist* (April 14, 2001, p. 65) for more details.

industry has had four leading platforms, Atari, Nintendo and Sega, and finally Sony. The business model that has emerged uses consoles as the loss leader and draws platform profit from applications development. To be certain, history has repeatedly shown that technically impressive platforms (e.g., Mattel in 1981, Panasonic in 1993, and Sega in 1985 and after 1995) fail when few quality games are written for them. But attracting game developers is only a necessary condition. In fact, the business model currently employed by Nintendo, Sega and Sony is to charge software developers a fixed fee together with a per-unit royalty on the games they produce.[25] Microsoft released in the fall of 2001 the Xbox in competition with Sony's dominant PlayStation 2. Interestingly, Microsoft manufactures the Xbox console and uses it as a loss leader. While courting the developers[26] by using the familiar X86 chip and Windows platform and by not charging for the Xbox Prototype kit, Microsoft has stated that it intends to draw revenue from royalties.

Although the industry's business model involves drawing revenue from developers, platforms can only go so far in taxing the latter. A key factor in Sony's PlayStation's victory over the Sega Saturn and Nintendo 64 was that Sony offered a development platform and software application that was much cheaper (about $10,000 per seat) and much easier to use (as it was PC based) that its two rivals'.[27]

7.5 Streaming-Media Technology

Streaming-media platforms incorporate encoding, compression, scripting and delivery technologies to allow the delivery of streaming content, facilitate content creation and permit interactivity; for example, it is central to conferencing and Webcast. The current competition is among the RealNetworks, Microsoft, and Apple platforms.

The streaming-media industry is still in its infancy and it is probably too early to point at "the" business model. The current business mostly, but not exclusively, subsidizes the client side. RealNetworks and Apple offer two clients: a basic, free client and a better, nonfree one. RealNetworks, the current leader charges significant amounts on the server side for RealServer Plus and its upgrades (see Cowell 2001 for details). Apple in contrast is currently free on the server side, but has the disadvantage on running only on Macs. Microsoft's Windows Media is free (bundled with the operating systems).

25. Initially, Nintendo placed a chip in its console. The console would not work unless an authenticating chip was present in the game cartridge. Encryption techniques allow platform manufacturers to meter game sales.

26. In September 2000, 157 developers were working on Xbox games. The Xbox is launched with 26 games. Interestingly, Electronic Arts (the maker of Fifa, SimCity, and James Bond) was able to impose special conditions on Microsoft.

27. See Cringely (2001) for more detail. Sony sold its console below cost and made the money back on game royalties.

7.6 Operating Systems

Both sides in the Microsoft browser case agreed that a key competitive advantage of Windows over competing operating systems is its large installed base of applications. Windows' business model is basically the opposite of that of videogame platforms. Windows makes money on users and as a first approximation does not make or lose money on applications developers.[28] It fixes the Applications Programming Interfaces three or four years in advance (a costly strategy) and invests heavily in developer support. This strategy proved profitable in its contest with Apple and IBM's OS/2. Apple has no integrated developers system tools allowing developers to test their programs; the latter had to buy an IBM workstation and a compiler. IBM viewed developer kits as a profit center.[29] While other factors undoubtedly played a role in the competition among the three platforms, observers usually agree that Microsoft's choice of business model helped Windows establish dominance.

7.7 Text Processing

A key issue confronting purchasers of text processing software is whether they will be able to "communicate" with people who don't make the same choice. Commercial software vendors have in this respect converged on the following business model: They offer a downgraded version of the paying software as "freeware." This free version allows "nonusers" to open, view, and print, but not edit documents prepared with the paying software, and copy information from those documents to other applications. Examples of such free viewers are Word Viewer, PDF Viewer, and Scientific Viewer.[30]

8. Final Thoughts about Two-Sided Markets

Our premise is that many (probably most) markets with network externalities are two- (or multiple-) sided markets. A market with network externalities is a two-sided market it platforms can effectively cross-subsidize between different

28. We are unaware of "hard data" on this and just report the industry's conventional wisdom. Nor do we have any hard data for handheld computer operating systems. Handheld computers operating systems, dominated by Palm's platform (75 percent market share in the United States) and Microsoft's Pocket PC software, have adopted a business model that is similar to Windows for PC operating systems. Palm and Microsoft apparently charge about 10 percent of the hardware's wholesale price ($5 to $15) to hardware manufacturers. Both provide standard user interfaces and central support and development tools for developers of third-party software. For more detail, see ⟨http://www5.zdnet.com/zdnn/stories/news/0,4586,2714210,00.html?chkpt-zdhpnews01 ⟩.

29. Software developer kits were sold at about $600.

30. For Scientific Word, a mathematics software program adding a user interface and various other functions on to LaTeX.

categories of end users that are parties to a transaction. That is, the volume of transactions on and the profit of a platform depend not only on the total price charged to the parties to the transaction, but also on its decomposition. There are two reasons why platforms may be unable to perform such cross-subsidization:

a) *Both sides of the market coordinate their purchases.* A debit card platform negotiating with a government for the handling of interagency financial transactions, an Internet operator offering an Intranet solution to a company, or a streaming-media platform offering streaming audio and video to a firm primarily for internal use all deal with a single party. A subsidization of the client side by the server side for example does not affect the total price of the software service and, ceteris paribus, does not affect the demand for the platform.[31]
b) *Pass-through and neutrality.* Even when end users on the two sides of the market act independently, monetary transfers between them may undo the redistributive impact and prevent any cross-subsidization. The value-added tax is an epitome of the possibility of neutrality. First-year economic students are taught that it really does not matter whether the seller or the buyer pays the tax. In the end, prices adjust so that any tax paid by the seller is passed through to the consumer. If such neutrality holds, then the markets discussed above should be treated as one-sided markets, that is markets in which only the total per transaction price charged by the platform matters and not its decomposition between end users.[32]

Yet thinking of the markets discussed in this paper as one-sided markets just runs counter to evidence. First, the platform owners in all these industries devote much attention to their price-allocation business model: Is it more important to woo one side or the other? The quest for "getting both sides on board" makes no sense in a world in which only the total price for the end user interaction, and not its decomposition, matters. And the trade press would not contain so many descriptions of "chicken-and-egg" problems. Second, the end users themselves are also very sensitive to the allocation of cost between them, indicating that

31. Mobile and fixed telephone services, for which most users are both callers and receivers, cannot be treated as one-sided markets. A high termination charge raises the marginal cost of calls and lowers the marginal cost of call receptions. In other words, the termination charge is an instrument of cross-subsidization similar to the interchange fee in credit card markets. Telephone users are on both sides of the market for different communications only. For a given communication, end users are on a single side and (unless they are engaged in a repeated relationship) their consumption behaviors depend on their own price (calling price for the caller, receiving price for the receiver). As a consequence, the choice of termination charge is not neutral. See Jeon et al. (2003) for more detail.

32. "Neutrality" refers to the pass-through property and a priori bears no connotation with respect to the well-being of end users and platforms and to social welfare. While neutrality reduces the number of instruments at the disposal of a given platform, it is not clear whether it helps or hurts the platforms in their rivalry. Similarly, neutrality a priori may be good or bad for end users and social welfare.

some actual redistribution is taking place. Merchants vocally object to increases in interchange fees, and website operators will do so if settlement charges are introduced in the Internet. End users would not react in this way if charges were passed through. There are three broad reasons why neutrality does not hold in practice:

a) *Transaction costs.* "Transaction costs" refer to a broad range of frictions that make it costly for one side of the market to pass through a redistribution of charges to the other side. Often, these transaction costs are associated with small stakes for individual transactions (which can become substantial when applied to a large number of transactions). The cost of thinking about including the pass-through, writing it into a contract, advertizing it to the other side and enforcing the covenant may then be prohibitive. For example, contractural relationships between a supplier, a buyer and their banks may not specify on which payment system the settlement of the transaction will occur.

   A second type of transaction cost has to do with the absence of a low-cost billing system. Suppose that an academic downloads a PDF file of another academic's paper. The micropayments that would be required for pass-through would probably require a costly third-party billing system to be developed cooperatively by Internet backbones and service providers.[33]

   A third transaction cost is the impossibility of monitoring and recording the actual transaction or interaction. In the portal and media example, neutrality would imply that when the platform (portal, TV network, newspaper) raises the price of advertizing, this price increase translates into a smaller amount of money given by the advertizer to the consumer for "listening to the ad." But "listening" is not easily measurable (except for the monitoring of clicks in the Internet, and even then it is impossible to measure whether the consumer pays genuine attention[34]). In practice therefore, the viewer/reader receives no compensation from the advertizer and neutrality does not obtain.

b) *Volume-insensitive costs.* Neutrality also fails when at least one side of the market incurs costs that a) are influenced by the platform and b) are not proportional to the number of transactions on the platform. For example, while software developers incur some costs, such as the per-game royalties paid by game developers, that are proportional to sales, many costs are insensitive and affected by the platform: The fixed

33. See Laffont et al. (2003) for a demonstration that termination charges are neutral in the Internet in the absence of the frictions considered in this section.

34. Such ways of charging consumers have been considered. For example, a startup called CyberGold devised a way to pay viewers of ads on the Web provided they peruse the Web ad to its last page. Advertisers were concerned about both moral hazard (clicking through ads without being really interested) and adverse selection (clickers would not be the high-demand consumers): See B. Ziegler's "Are Advertisers Ready to Pay Viewers," *Wall Street Journal,* 11/14/1996.

development cost is influenced by platform through software design, and so is the fixed charge for the development kit. On the user side, getting familiar with the platform's user interface may also involve some fixed costs.[35] End user transaction-insensitive prices and non-price attributes of a platform affect the number of end users or applications, but not directly the terms of the transactions between the end users.

The portal-and-media and real estate example offers another illustration of this phenomenon. The advertising cost of locating a potential buyer ought to be treated by the seller as a sunk cost when choosing the price to offer to the potentially interested buyer.

c) *Platform-determined constraints on pass-through.* The platform may also take steps that limit the extent of pass-through. A case in point is the no-discrimination-rule adopted by credit card associations (Visa, MasterCard) and for-profits (Amex).[36] Merchants do not pass the merchant discount through only to cardholders. So there is only a partial pass-through between the two sides.

These reasons, which have been embodied in our model, explain why markets with network externalities are predominantly two-sided markets.

Appendix

PROOF OF PROPOSITION 5

(i) Monopoly. Total volume is given by

$$Q = D^B(p^B, \theta) D^S(p^S, \theta).$$

We assume that D^B and D^S are (strictly) log concave with respect to prices, so that the first-order conditions are sufficient for the maximization of volume under a constant margin (case of an association) and for the maximization of profit (case of a for-profit monopoly).

(i1) Monopoly Association. The buyer price p^B induced by an association is characterized by

$$\varphi(p^B, \theta) = \lambda^B(p^B, \theta) - \lambda^S(c + m - p^B, \theta) = 0,$$

35. Similarly, end users seem to be averse to being "nickelled and dimed" by Internet portals (perhaps because they have a hard time thinking through the total amounts at stake) and flat fees are still quite popular in that industry.

36. In the United States, the associations' no-discrimination-rule takes a weaker form. Namely, merchants are not allowed to impose surcharges on card payments; but they can offer discounts for cash purchases! That very few do is an interesting fact, that is probably related to the transaction costs category. In Rochet and Tirole (2002), we abstract from such transaction costs and show that the level of the interchange fee is irrelevant if the no-discrimination rule is lifted.

where $\lambda^B(p^B, \theta) = -(D^B(p^B, \theta))'/(D^B(p^B, \theta))$ and $\lambda^S(p^S, \theta) = -(D^S(p^S, \theta))'/(D^S(p^S, \theta))$ denote the "sensitivities" of demands and $c + m$ is the (fixed) total price. We can apply the implicit function theorem to φ, given that $\partial\varphi/\partial p^B > 0$. This is because the strict log concavity of demands implies that sensitivities are increasing. Thus p^B is differentiable in θ and $dp^B/d\theta$ has the same sign as $-(\partial\varphi/\partial\theta)$.

We just have to compute $\partial\varphi/\partial\theta$ in our two examples:

Marquee buyers:

$$\varphi(p^B, \theta) = \lambda^B(p^B) - \lambda^S(c + m - \theta - p^B)$$

$$\frac{\partial\varphi}{\partial\theta} = (\lambda^S)' > 0 \quad \text{(since } D^S \text{ is log concave).}$$

Thus $dp^B/d\theta$ is negative.

Captive buyers:

$$\varphi(p^B, \theta) = \frac{-(D^B)'}{D^B + \theta} - \lambda^S(c + m - p^B)$$

$$\frac{\partial\varphi}{\partial\theta} = \frac{(D^B)'}{(D^B + \theta)^2} < 0.$$

And so $dp^B/d\theta$ is positive.

(i2) For-Profit Monopoly. The maximum of the (log) profit is characterized by two conditions:

$$\begin{cases} \lambda^B(p^B, \theta) - \dfrac{1}{p^B + p^S - c} & = 0 \\ \lambda^S(p^S, \theta) - \dfrac{1}{p^B + p^S - c} & = 0. \end{cases}$$

Denoting by $\varphi(p, \theta)$ the (vector) function on the left-hand side, we can apply the implicit function theorem (this time in $\mathbb{R}^2$) given that the Jacobian $D\varphi/Dp$ is nonsingular (by the strict concavity of $\log \pi$ the determinant of $D\varphi/Dp$ is positive). Thus we obtain:

$$\frac{dp}{d\theta} = -\left(\frac{D\varphi}{Dp}\right)^{-1} \frac{\partial\varphi}{\partial\theta},$$

where

$$\left(\frac{D\varphi}{Dp}\right)^{-1} = \frac{1}{\det \dfrac{D\varphi}{Dp}} \begin{bmatrix} \dfrac{\partial\lambda^S}{\partial p^S} + \dfrac{1}{(p-c)^2} & -\dfrac{1}{(p-c)^2} \\ -\dfrac{1}{(p-c)^2} & \dfrac{\partial\lambda^B}{\partial p^B} + \dfrac{1}{(p-c)^2} \end{bmatrix}$$

and

$$\frac{\partial \varphi}{\partial \theta} = \begin{pmatrix} \dfrac{\partial \lambda^B}{\partial \theta} \\ \dfrac{\partial \lambda^S}{\partial \theta} \end{pmatrix}.$$

Marquee Buyers: $\partial \lambda^B / \partial \theta = 0$, $\partial \lambda^S / \partial \theta < 0$

$$\begin{pmatrix} \dfrac{dp^B}{d\theta} \\ \dfrac{dp^S}{d\theta} \end{pmatrix} = \frac{-\dfrac{\partial \lambda^S}{\partial \theta}}{\det \dfrac{D\varphi}{Dp}} \begin{bmatrix} -\dfrac{1}{(p-c)^2} \\ \dfrac{\partial \lambda^B}{\partial p^B} + \dfrac{1}{(p-c)^2}. \end{bmatrix}.$$

Thus $dp^B/d\theta < 0$ and $dp^S/d\theta > 0$.

Captive Buyers: $\partial \lambda^B / \partial \theta > 0$, $\partial \lambda^S / \partial \theta = 0$, and so:

$$\begin{pmatrix} \dfrac{dp^B}{d\theta} \\ \dfrac{dp^S}{d\theta} \end{pmatrix} = \frac{-\dfrac{\partial \lambda^B}{\partial \theta}}{\det \dfrac{D\varphi}{Dp}} \begin{bmatrix} \dfrac{\partial \lambda^S}{\partial p^S} + \dfrac{1}{(p-c)^2} \\ -\dfrac{1}{(p-c)^2}. \end{bmatrix}$$

Thus $dp^B/d\theta > 0$ and $dp^S/d\theta < 0$.

(ii) Competing Associations. In the case of associations, the equilibrium buyer price is characterized by:

$$\lambda_0^B(p^B, \theta)\sigma(p^B, \theta) - \lambda^S(c + m - p^B, \theta) = 0,$$

where

$$\lambda_0^B(p^B, \theta) = \frac{-\dfrac{\partial d_1^B}{\partial p_1^B}}{d_1^B}(p^B, p^B, \theta)$$

is the "own-price sensitivity" of buyer demand and

$$\sigma(p^B, \theta) = 2 - \frac{\hat{D}^B(p^B, \theta)}{d^B(p^B, \theta)}.$$

In order to determine the monotonicity properties of p^B with respect to θ, we apply the implicit function theorem to the left-hand side of the above equation:

$$\psi(p^B, \theta) = \lambda_0^B(p^B, \theta)\sigma(p^B, \theta) - \lambda^S(c + m - p^B, \theta).$$

However we need additional assumptions to ensure that $\partial\psi/\partial p^B > 0$, so that $p^B(\theta)$ is indeed (locally) unique and differentiable, for two reasons:

- Possible nonexistence of equilibrium, due to the fact that the volume on system i is not necessarily quasiconcave with respect to (p_i^B, p_i^S). The proof of Proposition 6 will observe that the candidate for equilibrium (i.e., the solution of $\psi = 0$) may sometimes be destabilized by "double deviations" of the form $(p^B + \varepsilon, p^S - \varepsilon)$.
- The possible presence of strategic complementarities that may generate a multiplicity of equilibria.

We will assume away these difficulties and postulate that $\partial\psi/\partial p^B > 0$ (regularity condition).[37] In this case, $p^B(\theta)$ is (locally) unique, differentiable, and $dp^B/d\theta$ has the same sign as $-(\partial\psi/\partial\theta)$. We then just have to determine the sign of $\partial\psi/\partial\theta$.

Marquee Buyers:

$$\psi(p^B, \theta) = \lambda_0^B(p^B)\sigma(p^B) - \lambda^S(c + m - \theta - p^B)$$

$$\frac{\partial\psi}{\partial\theta} = (\lambda^S)' > 0.$$

Captive Buyers:

$$\psi(p^B, \theta) = \left(\frac{-\dfrac{\partial d_1^B}{\partial p_1^B}}{d_1^B + \theta}\right)\left(2 - \frac{D^B + \theta}{d^B + \theta}\right) - \lambda^S(c + m - p^B)$$

$$\frac{\partial\psi}{\partial\theta} = -\frac{\lambda_0^B\sigma}{d^B + \sigma} - \frac{d^B - D^B}{(d^B + \theta)^2}\lambda_0^B = \frac{-\lambda_0^B}{d^B + \theta}\left[\sigma + \frac{d^B - D^B}{d^B + \theta}\right] < 0.$$

An increase in the number of captive buyers has two opposite effects. First, and as in the monopoly case, the captive customers reduce the elasticity of buyer demand, calling for a higher buyer price. Second, captive customers make steering more attractive, which pushes toward a higher seller price. The first effect dominates the second.

(iii) Increase in Singlehoming. Again, we focus on competing associations. The buyer price at equilibrium is determined by:

$$\psi(p^B, \theta) = \lambda_0^B(p^B)\sigma^B(p^B, \theta) - \lambda^S(c + m - p^B) = 0.$$

37. This regularity condition is satisfied when $\partial\lambda_o^B/\partial p^B$ and $\partial\sigma/\partial p^B$ are positive.

By the same reasoning as above, $\partial\sigma^B/\partial\theta > 0$ implies $\partial\psi/\partial\theta > 0$ and $dp^B/d\theta > 0$. ■

PROOF OF PROPOSITION 6

(a) Price Structure. Letting $T = p_0 + \delta(\Delta + \delta)/2$, the quasi-demands are given by:[38]

$$d_1^B(p_1^B, p_2^B) = \frac{p_2^B - p_1^B}{\Delta} + \frac{T - p_1^B}{\delta}, \tag{25}$$

and

$$D_1^B(p_1^B) = (T - p_1^B)\left(\frac{1}{\delta} + \frac{1}{\Delta + \delta}\right). \tag{26}$$

The expressions of d_2^B and D_2^B are obtained by symmetry. Due to the linearity of these expressions, several simplifications appear. For example, the singlehoming index is price independent:

$$\sigma = 2 - \frac{D_1^B(p^B)}{d_1^B(p^B, p^B)} = \frac{\Delta}{\Delta + \delta}.$$

Similarly the expression of the marginal seller (who is indifferent between multihoming and singlehoming with the cheapest platform), does not depend on buyers' prices. For example, when $p_1^S \leq p_2^S$, formula (12) gives:

$$\hat{b}_{12} = p_2^S \frac{d_2^B}{d_1^B + d_2^B - D_1^B} + p_1^S \frac{d_1^B - D_1^B}{d_1^B + d_2^B - D_1^B}.$$

Hence (for $p_1^B = p_2^B = p^B$):

$$\hat{b}_{12} = p_2^S + \frac{\delta}{\Delta}(p_2^S - p_1^S).$$

38. The expressions of quasi-demands are easily deduced from the locations of marginal buyers:

- x_1 is indifferent between 1 and 1′: $p_1^B + ½(x_1 + (\Delta/2))^2 = p_0 + ½(x_1 + (\Delta/2) + \delta)^2$, which gives:

$$x_1 = \frac{p_1^B - p_0}{\delta} - \frac{\Delta + \delta}{2};$$

- x_2 is indifferent between 1 and 2: $p_1^B + ½(x_2 + (\Delta/2))^2 = p_2^B + ½(x_2 - (\Delta/2))^2$, which gives:

$$x_2 = \frac{p_2^B - p_1^B}{\Delta};$$

- x_3 is indifferent between 1 and 2′: $p_1^B + ½(x_3 + (\Delta/2))^2 = p_0 + ½(x_3 - (\Delta/2) - \delta)^2$, which gives:

$$x_3 = \frac{p_0 - p_1^B}{\Delta + \delta} + \delta/2.$$

and so $\hat{b}_{12}$ does not depend on p^B. Furthermore, steering is particularly powerful (in that undercutting induces many sellers to stop multihoming) when most consumers multihome, that is when σ is low.

Another simplification that appears when buyers' quasi-demand is linear is that the ratio of total elasticity to own-brand elasticity is equal to the single homing index σ:

$$\frac{\eta^B}{\eta_o^B} = \frac{\dfrac{\partial d_1^B}{\partial p_1^B} + \dfrac{\partial d_1^B}{\partial p_2^B}}{\dfrac{\partial d_1^B}{\partial p_1^B}} = \frac{\dfrac{1}{\delta}}{\dfrac{1}{\Delta} + \dfrac{1}{\delta}} = \frac{\Delta}{\Delta + \delta} = \sigma.$$

This property implies that the price structure under platform competition (between for-profits or between associations) is the same as under a monopoly platform:

$$\frac{p^B}{\eta^B} = \frac{p^S}{\eta^S}.$$

Consider for example a decrease in Δ. As the platforms become more substitutable, buyer multihoming increases (σ falls); this induces platforms to steer, resulting in low prices on the seller side. However, competition also becomes more intense on the buyer side, resulting in lower buyer prices (p^B falls) and thereby in a higher opportunity cost ($c - p^B$) of servicing sellers. For linear demand on the buyer side, these two effects offset.

Last, let us compare the common price structure with that of the Ramsey optimum. A useful property of linear demands is that the revenue (price times quantity) is equal to twice the product of the net surplus and the elasticity of demand. This property implies that if seller's quasi-demand is linear as well, (7) is equivalent to (5), and so the common price structure is Ramsey optimal.

(b) Second-Order Conditions. In the Hotelling model:

$$\pi_i = (p_i^B + p_i^S - c)Q_i;$$

$$\frac{\partial \pi_i}{\partial p_i^B} = Q_i + (p_i^B + p_i^S - c)\frac{\partial Q_i}{\partial p_i^B};$$

$$\frac{\partial \pi_i}{\partial p_i^S} = Q_i + (p_i^B + p_i^S - c)\frac{\partial Q_i}{\partial p_i^S};$$

$$\frac{\partial^2 \pi_i}{(\partial p_i^B)^2} = 2\frac{\partial Q_i}{\partial p_i^B}; \qquad \frac{\partial^2 \pi_i}{(\partial p_i^S)^2} = 2\frac{\partial Q_i}{\partial p_i^S};$$

$$\frac{\partial^2 \pi_i}{\partial p_i^B \partial p_i^S} = \frac{\partial Q_i}{\partial p_i^B} + \frac{\partial Q_i}{\partial p_i^S} + (p_i^B + p_i^S - c)\frac{\partial^2 Q_i}{\partial p_i^B \partial p_i^S}.$$

At a symmetric equilibrium of the game between competing proprietary platforms, we have

$$\frac{\partial Q_i}{\partial p_i^B} = \frac{\partial Q_i}{\partial p_i^S} = -(p - c)\frac{(\Delta + \delta)^2}{\Delta^2 \delta} \equiv \alpha < 0.$$

Therefore the second-order condition is satisfied whenever the Hessian determinant of π_i is nonnegative:

$$H = \frac{\partial^2 \pi_i}{(\partial p_i^B)^2} \cdot \frac{\partial^2 \pi_i}{(\partial p_i^S)^2} - \left(\frac{\partial^2 \pi_i}{\partial p_i^B \partial p_i^S}\right)^2$$

$$H = 4\alpha^2 - (2\alpha + \beta)^2 = -\beta(4\alpha + \beta),$$

where

$$\beta \equiv (p_i^B + p_i^S - c)\frac{\partial^2 Q_i}{\partial p_i^B \partial p_i^S}$$

has a different expression in the two regions:

$$\beta_1 = (p - c)\frac{(\Delta^2 + \delta\Delta - \delta^2)}{\Delta^2 \delta} \qquad \text{when } p_1^S < p_2^S, \qquad \text{and}$$

$$\beta_2 = (p - c)\frac{(\Delta + \delta)^2}{\Delta^2 \delta} \qquad \text{when } p_1^S > p_2^S.$$

The second-order condition is always satisfied in the second region, since $\beta_2 = -\alpha > 0$ so that $H = 3\alpha^2 > 0$. In the first region, it is easy to see that $\beta_1 + 4\alpha$ is always negative. Thus the second-order condition is satisfied if and only if $\beta_1 \geq 0$, which is equivalent to

$$\delta^2 - \delta\Delta - \Delta^2 \leq 0$$

or

$$\frac{\delta}{\Delta} \leq \frac{1 + \sqrt{5}}{2}.$$

When this condition is not satisfied, there is no symmetric equilibrium in pure strategies. The only candidate equilibrium (p^B, p^S) can be destabilized by a "double-deviation," where one of the platforms (say Platform 1) increases p_1^B by ε and simultaneously decreases p_1^S by the same amount. The first order increase in profit is zero (as guaranteed by the first-order conditions) but the second-order increase is positive:

$$\Delta\pi_1 \sim \left[\frac{\partial^2\pi_1}{(\partial p_1^B)^2} + \frac{\partial^2\pi_1}{(\partial p_1^S)^2} - 2\,\frac{\partial^2\pi_1}{\partial p_1^B \partial p_1^S}\right]\varepsilon^2 = -2\beta_1\varepsilon^2 > 0.$$

Finally, equilibrium prices can be obtained explicitly if we assume that the sellers' quasi-demand is also linear:

$$D^S(p^S) = A - p^S. \tag{27}$$

The volume on Platform 1 when $p_1^S \leq p_2^S$ is:

$$\begin{aligned} Q_1 = {} & \left(\frac{p_2^B - p_1^B}{\Delta} + \frac{T - p_1^B}{\delta}\right)\left[A - p_2^S - \frac{\delta}{\Delta}(p_2^S - p_1^S)\right] \\ & + \left(\frac{1}{\delta} + \frac{1}{\Delta + \delta}\right)(T - p_1^B)(p_2^B - p_1^S)\left(1 + \frac{\delta}{\Delta}\right). \end{aligned} \tag{28}$$

When $p_1^S > p_2^S$, the expression is simpler:

$$Q_1 = \left[A - p_1^S - \frac{\delta}{\Delta}(p_1^S - p_2^S)\right]\left[\frac{p_2^B - p_1^B}{\Delta} + \frac{T - p_1^B}{\delta}\right]. \tag{29}$$

In Proposition 2 we have shown that a symmetric equilibrium between competing associations must satisfy condition (22):

$$\frac{p^B}{\sigma\eta_o^B} = \frac{p^S}{\eta^S}. \tag{22}$$

Using formulae (25), (26), and (27) and after simplifications, this condition becomes:

$$p^B - p^S = T - A. \tag{30}$$

Recall that this condition is necessarily satisfied in a symmetric equilibrium between competing platforms, independently of their governance structure. However the value of the total price is different:

$$p^B + p^S = c + m$$

for associations, and

$$p^B + p^S - c = \frac{\Delta}{\Delta + \delta}(T - p^B) = \frac{\Delta}{\Delta + \delta}(A - p^S).$$

for proprietary platforms. The resulting equilibrium prices are:

$$p_A^B = \tfrac{1}{2}(c + m + T - A), \qquad p_A^S = \tfrac{1}{2}(c + m - T + A), \tag{31}$$

for associations, and

$$p_P^B = \frac{c - A + T\left(1 + \dfrac{\Delta}{\Delta + \delta}\right)}{2 + \dfrac{\Delta}{\Delta + \delta}}, \qquad p_P^S = \frac{c - T + A\left(1 + \dfrac{\Delta}{\Delta + \delta}\right)}{2 + \dfrac{\Delta}{\Delta + \delta}} \tag{32}$$

for proprietary systems.

References

Armstrong, M. (2002). "Competition in Two-Sided Markets." Presentation at Econometric Society European meeting (ESEM) meeting, Venice, August.

Baumol, W. J., J. Panzar, and R. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace, Jovanovich.

Baye M., and J. Morgan (2001). "Information Gatekeepers on the Internet and the Competitiveness of Homogenous Product Markets." *American Economic Review,* 91, pp. 454–474.

Caillaud, B. and B. Jullien (2003). "Chicken and Egg; Competing Matchmakers." *Rand Journal of Economics,* 34(2), pp. 309–328.

Cowell, A. (2001). "Real Networks Skins Apple, Microsoft." ⟨http://www.networkcomputing.com/1204/1204f1b1.html⟩ (February 1).

Cringely, R. (2001). "The Plumber is Back: How Nintendo and Microscoft Are about to Open a Case of Whoop-Ass on Sony's Video Game Empire." ⟨http://www.pbs.org/cringely/pulpit/pulpit 20010322.html⟩ (March 22).

Elfenbein, D., and J. Lerner (2003). "Links and Hyperlinks: An Empirical Analysis of Internet Portal Alliances, 1995–1999," *Rand Journal of Economics*, 34(2), pp. 356–369.

Evans, D. (2003). "The Antitrust Economics of Two-Sided Markets." *Yale Journal of Regulation,* forthcoming.

Farrell, J. and G. Saloner (1985). "Standardization, Compatibility, and Innovation." *Rand Journal of Economics,* 16, pp. 70–83.

Farrell, J. and G. Saloner (1986). "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation." *American Economic Review,* 76, pp. 940–955.

Hausman, J., G. K. Leonard, and Jean Tirole (2003). "On Nonexclusive Membership in Competing Joint Venture." *Rand Journal of Economics,* 34(1), pp. 43–62.

Jeon, D. S., J. J. Laffont and Jean Tirole (2003). "On the Receiver Pays Principle," *Rand Journal of Economics*, forthcoming.

Jullien, B. (2001). "Competing with Network Externalities and Price Competition," mimeo, Institut D'economie Industrielle (IDEI), Toulouse, France.

Katz, M., and C. Shapiro (1985). "Network Externalities, Competition, and Compatibility." *American Economic Review,* 75, pp. 424–440.

Katz, M., and C. Shapiro (1986). "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy,* 94, pp. 822–841.

Laffont, J. J., S. Marcus, P. Rey, and Jean Tirole (2003). "Internet Interconnection and the Off-Net-Cost Pricing Principle," *Rand Journal of Economics,* 34(2), pp. 370–390.

Parker, G. and M. Van Alstyne (2000). "Information Complements, Substitutes, and Strategic Product Design," mimeo. New Orleans, Louisiana: Tulane University and University of Michigan.

Rochet, Jean-Charles, and Jean Tirole (2002). "Cooperation among Competitors: The Economics of Payment Card Associations." *Rand Journal of Economics,* 33(4), pp. 1–22.

Rohlfs, J. (1974). "A Theory of Interdependent Demand for a Communications Service." *Bell Journal of Economics,* 5(1), pp. 16–37.

Rysman, M. (2000). "Competition between Networks: A Study of the Market for Yellow Pages," IPS DP 104, Boston University.

Schmalensee, R. (2002). "Payment Systems and Interchange Fees." *Journal of Industrial Economics,* 50, pp. 103–122.

Stahl, D. (1988). "Bertrand Competition for Inputs and Walrasian Outcomes," *American Economic Review,* 78, pp. 189–201.

Wilson, R. (1993). *Nonlinear Pricing*. New York: Oxford University Press.

Yanelle, M. O. (1997). "Banking Competition and Market Efficiency." *The Review of Economic Studies,* 64(219), pp. 215–240.

Ziegler, B. (1996). "Are Advertisers Ready to Pay Viewers?" *Wall Street Journal,* November 14.

Mathematical Appendix

Ramsey Pricing

PROOF OF THEOREM 1:

I maximize social welfare V subject to $\pi > K$. The Lagrangian is

$$V + \kappa(\pi - K)$$

the associated first-order conditions are that for each I

$$(1+\kappa)\left(P^{\mathcal{I}} - C^{\mathcal{I}} - cN^{\mathcal{J}}\right) - \kappa\mu^{\mathcal{I}} + \left(\overline{b^{\mathcal{J}}} + \kappa\widetilde{b^{\mathcal{J}}}\right)N^{\mathcal{J}} = 0$$

or

$$P^{\mathcal{I}} - C^{\mathcal{I}} - cN^{\mathcal{J}} - \left(\frac{\kappa}{1+\kappa}\right)\mu^{\mathcal{I}} + \left(\left[\frac{1}{1+\kappa}\right]\overline{b^{\mathcal{J}}} + \left[\frac{\kappa}{1+\kappa}\right]\widetilde{b^{\mathcal{J}}}\right)N^{\mathcal{J}} = 0$$

or letting $\lambda \equiv \frac{\kappa}{1+\kappa}$

$$P^{\mathcal{I}} = C^{\mathcal{I}} + cN^{\mathcal{J}} + \lambda\mu^{\mathcal{I}} - \left([1-\lambda]\overline{b^{\mathcal{J}}} + \lambda\widetilde{b^{\mathcal{J}}}\right)N^{\mathcal{J}}$$

Then solve for λ by substituting into the desired profit condition $\pi = K$:

$$\lambda\left(\mu^{\mathcal{A}}N^{\mathcal{A}} + \mu^{\mathcal{B}}N^{\mathcal{B}} + \left[\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{B}}} - \widetilde{b^{\mathcal{A}}} - \widetilde{b^{\mathcal{B}}}\right]N^{\mathcal{A}}N^{\mathcal{B}}\right) = K + \left(\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{B}}} - c\right)N^{\mathcal{A}}N^{\mathcal{B}}$$

I now turn to the two alternative forms of the Ramsey problem mentioned in the text.

PROPOSITION 8: *Interior Ramsey prices maximizing user surplus subject to the constraint that the firm makes a rate of return r on her variable costs must solve*

$$\text{(A 1)} \qquad P^{\mathcal{I}} = \underbrace{(1+r)\left(C^{\mathcal{I}} + cN^{\mathcal{J}}\right)}_{\textit{cost-plus}} - \underbrace{N^{\mathcal{J}}\left(\lambda\widetilde{b^{\mathcal{J}}} + [1-\lambda]\overline{b^{\mathcal{J}}}\right) + \lambda\mu^{\mathcal{I}}}_{\textit{two-sided classical Ramsey pricing (Theorem 1}}$$

$$\textit{where } \lambda \equiv \frac{\overbrace{\left(\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{B}}} - [1+r]\,c\right)N^{\mathcal{A}}N^{\mathcal{B}}}^{\textit{subsidy (in cost-plus terms) required for Pigouvian prices}}}{\underbrace{N^{\mathcal{A}}\mu^{\mathcal{A}} + N^{\mathcal{B}}\mu^{\mathcal{B}} + \left(\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{A}}} - \widetilde{b^{\mathcal{A}}} - \widetilde{b^{\mathcal{B}}}\right)N^{\mathcal{A}}N^{\mathcal{B}}}_{\textit{demoninator from Theorem 1}}}$$

PROOF:

The Lagrangian is

$$V - (1-\kappa)\pi - r\kappa\left(cN^{\mathcal{A}}N^{\mathcal{B}} + C^{\mathcal{A}}N^{\mathcal{A}} + C^{\mathcal{B}}N^{\mathcal{B}}\right)$$

with corresponding first-order conditions

$$\kappa P^{\mathcal{I}} - \kappa(1+r)(C^{\mathcal{I}} + cN^{\mathcal{J}}) + (1-\kappa)\mu^{\mathcal{I}} + \left(\overline{b^{\mathcal{J}}} - [1-\kappa]\widetilde{b^{\mathcal{J}}}\right)N^{\mathcal{J}} = 0$$

or letting $\lambda \equiv \left(1 - \frac{1}{\kappa}\right)$

$$P^{\mathcal{I}} = (1+r)\left(C^{\mathcal{I}} + cN^{\mathcal{J}}\right) + \lambda\mu^{\mathcal{I}} - N^{\mathcal{J}}\left(\lambda\widetilde{b^{\mathcal{J}}} + [1-\lambda]\overline{b^{\mathcal{J}}}\right)$$

Plugging into the rate-of-return constraint $\pi = rC$ yields

$$\alpha\left(\mu^{\mathcal{A}}N^{\mathcal{A}} + \mu^{\mathcal{B}}N^{\mathcal{B}} + \left[\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{B}}} - \widetilde{b^{\mathcal{A}}} - \widetilde{b^{\mathcal{B}}}\right]\right) = \left(\overline{b^{\mathcal{A}}} + \overline{b^{\mathcal{B}}} - [1+r]c\right)N^{\mathcal{A}}N^{\mathcal{B}}$$

which yields the desired solution.

Thus user Ramsey pricing with a required rate of return over variable costs is a natural combination of classical cost-plus regulation and the classical two-sided Ramsey pricing of Theorem 1. Prices mark-up costs by the required rate of return, incorporating two-sided internalization of cross-benefits and market power in the weighted-average fashion of classical Ramsey pricing above. Now that there is no required profit threshold, the weighting is given the amount necessary to offset the subsidy needed to achieve cost-plus Pigouvian pricing while maintaining the required rate of return. As noted earlier, if $r = 0$ and the platform must simply break even, the solution is identical to the break-even classical OT pricing.

PROPOSITION 9: *Interior Ramsey prices maximizing social surplus subject to the constraint that the firm makes a rate of return r on her variable costs must solve*

$$P^{\mathcal{I}} = \underbrace{(1+\lambda r)(C^{\mathcal{I}} + cN^{\mathcal{J}})}_{\textit{cost-plus}} - \underbrace{N^{\mathcal{J}}\left(\lambda\overline{b^{\mathcal{J}}} + [1-\lambda]\widetilde{b^{\mathcal{J}}}\right) + \lambda\mu^{\mathcal{I}}}_{\textit{two-sided classical Ramsey pricing (Theorem 1)}} \tag{A 2}$$

$$\textit{where } \lambda \equiv \underbrace{\frac{\overbrace{r\left(C^{\mathcal{A}}N^{\mathcal{A}}+C^{\mathcal{B}}N^{\mathcal{B}}+cN^{\mathcal{A}}N^{\mathcal{B}}\right)}^{\textit{target profits}}+\overbrace{\left(\overline{b^{\mathcal{A}}}+\overline{b^{\mathcal{B}}}-c\right)N^{\mathcal{A}}N^{\mathcal{B}}}^{\textit{Pigouvian subsidy}}}{\left(rC^{\mathcal{A}}+\mu^{\mathcal{A}}\right)N^{\mathcal{A}}+\left(rC^{\mathcal{B}}+\mu^{\mathcal{B}}\right)N^{\mathcal{B}}+\left(\overline{b^{\mathcal{A}}}+\overline{b^{\mathcal{B}}}-\widetilde{b^{\mathcal{A}}}-\widetilde{b^{\mathcal{B}}}+2rc\right)N^{\mathcal{A}}N^{\mathcal{B}}}}_{\textit{(local) profit gain from moving towards rate-of-return maximization}}$$

PROOF:

The Lagrangian is now

$$V+\kappa\left(\pi-r\left[C^{\mathcal{A}}N^{\mathcal{A}}+C^{\mathcal{B}}N^{\mathcal{B}}+cN^{\mathcal{A}}N^{\mathcal{B}}\right]\right)$$

giving first-order conditions

$$(1+\kappa)P^{\mathcal{I}}-(1+\kappa[1+r])\,C^{\mathcal{I}}-\kappa\mu^{\mathcal{I}}+\left(\overline{b^{\mathcal{I}}}+\kappa\widetilde{b^{\mathcal{I}}}\right)N^{\mathcal{J}}-(1+\kappa[1+r])\,cN^{\mathcal{J}}$$

Letting $\lambda \equiv \frac{\kappa}{1+\kappa}$ this takes the form in the text. Plugging this back into $\left(P^{\mathcal{A}}-[1+r]C^{\mathcal{A}}\right)N^{\mathcal{A}}+\left(P^{\mathcal{B}}-[1+r]C^{\mathcal{B}}\right)N^{\mathcal{B}}-(1+r)cN^{\mathcal{A}}N^{\mathcal{B}}=0$ yields the definition for λ.

The social Ramsey pricing problem with a required rate of return over variable costs is a slight modification of user-optimal Ramsey pricing in Proposition 9.

Generalization

PROOF OF THEOREM 2:

I compute $V^{\mathcal{I}}_{\mathcal{J}}$. By definition

$$\begin{aligned} V^{\mathcal{I}}(\mathbf{N})) &= \int_{\theta^{\mathcal{I}}\in\overline{\Theta^{\mathcal{I}}}(\mathbf{N})} u^{\mathcal{I}}\left(\mathbf{N};\theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}} \\ &= \int_{\theta^{\mathcal{I}}:u^{\mathcal{I}}\left(\mathbf{N};\theta^{\mathcal{I}}\right)\geq P^{\mathcal{I}}(\mathbf{N})} u^{\mathcal{I}}\left(\mathbf{N};\theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}} \end{aligned}$$

Because $\widetilde{\Theta^{\mathcal{I}}}$ is the boundary of $\overline{\Theta^{\mathcal{I}}}$, by the Leibnitz Integral Rule

$$V^{\mathcal{I}}_{\mathcal{J}}=\int_{\theta^{\mathcal{I}}\in\overline{\Theta^{\mathcal{I}}}} u^{\mathcal{I}}_{\mathcal{J}}\left(\mathbf{N};\theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}}+$$
$$\int_{\theta^{\mathcal{I}}\in\widetilde{\Theta^{\mathcal{I}}}}\left(u^{\mathcal{I}}_{\mathcal{J}}\left[\mathbf{N};\theta^{\mathcal{I}}\right]-P^{\mathcal{I}}_{\mathcal{J}}[\mathbf{N}]\right)u^{\mathcal{I}}\left(\mathbf{N};\theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}}=$$

$$\overline{u^{\mathcal{I}}_{\mathcal{J}}}N^{\mathcal{J}}+P^{\mathcal{I}}\int_{\theta^{\mathcal{I}}\in\widetilde{\Theta^{\mathcal{I}}}}\left(u^{\mathcal{I}}_{\mathcal{J}}\left[\mathbf{N};\theta^{\mathcal{I}}\right]-P^{\mathcal{I}}_{\mathcal{J}}[\mathbf{N}]\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}} \tag{A 3}$$

By the implicit function theorem for $\mathcal{J}\neq\mathcal{I}$ and letting $\widetilde{N^{\mathcal{I}}_{1}}\equiv\frac{\partial\widetilde{N^{\mathcal{I}}}}{\partial P^{\mathcal{I}}}$

$$P^{\mathcal{I}}_{\mathcal{J}}=-\frac{\widetilde{N^{\mathcal{I}}_{\mathcal{J}}}}{\widetilde{N^{\mathcal{I}}_{1}}}\underbrace{=}_{\text{The Leibnitz Integral Rule}}\frac{\int_{\theta^{\mathcal{I}}\in\widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} u^{\mathcal{I}}_{\mathcal{J}}\left(\mathbf{N};\theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}}}{\int_{\theta^{\mathcal{I}}\in\widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right)d\theta^{\mathcal{I}}}$$

Thus term second term in A 3 vanishes for $\mathcal{I} \neq \mathcal{J}$. On other other hand for $\mathcal{I} = \mathcal{J}$ by the implicit function theorem

$$P_{\mathcal{I}}^{\mathcal{I}} = \frac{\widetilde{N_{\mathcal{J}}^{\mathcal{I}}} - 1}{\widetilde{N_{1}^{\mathcal{I}}}} = \frac{\int_{\theta^{\mathcal{I}} \in \widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} u_{\mathcal{J}}^{\mathcal{I}}\left(\mathbf{N}; \theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right) d\theta^{\mathcal{I}} - 1}{\int_{\theta^{\mathcal{I}} \in \widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right) d\theta^{\mathcal{I}}}$$

So the second term of A 3 is simply $P^{\mathcal{I}}$ (as the gross utility of all users in $\widetilde{\Theta^{\mathcal{I}}}$) is $P^{\mathcal{I}}$. Thus the first-order condition that marginal social benefit equal marginal cost becomes

$$\sum_{\mathcal{J}} V_{\mathcal{J}}^{\mathcal{I}} = C_{\mathcal{I}} \iff$$

$$P^{\mathcal{I}} = C_{\mathcal{I}} - \sum_{\mathcal{J}} \overline{u_{\mathcal{I}}^{\mathcal{J}}} N^{\mathcal{J}}$$

which is the Pigouvian condition reported in the text.

PROOF OF THEOREM 3:

Here I begin by computing marginal revenues. For $\mathcal{J} \neq I$:

$$R_{\mathcal{J}}^{\mathcal{I}} = \frac{\partial P^{\mathcal{I}}(\mathbf{N}) N^{\mathcal{I}}}{\partial N^{\mathcal{J}}} = P_{\mathcal{J}}^{\mathcal{I}} N^{\mathcal{I}} = \frac{\int_{\theta^{\mathcal{I}} \in \widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} u_{\mathcal{J}}^{\mathcal{I}}\left(\mathbf{N}; \theta^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right) d\theta^{\mathcal{I}}}{\int_{\theta^{\mathcal{I}} \in \widetilde{\Theta^{\mathcal{I}}}(\mathbf{N})} f^{\mathcal{I}}\left(\theta^{\mathcal{I}}\right) d\theta^{\mathcal{I}}} N^{\mathcal{J}} = \widetilde{u_{\mathcal{J}}^{\mathcal{I}}} N^{\mathcal{J}}$$

while

$$R_{\mathcal{I}}^{\mathcal{I}} = P^{\mathcal{I}} + P_{\mathcal{I}}^{\mathcal{I}} N^{\mathcal{I}} = P^{\mathcal{I}} + \widetilde{u_{\mathcal{I}}^{\mathcal{I}}} N^{\mathcal{I}} - \frac{P^{\mathcal{I}}}{\epsilon^{\mathcal{I}}}$$

where $\epsilon^{\mathcal{I}} \equiv \frac{\partial \widetilde{N^{\mathcal{I}}}}{\partial P^{\mathcal{I}}} \frac{P^{\mathcal{I}}}{\widetilde{N^{\mathcal{I}}}}$ is the price elasticity of participation (holding fixed anticipated participation). As usual, let $\mu^{\mathcal{I}} \equiv \frac{P^{\mathcal{I}}}{\epsilon^{\mathcal{I}}}$ be the *market power*, that is the difference between price and classical marginal revenue. Then equating marginal revenue to marginal cost requires

$$P^{\mathcal{I}} - \mu^{\mathcal{I}} + \sum_{\mathcal{J}} \widetilde{u_{\mathcal{I}}^{\mathcal{J}}} N^{\mathcal{J}} = C_{\mathcal{I}}$$

which can easily be re-arranged to yield either of the formulae in the theorem.

Comparative Statics

I begin by showing that $\rho^{\mathcal{I}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}} \frac{\partial^2 \pi}{\partial N^{\mathcal{I}2}}}$ and that local strict concavity of profits is equivalent to the local version of 2SC, while violation of local weak 2SC implies local non-concavity. Then

taking these facts as given, I establish Theorem 4.

$$\rho^{\mathcal{I}} \equiv \left.\frac{dP^{\mathcal{I}}}{dC^{\mathcal{I}}}\right|_{N^{\mathcal{J}}} = P_1^{\mathcal{I}} \left.\frac{dN^{\mathcal{I}}}{dC^{\mathcal{I}}}\right|_{N^{\mathcal{J}}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}} \frac{\partial^2 \pi}{\partial N^{\mathcal{I}2}}}$$

Thus $\rho^{\mathcal{I}} > (\geq) 0 \iff \frac{\partial^2 \pi^2}{\partial N^{\mathcal{I}2}} < (\leq) 0$ and

$$\frac{\partial^2 \pi}{\partial N^{\mathcal{A}2}} \frac{\partial^2 \pi}{\partial N^{\mathcal{B}2}} > (\geq) \frac{\partial^2 \pi^2}{\partial N^{\mathcal{A}} \partial N^{\mathcal{B}}} \iff \frac{\mu^{\mathcal{A}} \mu^{\mathcal{B}}}{N^{\mathcal{A}} N^{\mathcal{B}}} > (\geq) \rho^{\mathcal{A}} \rho^{\mathcal{B}} \chi^2$$

which shows that strict local 2SC implies local concavity (by the classical Hessian test) and weak 2SC is necessary for local concavity. Based on these, Theorem 4 is easy to establish.

Two-sided contraction

PROOF OF THEOREM 4:

The forward direction of the proof follows directly from above: global concavity of profits (which is equivalent to the negative definiteness of profits, which is 2SC) in $\langle N^{\mathcal{A}}, N^{\mathcal{B}} \rangle$ is independent of $\langle C^{\mathcal{A}}, C^{\mathcal{B}} \rangle$ by inspection. For the converse, suppose that 2SC were violated at some pair of participation rates $\langle \widetilde{N^{\mathcal{A}}}, \widetilde{N^{\mathcal{B}}} \rangle$. Let

$$\widetilde{C^{\mathcal{I}}} \equiv P^{\mathcal{I}} - \mu^{\mathcal{I}} + \widetilde{b^{\mathcal{J}}} N^{\mathcal{J}} - c N^{\mathcal{J}}$$

be evaluated at $\langle \widetilde{N^{\mathcal{A}}}, \widetilde{N^{\mathcal{B}}} \rangle$. Then clearly equation (2) given costs $\langle \widetilde{C^{\mathcal{A}}}, \widetilde{C^{\mathcal{B}}} \rangle$ has a solution at $\langle \widetilde{N^{\mathcal{A}}}, \widetilde{N^{\mathcal{B}}} \rangle$. However because 2SC is violated here, this must be a local minimum or saddle-point of the firm's profits. Therefore it cannot be the optimal participation pair.

In the Armstrong case, from equation (12) $\chi = b^{\mathcal{A}} + b^{\mathcal{B}}$ so 2SC becomes

$$\frac{\mu^{\mathcal{A}} \mu^{\mathcal{B}}}{N^{\mathcal{A}} N^{\mathcal{B}}} > \rho^{\mathcal{A}} \rho^{\mathcal{B}} \left(b^{\mathcal{A}} + b^{\mathcal{B}} \right)^2$$

In the special case of linear demand that is typically analyzed, $\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}} = \frac{1}{\phi^{\mathcal{I}}}$ where $\phi^{\mathcal{I}}$ is the (constant) slope of demand. If we parameterize demand by $N^{\mathcal{I}} \left(P^{\mathcal{I}}, N^{\mathcal{J}} \right) = \frac{Q^{\mathcal{I}} \left(2\widetilde{P^{\mathcal{I}}} - P^{\mathcal{I}} - b^{\mathcal{I}} N^{\mathcal{J}} \right)}{2\widetilde{P^{\mathcal{I}}}}$ so that $Q^{\mathcal{I}}$ demand at price 0 and $\widetilde{P^{\mathcal{I}}}$ is the optimal price when there is no two-sidedness and

$C^{\mathcal{I}} = 0$ then $\frac{1}{\sigma^{\mathcal{I}}} = \frac{Q^{\mathcal{I}}}{2\widetilde{P^{\mathcal{I}}}}$ so the condition simplifies, after a bit of algebra, to

$$\frac{\left(b^{\mathcal{A}} + b^{\mathcal{B}}\right) Q^{\mathcal{A}} Q^{\mathcal{B}}}{\sqrt{\frac{Q^{\mathcal{A}} Q^{\mathcal{B}} \widetilde{P^{\mathcal{A}}} \widetilde{P^{\mathcal{B}}}}{4}}} < 8$$

which is that the surplus generated by two-sidedness at zero price should not be more than 8 times the geometric mean of optimal profits on the two sides if there were no two-sidedness, as optimal profits for linear demand with this parameterization are $\frac{Q^{\mathcal{I}} \widetilde{P^{\mathcal{I}}}}{2}$.

Pigouvian second-order conditions

Now I derive the second-order conditions for the Pigouvian problem. The first derivative of social welfare V with respect to participation on side $\mathcal{I}$ is

$$P^{\mathcal{I}} - C^{\mathcal{I}} - cN^{\mathcal{J}} + \overline{b^{\mathcal{J}}} N^{\mathcal{J}}$$

Taking the derivative with respect to $N^{\mathcal{I}}$ yields

$$P_1^{\mathcal{I}} + \overline{b^{\mathcal{J}}}_2 N^{\mathcal{J}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}} + \frac{\partial \int_{-\infty}^{\infty} \int_{P^{\mathcal{J}}(N^{\mathcal{J}}, N^{\mathcal{I}}) - b^{\mathcal{J}} N^{\mathcal{I}}}^{\infty} b^{\mathcal{J}} f^{\mathcal{J}}(B^{\mathcal{J}}, b^{\mathcal{J}}) dB^{\mathcal{J}} db^{\mathcal{J}}}{\partial N^{\mathcal{I}}} =$$

$$-\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}} + \int_{-\infty}^{\infty} \left(b^{\mathcal{J}} - \widetilde{b^{\mathcal{J}}}\right) b^{\mathcal{J}} f^{\mathcal{I}} \left(P^{\mathcal{J}}[N^{\mathcal{J}}, N^{\mathcal{I}}] - b^{\mathcal{J}} N^{\mathcal{I}}, b^{\mathcal{J}}\right) db^{\mathcal{J}}$$

$$= -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}} + \frac{N^{\mathcal{J}}}{\mu^{\mathcal{J}}} \left(\widetilde{b^{\mathcal{J}^2}} - \widetilde{b^{\mathcal{J}}}^2\right) = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}} + \frac{N^{\mathcal{J}}}{\mu^{\mathcal{J}}} \left(\widetilde{b^{\mathcal{J}^2}} - \widetilde{b^{\mathcal{J}}}^2\right)$$

Thus the own-second derivative condition for concavity is that on both sides I

$$\frac{\mu^{\mathcal{A}} \mu^{\mathcal{B}}}{N^{\mathcal{A}} N^{\mathcal{B}}} > \widetilde{\sigma^{\mathcal{I}}}$$

where $\widetilde{\sigma^{\mathcal{I}}}$ is the variance of interaction benefits among marginal users on side $\mathcal{I}$. The cross partial is then

$$P_2^{\mathcal{A}} - c + \frac{\partial \widetilde{b^{\mathcal{B}}} N^{\mathcal{B}}}{\partial N^{\mathcal{B}}} = \widetilde{b^{\mathcal{A}}} - c + \frac{\partial \int_{-\infty}^{\infty} \int_{P^{\mathcal{B}}(N^{\mathcal{B}}, N^{\mathcal{A}}) - b^{\mathcal{B}} N^{\mathcal{A}}}^{\infty} b^{\mathcal{B}} f^{\mathcal{B}}(B^{\mathcal{B}}, b^{\mathcal{B}}) dB^{\mathcal{B}} db^{\mathcal{B}}}{\partial N^{\mathcal{B}}} =$$

$$\widetilde{b^{\mathcal{A}}} - c + \frac{\mu^{\mathcal{B}} \int_{-\infty}^{\infty} b^{\mathcal{B}} f^{\mathcal{I}} \left(P^{\mathcal{B}}[N^{\mathcal{B}}, N^{\mathcal{A}}] - b^{\mathcal{B}} N^{\mathcal{A}}, b^{\mathcal{B}}\right) db^{\mathcal{B}}}{N^{\mathcal{B}}} = \widetilde{b^{\mathcal{A}}} + \widetilde{b^{\mathcal{B}}} - c$$

So the positive determinant condition requires

$$\frac{\mu^{\mathcal{A}}\mu^{\mathcal{B}}}{N^{\mathcal{A}}N^{\mathcal{B}}} > \frac{\left(\widetilde{b^{\mathcal{A}}} + \widetilde{b^{\mathcal{B}}} - c\right)^2}{\left(\frac{N^{\mathcal{A}}N^{\mathcal{B}}\widetilde{\sigma^{\mathcal{A}}}}{\mu^{\mathcal{A}}\mu^{\mathcal{B}}} - 1\right)\left(\frac{N^{\mathcal{A}}N^{\mathcal{B}}\widetilde{\sigma^{\mathcal{B}}}}{\mu^{\mathcal{A}}\mu^{\mathcal{B}}} - 1\right)} \tag{A 4}$$

If imposed globally, just as with 2SC, these ensure sufficiency of first-order conditions. By the same argument as in Section III for 2SC, these conditions are weak in the sense of Theorem 4. In the RT2003 case, heterogeneity is unidimensional so all marginal users have the same preferences and $\widetilde{\sigma^{\mathcal{A}}} = \widetilde{\sigma^{\mathcal{B}}} = 0$. Therefore inequality A 4 becomes

$$m^{\mathcal{A}}m^{\mathcal{B}} > \left(p^{\mathcal{A}} + p^{\mathcal{B}} - c\right)^2$$

as $p^{\mathcal{I}} = \widetilde{b^{\mathcal{I}}}$ as there are only interaction values. By the first-order condition for the Pigouvian problem in the RT2003 case $p^{\mathcal{A}} + p^{\mathcal{B}} - c = -\overline{s^{\mathcal{A}}} = -\overline{S^{\mathcal{B}}}$ so this becomes

$$m^{\mathcal{A}}m^{\mathcal{B}} > \overline{s^{\mathcal{A}}} \cdot \overline{s^{\mathcal{B}}} = \overline{\rho^{\mathcal{A}}} \cdot \overline{\rho^{\mathcal{B}}} m^{\mathcal{A}}m^{\mathcal{A}} \iff 1 > \overline{\rho^{\mathcal{A}}} \cdot \overline{\rho^{\mathcal{B}}}$$

which is my social second-order condition from E. Glen Weyl (2009). In the Armstrong case

$$\frac{\mu^{\mathcal{A}}\mu^{\mathcal{B}}}{N^{\mathcal{A}}N^{\mathcal{B}}} > \left(b^{\mathcal{A}} + b^{\mathcal{B}}\right)^2$$

In the linear case, when $\rho^{\mathcal{I}} = \frac{1}{2}$ this is exactly the private 2SC, except four times stronger.

Positive comparative statics

PROOF OF THEOREM 5:

Letting $R^{\mathcal{I}}\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right) \equiv P^{\mathcal{I}}\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right) N^{\mathcal{I}}$ be revenue on side $\mathcal{I}$ the cross partial is

$$\chi = R^{\mathcal{A}}_{12} + R^{\mathcal{B}}_{12} - c = \frac{\partial MR^{\mathcal{A}}}{\partial N^{\mathcal{B}}} + \frac{\partial MR^{\mathcal{B}}}{\partial N^{\mathcal{A}}} - c = \frac{\partial\left(P^{\mathcal{A}} + P^{\mathcal{A}}_1 N^{\mathcal{A}}\right)}{\partial N^{\mathcal{B}}} + \frac{\partial\left(P^{\mathcal{B}} + P^{\mathcal{B}}_1 N^{\mathcal{B}}\right)}{\partial N^{\mathcal{A}}} - c =$$

$$P^{\mathcal{A}}_2 + P^{\mathcal{A}}_{12}N^{\mathcal{A}} + P^{\mathcal{B}}_2 + P^{\mathcal{B}}_{12}N^{\mathcal{B}} - c = \widetilde{b^{\mathcal{A}}} + \widetilde{b^{\mathcal{A}}_1}N^{\mathcal{A}} + \widetilde{b^{\mathcal{B}}} + \widetilde{b^{\mathcal{B}}_1}N^{\mathcal{B}} - c = \widetilde{b^{\mathcal{A}}} + \widetilde{b^{\mathcal{B}}} - c - \frac{\mu^{\mathcal{A}}_{\widetilde{b}}}{N^{\mathcal{B}}} - \frac{\mu^{\mathcal{B}}_{\widetilde{b}}}{N^{\mathcal{A}}}$$

where $MR^{\mathcal{I}}$ is the classical marginal revenue on side $\mathcal{I}$. Thus there are substitutes (complements)

if $\widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}} > (<)\mu_{\widetilde{b}}$. But by equation (7), $\mu^{\mathcal{I}} = P^{\mathcal{I}} - C^{\mathcal{I}} + \left(\widetilde{b^{\mathcal{J}}} - c\right) N^{\mathcal{J}}$ thus

$$\mu = 2\widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}} + \left(\widetilde{B^{\mathcal{A}}} - C^{\mathcal{A}}\right) N^{\mathcal{A}} + \left(\widetilde{B^{\mathcal{B}}} - C^{\mathcal{B}}\right) N^{\mathcal{B}} = \pi + \widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}$$

Clearly $\mu, \pi > 0$ so dividing through there are substitutes (complements) if

$$\frac{\widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}}{\pi + \widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}} < (>) \frac{\mu_{\widetilde{b}}}{\mu}$$

$$\iff 1 - \frac{\widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}}{\pi + \widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}} > (<) 1 - \frac{\mu_{\widetilde{b}}}{\mu} \iff \frac{\pi}{\pi + \widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}} > (<) \frac{\mu_{\widetilde{B}}}{\mu}$$

If $\mu_{\widetilde{B}} < 0$ then we clearly have substitutes. Otherwise this is equivalent to

$$\frac{\pi + \widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}}{\pi} < (>) \frac{\mu_{\widetilde{b}} + \mu_{\widetilde{B}}}{\mu_{\widetilde{B}}} \iff \frac{\widetilde{b}N^{\mathcal{A}}N^{\mathcal{B}}}{\pi} < (>) \frac{\mu_{\widetilde{b}}}{\mu_{\widetilde{B}}}$$

Normative comparative statics

PROOF OF THEOREM 6:

By implicit differentiation of equation (1)

$$\frac{dN^{\mathcal{I}}}{dC^{\mathcal{J}}} = -\frac{N^{\mathcal{I}}\rho^{\mathcal{I}}\chi \frac{dN^{\mathcal{J}}}{dC^{\mathcal{J}}}}{\mu^{\mathcal{I}}}$$

So

$$\frac{\frac{dN^{\mathcal{I}}}{dC^{\mathcal{J}}}}{\frac{dN^{\mathcal{J}}}{dC^{\mathcal{J}}}} = -\frac{N^{\mathcal{I}}\rho^{\mathcal{I}}\chi}{\mu^{\mathcal{I}}}$$

$$\frac{dV^{\mathcal{I}}}{dC^{\mathcal{J}}} = V_1^{\mathcal{I}} \frac{dN^{\mathcal{I}}}{dC^{\mathcal{J}}} + V_2^{\mathcal{I}} \frac{dN^{\mathcal{J}}}{dC^{\mathcal{J}}} \propto -N^{\mathcal{I}} \left(\rho^{\mathcal{I}}\chi + \overline{b^{\mathcal{I}}} - \widetilde{b^{\mathcal{I}}}\right)$$

where the constant of proportionality is positive by concavity of profits. Clearly the effect of an exogenous increase in side $\mathcal{J}$ participation, driven by cost, is the opposite of the effect of an increase in cost (in sign, which is all I am concerned with).

The Scale-Income Model

In the SI model $b_i^{\mathcal{I}} = \beta^{\mathcal{I}} B_i^{\mathcal{I}}$. I therefore index a user's type by her membership benefit $B_i^{\mathcal{I}}$. A user i on side $\mathcal{I}$ will participate if

$$B_i^{\mathcal{I}} + \beta^{\mathcal{I}} B_i^{\mathcal{I}} N^{\mathcal{J}} > P^{\mathcal{I}}$$

that is if $B_i^{\mathcal{I}} > \frac{P^{\mathcal{I}}}{1+\beta^{\mathcal{I}} N^{\mathcal{J}}} = \frac{\nu^{\mathcal{I}}(N^{\mathcal{J}})P^{\mathcal{I}}}{\beta^{\mathcal{I}}}$ if $\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}$ is positive and if $B_i^{\mathcal{I}} < \frac{\nu^{\mathcal{I}}(N^{\mathcal{J}})P^{\mathcal{I}}}{\beta^{\mathcal{I}}}$ if $\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}$ is negative. Intuitively, $\nu^{\mathcal{I}}$ (roughly) represents the fraction of heterogeneity that is along the interaction dimension; if $\nu^{\mathcal{I}}$ is negative, then as prices rise interaction values tend to fall. Dividing by $\beta^{\mathcal{I}}$ converts this into membership benefits. It may be either the large or small scale users that participate and they may have either positive or negative membership benefits. In what follows, I derive everything from first principles, as in unidimensional models my assumption of smoothness of $f^{\mathcal{I}}$ fails. All the results would follow identically from directly specializing my more general results, as they did in the RT2003 and Armstrong cases; however, for expositional purposes it is useful to show how direct analysis of a unidimensional model leads to the same results as specializing general results from the RT2006 model.

$N^{\mathcal{I}}(P^{\mathcal{I}}, N^{\mathcal{J}}) = 1 - F^{\mathcal{I}}\left(\frac{\nu^{\mathcal{I}}[N^{\mathcal{J}}]P^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right)$ if $\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}} > 0$ and $N^{\mathcal{I}} = F^{\mathcal{I}}$ if $\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}} < 0$ where $F^{\mathcal{I}}$ is the c.d.f. of membership values. Let $D^{\mathcal{I}}(B^{\mathcal{I}}) \equiv 1 - F^{\mathcal{I}}$ or $F^{\mathcal{I}}$ as appropriate. Then $N_1^{\mathcal{I}} = -\left|\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right| f^{\mathcal{I}}$, where $f^{\mathcal{I}} = F^{\mathcal{I}\prime}$, and $N_2^{\mathcal{I}} = -\left(\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right)\frac{\nu^{\mathcal{I}\prime}P^{\mathcal{I}}}{\beta^{\mathcal{I}}} f^{\mathcal{I}}$. By the definition, $\nu^{\mathcal{I}\prime} = -\left(\nu^{\mathcal{I}}\right)^2$, yielding

$$N_2^{\mathcal{I}} = \left(\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right)\frac{P^{\mathcal{I}}\left(\nu^{\mathcal{I}}\right)^2}{\beta^{\mathcal{I}}} = \nu^{\mathcal{I}} P^{\mathcal{I}}\left|\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right|$$

Letting $P^{\mathcal{I}}(N^{\mathcal{I}}, N^{\mathcal{J}})$ be the inverse of $N^{\mathcal{I}}$ with respect to its first argument, by the inverse and implicit function theorems we have $P_1^{\mathcal{I}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}}$ where $\mu^{\mathcal{I}} \equiv \frac{N^{\mathcal{I}}}{f^{\mathcal{I}}}\left|\frac{\beta^{\mathcal{I}}}{\nu^{\mathcal{I}}}\right|$ and $P_2^{\mathcal{I}} = \nu^{\mathcal{I}} P^{\mathcal{I}} = \widetilde{b^{\mathcal{I}}} = \beta^{\mathcal{I}} D^{\mathcal{I}^{-1}}\left(N^{\mathcal{I}}\right) = \beta^{\mathcal{I}}\widetilde{b^{\mathcal{I}}}$. Surplus on side $\mathcal{I}$ is

$$\left(\frac{\nu^{\mathcal{I}}\left[N^{\mathcal{J}}\right]}{\beta^{\mathcal{I}}}\right)\int_{\frac{\nu^{\mathcal{I}}[N^{\mathcal{J}}]P^{\mathcal{I}}}{\beta^{\mathcal{I}}}}^{\left(\frac{\nu^{\mathcal{I}}[N^{\mathcal{J}}]}{\beta^{\mathcal{I}}}\right)\cdot\infty}\left(B^{\mathcal{I}} + \beta^{\mathcal{I}} B^{\mathcal{I}} N^{\mathcal{J}} - P^{\mathcal{I}}\right) f^{\mathcal{I}}(B^{\mathcal{I}}) dB^{\mathcal{I}}$$

Using the change of variables $\phi^{\mathcal{I}} = \frac{B^{\mathcal{I}}\beta^{\mathcal{I}}}{\nu^{\mathcal{I}}}$ this becomes

$$V^{\mathcal{I}} = \left|\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right|\int_{P^{\mathcal{I}}}^{\infty}\left(\phi^{\mathcal{I}} - P^{\mathcal{I}}\right) f^{\mathcal{I}}\left(\frac{\nu^{\mathcal{I}}\phi^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right) d\phi^{\mathcal{I}} = \int_{P^{\mathcal{I}}}^{\infty} D^{\mathcal{I}}\left(\frac{\nu^{\mathcal{I}}\phi^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right) d\phi^{\mathcal{I}}$$

by the standard integration by parts for surplus. Therefore $V_1^{\mathcal{I}} = \mu^{\mathcal{I}}$ and

$$V_2^{\mathcal{I}} = -\int_{P^{\mathcal{I}}}^{\infty} \frac{\nu^{\mathcal{I}\prime}\phi^{\mathcal{I}}}{\beta^{\mathcal{I}}}\left(\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right) f^{\mathcal{I}} d\phi^{\mathcal{I}} - \nu^{\mathcal{I}} P^{\mathcal{I}} N^{\mathcal{I}} =$$

$$\int_{P^{\mathcal{I}}}^{\infty} \nu^{\mathcal{I}}\phi^{\mathcal{I}} \left|\frac{\nu^{\mathcal{I}}}{\beta^{\mathcal{I}}}\right| f^{\mathcal{I}} d\phi^{\mathcal{I}} - \nu^{\mathcal{I}} P^{\mathcal{I}} N^{\mathcal{I}} = \nu^{\mathcal{I}} V^{\mathcal{I}}$$

By E. Glen Weyl and Michal Fabinger's (2009)'s Theorem 3, $V^{\mathcal{I}} = N^{\mathcal{I}}\widehat{\mu^{\mathcal{I}}}\widehat{\rho^{\mathcal{I}}}$ where $\widehat{\mu^{\mathcal{I}}}$ is the inverse hazard rate of $D^{\mathcal{I}}$ with respect to ϕ at $\phi = P^{\mathcal{I}}$ and $\widehat{\rho^{\mathcal{I}}}$ is an average of $\frac{1}{1-\frac{\partial\widehat{\mu^{\mathcal{I}}}}{\partial N^{\mathcal{I}}}\frac{1}{P_1^{\mathcal{I}}}}$ over $\phi > P^{\mathcal{I}}$. Therefore to establish the formulae given in the text, I must show that, in fact, $\widehat{\mu^{\mathcal{I}}} = \mu^{\mathcal{I}}$ and $\widehat{\rho^{\mathcal{I}}} = \overline{\rho^{\mathcal{I}}}$.

$$\widehat{\mu^{\mathcal{I}}} = -\frac{N^{\mathcal{I}}}{\left.\frac{\partial D^{\mathcal{I}}}{\partial\phi^{\mathcal{I}}}\right|_{\phi^{\mathcal{I}}=P^{\mathcal{I}}}} = \frac{N^{\mathcal{I}}}{f^{\mathcal{I}}}\left|\frac{\beta^{\mathcal{I}}}{\nu^{\mathcal{I}}}\right| = \mu^{\mathcal{I}}$$

For the equivalence of $\widehat{\rho^{\mathcal{I}}}$ to an average of $\rho^{\mathcal{I}}$, let me calculate $\rho^{\mathcal{I}}$.

$$\rho^{\mathcal{I}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}\frac{\partial^2\pi}{\partial N^{\mathcal{I}2}}} = -\frac{\mu^{\mathcal{I}}}{N^{\mathcal{I}}\left(P_1^{\mathcal{I}} - \mu_1^{\mathcal{I}}\right)} = \widehat{\rho^{\mathcal{I}}}$$

Following a similar logic

$$\chi = P_2^{\mathcal{A}} + \nu^{\mathcal{B}} N^{\mathcal{B}} P_1^{\mathcal{B}} + \nu^{\mathcal{B}} P^{\mathcal{B}} - \mu_2^{\mathcal{I}} - c = \nu^{\mathcal{A}} P^{\mathcal{A}} + \nu^{\mathcal{B}} P^{\mathcal{B}} - c - \nu^{\mathcal{B}} N^{\mathcal{B}} \frac{\mu^{\mathcal{B}}}{N^{\mathcal{B}}} - \mu_2^{\mathcal{I}}$$

$$\mu_2^{\mathcal{I}} = \frac{\partial \frac{N^{\mathcal{I}}|\beta^{\mathcal{I}}|}{f^{\mathcal{I}}(N^{\mathcal{I}})}\frac{1}{|\nu^{\mathcal{I}}[N^{\mathcal{J}}]|}}{\partial N^{\mathcal{J}}} = \frac{N^{\mathcal{I}}\left|\beta^{\mathcal{I}}\right|}{f^{\mathcal{I}}}\frac{\partial\left|\frac{1}{\beta^{\mathcal{I}}} + N^{\mathcal{J}}\right|}{\partial N^{\mathcal{J}}} = (\nu^{\mathcal{I}})\frac{N^{\mathcal{I}}\left|\beta^{\mathcal{I}}\right|}{f^{\mathcal{I}}} = \nu^{\mathcal{I}}\mu^{\mathcal{I}}$$

therefore

$$\chi = \sum_{\mathcal{I}=A,B}\left(\nu^{\mathcal{I}}\left[P^{\mathcal{I}} - \mu^{\mathcal{I}}\right]\right) - c$$

Because the natural analogs of the naïve simplifications from above are now apparent, I revert to directly using these, but direct derivations are available on request.

The sign of equilibrium network effects from side $\mathcal{I}$ to side $\mathcal{J}$ are now determined by

$$\nu^{\mathcal{J}}\mu^{\mathcal{J}}\overline{\rho^{\mathcal{J}}} + \rho^{\mathcal{J}}\left(\sum_{K=A,B}\left[\nu^{K}\left(P^{K}-\mu^{K}\right)\right]-c\right)$$

As stated in the text, if the second term is positive (as in the newspaper case) and the first term also is (as on the advertiser side, because $\nu^{\mathcal{B}} > 0$) then clearly equilibrium externalities from readers to advertisers are positive. On the reader's side the second term is positive, but the first is negative as $\nu^{\mathcal{A}} < 0$. The equilibrium network effect will tend to be positive when complementarity is large relative to reader-side market power and when pass-through is decreasing, so that inframarginal harms are small relative to the pass-through of marginal cost changes.

The own-side cost-welfare effect on side $\mathcal{I}$

$$-\left(\mu^{\mathcal{A}}\mu^{\mathcal{B}} + \left[\sum_{K=A,B}\left(\nu^{K}\left[P^{K}-\mu^{K}\right]\right)-c\right]\rho^{\mathcal{J}}\nu^{\mathcal{I}}\mu^{\mathcal{I}}\overline{\rho^{\mathcal{I}}}N^{\mathcal{A}}N^{\mathcal{B}}\right) \propto$$

$$-\left(\mu^{\mathcal{J}} + \nu^{\mathcal{I}}\rho^{\mathcal{J}}\overline{\rho^{\mathcal{I}}}N^{\mathcal{A}}N^{\mathcal{B}}\left[\sum_{K=A,B}\left(\nu^{K}\left[P^{K}-\mu^{K}\right]\right)-c\right]\right)$$

For advertisers this is clearly negative as both terms are. For readers this is again ambiguous: while they lose from the higher prices the gain from resulting reduction in the Spence distortion and therefore advertising. Thus it is not clear that readers will be worse off as the newspaper business becomes less competitive, because on both sides of the market reduced competition is likely to lead to higher advertising rates and this may improve reader welfare.

Applications

In this appendix I describe formally the GASH model of duopoly, derive equilibrium conditions for it and use these to state formally and prove the results I referred to in Subsection VI.C. Armstrong only considers a linear demand, product differentiated model where all users must choose one platform in the spirit of Harold Hotelling (1929), making it impossible to compare monopoly to competition as under monopoly the platform would charge infinite prices.

Demand

Users on each side $\mathcal{I}$ have a usage benefit $b^{\mathcal{I}}$ which depends only on which side of the market they are on, not which platform they use nor their identity. users have a membership benefit for

each platform 1 and 2, $B_i^{\mathcal{I},1}$ and $B_i^{\mathcal{I},2}$ respectively, drawn from a symmetric density function $g^{\mathcal{I}} : R^2 \to R$. Given a vector, consisting of a pair of side $\mathcal{I}$ prices and a pair of participation levels on side $\mathcal{J}$, $\left(P^{\mathcal{I},i}, P^{\mathcal{I},-i}, N^{\mathcal{J},i}, N^{\mathcal{J},-i}\right)$, we can write the number of side $\mathcal{I}$ agents that choose to participate on platform i. Let us call this $N^{\mathcal{I},i}$, where

$$N^{\mathcal{I},i}(P^{\mathcal{I},i}, P^{\mathcal{I},-i}, N^{\mathcal{J},i}, N^{\mathcal{J},-i}) = \int_{P^{\mathcal{I},i}-b^{\mathcal{I}}N^{\mathcal{J},i}}^{\infty} \int_{-\infty}^{B^{\mathcal{I},i}+b^{\mathcal{I}}N^{\mathcal{J},i}-P^{\mathcal{I},i}+P^{\mathcal{I},-i}-b^{\mathcal{I}}N^{\mathcal{J},-i}} g^{\mathcal{I}}(B^{\mathcal{I},i}, B^{\mathcal{I},-i}) dB^{\mathcal{I},-i} dB^{\mathcal{I},i}.$$

Let $R^{\mathcal{I},i} \equiv b^{\mathcal{I}}N^{\mathcal{J},i} - P^{\mathcal{I},i}$ be the payoff an agent on side $\mathcal{I}$ receives from joining platform i, net of membership benefits. We can then express the number of side $\mathcal{I}$ agents that participate of platform i as $N^{\mathcal{I},i} = N^{\mathcal{I},i}(P^{\mathcal{I},i}, N^{\mathcal{J},i}, R^{\mathcal{I},-i})$, where

$$N^{\mathcal{I},i}(P^{\mathcal{I},i}, N^{\mathcal{J},i}, R^{\mathcal{I},-i}) = \int_{P^{\mathcal{I},i}-b^{\mathcal{I}}N^{\mathcal{J},i}}^{\infty} \int_{-\infty}^{B^{\mathcal{I},i}+b^{\mathcal{I}}N^{\mathcal{J},i}-P^{\mathcal{I},i}-R^{\mathcal{I},-i}} g^{\mathcal{I}}(B^{\mathcal{I},i}, B^{\mathcal{I},-i}) dB^{\mathcal{I},-i} dB^{\mathcal{I},i}.$$

Note that as $N^{\mathcal{I},i}$ is strictly decreasing in it's first argument, it can be inverted, yielding a well defined inverse demand $P^{\mathcal{I},i}\left(N^{\mathcal{I},i}, N^{\mathcal{J},i}, R^{\mathcal{I},-i}\right)$.

Duopoly equilibrium

Supposing that both firm 1 and firm 2 offer insulating tariffs on both sides of the market; that is, $P^{\mathcal{I},i}\left(P^{\mathcal{J},i}\right) = b^{\mathcal{I}}N^{\mathcal{J},i} + h^{\mathcal{I},i}$ for some chosen $h^{\mathcal{I},i}$ as the insulating tariff for the Armstrong model is just full insurance. We can the consider each firm i to choose participation levels on each side, taking as given $R^{\mathcal{A},-i}$ and $R^{\mathcal{B},-i}$. Thus, in equilibrium, each firm maximizes, with respect to $N^{\mathcal{I},i}$, $I = A, B$,

$$[P^{\mathcal{A},i}(N^{\mathcal{A},i}, N^{\mathcal{B},i}, R^{\mathcal{A},-i}) - C^{\mathcal{A}}]N^{\mathcal{A},i} + [P^{\mathcal{B},i}(N^{\mathcal{B},i}, N^{\mathcal{A},i}, R^{\mathcal{B},-i}) - C^{\mathcal{B}}]N^{\mathcal{B},i} \tag{A 5}$$

The first-order condition for maximization of (A 5), with respect to $N^{\mathcal{I},i}$, is

$$P^{\mathcal{I},i} - C^{\mathcal{I}} + P_1^{\mathcal{I},i}N^{\mathcal{I},i} + P_2^{\mathcal{J},i}N^{\mathcal{J},i} = 0 \tag{A 6}$$

Note that, by the inverse function theorem, $P_1^{\mathcal{I},i} = \frac{1}{N_1^{\mathcal{I},i}}$, and that by the implicit function theorem, $P_2^{\mathcal{I},i} = -\frac{N_2^{\mathcal{I},i}}{N_1^{\mathcal{J},i}}$. Calculating $N_1^{\mathcal{I},i}$, we find

$$N_1^{\mathcal{I},i} = \underbrace{-\int_{-\infty}^{-R^{\mathcal{I},-i}} g^{\mathcal{I}}(P^{\mathcal{I},i}(N^{\mathcal{J},i}) - b^{\mathcal{I}}N^{j,i}, B^{\mathcal{I},-i})dB^{\mathcal{I},-i}}_{\text{market expansion}}$$
$$\underbrace{-\int_{P^{\mathcal{I},i}(N^{\mathcal{J},i})-b^{\mathcal{I}}N^{\mathcal{J},i}}^{\infty} g^{\mathcal{I}}(B^{\mathcal{I},i}, B^{\mathcal{I},i} + b^{\mathcal{I}}N^{\mathcal{J},i} - P^{\mathcal{I},i}(N^{\mathcal{J},i}) - R^{\mathcal{I},i})dB^{\mathcal{I},i}}_{\text{cannibalization}}.$$

Calculating $N_2^{\mathcal{I},i}$, we find

$$N_2^{\mathcal{I},i} = \int_{-\infty}^{-R^{\mathcal{I},-i}} b^{\mathcal{I}}g^{\mathcal{I}}(P^{\mathcal{I},i}(N^{\mathcal{J},i}) - b^{\mathcal{I}}N^{j,i}, B^{\mathcal{I},-i})dB^{\mathcal{I},-i}$$
$$+\int_{P^{\mathcal{I},i}(N^{\mathcal{J},i})-b^{\mathcal{I}}N^{\mathcal{J},i}}^{\infty} b^{\mathcal{I}}g^{\mathcal{I}}(B^{\mathcal{I},i}, B^{\mathcal{I},i} + b^{\mathcal{I}}N^{\mathcal{J},i} - P^{\mathcal{I},i}(N^{\mathcal{J},i}) - R^{\mathcal{I},i})dB^{\mathcal{I},i}$$
$$= -b^{\mathcal{I}}N_1^{\mathcal{I},i}$$

Thus, (A 6) can be rewritten

$$P^{\mathcal{I},i} - C^{\mathcal{I}} + b^{\mathcal{J}}N^{\mathcal{J},i} = \frac{N^{\mathcal{I},i}}{-N_1^{\mathcal{I},i}} \equiv \mu_o^{\mathcal{I},i} \tag{A 7}$$

where $\mu_o^{\mathcal{I},i}$ is firm i's *own-price market power* on side $\mathcal{I}$. Equation (A 7) can be thought of as governing firm i's best-response to firm $-i$'s choices of $R^{\mathcal{I},-i}$. Thus the conditions for a symmetric-across-firms (SAF) equilibrium in insulting tariffs, what I call an *symmetric insulated equilibrium* (SIE), it must be the case that for $I = A, B$ and $i = 1, 2$

$$P^{\mathcal{I}} - C^{\mathcal{I}} + b^{\mathcal{J}}\frac{N^{\mathcal{J}}}{2} = \mu_o^{\mathcal{I}} \tag{A 8}$$

where $P^{\mathcal{I}}$ is the SIE price on side $\mathcal{I}$, $N^{\mathcal{J}}$ is the total market participation on side $\mathcal{J}$ and $\mu_o^{\mathcal{I}}$ is the SAF own-price market power on side $\mathcal{I}$.

Monopoly

Suppose firms i and $-i$ can act, via merger or some other means, in such a way so as to maximize joint profits. Then, assuming the platforms choose a SAF strategy, they solve

$$\max_{N^{\mathcal{I},m}} \left\{ 2 \sum_{\mathcal{I}} \left[P^{\mathcal{I},m}(N^{\mathcal{I},m}, N^{\mathcal{J},m}) - C^{\mathcal{I}} \right] N^{\mathcal{I},m} \right\}$$

where $P^{\mathcal{I},m}(N^{\mathcal{I},m}, N^{\mathcal{J},m})$ is the inverse, with respect to the first argument, of $N^{\mathcal{I},m}\left(P^{\mathcal{I},m}, N^{\mathcal{J},m}\right)$, the number of side $\mathcal{I}$ agents who join any one of the two symmetric platforms

$$N^{\mathcal{I},m}(P^{\mathcal{I},m}(N^{\mathcal{J},m}), N^{\mathcal{J},m}) = \int_{P^{\mathcal{I},m}(N^{\mathcal{J},m})-b^{\mathcal{I}}N^{\mathcal{J},m}}^{\infty} \int_{-\infty}^{B^{\mathcal{I},m}} g^{\mathcal{I}}(B^{\mathcal{I},m}, B^{\mathcal{I},-m}) dB^{\mathcal{I},-m} dB^{\mathcal{I},m}$$

The first-order condition of the platforms is

$$P^{\mathcal{I}} - C^{\mathcal{I}} + b^{\mathcal{J}} \frac{N^{\mathcal{J}}}{2} = \frac{N^{\mathcal{I},m}}{-N_1^{\mathcal{I},m}} \equiv \mu^{\mathcal{I}}$$

where $\mu^{\mathcal{I}}$ is the *joint market power* and

$$N_1^{\mathcal{I},m} = -\int_{-\infty}^{P^{\mathcal{I},m}-b^{\mathcal{I}}N^{\mathcal{J},m}} g^{\mathcal{I}}(P^{\mathcal{I},m} - b^{\mathcal{I}}N^{\mathcal{J},m}, B^{\mathcal{I},-m}) dB^{\mathcal{I}-m}$$

Note that $\mu^{\mathcal{I}}\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right) > \mu_o^{\mathcal{I}}\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right)$ for any pair of participation rates $\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right)$ as the denominator of $\mu^{\mathcal{I}}$ does not include the (strictly positive) cannibalization term of $\mu_o^{\mathcal{I}}$.

Merger effects

PROPOSITION 10: *Assume 2SC for the Armstrong model applied to the monopoly optimum in this context: $\rho^{\mathcal{I}} > 0$ for both I and $\frac{\mu^{\mathcal{A}}\mu^{\mathcal{B}}}{N^{\mathcal{A}}N^{\mathcal{B}}} > \rho^{\mathcal{A}}\rho^{\mathcal{B}}\left(b^{\mathcal{A}} + b^{\mathcal{B}}\right)^2$. Then participation and welfare are higher on both sides of the market under duopoly than under monopoly.*

PROOF:

Rewrite condition $(A8)$ as

$$P^{\mathcal{I}} - C^{\mathcal{I}} + b^{\mathcal{J}}N^{\mathcal{J}} = \mu^{\mathcal{I}} - \mu^{\mathcal{I}}\left(1 - \frac{M_o^{\mathcal{I}}}{M^{\mathcal{I}}}\right)$$

Let $G^{\mathcal{I}}\left(N^{\mathcal{I}}, N^{\mathcal{J}}\right) = 1 - \frac{M_o^{\mathcal{I}}}{M^{\mathcal{I}}}$ and let the value of $G^{\mathcal{I}}$ at SIE participation rates $\left(N^{\mathcal{A}^\star}, N^{\mathcal{B}^\star}\right)$ be $G^{\mathcal{I}^\star}$. Note that $G^{\mathcal{I}}, G^{\mathcal{I}^\star} > 0$ as noted above. Inthe monopoly problem costs are given by

$$C^{\mathcal{I}} - tG^{\mathcal{I}}$$

for $t \in [0, 1]$. Clearly for $t = 0$ the solution is the same as the SAF monopoly optimum (which is unique by the assumed second-order conditions). For $t = 1$, the allocation must be the same as at the SIE equilibrium. But by my reasoning in Section III.A, participation rates on the two sides of the markets are complements, so a fall in either cost raises both participation rates. Thus an increase in t must raise both participation rates. Therefore participation on both sides of the market is higher under the SIE equilibrium than under the SAF monopoly optimum. Furthermore in the Armstrong model welfare on each side of the market depends (directly) only on participation on that side as there is no Spence distortion, so both groups of users must be better off. Finally, because participation rates are below their optimal level for any amount of market power on both sides, the resulting increase in participation on both sides must be social welfare enhancing. Thus a merger without efficiencies from a SIE equilibrium is harmful to participation, both groups of users' welfare and overall social welfare.

This result contrasts significantly with the RT2003 model, where competition can easily harm one or even both groups of users. Of course the result depends on a particular assumption about tariffs, namely that firms use insulating tariffs. In a previous version of this paper I proved that this result was significantly more robust and holds in a very wide range of reasonable cases with non-insulating tariffs. Those results, omitted for the sake of brevity and clarity, are available on request.

REFERENCES

Hotelling, Harold. 1929. "Stability in Competition." *Economic Journal*, 39(153): 41–57.

Weyl, E. Glen. 2009. "Monopoly, Ramsey and Lindahl in Rochet and Tirole (2003)." *Economics Letters*, 103(2): 99–100.

Weyl, E. Glen, and Michal Fabinger. 2009. "Pass-Through as an Economic Tool." http://www.fas.harvard.edu/∼weyl/research.htm.

American Economic Review 100 (September 2010): 1642–1672
http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.4.1642

A Price Theory of Multi-Sided Platforms

By E. Glen Weyl*

I develop a general theory of monopoly pricing of networks. Platforms use insulating tariffs to avoid coordination failure, implementing any desired allocation. Profit maximization distorts in the spirit of A. Michael Spence (1975) *by internalizing only network externalities to marginal users. Thus the empirical and prescriptive content of the popular Jean-Charles Rochet and Jean Tirole* (2006) *model of two-sided markets turns on the nature of user heterogeneity. I propose a more plausible, yet equally tractable, model of heterogeneity in which users differ in their income or scale. My approach provides a general measure of market power and helps predict the effects of price regulation and mergers.* (*JEL* D42, D85, L14)

The pricing problems of payment and advertising platforms have much in common. Both seek to attract two distinct groups of users: AmEx needs cardholders and merchants, while the *New York Times* recruits readers and advertisers. Because the value each group takes from using these services depends on the size of the *other side of the market*, the platform's pricing and marketing strategies to each group are closely linked. Therefore policy directed at alleviating distortions caused by market power in these industries must take account of how interventions on one side affect welfare and platform behavior on the other.

Yet despite credit cards and newspapers both being canonical *two-sided markets*, the economics of these industries seem intuitively quite different. Consumers most likely to carry AmEx are those who most value the opportunity to use the card. These loyal cardholders therefore value the participation of merchants more than those indifferent between AmEx and another payment form do. Given its limited ability to price discriminate, AmEx fails to fully internalize the preferences of loyal users, putting too little effort into attracting merchants and charging them a higher price than would be socially optimal. However, when the costs of attracting cardholders rise and therefore cardholder incentives fall, AmEx will tend to serve only users who value merchant participation more strongly, leading them to attract more merchants with lower fees. This logic is the basis of the burgeoning literature on two-sided markets pioneered by Rochet and Tirole (2003).

* Harvard Society of Fellows and Toulouse School of Economics: 78 Mount Auburn Street, Cambridge, MA 02138 (e-mail: weyl@fas.harvard.edu). This paper was split off as one part of a previous working paper "Monopolies in Two-Sided Markets: Comparative Statics and Identification." I am grateful to el Ministerio de Hacienda de Chile and the Centro de Investigación Económica at the Instituto Tecnológico Autonomo de México (CIE-ITAM), which hosted me on a visit while I conducted parts of this research. The Harvard Milton Fund supported the last stage of this project and financed the excellent research assistance provided by Will Weingarten, Alex White, and especially Stephanie Lo. Ahmed Jaber and Dan Sacks also acted as research assistants with the permission of their employer Amy Finkelstein, to whom I am grateful for her generosity. I also appreciate the helpful comments and advice on this research supplied by Mark Armstrong, David Evans, Jeremy Fox, Andrea Hawkley, Alisha Holland, Scott Kominers, Luis Rayo, Bill Rogerson, Dick Schmalensee, Jesse Shapiro, and especially Maxim Engers, and seminar participants at the Banco de México, the Barcelona Graduate School of Economics, CIE-ITAM, Harvard University, IESE Business School, Northwestern University, Télécom ParisTech Conference on the Economics of Information and Communications Technology, the Toulouse School of Economics (TSE), University of Chicago, and University of California at Berkeley. I owe a special debt to Pai-Ling Yin for detailed discussions of the paper. All confusion and errors are my own.