

ACCESS \neq ACCESS₁ + ACCESS₂

Gerald R. Faulhaber*

Abstract

“Access” has been and continues to be an important concept in regulation and antitrust. In this paper, I consider two interrelated access concepts: access to essential facilities (access₁) and access via interconnection to customers (access₂). Neither concept is new; some industries are characterized by one or the other, some industries characterized by both. I argue that the public policy implications of each are rather different, and relate this difference to antitrust treatment of the “new” economy.

* Professor, Business and Public Policy Department, Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Address inquiries to faulhaber@wharton.upenn.edu.

ACCESS \neq ACCESS₁ + ACCESS₂

Gerald R. Faulhaber

Introduction

The essential facilities doctrine (Access₁) has been at the heart of antitrust since *United States v. Terminal Railroad Association* in 1912,¹ in which the Court held that a “bottleneck” (i.e., essential) facility that could not feasibly be duplicated must be shared among rivals. This doctrine has been upheld over the intervening years, for example *Otter Tail Power Co. v. United States*² and *MCI Communications Corp. v. AT&T Corp.*,³ although the doctrine is not beyond dispute.⁴ Most famously, *US v. AT&T*⁵ was based entirely on the concept that the telephone local access line (the local loop) was a bottleneck facility; divesting the Operating Companies that owned the local loop from the AT&T long distance company would thus remove the incentive to use the local loop to disadvantage AT&T’s long distance competitors. Most recently, the 1996 Telecommunications Act⁶ requires the Regional Bell Operating Companies to offer to all potential competitors the constituent parts of its local loops (unbundled network elements) at cost-based prices, based on the premise that the local loop is an essential facility.

Interconnection (Access₂) has been at the heart of the regulation of network industries since the early decades of the twentieth century. The end of the nineteenth century saw somewhat disruptive practices in railroad and telephone, in which dominant players refused to transit traffic to or from their competitors’ customers. For example, customers of Keystone Telephone Company in Philadelphia, a local exchange provider, could neither receive nor complete calls to customers of Bell of Pennsylvania, the dominant local exchange provider, which in turn was connected to AT&T’s unique long distance service. The telephone industry developed into local monopolies dominated by the Bell System and subject to state regulation. Regulators required Bell to interconnect with the

¹ 224 U.S. 383 (1912). “A consortium of 14 of the 24 railroads that shipped freight across the Mississippi River at St. Louis got control of the terminal facilities at each side of the river. The Supreme Court, while assuming that the operation of these facilities as a single entity was the most efficient way to operate them (that is, they comprised a natural monopoly), held that the Sherman Act required the consortium to provide access to the terminal facilities to the 10 other railroads on nondiscriminatory terms.” Richard Posner, in *Blue Cross et al v. Marshfield et al*, 65 F.3d 1406 (7th Cir. 1995)

² 410 U.S.366 (1973)

³ 708 F.2d 1081, 1132-33 (7th Cir. 1983).

⁴ See, for example, H. Hovenkamp, *Federal Antitrust Policy: the Law of Competition and its Practice* §5.6c (1994).

⁵ 552 F. Supp. 131 (D.D.C., 1982).

⁶ Pub. L. 104-104, 110 Stat. 56 (1996) (codified at 47 U.S.C. § 151 et seq.).

remaining Independent telephone companies, and regulated the interconnection prices as well. Railroad regulators encouraged railroads to enter into “interlining” agreements whereby each party agreed to accept traffic from other railroads destined for their customers. Central to the interconnection requirement is that such networks exhibit “network effects” (or “network externalities”) in which the value of the network to each customer increases as the number of customers increases.⁷ In the telephone network, this is very easy to see: the more customers I can call (or be called by), the more valuable is the network to me. Similarly with rail transport: the more customers I can ship to (or receive from), the more valuable is the network to me. Should a firm decide not to interconnect with its competitors, then all customers lose value as they now have fewer customers they can reach. Of course, smaller competitors are disadvantaged more than large competitors, suggesting that refusal to interconnect may be a strategic decision by larger firms.

Network effects were extensively studied in the mid-1980s;⁸ initially, network effects were thought limited to two-way communications systems (“direct” network effects) but this later work identified “indirect” or “virtual” network effects based on complementary goods. This latter category includes videotape players, in which the more customers own a particular format (say, VHS), the more tapes are produced in that format and therefore the more benefit each customer derives from the service. In this paper, I am concerned only with direct network effects, most relevant to interconnection and access issues.

Some industries exhibit both essential facilities and network effects; the wireline telephone industry is a case in point. Other industries may have network effects without an essential facility (such as Instant Messaging) and some have essential facilities without network effects (such as railroads). In each case, possible anticompetitive actions can lead these and related industries to (near-) monopoly, and may require some form of government intervention (either regulatory or antitrust) to prevent these anticompetitive actions.

The focus of this paper is simple: the two forms of access ($access_1$ and $access_2$) result from very different market failures and call for very different remedies. Even in industries in which both forms of access problems exist, solving one will not solve the other. For example, solving the interconnection problem (by regulatory fiat) in telephone has not solved the bottleneck local loop problem. Similarly, if the bottleneck local loop problem could be solved, a solution to the interconnection problem would still be required (that is, even if the local loop had competitive alternatives, some regulation of interconnection will still be required).

⁷ The network externality was first introduced into the economics literature by Jeffrey Rohlfs, “A Theory of Interdependent Demand for a Communications Service,” *Bell Journal of Economics and Management Science*, 5(1), 16-37, Spring, 1974. For a recent survey with real-world examples, see Jeffrey Rohlfs, *Bandwagon Effects in High-Technology Industries*, Cambridge, MA: MIT Press, 2001.

⁸ For an excellent and accessible summary of this research, see Katz, Michael L., and Carl Shapiro, “Systems Competition and Network Effects.” *Journal of Economic Perspectives* 8, 93–115, Spring, 1994.

Essential Facilities (Access₁)

Traditionally, an essential facility is in fact a physical facility.⁹ The telephone local loop, terminal facilities in St. Louis, a rail line to the mouth of a coal mine, all constitute essential facilities. Also, essential facilities cases tend to be vertical cases; a vertically integrated firm may own an essential facility through which upstream firms with whom it competes must reach their customers. The question at the heart of essential facilities cases is what makes the facility so essential and how does it differ from simple monopoly?

A simple example is that of a harbor suitable for shipping for which one individual owns all land surrounding the harbor. Any firm that wished to use the harbor must use this individual's land and any improvements, such as roads and docks. Assume that this harbor owner also owns firms in the neighborhood that ship from this harbor. The owner of the harbor could be expected to charge a monopoly price, more likely discriminatory prices, to those who wished to use the harbor and may refuse to deal with firms that are its direct upstream competitors. The pricing of the monopolist is limited only by (i) the value of the service he is providing to his customers; and (ii) the cost of alternative methods to his customers of obtaining equivalent services. If the nearest harbor is hundreds of miles away over unimproved roads, then that alternative cost could be quite high, allowing the monopoly harbor owner to extract substantial rents from his customers. Should the value of harbor use to his customers be very high, then the rents will be correspondingly high, particularly if the monopolist is adept at price discrimination.

In this case, the essential facility is truly non-duplicable; there is no feasible alternative to ocean shipping beyond this unique resource. However, it raises a fundamental issue: how is this different from simple monopoly? Since antitrust law does not forbid monopolies that are earned (as opposed to merger to monopoly, or monopolization via anticompetitive practices), why should the ownership and exploitation of this essential facility not be treated as a simple monopoly?

Of course, it may be judged that the exploitation of a particular monopoly has pernicious spillovers, such as drastically reduced economic development within the harbor's natural economic zone due to high pricing of access to shipping. If so, granting access to the harbor for the harbor owner's upstream competitors need not help with these spillovers, if the harbor owner simply charges a monopoly price to all. In order to solve this broader problem, both access and price would have to be controlled,¹⁰ a job more suited to

⁹ This has not stopped antitrust cases based on non-physical "essential" facilities. In *Blue Cross et al v. Marshfield et al*, 65 F.3d 1406 (7th Cir. 1995), Blue Cross claimed, *inter alia*, that the Mansfield Clinic and HMO had monopolized medical services in northern Wisconsin and was therefore an essential facility to which they and their subsidiary CompCare should be provided access. The trial judge held in their favor, but this portion of the verdict was overturned on appeal in the 7th Circuit.

¹⁰ More than just price is at issue. A number of difficult questions must be addressed if access to an essential facility is granted: (i) do all competitors have such access? How about customers? If only a limited number of competitors are granted access, how is it decided which ones get access? (ii) The quality of service to the competitors must be monitored; are competitors getting service as good as the bottleneck

regulators and less suited to antitrust authorities. We note that there is a long and extensive literature which calls into question the efficacy of such regulation in promoting broader economic and social aims.¹¹ Unless antitrust authorities are willing to pursue such regulatory solutions, the imposition of access requirements on owners of essential facilities would appear to have only the effect of re-dividing the monopoly rents that accrue from these facilities among competitors, with little positive impact on other agents such as customers.

Even in the case of a truly essential facility such as the harbor example, the question of how essential is the facility still has teeth. If the surrounding economic zone produces low-value bulk goods for export such as grain or coal, then sea transport is indeed essential. However, is it essential that grain or coal have to be produced here? If production shifted to high-value small goods such as microprocessors or drugs, then the preferred mode of transport is air freight, not ocean shipping, and the essential nature of the harbor facility fades. If it is essential that grain and coal be produced here (because of natural endowments, say), is it essential that they be sold overseas? If trains or coal slurry pipelines could reach domestic markets easily, then the essential nature of the harbor facility fades.

In examples more realistic than the harbor owner, other factors come into play in an essential facilities analysis. In the case of the telephone local loop, it is not the case that this facility is non-duplicable. It is actually quite easy to build a parallel local loop; the technology is well known and indeed quite old. The problem is that the local loop is subject to extreme economies of scale, and is thus thought to be a “natural monopoly.” Having two or more local loops serving the same homes and businesses was thought to be unnecessarily duplicative and wasteful, because of the extreme scale economies.¹² The premise is that the market could not support more than one local loop.

The market evidence for this is dubious. In the 1890s, most towns and cities in the US were served by two or more local exchange companies, each with their own local loops. Until the Bell System introduced long-distance service and refused to interconnect with its independent competitors, the market easily supported multiple local exchange carriers, using a technology (the twisted pair) that differs little from that in place today. Further, until the 1996 Telecommunications Act, almost all states granted monopoly franchises to their incumbent telephone companies, suggesting that legal protection against potential entrants may well have been necessary. And lastly, telephone prices to residential users

owners' upstream operator? (iii) Do competitors receive installation and maintenance services as promptly and thoroughly as the bottleneck owners' upstream operator? (iv) In periods of congestion, are competitors treated the same as the integrated operator? (v) Do all competitors receive the same choices of price and service quality? The authority that regulates access must monitor these issues and enforce its regulations in a contentious, litigious environment.

¹¹ Another solution is public ownership of the essential facilities, an option which may have even more practical difficulties than public regulation of private essential facilities. However, I note that many port facilities in the US are run by publicly managed Port Authorities (such as New York and Philadelphia), apparently with some success.

¹² In the sense that having two gas stations at an intersection is not considered duplicative and wasteful, but simply the result of healthy competition (and no scale economies).

have traditionally been subsidized by higher rates to business users and long distance access charges, a situation that kept local access prices below cost and therefore unattractive to entry.

In addition, the role of technology in assessing how essential is the facility is important. For example, at the time of the AT&T cases, wireless services were just beginning to emerge with their potential to substitute for the local loop. More recently, wireless has become far more ubiquitous; should wireline services (which use the local loop) become substantially more expensive, it would be reasonable to expect customers to abandon wireline service for wireless only. This phenomenon is already occurring in Europe, in which wireline marginal prices are considerably higher than in the US.

In both *MCI Communications v. American Tel. & Tel. Co.*¹³ and *United States v. American Tel. & Tel. Co.*¹⁴ the Courts found the local loop to indeed be a bottleneck facility. In the latter case, the remedy was the most extensive divestiture in US history, up to the date of this writing. Most industry analysts and observers view this divestiture as a success in bringing competition to the long distance market. This success supports the hypothesis that the local loop is (or at least was) an essential facility, and that separating ownership of the bottleneck from ownership of the competing upstream long distance firm achieved the policy objectives of the case.

In the AT&T case, the monopoly over the bottleneck local loop was not earned but rather granted by regulators as a franchise to the Bell System. This was not a situation, in which the monopolist won its favored position fair and square. Further, the failure of regulation to control the vertical power of the Bell System to leverage its local access monopoly as been documented elsewhere.¹⁵ The remedy of divestiture, radical though it was, seems in this case to justify antitrust intervention based on an essential facilities premise.

Nevertheless, the burden of proof for essential facilities cases should remain high. In order for the Court to find an essential facility and grant access to competitors, I suggest that each of the following questions be answered affirmatively:

1. The monopoly was not earned, or there is a compelling external social or economic reason (beyond mere monopoly exploitation) for forcing access.
2. There exists a mechanism in place for monitoring and regulating the price and terms of trade of access (see footnote 10).
3. The final customers must have no feasible way of circumventing the use of the facility, such as changing production plans, accessing new markets, etc.
4. There are no new technologies for the foreseeable future that could substitute for the essential facility.

¹³ 708 F.2d1081 (1983)

¹⁴ 552 F. Supp. 131 (D.D.C., 1982).

¹⁵ See, for example, Gerald R. Faulhaber, "Policy-Induced Competition: the Telecommunications Experiments," to appear in *Information Economics and Policy*, 2003, available at <http://rider.wharton.upenn.edu/~faulhabe/Policy-Induced%20Competition.pdf>

In a world in which capital is costly, economic options are few, and technology unlikely to provide new alternatives, the essential facilities doctrine may have a place. But this best describes the world of *United States v. Terminal Railway Association*; the AT&T case represents a unique combination of facts that for this case appears to justify the use of the essential facilities doctrine. Looking forward to a world of inexpensive and readily available capital, temporary technology-based monopolies that could be overturned by next generation systems, customers with lots of options, it is difficult to see a justifiable essential facilities case being successfully prosecuted.

A recent case in point is the Federal Trade Commission's conditional approval of the AOL-Time Warner merger in December 14, 2000.¹⁶ The Commission imposed conditions on the parties that required "open access" to the IP channel¹⁷ on Time-Warner's cable systems to Internet Service Providers (ISPs) other than AOL. The FTC's analysis of the merger focused on increased market power in the broadband ISP market and in the broadband transport market. It states "...the merger will increase the ability of the combined firm to unilaterally exercise market power in Time Warner cable areas..."¹⁸ It also notes that the only viable competitor to cable modems was DSL, a similar transport service offered by telephone companies, and that the merger would inhibit DSL's deployment as a competitive alternative to cable modem in the transport market. While not explicitly mentioning the essential facilities doctrine, much of the order reflects a view of the cable IP channel as an essential facility. Certainly the decision to force open access is an essential facility-type remedy.

In fact, the remedy is not quite open access as envisioned by open access's most fervent proponents. AOL-Time Warner is not required to open its IP channel to all comers; it is required to open to at least three other ISPs, one of which is Earthlink (that signed a contract with Time Warner prior to FTC approval of the merger) and the other two must be approved by the FTC. Prices and terms of trade are to be "similar" to those embodied in the Earthlink agreement, which was thoroughly vetted by the FTC. To ensure compliance with these regulations, the FTC appointed a "monitor trustee" whose ongoing job is to monitor AOL-Time Warner's performance and report to the FTC. In short, the FTC has taken on the job of regulating AOL-Time Warner's open access going forward.

I take the liberty of considering this condition an essential facilities condition, and ask if it would meet the criteria set forth above; can we answer all of the questions in the affirmative?

¹⁶ Federal Trade Commission, 2000, "Decision and Order," Docket No. C-3989, at <http://www.ftc.gov/os/2000/12/aoldando.pdf>

¹⁷ Cable systems that offer broadband Internet access using cable modems do so by committing some portion of their cable capacity (typically, 6 Mhz out of a total of 750 Mhz) to a two-way IP (Internet Protocol) channel for transmitting and receiving Internet traffic to the cable modem. The cable modem is connected to the customer's PC which interprets the information and commands and displays the results on the customer's screen (either a browser, e-mail program or other client software). I refer to this cable capacity as the IP channel.

¹⁸ Federal Trade Commission, 2000, "Analysis of Proposed Consent Order to Aid Public Comment," 2; at <http://www.ftc.gov/os/2000/12/aolanalysis.pdf>

1. It is certainly the case that the monopoly¹⁹ of Time Warner cable systems were not “earned” by superior performance but rather by being granted local franchises in each of their service territories, so the first question can be answered in the affirmative.
2. There does not exist a mechanism for monitoring and regulating the price and terms of trade of open access. The most likely candidate for this job was the Federal Communications Commission, which had no interest in (in fact, recoiled at the mere suggestion of) this job. As a result, the FTC took on this job itself. This question cannot be answered in the affirmative.
3. Final customers have other options; they can continue to use narrowband services and they can use broadband services at their place of employment. Individual municipalities can deploy broadband systems, as has been done in a number of US cities.²⁰ This question cannot be answered in the affirmative.
4. New technologies for providing broadband access to customers are legion. Not only is DSL a real competitive threat but also wireless could well evolve into a threat as could satellite providers such as Starband. Providing open access may thus reduce the incentives to deploy other technologies to bypass this essential facility. This question cannot be answered in the affirmative.

This suggests the FTC’s case, considered as an essential facilities case, was not a good one. I note, however, that the FTC did not explicitly consider this an essential facilities case, though they did impose an access condition, redolent of the essential facilities doctrine.

Network Effects and Interconnection (Access₂)

Direct network effects industries such as two-way communications present unique public policy problems for both regulators and antitrust authorities. The earliest research on network effects²¹ was performed in the context of the old Bell System monopoly, in which a single firm controlled all telephone customers. The issue addressed was pricing telephone service to reflect the presence of the network externality: the total value of a customer joining the network was not only the private benefits (upon which the customer based his/her decision to join) but on external benefit of being able to call this customer that is conferred on others by his action. The result was a justification for charging a below-cost price for access to encourage subscribership²² to achieve “universal service,” the regulatory sobriquet sometimes used to describe this phenomenon. Initially, the link between network effects and interconnection was not an issue of economic study.

¹⁹ Ignoring for the moment the potential competition from DSL for data and the real competition from Direct Broadcast Satellite for video.

²⁰ See, for example, Andrew Ratner, “Shoofly Pie, and Broadband, Too,” *Baltimore Sun*, March 10, 2002, at <http://www.baltimoresun.com/technology/bal-bz.rural10mar10.story> .

²¹ See Rohlfs, *op. cit.* 1974.

²² Of course, the issue is more complicated; below-cost pricing of access must be compensated for by above-cost pricing elsewhere in the system such as long-distance prices, which could itself discourage subscribership.

Interconnection was a staple of telephone regulation post-World War I. The newly dominant Bell System had agreed to stop acquiring independent telephone companies and to interconnect with them, in return for which states moved toward a regulated monopoly model for the industry.²³ Local exchange companies no longer competed for local customers, and franchise monopoly, controlled by a state (and after 1934) and Federal regulator, became the industry standard. The requirement to interconnect grew out of the brutal competitive period during which Bell refused interconnection with independent companies, driving down their market value and eventually acquiring them on the cheap. At that time, refusal to interconnect was perceived as an aggressive even anticompetitive strategy. However, early antitrust authorities did not prosecute the early Bell System for monopolization; rather, the problem was viewed as a regulatory problem. Similar issues arose in the context of railroad regulation as well.

As wireless voice networks developed, the regulatory mandate that telephone companies had to interconnect was extended to include the new wireless carriers. Local wireline carriers, such as Verizon, are required to interconnect with wireless firms that compete with their own wireless carriers. Customers of AT&T, Sprint and Nextel wireless can complete calls to Verizon customers, both wireless and wireline. More recently, competitive local exchange carriers (CLECs) have entered local exchange markets, and the incumbent carriers (such as Verizon) are required to interconnect with the CLECs as well. This is, of course, a mutual obligation; the CLECs and wireless firms must complete the calls of Verizon wireline customers to their customers as well. Clearly, the regulators were aware of the anticompetitive potential of a large wireline monopoly refusing to interconnect to a small CLEC or fledgling wireless firm.

The price of interconnection is clearly as important as the requirement to interconnect, and interconnection pricing has occupied the FCC for more than two decades. Initially, interconnection pricing focused on the price charged by incumbent local exchange carriers (ILECs) to long-distance carriers such as AT&T, Sprint and MCI to originate/complete their customers' calls. More recently, interconnection prices to wireless firms, paging firms and CLECs have occupied center stage at the FCC, which has launched a proceeding aimed at achieving a unified approach to "intercarrier compensation."²⁴

The rise of the Internet as an alternative communications medium in the early 1990s brought new networks into being that were not subject to telecommunications regulation and therefore not subject to the injunction to interconnect. After the National Science

²³ This history has been recounted often; see for example, Gerald Faulhaber, *Telecommunications in Turmoil*, Boston: Ballinger, 1986.

²⁴ Federal Communications Commission 01-132, *Developing a Unified Intercarrier Compensation Regime* CC Docket 01-92, at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-01-132A1.pdf. See also "Bill and Keep at the Central Office as the Efficient Compensation Regime," Patrick DeGraba, OPP Working Paper 33, December, 2000, at http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp33.pdf and "A Competitively Neutral Approach to Interconnection," Jay Atkinson and Christopher Barnekov, OPP Working Paper 34, December, 2000, at http://www.fcc.gov/Bureaus/OPP/working_papers/oppwp34.pdf

Foundation withdrew from the Internet backbone network business, a number of private firms took up the function, collectively providing the core of the Internet's non-local transport function of shuttling messages from clients to servers and back again. Each of these Internet backbone providers had agreements to interconnect (or "interoperate") with each other. The largest five did so on the basis of "peering," in which traffic was exchanged and no money changed hands. Since traffic was relatively balanced, it apparently was not necessary to incur the expense of actually counting minutes and charging; the large backbone firms simply exchanged traffic.²⁵

The growth of the Internet outside the traditional strictures of regulation and its market-based solution to interconnection could well lead one to believe such restrictions aren't necessary to ensure interconnection, despite the experience of the early twentieth century in telephone and regulations requiring interconnection in that industry. However, the explosive growth of Instant Messaging, first introduced by AOL, suggests some caution is in order.

Instant messaging is a text-based means of near-real-time communication between customers who have signed up for the service. Customer *A* can send an "instant message" to customer *B* and receive an immediate reply, thus carrying on a text conversation. Instant Messaging fits the classic definition of a network effects business, in that the value to each customer depends upon how many other customers can be reached. This service has roots in the original Unix system, in which users on the same server could exchange messages in conversation mode. Instant messaging differs from e-mail in that it is a true dialog, operating in synchronous rather than asynchronous mode. AOL introduced the service as a feature for its customers in 1989. However, it became wildly popular when AOL added the "buddy list" feature in 1996. This feature displays a small window that lists all the customer's (self-designated) "buddies" with an indication of whether or not each buddy is online or not. This feature ensured that if the customer sent an IM to a particular buddy, the customer knew the buddy would receive it immediately and could reply. With the advent of the buddy list, Instant Messaging became one of the most popular features of AOL. In 1997, AOL took the unusual step of offering AIM (AOL Instant Messenger) as a stand-alone free download on the Web for non-AOL subscribers.²⁶ AIM customers could IM both other AIM customers and AOL subscribers using IM, and vice-versa; this is referred to as interoperation, equivalent to interconnection.

In 1999, several firms established competitive IM services, including Microsoft, Yahoo!, Otigo and Tribal Voice. However, these competitors quickly learned that few customers wanted to sign up for a service that couldn't IM with AOL's huge customer base, reported at the time to be over 30 million subscribers. These competitors designed their IM clients to be compatible with AOL's IM protocols (which they earlier had published

²⁵ However, the larger backbone firms have "transit" fees for smaller networks that are charged a per-message fee for use of the larger backbone network.

²⁶ All other features of AOL, such as chat rooms and proprietary content, are provided only to AOL subscribers. AIM represents the first and only (to the author's knowledge) departure from this policy.

on the Web) to ensure their IM clients interoperated with AOL's AIM/IM.²⁷ AOL chose to interpret these attempts to interoperate as "hacking," and took immediate steps to block such attempts. During the summer of 1999, various attempts by competitors were temporarily successful but quickly blocked by AOL. By year's end, AOL had demonstrated that it had both the will and the capability of blocking all such competitor attempts to interoperate with its AIM/IM services.

AOL was criticized in the press for refusing to interoperate with other IM services. Indeed, many contrasted AOL's call for open access in the pending AT&T-MediaOne merger with their refusal to interoperate their AIM/IM services with competitors. AOL's response was that they had safety and security concerns on behalf of their customers, and were unwilling to expose them to nuisance or even pornographic IMs from non-AOL sources. They did indicate that they would interoperate with competitors, but only when technical protocols were agreed to that ensured the safety and security of their IM customers.

Is the continuing refusal to interoperate/interconnect with competitors spring from a heartfelt desire to protect AOL's customers from evil messages? Or is perhaps a return to the early days of the telephone industry when refusal to interconnect was used aggressively by the Bell System against its independent rivals?

Network Effects and Corporate Strategy

In general, there may be many producers of a network effects good, competing amongst themselves for customers and perhaps cooperating to help grow the overall market. However, strategic competition is somewhat different in a network effects market than in normal markets, in several important ways:

1. Firms will often choose to adhere to a common standard and interconnect with each other. In this way, customers achieve maximum benefit, as they are in the largest possible network. An example is the current arrangement among the largest Internet backbone providers to interconnect (without cost) to each other to carry Internet traffic. However, firms may have difficulty differentiating their product or service in such a market. Further, innovation may be difficult, in that it is constrained by the existing standard and existing interconnection agreements.
2. Firms may choose a different standard or not interconnect with their competitors, perhaps believing that the superiority of their product more than offsets the network effects loss for customers. Firms with a technologically advanced product may opt for this strategy, in order to supplant an existing standard or

²⁷ The success of the AOL IM model was replicated closely by the new competitors, who introduced services almost identical to that of AOL's. The feature sets of AOL, Microsoft, Yahoo and other IM services differ only minimally. Incremental feature improvements are built into each upgrade, but competitors immediately match these new features. Absent the future development of a proprietary "killer feature" by one of the competitors, IM appears to be a market in which product differentiation does not play a significant role.

system for their own advantage.

3. In some cases, other firms can provide “converters” which permit customers to access an otherwise incompatible system. For example, in the mid-1980s, owners of Macintosh computers were unable to read floppy disks created on DOS systems. As the popularity of the PC grew, Apple built DOS-compatible disk drives into their systems. Another example is the software template built into Microsoft Word that permits it to exhibit the precise keyboard behavior of the rival WordPerfect word processor. WordPerfect customers were used to a complex series of keyboard combinations for performing actions; switching to another word processor such as Word could render this human capital valueless. To overcome this asset barrier, Microsoft built in a “converter” to make Word behave just like WordPerfect. Note that converters are one-sided, in that Microsoft could unilaterally implement its converter for WordPerfect without first obtaining permission from the producers of WordPerfect.
4. In other cases, interconnection requires consent of both parties, and often a formal business arrangement. It is technically possible (though currently illegal) for one telephone company to refuse to interconnect with another telephone company, refusing to accept and complete calls originated by customers of the other firm or hand off calls originating with its own customers.
5. For many markets involving network effects, interconnection and standards are mandated by regulation. For example, in the telephone system, interconnection is mandated by law and interconnection prices are regulated by the Federal Communications Commission and state regulators.

The *scope* of the network effect matters greatly. If I join a communications system in which I want to communicate with (say) only people in my work group, then adding other customers to my communications network has no value to the work group, and may actually decrease value. In this case, the network effect spans only a fairly small group; in such cases, the communications system may be designed and/or chosen by a representative of that group (such as the telecommunications manager or the IT manager), so everyone who ought to be on the system is on the system. In systems that have a local scope, interconnection among the relevant customer group can often be assured by the customer group itself (such as a firm, a club, or a school). Economists refer to this as *internalizing* the network externality. If the full scope of the network effect is a small otherwise coherent group, then the externality can be internalized by the group.

In fact, most communication systems customers focus most of their usage on a few other people who are friends or associates. However, they may also value the ability to communicate with merchants and people anywhere in the world in addition to their list of “friends and family,” in which case the scope of the network effect is quite large, even universal.

Indirect network effects are almost always broad in scope, as they depend upon a complementary good, service or asset. The value to me of having lots of other customers for my VHS video player is that there are many video titles produced for the VHS format, which is a national or even global phenomenon. Therefore, the scope of the video player network effect is national or global, and not limited to just my immediate friends and associates. On the other hand, the value of purchasing a particular brand of automobile depends on how many *local* service centers can repair this auto, which depends upon the number of customers in my local area, not in the nation as a whole.

The presence of network effects would appear at first glance to be very helpful in determining market definition in antitrust analysis, and in some cases this may be true. For example, there are several providers of Internet messaging service, including Microsoft, AOL, and Yahoo! *Ceteris paribus*, interconnecting these systems would almost surely raise the value of each system to its customers, and this strongly suggests that these messaging systems are in the same market. However, network effects are usually only one piece of the market definition. For example, customers of wireless telephones benefit greatly from being able to send and receive calls to and from customers of traditional wireline telephones, suggesting strong network effects. But it is likely that wireless and wireline telephone service are distinct markets with (at present) fairly low cross-elasticities of demand,²⁸ even though network effects between the two services are apparently strong.

As a general rule, interconnecting competing systems with network effects adds value for customers, at least up to the limit of their scope. Network effects industries may have a “start up” problem; the initial networks may be so small that they are not sufficiently attractive to potential customers; so few people join, resulting in a market equilibrium with a small network. In start-up markets, competing firms may have an incentive to interconnect so that the industry as a whole is more attractive to customers. They may also have incentive to convince customers that this market is growing, and that in the future there will be many more customers on the network. For example, if the start-up industry has several well-established firms with well-known brands that are unlikely to suddenly exit the market, customers may be more willing to sign up now in expectation of “getting in on the ground floor” of a sure-to-be-successful network effects service.

What is a firm’s incentive to interconnect? If the technology of the service in question permits converters, then interconnection is a unilateral decision for each firm: does the customer base of my competitor justify investing in a converter? In the case of PC disks mentioned above, it paid Apple to invest in a converter in the form of a DOS disk drive, as the base of DOS customers grew much larger than the Macintosh customer base. But these facts also suggest that it would not pay a PC producer of DOS machines to invest in a converter for Macintosh disks, and indeed they didn’t.

²⁸ Only about 1-2% of US cellular phone customers use this phone as their only phone; most prefer to have both a wireless and a wireline telephone. However, there are apparently more customers in Europe that use wireless exclusively. It is possible that as the price of wireless is reduced and its quality improves, it will become a true substitute for wireline telephony.

However, if converters are not possible (or can be thwarted), then interconnection can only be by consent of each firm. In telecommunication systems, physical interconnection and shared protocols are required, so that interconnection is a mutual business decision (although regulation may require such interconnection). In this case, the decision to interconnect is a crucial strategic decision for each firm, perhaps the most important decision they will make in their corporate lives.²⁹ It is also a decision that may determine the degree of competition in the market, and so is of great interest to antitrust authorities.

We first consider a mature industry with network effects that are broad in scope, in which new customer growth is modest or nil, with a small number of firms of roughly the same size. Is it likely that only some (or even none) of the market participants will interconnect?³⁰ No; if two such firms interconnect, they will both offer their customers a higher value than the remaining non-interconnecting firms. These two firms will be in the enviable position of being able to charge higher prices and attracting customers from the non-interconnecting firms! In this case, an interconnection arrangement helps each firm grow and increases its profitability. Non-interconnecting firms face a choice of interconnecting with the other firms or losing their customers to the more valuable interconnected network. In this case, the only stable outcome (that is, the market equilibrium) is for all firms to interconnect.

A variant of this case is a mature market with a small number of large-ish players and a large number of very small players. In this type of market, the large-ish players have an incentive to interconnect with each other, but not with the smaller players. They may interconnect with the smaller players but on terms much less favorable than the terms on which they interconnect with each other. It appears this is the situation in the previously referenced Internet backbone network business, in which about five firms of roughly equal size exist with a large number of much smaller players. The five firms interconnect with each other as “peers;” since the exchange of traffic is roughly equal, they do not charge each other for carriage of Internet traffic. However, they charge “transit” fees to smaller backbone networks.

We next consider a mature market in which a single firm has a dominant position, such as a market share well in excess of 50%. It is clear that all other firms have an incentive to interconnect with the dominant firm, as its customer base adds much value to the smaller firms’ networks. The dominant firm has a lesser incentive to interoperate, as adding the smaller firms’ customers to its network adds much less value to the larger firm. But it

²⁹ Interconnection may be costly, in at least two ways: (i) it may require investment in physical assets, such as switches and routers, as well as ongoing variable costs; (ii) it may require monitoring and/or blocking of non-conforming uses. For example, if one firm produces an excessive number of abusive or obscene communications to customers of firms with which it interconnects, then costs are imposed on both these firms and their customers, a claim made by AOL to explain its refusal to interconnect.

³⁰ The economic argument that follows is an “equilibrium” argument. Of course, it is possible in the short run that just about any industry configuration is possible. But if any firm would find it profitable to change its strategy (such as whether or not to interconnect), then that configuration cannot be an equilibrium market structure, since at least one firm would wish to choose a different strategy. A market equilibrium (in strategies) is characterized by the fact that no firm finds it profitable to adopt a different strategy than the equilibrium strategy (formally, a “Nash equilibrium,” after its inventor, mathematician John Nash).

also has an incentive to refuse interoperation, since it may expect that customers of the smaller firms will soon join their larger network to obtain the higher value of more network customers. This would increase the size of its network even more, making it even more attractive to the customers that chose to remain with smaller firms. It is possible that market-driven feedback loop is so compelling that the larger firm eventually drives out the smaller firm and monopolizes the industry, simply by refusing to interoperate. Thus, refusal to interoperate, followed by full or partial monopolization, is a possible market equilibrium in network effects markets. The short-term loss in value from refusing interconnection could be more than made up by the long-term gain in firm value from increasing market share. Refusal to interoperate can thus become a strategy for precluding competition with no offsetting benefit to customers. It is this anticompetitive outcome that is of concern to antitrust authorities.

By way of example, consider a market with one firm at 70% market share and two other firms at 15% share each, and suppose the scope of the network effect included all customers. Even if the two smaller firms interconnected, their combined share at 30% may not be sufficient to keep customers attracted by the value of the network effect of the much larger firm. It is true that the value of being able to reach 70% of the population is less than can be obtained by full interconnection with 100% of the population. However, if the large firm refuses to interconnect, then the appropriate comparison is between reaching 70% of the population vs. reaching 15% or even 30% of the population. Clearly, the larger firm creates the greater value, and as more customers switch from the smaller firms to the larger firm, the value becomes even greater. Of course, it could be that the smaller firms may have “diehard” customers who will never desert them, even if the network effects are very large. For example, even though Microsoft Windows appears to have over 90% share in the OS market and the “application barrier to entry” as found by the antitrust court is strong (indirect network effect), Macintosh computers still command the loyalty of their few remaining diehard fans. Thus, a complete 100% Windows monopoly is unlikely to occur.

How likely is it that the anticompetitive refusal to interconnect equilibrium will result from the interplay of market forces? Several factors determine if the market leader adopts a non-interconnection strategy:

1. The largest player must be substantially larger than its competitors, perhaps bigger than all its competitors combined (market share greater than 50%).
2. The network effect must be strong, so that switching to the largest provider adds substantial value for customers.
3. Customer switching costs (“stickiness”) must be low, so that switching to the largest provider is not too costly for customers.
4. It must be difficult for smaller competitors to agree to interconnect amongst themselves. For example, in a market with a leader at 40% share and sixty small firms each with 1% share that don’t interconnect, the leader is likely to gain customers fairly quickly. However, if the sixty firms can agree to interconnect, then their total network share is 60%, and the leader is likely to be the one losing share by refusing to interconnect, not the smaller firms. Even if the leader has

over 50% market share, its competitors stand a better chance of surviving (but no guarantee) if they combine forces through interconnection.

If all four factors are in place, then the market has “tipped,” in that the largest firm will implement a non-interconnection strategy and drive the market toward monopoly. Tipping does not imply that the market is at monopoly or near-monopoly; merely that market conditions are such that this will be the ultimate outcome.

We next consider a growing market of broad scope, in which some potential customers have joined a network but there are still many more potential customers as yet not signed up. The same factors that matter in a mature market still matter in a growing market: market share of largest firm, strength of the network effect, customer switching costs, and the ease with which smaller firms can reach a mutual interconnection agreement. In addition, several other factors come into play in growing markets:

- How costly is it to acquire customers through marketing efforts? Of particular importance is the difference in acquisition costs among firms. For example, a firm with an established brand name may find it easier to acquire customers. Also, if a firm controls a complementary good with a large market share, it may find it particularly easy to attract customers for its new service. For example, many have claimed that since Microsoft controls the Windows OS, it has an advantage in introducing new services such as its Microsoft Messenger service; others claim that since Yahoo! is the most popular web portal (in terms of number of page views) that it has an advantage in introducing new services such as Yahoo! Messenger.³¹ Firms with particularly low customer acquisition costs are sometimes able to overcome a rival’s early lead in a network effects market by using its competitive advantage and “out-marketing” its rival to surpass it in market share and thus gain the network effects advantage. In this case, it is possible that the early leader that does not possess a customer acquisition advantage may prefer to interconnect, while the “fast follower” firm with the ability to enlist a mass of current customers of a related service into this new service may refuse to interconnect, expecting to surpass the current market leader and tip the market itself.³²

³¹ Reports in the business press often assess the prospects for success of software/Internet firms in new markets by reference to their current strengths or weaknesses in complementary markets. For example, Microsoft is viewed as particularly strong in that it is able to use its Windows OS platform to “make it easy for customers to use MS Messenger.” In economic terms, these are statements about relative customer acquisition costs. When Microsoft bundled the Internet Explorer browser into Windows 98, it reduced the cost of acquiring a customer for Internet Explorer close to zero; customers could still obtain Netscape, the competing browser, free of charge, but it required a rather lengthy download, so customer acquisition costs were higher for Netscape.

³² This is likely to be an unnecessarily risky strategy on the part of the fast follower; it may fail to surpass its rival and thus find itself on the wrong end of a non-interconnection strategy by the leader. On the other hand, if its strength in complementary markets is strong, it may find that interconnection is the better strategy, as it can count on brand name and customer loyalty to dominate this new market rather than relying on counting on a “surpass and then refuse interconnection” strategy.

- What is the *shape* of the network effects value as a function of share of potential market? As a practical matter, there appears to be two generic types of the value function; the first is the *S-curve*, in which the value per customer is very low for network sizes up to perhaps 20-30% of the potential market; the value increases sharply for network sizes from 20-30% up to perhaps 70%. Adding further customers increases the value, but at a decreasing rate. Figure 1 is a graphical representation of this type of value function:

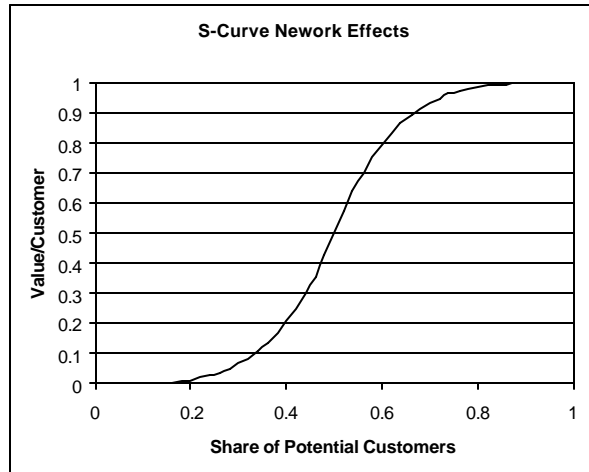


Figure 1

In this particular example, a non-interconnecting firm with market share less than 20% offers virtually no value to customers. If no strategy can be found to grow the network beyond 20%, it is likely that the industry will either die, or be limited to markets with very narrow scope. One such strategy is, of course, interconnection. If the start-up firms agree to interconnect, they could well grow the total network beyond the 20%, and the industry as a whole could then take off. With S-curve network effects, early interconnection agreements are likely.

The second generic type of value of the network effects is *concave*, in that the value as a function of share of potential market, in which the earliest growth produces the greatest value and later growth adding value but at a diminishing rate. Figure 2 is a graphical representation of a concave network effects value function:

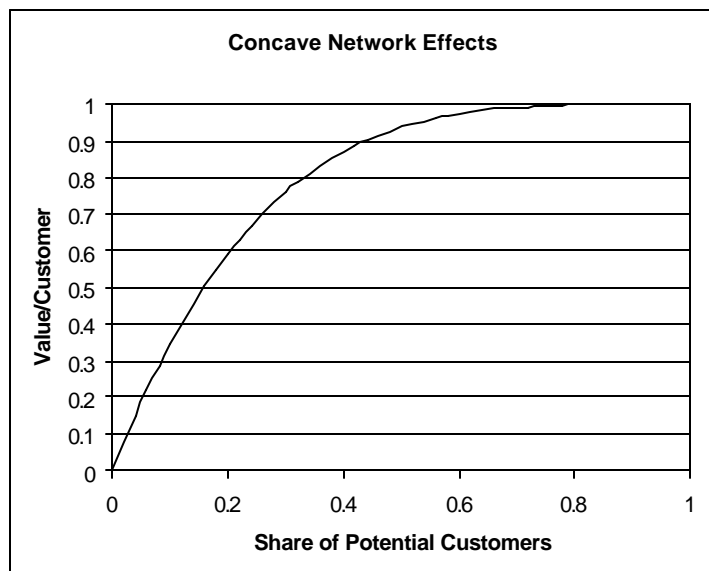


Figure 2

In this case, a firm that grows rapidly gains substantial customer value through its network effects, and has potential to dominate the market at an early stage. In this example, a firm with 20% market share has already captured almost 60% of the network effects value. With concave network effects, refusal to interoperate at an early growth stage is likely, particularly if one firm obtains an early lead while experiencing rapid growth.

In sum, the likelihood of a non-interconnection strategy by the largest player in a growing market of broad scope depends not only on the factors identified for mature markets (market share, network effects strength, customer “stickiness”, and ease of smaller competitors interconnecting), but also two additional factors:

5. The relative ease of customer acquisition, taking account of competitors’ positions in complementary markets; and
6. The shape of the network effects value as a function of share of potential market (S-curve or concave).

Is Tipping Irrevocable? The “Serial Monopoly” Hypothesis

The models above suggest that a dominant firm in a market with broad network effects and favorable conditions can refuse to interconnect and thus “tip” the market, causing customers to leave smaller suppliers and join its network until near-monopoly results. This might be taken to imply that once the dominant firm refuses to interoperate (for strategic reasons) that the process is irrevocable and near-monopoly is a foregone conclusion.

This is not the case; even markets that have tipped to monopoly may in fact be subject to successful competitive attack by new products that are so superior that customers are willing to undertake switching costs, including network effects, to buy the superior product. Of course, this could also occur after the market has tipped and the largest provider is gaining market share. A “killer app” changes the market dynamic and could well reverse the fortunes of the previous market leader. However, the presence of network effects almost always makes this more difficult; it is not sufficient that the new product be superior to the existing product in which the largest provider has, or is about to have, a near-monopoly; it must be so much better as to induce customers to forego the network effect of the old product to try the new. The network effect acts as a barrier to entry; it is not insuperable, nor is the network effect irrevocable. But it requires more than simply a superior product to traverse this barrier to entry; it requires a product so superior that the added benefits of it not only overcome normal customer inertia but also the cost of foregoing network effects as well.

Several authors³³ have argued that in the new economy, this “leapfrogging” is a natural outcome of technology-driven markets. Innovation in these markets is so rapid and revolutionary, the argument goes, that no market leader, even with strong network effects, can defend its position for long against hordes of new entrants with revolutionary products and services. The high-fixed-cost/low-variable-cost nature of the high technology sector, coupled with strong network effects, suggests that at any point in time, monopoly may be the only sustainable market structure. Of course, this is a variant of the “natural monopoly” argument of J.S. Mill, typically used as a justification for regulated monopoly in public utilities. However, the proponents argue that the rapid pace of technical innovation in these industries will challenge and overturn the monopoly in favor of a new technology and eventually a new monopoly. The competitive struggle is *for* the market, not *in* the market. Because the monopoly is never secure in its position, it must always struggle against new entrants and new technologies by constant improvement of its product and service, and respond to constant price pressure. It does this not because of existing competitors in the market, but rather the threat of new competitors with new technologies who may enter the market. This is the “serial monopoly” hypothesis:³⁴ at any time, it is likely that only one firm is in the market; but the threat that they could be overtaken at any time by any entrant disciplines their price, quality and innovation behavior.

³³ The most recent and persuasive article in this literature is David S. Evans & Richard L. Schmalensee, 2001, “Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries,” NBER Working Paper No. W8268, May, at <http://www.nber.org/books/innovation2/evans5-1-01.pdf>. A popular version of the hypothesis is presented in Liebowitz, Stan & Stephen Margolis, *Winners, Losers and Microsoft*, Oakland CA: The Independent Institute, 1999. as well as earlier work of these authors.

³⁴ The serial monopoly hypothesis bears a strong resemblance to contestable market theory of the early 1980’s; in that theory, monopoly was perfectly acceptable as long as barriers to entry were low and new firms could enter at low cost. The monopolist would then be required to maintain low prices and high quality for fear of new entry. In the serial monopoly version of the story, high barriers to entry (in the static sense) are acceptable as long as technological innovation is fast enough so that such barriers to entry are short-lived.

Unfortunately, there is essentially no empirical support for the hypothesis that potential entry by new previously unknown firms either does actually occur or its threat disciplines incumbents. Evans and Schmalensee³⁵ and Leibowitz and Margolis³⁶ quote a number of compelling examples in the recent computer hardware and software markets. But their statements that firms dominating markets that depend upon intellectual property are “fragile,” and subject to being overtaken on a moment’s notice has little empirical support. Of course, this is not to say their hypothesis is incorrect; it is simply unproven.

More compellingly, Evans and Schmalensee³⁷ argue that traditional market definition/market share antitrust analysis is not appropriate for dynamically competitive industries characterized by high rates of innovation. They make the point that markets subject to strong economies of scale and network effects based on risky innovation investments *require* high operating margins protected by short-term barriers to entry, or else investment would dry up. Therefore, static market power analysis based on market definition/market share is bound to find such barriers even when they are absolutely necessary to fuel investment in innovation. They argue, correctly in our view, that “Unlike price/output decisions, analysis of dynamic competition requires evidence about, among other things, the pattern of investment in developing new products (and complements thereof), the control of critical assets (particularly intellectual property and distribution channels), and the beliefs (preferably as revealed by behavior) of market participants and informed observers about the nature and pace of innovation.” (p. 47).

The simplistic serial monopoly hypothesis suggests that in the “new economy,” monopoly is the natural market structure, but technological advance ensures that all monopolies are temporary, soon to be overtaken or at least disciplined by the Next Big Thing. In Evans and Schmalensee’s more nuanced view of the serial monopoly hypothesis, antitrust authorities are not asked to have blind faith in the ability of rapidly advancing technology to solve all problems of monopoly; rather, antitrust authorities are asked to move away from blind faith in static market power/market share analysis to an approach more consonant with dynamic competition. We are in agreement with this view. We also note, however, that developing comprehensive principles and guidelines for antitrust authorities in the area of dynamic competition is extremely difficult, and likely beyond the current capabilities of economists. This paper should be viewed as one small step in this much broader research program.

Inherent in the serial monopoly hypothesis is the argument that an innovator needs a period of monopoly in order to recoup its investment in innovation (including investment in failed ideas as well successful ones). Should competitors be able to immediately enter and the market become fully competitive (in the static sense), then prices would drop and profits driven to zero, thus eliminating future incentives to innovate. Indeed, this is precisely the economic argument that underpins intellectual property law: a limited monopoly is granted innovators as a reward for innovation.

³⁵ *op. cit.*

³⁶ *op. cit.*

³⁷ *op. cit.*

Of course, this raises the question as to how best to protect both innovation and customers; should we encourage patenting such inventions as operating systems and instant messaging services, or should we stay the hand of antitrust authorities, permitting monopolies for what we hope is the short term?

There is no doubt that intellectual property protection is costly to obtain and defend, and often requires years of litigation before a truly clear title to a patent is obtained. In the fast-paced “new economy,” one’s innovation is likely obsolete long before strong patent protection can be obtained. And even once obtained, it could well be very easy to “invent around,” so the innovator’s source of rent can still be eroded by competitive forces. The usefulness of the intellectual property system varies greatly from industry to industry. For example, in pharmaceuticals, and industry known for rapid technological advance, patents are the foundation of the business models of both incumbents and start-ups, although costly to defend. In computer hardware and software as well as Internet services, practices to protect of intellectual property vary enormously, from copyrighting software to patenting business models to the use of trade secrets to proprietary network effects to tying and bundling. Evans and Schmalensee³⁸ argue that a dynamically competitive industry may only be sustainable via actions such as tying that appear highly anticompetitive from a static perspective.

If no intellectual property regime can protect an innovator’s rent, then how should antitrust authorities treat apparent monopolization? The key to resolving this dilemma comes from intellectual property law itself, which grants an innovator a *limited* monopoly for his invention, and requires that knowledge of the technology be shared. The Department of Justice implicitly recognized this in *U.S. v. Microsoft*, objecting not to Microsoft’s Windows monopoly, but only to (i) Microsoft’s exclusionary contracts with OEMs and (ii) Microsoft’s attempts to *leverage* its market power from the Windows OS into new markets such as the browser market (by tying). The government’s case can be understood as attempting to limit the Windows monopoly in product space, much as intellectual property law limits innovators’ monopolies in product space and in time.

The appropriate question for antitrust authorities, in addition to those posed by Evans and Schmalensee, is whether or not a proposed remedy to a perceived anticompetitive practice could reasonably be expected to chill future innovation. This question has both short run and long run components. For example, in its review of the AOL-Time Warner merger, the FCC imposed a condition requiring AOL to interconnect its Instant Messenger service with competitors prior to offering advanced services. Does this order chill future innovation? Clearly, it has two short run effects: discouraging AOL from innovating into such advanced services, and encouraging competitors to do precisely this form of innovation. Its long run effect is more problematic: will firms in the future be more or less likely to introduce new communications-type services knowing that the regulators or antitrust authorities may limit the scope of the resulting monopoly be requiring interconnection? If the fairly extensive period during which AOL was able to refuse interconnection with competitors is viewed as sufficient for AOL to have realized an above-normal return on its innovation investment, then the answer to this question is

³⁸ *op. cit.*

that no chill on future innovation should occur. If future innovators fear that the FCC has moved too quickly, before AOL realized its innovation returns, then the answer is that the long-term effect of this order is to reduce innovation in this field.

For services with network effects, it is often difficult to assess future or even present returns to these services, as new firms often provide the service at low or even zero price. This was the case in both *U.S. v. Microsoft*, in which Microsoft made its browser available for free to Windows customers, and in the AOL-Time Warner Instant Messenger service, in which AOL made IM available free to its customers and free to all others as an Internet download. In such cases, firms can credibly claim that since they give away the service, they earn nothing on it and so the product cannot have monopoly characteristics. This claim is disingenuous, for at least three reasons:

1. The provision of the service as a “feature” of the firm’s product, the service certainly enhances the value of the product and permits the firm to either increase its market share, raise its price, or both. For example, AOL established Instant Messenger as a feature of its basic service package; the popularity of IM no doubt contributed to the rapid growth of the AOL customer base. Similarly, adding the Internet Explorer browser to the Windows OS no doubt increased the value of the Windows OS for customers. In both cases, the firms’ service introductions increased profitability, even if not via pricing the service on a stand-alone basis.
2. The critical profit impact of a network effects market occurs after the firm has rapidly built up its market. One part of the strategy to accomplish this rapid build-up is to give the product away at the early stages, thus minimizing the cost to the customer of becoming loyal to the firm. Giving away the service early, akin to introductory pricing, helps build the customer base upon which the network effects rely. Giving away the service can have a substantial effect on future profits; it is best understood as an investment in proprietary network effects, whose value is realized after the firm dominates the market.
3. In some cases, the firm gives away its service to capture a related market from which it perceives an entry threat to its main market. In *U.S. v. Microsoft*, the government alleged that Microsoft perceived a threat to its Windows monopoly from browsers running Java programs, which could be a platform alternative to the Windows OS. In this scenario, Microsoft sought to control the browser market by giving away its Internet Explorer because it perceived this market to be an avenue of competition to its Windows monopoly and was anxious to shut down that avenue.

In sum, the fact that a firm gives away a service in no way removes the need for antitrust scrutiny. Indeed, it suggests that heightened antitrust scrutiny is called for, especially in a market with very limited extant competition. It is unlikely that the firm is acting out of charity, and unless existing or likely entrants are forcing the firm to offer these services

via market forces, it is likely that such actions signal anticompetitive intent, particularly in the presence of network effects.³⁹

The FCC's Instant Messaging Condition in the AOL-Time Warner Merger

The FCC's analysis of the merger was principally focused on AOL's Instant Messaging service; in particular, the FCC was concerned that merging AOL's IM service in which the market is (likely) tipped with Time Warner's broadband conduits to 20% of US cable homes would enable the merged firm to leverage these assets into next generation IM services to obtain an "instant monopoly" in these new IM services. The FCC therefore imposed a condition that required AOL-Time Warner to interoperate with other advanced IM competitors prior to offering next generation IM services.

Specifically, the FCC imposed the condition that AOL cannot offer advanced IM services until it satisfies one of the following three "grounds for relief:"

1. AOL demonstrates it has adopted a public standard of interoperability established by IETF or official standard-setting body (interoperability by standard).
2. AOL enters into interoperability contracts with at least two IM providers, and stands ready to negotiate in good faith with other IM providers under the identical technical interoperability standards (interoperability by contract).
3. AOL demonstrates that it is no longer dominant in IM, either using market share data or other evidence.

Advanced IM services are defined to entail the transmission of one- or two-way streaming video over Time Warner facilities using AOL's NPD. In addition, this condition sunsets in five years.

The condition suggests the struggle of attempting to deal with future problems in a fast-moving industry. The FCC had no wish to insist that AOL adopt an industry standard when such a standard does not exist now and may never exist, not least as competitors may use the standard-setting process strategically. The FCC did not establish any technical standards of interoperability, preferring to leave that to independent technical standards boards or to the market. Additionally, the FCC wished to account for changing conditions; it could well be that a competitor could come to dominate the IM market as

³⁹ Giving away a product or service for free can be viewed as the ultimate in predatory pricing, long a staple of traditional antitrust analysis. Evans and Schmalensee (*op. cit.*) question the government's definition of predation in the Microsoft case, which relies on the increase in profits from driving one's competitors from the market. They note that in a network effects business, this is precisely the effect of vigorous (and legal) competition. Making this an antitrust violation risks placing a chill on legitimate competitive actions. It is clear that more research on predatory pricing in dynamically competitive industries is urgently required in order that antitrust authorities have the tools they need to deal with the "new economy."

advanced IM services became available, and AOL would not then have market power. If this could be proved, then AOL would be relieved of the interoperability requirement.⁴⁰

Perhaps most important, if there really were no next generation IM services (or none that AOL wished to offer), then there was no interoperability requirement. The FCC gave tacit assent to the “earned monopoly” hypothesis, that AOL was entitled to the rents from its text-based IM innovation.

In contrast to the FTC, the FCC directly addressed this as a “new economy” merger in which interoperation/interconnection of a network effects communications service was central. It therefore faced all the problems discussed in Evans and Schmalensee⁴¹ plus some even more basic evidentiary issues:

1. What evidence is required to prove the service is imbued with network effects?
2. What evidence is required to prove the market has tipped?
3. What evidence is required to prove that the firm refuses to interoperate/interconnect for strategic reasons rather than cost or operational reasons?
4. If the remedy is oriented toward future services, then how can the remedy be structured so that it is effective if the forecasted conditions obtain and has no effect if these conditions do not obtain.

Lessons of the AOL-Time Warner Case

Without question, the two salient issues of this merger were open access (at the FTC) and advanced IM services (at the FCC). Both represent challenges to antitrust analysis that derive from the “new economy:” should we mandate access to new economy essential facilities or do we let the market sort it out? Should we seek to negate a network effects barrier to entry to new entry or recognize such barriers as a necessary part of investments incentive in the new economy?

The FTC answered the first in the affirmative; it accepted the prima facie case that more competition must be good, and adopted what appears to be an aggressively regulatory remedy. The remedy was a popular one; it was well received by consumer groups and the press. Whether customers are now better off than they otherwise would be remains unknown. The costs of this regulatory approach are not likely to be large; however, the precedent of having a monitor trustee provide ongoing supervision of an antitrust conduct remedy is somewhat disturbing. It is reminiscent of Judge Harold Greene’s decade-long supervision of the AT&T breakup, essentially creating another layer of regulation on the

⁴⁰ Of course, if AOL really were not dominant, then economic theory would predict that it’s optimal strategy would be to interoperate. Should they apply to the Commission for relief from interoperability on the basis that they are not dominant, the act of applying appears to refute the premise of the application: they must still be dominant if they wish to continue to refuse interoperation.

⁴¹ *Op. cit.*

telephone industry that quickly became tiresome to even the most ardent supporters of the Court.

The FCC answer to the second question was in the affirmative to both parts. While there was economic analysis aplenty, the case raised troubling issues more generally for antitrust in the new economy.

Proactive vs. Reactive A merger case is necessarily forward-looking (as compared to a conduct case), but in this case the FCC appeared to be reaching to forestall potential harms in as-yet-unimagined service markets. Such a strategy always runs the risk of hyperactive government intervening to prevent imaginary villainy, and some critics thought this was the case here. The FCC was aware of this problem and attempted to structure the remedy so that it would have no effect (and hence no cost) if the hypothesized problems never materialized. Some critics thought that this resulted in a very weak remedy with too many loopholes. This is a trade-off that antitrust authorities acting proactively must resolve: heading off future problems without causing costs if their predictions are wrong, without gutting their remedies.

The alternative is to only act on problems that are immediate and plainly evident. The FTC's remedy on open access is clearly in this vein. But often in fast-moving markets, by the time the antitrust agency recognizes a problem and gears up to solve it, the market has moved on, the remedy based on past actions is no longer appropriate and the damage is already done. Clearly, antitrust agencies must strike this balance as well.

Serial Monopoly and the New Economy Scholars working in the area of antitrust in dynamically competitive industries have argued that the competitive norm in such industries is temporary "winner take all" innovators, succeeded in rapid order by a new "winner take all" innovator, and so forth. The period of monopoly for each innovator is in fact a reward to such innovators and the temporary monopoly rents are merely the quasi-rents to a social beneficial activity. Imposing "old economy" antitrust to break that monopoly in fact will lesson innovation as it lessens the legitimate rewards to that innovation. In this view, temporary monopolies generate the quasi-rents to innovation much as patent protection helps generate quasi-rents for a limited period of time. The unstated assumption of the serial monopoly theorists is that somehow intellectual property protection is not available in these fast-moving markets. But even these scholars admit that anticompetitive behavior can still exist: for example, exclusive distribution contracts with today's monopolist can inhibit the rise of the next monopolist. The current serial monopolist, in this view, cannot use its temporary market power to forestall new innovators and thus turn itself into a permanent monopolist.

The FCC specifically recognized that AOL had earned whatever rents it could garner from its IM innovation, and that network effects barriers to entry could well be part of this rent generation. It also recognized that a temporary monopoly should not be used to foreclose new innovators and thus create a permanent monopoly. Its judgment was that the merger would create just such conditions and sought to forestall such actions. Its remedy was designed to ensure that the next-in-line serial monopolist would not be

foreclosed by the use of merger assets. In essence, it argued that network effects combined with merger assets that dramatically improve deployment capabilities is similar in effect to exclusive distribution contracts; it makes entry by the next serial monopolist almost impossible. Whether the factual case supports the remedy can be argued; but the theory of the case seems well within the confines of antitrust in the new economy. It also illustrates the difficulty of conducting the analysis and the case with those confines.

It also illustrates the dilemma of substituting antitrust enforcement (or non-enforcement) for intellectual property law. It may well be true that IP affords little protection to innovators in fast-moving markets, and perhaps antitrust enforcement needs to recognize the need for temporary quasi-rents to innovators. But the problem appears to be with the inability of IP law to afford protection in important fast-moving markets. This suggests the appropriate policy conclusion is not to alter antitrust enforcement but rather to fix intellectual property law to provide such protection.

Conclusions

The AOL-Time Warner case presented a textbook case of both Access₁ and Access₂, and how two different agencies reacted to each. The FTC took an essential facilities-like approach to the cable conduit, requiring open access as a condition of the merger. The FCC took a network effects approach to advanced IM, requiring (future) interoperability as a condition of the merger. Both conditions can legitimately be called “access” conditions; neither has much to do with the other; both problems exist simultaneously in the merger firm. Yet the problems are very different and their solutions are very different as well.

In the technology-intensive industries of the future, it is likely that essential facilities cases will be rare; technology has a way of finding alternatives to bottlenecks. It is also likely that network effects will play a much larger role than in previous cases. Prior to the FCC’s findings in AOL-Time Warner, there were few cases in which network effects played a role.⁴² In order to support antitrust enforcement in network effects industries, continued research is essential.

⁴² See A. Douglas Melamed, “Network Industries and Antitrust,” 23 *Harv J.L. & Pub Pol’y* 147 (1999).