Antitrust, the Internet, and the Economics of Networks

Christopher S. Yoo  
*University of Pennsylvania Carey Law School*

Daniel F. Spulber  
*Northwestern University*

Follow this and additional works at: [https://scholarship.law.upenn.edu/faculty_scholarship](https://scholarship.law.upenn.edu/faculty_scholarship)

Part of the Antitrust and Trade Regulation Commons, Communications Law Commons, Computer Law Commons, Economic Policy Commons, Internet Law Commons, Law and Economics Commons, Political Economy Commons, Science and Technology Policy Commons, and the Technology and Innovation Commons

**Repository Citation**

Yoo, Christopher S. and Spulber, Daniel F., "Antitrust, the Internet, and the Economics of Networks" (2013). *Faculty Scholarship at Penn Law*. 568.  
[https://scholarship.law.upenn.edu/faculty_scholarship/568](https://scholarship.law.upenn.edu/faculty_scholarship/568)

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.
Chapter 17
Antitrust, the Internet, and the Economics of Networks
Daniel F. Spulber and Christopher S. Yoo*

Network industries have been the subject of some of the most important cases brought under the antitrust laws.1 Government suits against AT&T led to the 1914 settlement following the Kingsbury Commitment, the 1956 settlement that barred AT&T from the computer industry, and most importantly the 1982 Modification of Final Judgment that broke up the largest company in the world. Other cases involving telecommunications companies, such as Trinko and linkLine, have reshaped monopolization doctrine and redefined how antitrust law fits with other parts of the legal regime. The government case against Microsoft spawned the first major use of network economic effects in an antitrust case.2 More recently, enforcement authorities have begun to focus attention on the most recent wave of network-based companies, including Google, Apple, and Facebook.

This Chapter examines the relationship between antitrust and network industries. It begins with an overview of the types of networks before examining the economic considerations generally thought to play a key role in antitrust review of network industries: natural monopoly, network economic effects, vertical exclusion, and dynamic efficiency. It then analyzes the

* Professor Spulber would like to thank the Ewing Marion Kauffman Foundation for research support. Professor Yoo would like to thank the Milton and Miriam Handler Foundation, the New York Bar Foundation, and the University of Pennsylvania Center for Technology, Innovation and Competition for their financial support for this project.

1 For a general study of network industries and the Internet see Spulber and Yoo (2009).

2 Network effects arguments also appeared in Microsoft v. Commission. See Spulber (2008c) for additional discussion.
implications for antitrust policy, focusing on structural separation and the essential facilities doctrine.

A. Types of Networks

There are many types of networks (Spulber and Yoo 2009). Although many of the examples in this Chapter are drawn from communication networks, for the most part the same conclusions apply to other types of networks as well. We can distinguish between physical networks with physical transmission facilities and virtual networks, which include two-sided platforms with technological compatibility.3

Network industries in the Transportation, Utilities, and Information sectors represent 9.1% of GDP. The components of Transportation are air transportation, railroad transportation, water transportation, truck transportation, transit and ground passenger transportation, pipeline transportation, other transportation and support activities, and warehousing and storage. Information is comprised of publishing industries (including software), motion picture and sound recording industries, broadcasting and telecommunications, and information and data processing services. Finally, Utilities includes power generation and supply, natural gas distribution, water, sewage, and sanitary services. Most of these industries are composed in large part of networks. Not all of these elements represent network contributions since communications includes production of content such as programming, and electricity services include the production of electric power. Although the Transportation and Utilities sectors include trucking and

3 We do not consider decentralized social and business networks. Such social and business networks are composed of sets of individuals in a society and the relationships between them. These networks are mechanisms that distribute wealth, transmit information, facilitate business transactions, and form personal relationships (Wasserman and Faust 1994). According to Polanyi (1944, p. 46), “man’s economy, as a rule, is submerged in his social relationships” (see also Granovetter 1973, 1985). The application of graph theory and network design to game theory is related to the theory of social networks (Myerson 1977, Dutta and Jackson 2003). For a discussion of graph theory in the context of the law and economics of communications networks, see Spulber and Yoo (2009).
warehousing, it might be instructive also to consider Wholesale Trade as operating distribution networks, which comprises another 5.6% of GDP. The Wholesale Trade sector includes intermediation activities and management of the distribution network, even though some trucking and warehousing are counted separately as part of transportation. Within the Retail sector, some large retail chains also engage in their own wholesale distribution activities (U.S. Department of Commerce 2012).

Virtual networks include networks of buyers and sellers connected to each other by technological compatibility and two-sided platforms. Such networks include search engines and Internet portals (Google, Bing), social networking platforms (Facebook, Twitter), online marketplaces (Amazon, eBay), computer operating system platforms (Microsoft’s Windows, Apple’s OS, and Google’s Chrome OS), smartphone operating system platforms (Apple’s iOS, Google’s Android, Samsung’s Bada), and video game platforms (Microsoft’s Xbox, Sony’s PlayStation, Nintendo’s Wii).4

B. Natural Monopoly

Perhaps the classic justification for regulating networks is that they have long been presumed to be natural monopolies. A given production technology is said to exhibit natural monopoly characteristics if it has a subadditive cost function, i.e., a single firm can supply the entire market demand at lower cost than could two or more firms. A sufficient condition for subadditivity is for the technology to exhibit scale economies, such as occurs when a production

technology requires the incurrence of joint and common fixed costs that can be spread across units of output or across multiple outputs.

Natural monopoly gives rise to two policy concerns that have served to justify regulation of both entry and prices in network industries. The first concern was that entry would result in cost inefficiencies due to duplication of facilities and the loss of economies of scale. The second concern was that entry regulations to protect a monopoly incumbent would result in monopoly pricing in the absence of additional price regulation. Several important developments in network industries generally have been mitigated or eliminated these policy concerns: technological changes in network architecture, increases in demand for network services, product differentiation, and innovation.

Changes in network architecture that have occurred in some network industries mitigate or eliminate these concerns. To the extent that the scale economies are the product of joint or common fixed costs, these economies can be dissipated either by a reduction in the fixed costs needed to create and operate a network or by an increase the total demand for the services provided by the network. Technological and economic forces have pushed from both of these directions to undermine the natural monopoly rationale.

A classic example is the reduction in fixed costs needed to provide local telephone service. Although the provision of local service involves a large number of components, the two that have required the greatest up-front investment in fixed costs have historically been (1) the wires needed to connect individual residences and businesses to the central office facility maintained by the local telephone company and (2) the switching equipment needed to route
individual calls to their destinations. Empirical scholars have long disagreed over whether local telephone service was in fact subadditive.\(^5\)

Technological change has in effect made this debate an anachronism. The fixed costs needed to provide both switching and transmission have dropped dramatically. The advent of first electronic and then digital switching has caused the fixed cost of switching technologies to plummet. The emergence of wireless alternatives to wireline transmission technologies has resulted in significant reductions in the fixed costs of transmission. The first significant deployment of wireless transmission technologies was the use of microwave relay and satellite systems as substitutes for wireline long distance services. This was followed by the widescale deployment of cellular telephony, personal communication services (PCS), and other wireless technologies that could serve as substitutes for local wireline telephone service. The net result is a dramatic reduction in the fixed costs needed to establish a telecommunications network sufficient to undercut the natural monopoly-based justifications for regulation.

Technological change also has made telecommunications markets contestable by reducing the *sunk costs* associated with market entry. This allows competition for the market so that even if there is a monopoly incumbent and the incumbent firm benefits from significant economies of scale, competitive pressures will drive prices toward average costs. So long as fixed costs are not also sunk costs, any attempt by an existing player to charge supracompetitive prices will only invite hit-and-run entry by firms that gather the available profits and depart as soon as competition drives prices down to competitive levels (Baumol, Panzar, and Willig 1988).

\(^5\) For studies concluding that local telephone service was subadditive, see Charnes, Cooper, and Sueyoshi (1988), Röller (1990a, 1990b), Wilson and Zhou (2001), and Gasmi, Laffont, and Sharkey (2002). For studies drawing the opposite conclusion, see Evans and Heckman (1983, 1984), Shin and Ying (1992), and Berg and Tschirhart (1995).
The effects of sunk costs are even lower in virtual networks, where suppliers have lower entry costs and customers have greater ease of switching in comparison to traditional network markets.

Contestability theory underscores a critical difference between wireline and wireless transmission technologies. Because telephone wires have historically been useless for any other purpose, fixed cost investments in telephone wires can properly be regarded as sunk costs and thus a potential source of barriers to entry. The same is not necessarily true for the infrastructure needed to construct a wireless transmission network. Wireless technologies require equipment located on transmission towers as well as the legal right to use particular portions of the electromagnetic spectrum. Since alternative uses exist for both of these assets (either by other wireless telephone providers or by providers of wireless broadband or other spectrum-based services), investments in wireless network technologies are less likely to be regarded as sunk costs and thus less likely to give rise to the market failures said to be associated with natural monopoly.

The weakening of the natural monopoly justification resulting from the reduction in fixed costs has been accompanied by an increase in the demand for the services provided by communications networks. The emergence of personal computing and the analog modem made it possible for subscribers to use their connections to telecommunications networks to send data as well as voice communications. The increase in functionality made possible by the deployment of digital technologies has mitigated the tendency of telecommunications markets to collapse into natural monopolies by greatly increasing the demand for the network services. These analog technologies are in the process of being replaced with digital technologies, such as digital subscriber lines, and by fiber optics, which are enhancing the value of the network connection still further.
The increasing packetization of communications technologies has also put pressure on the distinctions between transmission technologies, which once made sense when each medium of communications employed distinct analog encoding formats. Voice over Internet protocol (VoIP) made cable modem systems a viable competitor to the voice services provided by local telephone companies, and telephone companies are developing packet-based television services. The deployment of new transmission technologies, such as fourth-generation wireless technology Long Term Evolution (LTE) and WiMax, will increase the competitiveness of last-mile telecommunications services. Once the shift towards packetization is complete, all forms of communications will simply be different applications traveling on the same data network, and the distinctions between transmission technologies will completely collapse.

This combination of reductions in fixed costs and increases the demand for network services has caused much of the telecommunications network to lose its natural monopoly. Multiple facilities-based providers now vie to provide telecommunications to large business enterprises. In addition, intermodal competition from different wireline and wireless technologies has the same effect on the residential and small business market as well.6

Product differentiation also weakens or eliminates the natural monopoly justification for regulation (Spulber 1995). The concept of natural monopoly assumes that products are homogeneous so that cost efficiencies imply the need for a single provider. With product differentiation, there are consumer benefits from having multiple providers offering differentiated products. There is a tradeoff between the benefits of product variety and reductions in economies of scale. Regulatory entry barriers would favor economies of scale, but

---

6 Distribution networks for water, natural gas, and electricity have not benefited to the same extent from improvements in transmission technologies and convergence.
would eliminate the benefits of product differentiation. With multiple providers of network services offering differentiated products, consumers benefit from competition among providers that provide services to address different consumer needs. Competition among providers also involves price competition. Product differentiation allows markets to reach equilibrium with multiple producers each producing on the declining portion of the average cost curve, as was shown by Edward Chamberlin’s seminal work on monopolistic competition (1934). So long as products are differentiated, the existence of unexhausted economies of scale need not necessarily force a network to collapse into a natural monopoly (Yoo 2005). Over time, the introduction of new types of products and services by new entrants generates turnover of market leaders, further weakening the cost-based natural monopoly argument for regulation of network industry.

Innovation also can invalidate the natural monopoly argument for regulation (Spulber 1995). The traditional natural monopoly justification for entry and price regulation is based on a static industry with a given cost function. Technological change in network industries implies that the cost function of network firms changes over time. This weakens the argument that a protected monopoly incumbent generates cost benefits due to economies of scale. Entry of more efficient firms will generate cost efficiencies even if there are reduced economies of scale. There is a tradeoff between the cost efficiencies generated by entry of innovative firms and the potential cost inefficiencies from smaller scale firms. Competition among firms over time addresses the tradeoffs between cost reductions through innovative entry and economies of scale from fewer firms.

C. Network Economic Effects

Another economic characteristic often regarded as a source of market power in network industries is network effects. Network effects exist when the value of a network is determined
not only by the services it provides, but also by the number of users connected to the network.\(^7\) The notion of network effects springs from “Metcalf’s Law,” which states that a network with \(n\) nodes has \((1/2)n(n - 1)\) potential node-to-node connections. In other words, the number of possible connections increases quadratically with the number of users, and presumably so does the number of potential transactions. If each new connection adds value, larger networks will enjoy greater returns to scale than smaller ones (Gilder 1993).

The telephone system has long been regarded as a classic example of a network that exhibits such effects, since the value of a telephone network is determined in large part by the number of people with whom one can communicate through that network. The more people that an individual subscriber can reach through the telephone network, the more valuable the network becomes. Internet-based communications networks exhibit similar networks effects because of the benefits users derive from the ability to communicate with a larger number of other users. Internet-based transaction networks also may exhibit networks effects when users derive benefits from the number of other potential trading partners on the network.

Because the Internet is a network of networks, it is often said to exhibit network economic effects, although upon closer inspection the constantly increasing returns to scale may be limited to specific networks, such as social networks, rather than the Internet itself. The ability to access the entire universe of customers need not benefit specialized businesses serving market niches, although there are advantages to serving the long tail (Brynjolfsson et al., 2006). Moreover, increases in the number of network subscribers represent a mixed blessing for applications designed to provide connectivity in that broader usage also causes congestion. The

\(^7\) More generally, buyers and sellers derive cross-market benefits, which may be the result of product variety and scale effects, market liquidity, and connectivity of communications networks (cf. Spulber 2010).
existence of private networks based on the same protocols as the Internet that nonetheless do not interconnect with the public Internet underscores that many network participants do not derive significant benefits from a larger number of connections. As with natural monopoly, there is a tradeoff between the benefits of a larger network and the benefits of product variety from multiple networks (Yoo 2012b).

Of course, the potential transactions that a network offers are very different from actual transactions. The value of such a network would be weighted by the likelihood of a transaction occurring and the potential benefits of a randomly chosen successful transaction. Thus, while the value may be proportional to the squared number of traders, the proportional weight may be very small indeed. The constantly increasing returns to scale from network effects also implicitly presumes that each new connection has equal value. Empirical research suggests that the distribution of value is far from uniform, with network users placing a disproportionately high value on a small, easily identifiable group of other users (Driscoll 2009, Adams 2012). When that is the case, the marginal returns from increases in network size will be small indeed (Yoo 2012a).

The problem is further compounded, if the likelihood of a successful match decreases with the number of potential trading partners, for the usual reason that search costs are high and more traders can mean greater diversity and greater costs of finding a good match. If these costs increase rapidly, they can outweigh the benefits of having more members so that there can be diminishing returns to larger networks. The intensity and quality of meetings may be better at a small party than a large gathering for example. However, with many people connected to a network there are returns to mechanisms that reduce search costs. If such mechanisms exhibit economies of scale, this can restore the benefits of larger networks. For example, with many
people connected to a telephone network, there are benefits from telephone directories. With many people on the Internet, there are benefits from establishing search engines. However, these are subtle concepts that are very different from the idea of simply squaring the number of people with access to a network.

Some economists contend that network effects can give rise to a kind of *externality* that can be a source of market failure. This view draws on the fact that an individual subscriber’s decision to join a network creates benefits for those who have already joined the network that the new subscriber cannot internalize. New subscribers’ inability to capture all of the benefits generated by their adoption decisions arguably creates a wedge between private and social benefit that may cause subscribers to forego joining a network even though the social benefits of doing so would exceed the costs. The concept of network externality thus suggests that network industries may be uniquely susceptible to market failures that may prevent the price mechanism from playing its usual role in generating efficient outcomes.

The market failure argument is based on the notion that market actors cannot coordinate their network participation decisions. However, there are many mechanisms through which market actors can coordinate their participation decisions so as to realize the benefits of using the services of a particular network. In addition, market intermediaries including the network companies themselves provide coordination through pricing, marketing, and provision of incentives for participation such as first-party content (Spulber 2006, 2008a, 2010, Hagiu and Spulber 2012). Through direct coordination or intermediation, market participants realize the benefits of network effects and reduce or eliminate potential inefficiencies. This implies that networks effects do not involve externalities, that is, economic effects outside of market transactions.
Another policy concern expressed by some antitrust economists is the problem of technology lock-in. The argument is that network effects cause markets to become “locked in” to existing technologies long after the arrival of new, more efficient network technologies. However, market coordination by market participants themselves, by market intermediaries, and by network firms again addresses these network effects and mitigates concerns regarding the ability of markets to facilitate adoption of new technologies (Spulber 2008b). Understanding market coordination of technology adoption decisions helps to explain the apparent rapid technological change and the entry and exit of firms in network industries.

Some theorists have also suggested that network effects can turn network access into a competitive weapon. By refusing to interconnect with other networks, network owners can force subscribers to choose one network to the exclusion of others. The fact that each network’s value increases with the number of subscribers connected to it provides a powerful incentive for new subscribers to flock to the largest network. According to this view, network economic effects create demand-side scale economies that can be a source of monopoly power (See, e.g., Katz and Shapiro 1985, Melody 2002). One oft-cited example of this phenomenon is the attempt by the Bell System to use its refusal to interconnect to combat the emergence of competition in local telephone service following the expiration of the initial Bell telephone patents in 1893. Presumably, refusing to interconnect with independent local telephone system would protect the Bell System’s market share by ensuring that it would remain the largest (and thus the most valuable) local telephone provider.

A review of the theoretical literature reveals that arguments that network economic effects inexorably lead to market failure are too simplistic (Yoo 2002). Even if, for the sake of argument, network effects were to create externalities, a consumer’s decision to switch to a new
network actually gives rise to two distinct and countervailing effects. On the one hand, the
decision to join a network enhances the value of the new network for those already connected to
that network and those who will join that network in the future. The inability to capture all of the
benefits created by its network adoption decision would give rise to a positive externality that
can cause a consumer to refuse to join a new network even when it would be socially beneficial
for it to do so, a phenomenon sometimes called lock in or excess inertia.

At the same time, the decision to join a new network also lowers the value of the old
network by reducing the number of people using it. In effect, switching to a new network
imposes costs on those connected to the old network that the person switching networks does not
bear. This may make that individual willing to adopt a new technology even when the costs to
society exceed the benefits, a situation variously called excess momentum or insufficient friction.
It is thus possible that network effects could make users too reluctant as well as too willing to
change networks. Which is the case depends upon which of these two countervailing effects
dominatest (Farrell and Saloner 1986a, Katz and Shapiro 1992). As already argued, coordination
among market participants helps to address either of these potential consequences of network
effects.

In addition, the alleged market failures identified by some economic models depend on
the assumption that the relevant markets are either dominated by a single firm or highly
Conversely, in markets composed of a small number of similarly sized networks, individual
networks have strong incentives to interconnect (Katz and Shapiro 1985, Faulhaber 2005,
Vanberg 2009). An implication of these models is that policymakers should undertake a detailed
examination of the structure of the relevant market before relying on network economic effects to impose antitrust liability.

The economics literature indicates that private ordering may well prove quite robust in solving any problems created by network effects (Spulber 2008a). One major reason is that with respect to telecommunications networks, potential network effects arise through direct connections to a physical network that is established and owned (Katz and Shapiro 1985, Farrell and Saloner 1985). Thus, even though individual users may not be in a position to internalize all of the benefits created by their network adoption decisions, the network owner will almost certainly be in a position to do so. The existence of a single network owner allows potential problems associated with network effects to be solved by placing property in the hands of a single owner and protecting it with well-defined property rights (Coase 1960, Hardin 1968). Any benefits created by network participation can thus be internalized and allocated through the direct interaction between the network owner and network users (Liebowitz and Margolis 1994, Spulber 2008a). The benefits of property rights in addressing network effects extend to markets with competition among network owners (Liebowitz and Margolis 1994).

Even assuming for the sake of argument that circumstances permit network effects to give rise to the problems of monopolistic dominance and technological lock-in, other features of the market and the structure of consumer preferences might mitigate, if not eliminate, these adverse effects. For example, the market may also dislodge an existing network technology so long as the new network provides additional value that exceeds the value derived from the size of old network (Farrell and Saloner 1986b, Katz and Shapiro 1994). This particularly true, given that, after networks have captured a sufficient number of subscribers, the marginal benefit from
adding another subscriber is likely to be low, which would greatly reduce the magnitude of any potential externality.

Heterogeneity of buyer preferences and product differentiation affect the outcome in markets with network effects. As Michael Katz and Carl Shapiro (1994, 106) have noted, “Customer heterogeneity and product differentiation tend to limit tipping and sustain multiple networks. If the rival systems have distinct features sought by certain customers, two or more systems may be able to survive by catering to consumers who care more about product attributes than network size. Here, market equilibrium with multiple incompatible products reflects the social value of variety.” With network effects, expectations of market actors affect participation rates but coordination among market participants can shape expectations (Hagiu and Spulber 2012). In a growing market, participation will reflect anticipation of the extent of the network in the future rather than the current extent of the network (Katz and Shapiro 1992, Shapiro 1995, Liebowitz and Margolis 1996).

The existence of large users may further mitigate any problems caused by network effects. If a single user controls a significant portion of the network, that user would be able to internalize more of the benefits of its adoption decision, which would help minimize any slippage caused by the existence of the network externality. Furthermore, because large users are in a position to capture a disproportionate share of the benefit resulting from the adoption of a new technology, they have significant incentives to make the investments needed to shift towards the new technology (Katz and Shapiro 1994). Indeed, formal models of such market structures indicate that “the sponsor of a new technology earns greater profits than its entry contributes to social welfare. In other words, markets with network externalities in which new
technologies are proprietary exhibit a bias towards new technologies” (Katz and Shapiro 1992, 73).

Determining the optimal number of networks and the optimal timing of technological change requires a careful balance of the relevant costs and benefits. Antitrust policymakers should bear in mind the relevant tradeoffs. A single network can involve monopoly rents, but can also offer benefits from standardization, interoperability, complementary products, and lower transaction costs. Conversely, competition lowers firms’ market power while increasing transactions costs and reducing some of the benefits of standardization, interoperability, and complementary products. Accordingly, some delay in the introduction of new products may reflect efficiency, not market failure.

The fact that markets seem fully capable of resolving most of the potential market failures identified by the theoretical literature on network economics suggests that any attempt to remedy these supposed problems through regulation and antitrust should be approached with considerable caution. Indeed, it would seem appropriate to insist on empirical proof that such problems actually exist before considering governmental action to redress them. More detailed examinations of the facts surrounding the examples of anticompetitive problems stemming from network economic effects typically cited in the literature raise serious questions about their empirical foundations (Liebowitz and Margolis 2001, Spulber 2008a, 2008b).

Lastly, even proof of the existence of the necessary empirical preconditions for network-induced market failure would not necessarily support intervention. Consider, for example, the manner in which antitrust law would attempt to solve the problems of technological lock-in. Such intervention would necessarily require the government to replace winners in the real-world technological marketplace with what it believed represented the superior technology. Moreover,
in order to be effective, the government must do so at an early stage in the technology’s
development, when making such determinations is the most difficult. Courts would also
typically have to make such determinations on extremely thin information that in most cases
would be provided by parties with a direct interest in the outcome. In addition, decisionmakers
would have to insulate themselves from the types of systematic biases traditionally associated
with political decisionmaking processes. It is for these reasons that even supporters of network
externality theories caution that governmental intervention might well make the problem worse,
ot better (Katz and Shapiro 1994, Bresnahan 1999).

D. Vertical Exclusion

The possibility of market power in one market raises the possibility that a network
provider could use its market power in that market to harm competition in a vertically related
market. This could be done either through vertical integration or through a vertical contractual
restraint involving exclusive dealing, territorial restrictions, tying, or some similar restriction.

As noted above, technological developments have substantially reduced the likelihood
that any network provider will wield monopoly power in many network industries. The same
forces that are increasing the competitiveness of every portion of the telecommunications
industry eliminate the plausibility that any network provider will have a dominant market
position to use as leverage over an adjacent market.

At the same time, the conventional wisdom with respect to vertical exclusion has
undergone a sea change over the past half century. While economic theorists during the 1950s
and 1960s were quite hostile toward vertical integration, vertical integration is now generally
recognized to be less problematic than previously believed. (For overviews, see Chapter 30 and
Yoo 2002.)
The driving force behind this transformation is the emergence of the so-called “one monopoly rent theorem,” which holds that monopolists have little, if any, incentive to engage in vertical exclusion. Because there is only one monopoly profit available in any vertical chain of production, a monopolist can capture all of that profit without having to resort to vertical integration simply by charging the monopoly price (Director and Levi 1956, Bowman 1957). Moreover, it is impossible to state a coherent theory of vertical exclusion unless two structural preconditions are met. First, the firm possesses monopoly power in one market (typically called the primary market), since without such power the network owner would not have anything to use as leverage over the upstream market for complementary services (Director and Levi 1956). Second, the market into which the firm seeks to exercise vertical exclusion (called the secondary market) must be protected by entry barriers. If no such barriers to entry exist, any attempt to raise price in the secondary market will simply attract new competitors until the price drops back down to competitive levels (Posner 1976). Unless these structural preconditions are met, the most that vertical integration would do is rearrange distribution patterns.

Since that time, the “post-Chicago” school of antitrust law and economics has used game theory to identify exceptions to the one monopoly rent theorem under which vertical integration can harm competition. Most of these exceptions are the result of highly stylized assumptions that do not match well with these industries (Yoo 2008). Even more interestingly, these models presuppose the existence of dominant-firm and oligopoly market structures in the primary market, which necessarily presuppose that both the primary and secondary markets are highly concentrated and protected by entry barriers (See, e.g., Salinger 1988, Hart and Tirole 1990, Ordover, Salop, and Saloner 1990, Riordan 1998). In the absence of such structural features, these formal models recognize that vertical integration may be just as likely to lower prices and
increase welfare and that the ability of existing players or new entrants to expand their outputs will be sufficient to defeat any attempt to increase prices above competitive levels (Salinger 1991, Riordan and Salop 1995). The post-Chicago literature has thus done little to disturb the basic conclusions that vertical integration is unlikely to harm competition unless the relevant markets are concentrated and protected by entry barriers.

Economic theorists have increasingly recognized that vertical integration could yield substantial efficiencies. For example, if two layers of a vertical chain of distribution are monopolistic or oligopolistic, firms in each layer will have the incentive to try to extract the entirety of the available supracompetitive returns, which could lead to a final price that would be even higher than the monopoly price. Vertical integration can eliminate this so-called double marginalization problem, since a company that spans both layers would rationalize the decisionmaking between the two levels of production and would avoid the uncoordinated action that would make supracompetitive pricing even worse (Spengler 1950).

In addition, to the extent that the inputs can be used in variable proportions, any attempt to charge supracompetitive prices for one input creates incentives for firms to substitute alternative inputs whenever possible. The resulting substitution creates an alternative potential source of inefficiency, as production processes deviate from the most efficient input mix. Allowing the provider of the monopolized input to vertically integrate into manufacturing can allow it to prevent inefficient input substitution (Vernon and Graham 1971). The welfare implications of input substitution are ultimately ambiguous, since prohibiting input substitution enhances the monopolist’s ability to exercise market power, which can create welfare losses sufficient to offset the welfare gains from preventing customers from deviating from the most efficient input mix (Schmalensee 1973, Hay 1973, Warren-Boulton 1974).
Finally, scholars building on Coase’s seminal work on the theory of the firm (1937) have demonstrated how vertical integration can reduce transaction costs. One example is the elimination of free riding. For example, suppose that a firm manufactures a technically complicated product that requires significant presale services (such as the demonstration of the product). Telser (1960) argues that retailers will have the incentive to shirk in providing such services in the hopes that other retailers will bear the costs of providing such services. If all retailers respond to these incentives in the same way, the total amount of presale services will fall below efficient levels. A manufacturer facing the possibility of such free riding can either rely on a vertical contractual restraint that specifies the level of presale services that each retailer is required to offer or can vertically integrate into distribution. Either solution effectively aligns the retailers’ incentives with the manufacturers’. The Supreme Court embraced precisely this rationale in *Sylvania*.

Determining whether a particular form of vertical integration will enhance or reduce economic welfare is thus an empirical question that turns on the particular market structure and the nature of the available efficiencies. A recent survey of the empirical literature on vertical integration found that the existing studies overwhelmingly support the proposition that vertical integration and vertical restraints tend to promote, rather than harm, competition (LaFontaine and Slade 2007). The conventional wisdom has now largely abandoned its hostility toward vertical integration. The manner in which technology is in the process of increasing the competitiveness of network industries and the real efficiencies from vertically integration has effectively undercut the threat of vertical exclusion as a justification for imposing antitrust liability.

---

8 For surveys of the empirical literature on vertical restraints that draw similar conclusions, see Cooper et al. (2005) and LaFontaine and Slade (2008).
E. Dynamic Efficiency

All of the rationales discussed up to this point focus on the most efficient way to organize the network that already exists. In focusing on allocating the network that already exists today, these rationales overlook the equally (if not more) important question of how to create incentives to invest in new network technologies that will comprise the optimal network of tomorrow. In other words, antitrust may have to tolerate some degree of static inefficiency in order to promote dynamic efficiency.

As one noted treatise observes, U.S. antitrust law reflects “a uniquely American, market-affirming response to [market] power” that “assumes that strong incentives promote efficiency” and that in the absence of entry barriers, market power “will erode under the pressure of market developments.” Indeed, high prices can play a key role in promoting competition, because “where supracompetitive pricing accompanies power, erosion of the power is thought to be more likely because high prices signal the need, and promise a reward, for entry” (Sullivan and Grimes 2006, 84–85). In effect, short-run static efficiency losses must be tolerated if necessary to promote long-run dynamic efficiency gains from innovation and successive entry.\(^9\) Imposing antitrust liability whenever firms earn supracompetitive returns would eliminate the primary impetus for competitive entry, in which case the supply curve would never shift outwards in order to bring the market back into long-run equilibrium.

Moreover, forcing network owners to share the benefits of their investments with their competitors or limiting the prices they can charge their channel partners would dampen the incentives to invest in alternate network capacity. In effect, forcing a monopolist to share its

\(^9\) For a discussion of antitrust and dynamic efficiency, see Sidak and Teece (2009).
network rescues other firms from having to provide or obtain the relevant input for themselves. As a result, it can have the perverse effect of entrenching a network monopoly by forestalling the emergence of the substitute capacity. This is particularly problematic in technologically dynamic industries, in which the prospects of developing new ways either to circumvent or to compete directly with the alleged bottleneck are the highest. Such a surrender to the monopoly only makes sense if competitive entry is infeasible.

Although some scholars have asserted that because the dynamic efficiency gains will be compounded over time, they will necessarily exceed the short-run static efficiency losses (Ordover and Baumol 1988), this approach seems too simplistic. Whether the dynamic efficiency gains will dominate the static efficiency losses depends on the relative magnitude of the gains and losses, the speed of entry, and the appropriate discount rate. That said, a number of institutional considerations militate in favor of the dynamic efficiency side of the balance. For example, calibrating the prices needed to implement rate regulation and access regulation will necessarily require the government to engage in an exquisite exercise in line drawing that requires a careful and fact-intensive balance of opposing considerations. This is made all the more complicated by rapid changes in network technology and in consumer demand for network services.

Antitrust policy must carefully balance these offsetting considerations. The alternative would be to allow the short-run supracompetitive returns to stimulate entry by alternative last-mile providers whenever such entry is possible. It is for this reason that courts and policymakers have been reluctant to compel access to a resource is available from another source, even if it is only available at significant cost and in the relatively long run (Areeda and Hovenkamp 2006).
F. Implications for Antitrust Policy

The economic critique has important implications for the doctrines and remedies that antitrust courts have applied to network industries. In particular, these insights affect the reliance on structural separation as a remedy and on the essential facilities doctrine as a cause of action.

1. Structural Remedies/Vertical Separation

One of the most common remedies sought during antitrust litigation in network industries is structural separation. For example, the 1956 consent decree settling the second major case against AT&T abandoned the government’s initial request for divestiture of AT&T’s equipment subsidiary, opting instead to restrict AT&T to furnishing common carrier communications services. The 1982 court order that broke up AT&T required that the local telephone services that remained monopolized be structurally separated from the portions of the business in which competition had become possible: long distance, telephone equipment, and “information services” that combined transmission with data processing. More recently, the federal government initially asked the court hearing the case against Microsoft to require the company to spin off its applications businesses into a separate subsidiary.

The rationale generally given for imposing structural separation is that rate regulation may encourage firms to vertically integrate into unregulated lines of business that they can bundle with the regulated service and for which they can charge the monopoly prices denied them by regulation (See, e.g., U.S. Department of Justice 1984, *Jefferson Parish* (O’Connor, J., concurring in the judgment)). In such cases, it is arguably appropriate to impose what is sometimes called the “Bell Doctrine” or “Baxter’s Law,” which prohibits vertical integration in order to isolate and quarantine the monopolist (Joskow and Noll 1999, Farrell and Weiser 2003).
A more subtle version of this argument exists when a firm subject to cost-of-service rate regulation provides an unregulated service that shares joint costs with the regulated service. In that case, a monopolist can attempt to allocate a disproportionate amount of the joint costs onto the regulated service and recover those costs through its regulated rates, which is a possibility given the inherent arbitrariness and uncertainty surrounding methodologies for allocating joint costs. Shifting a disproportionate amount of the joint costs onto the regulated market also allows firms to reduce the size of the markup to cover joint costs included in the prices charged in the unregulated market. The firm’s nonvertically integrated competitors, however, charge prices that reflect costs and competition in the unregulated market (Farrell 1996, Sullivan 1996, Huber, Kellogg, and Thorne 1999, Joskow and Noll 1999).

The traditional solution to these problems is to prohibit network providers from offering any unregulated services. A less severe restriction is to require network providers to segregate their regulated services into a separate subsidiary and to require that subsidiary to offer carriage to others on the same terms that it offers carriage to itself. Structurally separating the regulated business from the unregulated business limits the network owner’s ability to shift costs from its unregulated service to its regulated service. Forcing the regulated subsidiary to negotiate interconnection agreements through arm’s length transactions also eliminates the ability to use bundling of regulated and unregulated services to avoid rate regulation and makes nondiscrimination easier to detect and enforce.

Structural separation requirements have proven exceedingly difficult to administer. For example, the court overseeing the implementation of the 1982 decree breaking up AT&T was bombarded with hundreds of requests to waive the decree’s line of business restrictions (Huber, Kellogg, and Thorne 1999). These requests often took from three to four years to process, with
estimates of the total welfare loss associated with these delays exceeding $1 billion (Rubin and Dezhbakhsh 1995, Hausman 1997).

The logic inherent in this approach suggests that regulation and the resulting quarantine should be limited only to those services that are inherently noncompetitive. Doing so would allow competition to determine outcomes to the greatest extent possible. This rationale thus presumes that the scope of regulation should constantly contract over time, as technological change and increases in demand open larger portions of the telecommunications industry to competition.

This dynamic quality of the scope of regulation explains much of the history of telecommunications policy. As noted earlier, initially the entire telephone system was regarded as a natural monopoly. As a result, the entire network was subjected to rate regulation. Over time, it became clear that equipment manufacturing was not subject to the high fixed costs traditionally associated with natural monopoly and was instead potentially open to competition. The emergence of microwave and satellite transmission also lowered the fixed costs of providing long distance service by allowing new long distance entrants like MCI to compete without having to establish a nationwide network of high-volume telephone lines. As it became clear that each of these markets was now open to competition, the FCC released them from rate regulation and prohibited regulated entities from offering them on an integrated basis. Eventually, the only portion of the telecommunications industry that remained subject to rate regulation was local telephone service. Over time, even that premise has come under fire, as wireless has reduced the fixed costs needed for transmission and as computer processing has reduced the fixed costs of switching. The growing constriction of the scope of rate regulation
and monopoly power has effectively curtailed the analytical foundations for mandating structural separation.

2. Behavioral Remedies/The Essential Facilities Doctrine

Perhaps the most commonly asserted basis under the antitrust laws for challenging network owners’ supposed attempts to engage in anticompetitive conduct is the essential facilities doctrine (see also Chapter 23, Spulber and Yoo 2009). The doctrine requires owners of bottleneck elements unavailable elsewhere in the marketplace to provide other firms with access to those elements on reasonable terms. Given the economic forces discussed above, it should come as no surprise that early leading cases associated with the doctrine (such as *Terminal Railroad* and *Otter Tail*) arose in network industries.\(^{10}\) In addition, both the 1956 settlement of the federal government’s antitrust suit against AT&T as well as the litigation that led to the 1984 breakup of AT&T were based on the essential facilities doctrine.

The central concern of the essential facilities doctrine is thus vertical exclusion (Werden 1987, Areeda and Hovenkamp 2006). As such, lower courts have incorporated into the doctrine the same economic considerations discussed above. Courts have applied the same structural preconditions to the essential facilities doctrine: concentration in the primary market (see, e.g., *MCI Communications Corp. v. AT&T Co.*) and concentration and entry barriers in the secondary market (see, e.g., *Alaska Airlines*). Moreover, courts applying the essential facilities doctrine consider whether exclusion from the monopoly might be justified by efficiencies (see, e.g., *City of Vernon v. Southern California Edison Co.*, *Illinois ex rel. Burris v. Panhandle Eastern Pipe Line Co.*, *Byars v. Bluff City News Co.*).

---

\(^{10}\) Although widely regarded as the seminal essential facilities case, *Terminal Railroad* is more properly regarded as arising from horizontal concentration rather than vertical exclusion (Spulber and Yoo 2009).
Finally, courts have recognized the importance of dynamic efficiency. In the words of the Ninth Circuit in *Alaska Airlines*, “[e]very time the monopolist asserts its market dominance” by denying rivals access to a bottleneck input gives the rival “more incentive to find an alternative supplier, which in turn gives alternate suppliers more reason to think that they can compete with the monopolist. Every act exploiting monopoly power to the disadvantage of the monopoly’s customers hastens the monopoly’s end by making the potential competition more attractive.”

The Supreme Court offered its most complete discussion of the tradeoff underlying vertical exclusion and the essential facilities doctrine claims in *Trinko*. Although the Court found “no need either to recognize . . . or to repudiate” the doctrine, even assuming *arguendo* that the doctrine applied, the Court found it inapposite. As an initial matter, the Court acknowledged that the essential facilities doctrine “serves no purpose” when the input to which access is sought is available through other means, in effect embracing the structural precondition requiring concentration in the primary market. The Court also held that any claim of vertical exclusion must show “a ‘dangerous probability of success’ in monopolizing [the] second market.” This language has been widely interpreted as requiring proof of market concentration and entry barriers in the secondary market before a monopolization claim will lie (Areeda and Hovenkamp 2006, Kauper 2005).

The *Trinko* court also explicitly recognized the importance of dynamic efficiency considerations, noting how the prospect of earning short-run supracompetitive returns can stimulate upgrades to the network infrastructure. In the words of the Court, “The mere possession of monopoly power, and the concomitant charging of monopoly prices, is not only not unlawful; it is an important element of the free-market system. The opportunity to charge
monopoly prices—at least for a short period—is what attracts ‘business acumen’ in the first place; it induces risk taking that produces innovation and economic growth.” Using antitrust law to require network owners to share that infrastructure “may lessen the incentive for the monopolist, the rival, or both to invest in those economically beneficial facilities.” Imposing antitrust liability on those who invest in such facilities “seem[ed] destined to distort investment.” The Court did not acknowledge that that denying others access to the facilities created with such investments could sometimes be unlawful. Its reasoning suggested that such remedies were limited to a narrow range of circumstances with preexisting business arrangements or when a network owner foregoes short-run profits by refusing to deal with competitors even when they are willing to pay full (presumably monopolistic) prices.

In addition, the *Trinko* Court weighed into the longstanding debate on the relative merits of structural and behavioral relief when it emphasized the problematic nature of the essential facilities doctrine in terms of administrability. In the words of the Court, “[e]nforced sharing also requires antitrust courts to act as central planners, identifying the proper price, quantity, and other terms of dealing.” Furthermore, because mandated access affects network elements “deep within the bowels” of a local telephone network, they can only be made available if “[n]ew systems [are] designed and implemented simply to make that access possible.” Additionally, requests for unbundled access “are difficult for antitrust courts to evaluate, not only because they are highly technical, but also because they are likely to be extremely numerous, given the incessant, complex, and constantly changing interaction of competitive and incumbent [local telephone companies] implementing the sharing and interconnection obligations.” The “uncertain virtue of forced sharing and the difficulty of identifying and remedying anticompetitive conduct by a single firm” suggested that the addition of an antitrust remedy
would do little to promote consumer welfare while posing a significant risk of deterring welfare-enhancing behavior. In so doing, Trinko echoed the criticism that the culmination of an antitrust case in such a regulatory decree is tantamount a confession that the case should never have been brought (Posner 1970, Posner and Easterbrook 1981).

Many commentators have concluded that little, if anything, remained of the essential facilities doctrine following Trinko (see, e.g., Fox 2005, Noll 2005, Geradin and O’Donoghue 2005, Shelanski 2007, Frischmann and Waller 2008, Hovenkamp 2008). Moreover, Trinko contradicted the assertions of many commentators that the essential facilities doctrine had greater relevance in the context of regulated industries (Werden 1987, Areeda and Hovenkamp 2006). To the extent that application of the doctrine depends on a monopoly in the primary market, the advent of competition in many network industries radically narrows its scope.

Conclusion

Network industries have exhibited significant growth, substantial competition, and rapid innovation, including Internet-based e-commerce. Some have argued that various economic theories, including natural monopoly, network effects, and vertical exclusion, suggest the need for heightened antitrust scrutiny of network industries in comparison with other industries. The present discussion suggests that other aspects of network industries argue for regulatory and antitrust forbearance, or at least suggest that such industries should not be the subject of heightened scrutiny. Technological changes in communications, product differentiation, and the effects of innovation imply that regulatory limits on entry into network industries do not promote consumer benefits. At the same time, natural monopoly need not create first-mover or incumbent advantages, so that natural monopoly should not justify greater antitrust scrutiny of network industries. Coordination among market participants, by the participants themselves, by
network intermediaries, and by network owners suggest that network effects are not a source of market failure and should not justify heightened antitrust scrutiny of network industries. The observation that vertical integration and vertical restraints tend to promote, rather than harm, competition applies to network industries. Again, vertical exclusion arguments should not justify heightened antitrust scrutiny of network industries.

Our discussion strongly suggests that antitrust policy makers should not presume that network industries are more subject to monopolization than other industries. Rapid innovation and firm turnover in network industries imply that antitrust policy makers should take into account dynamic efficiency considerations in evaluating competition in network industries.
REFERENCES


Dutta, Bhaskar and Matthew O. Jackson (2003), *Networks and Groups*, New York: Springer.


Granovetter, Mark (1973), Strength of Weak Ties, *American Journal of Sociology* 78: 1360–1380.


**CASES CITED**

*Alaska Airlines, Inc. v. United Airlines, Inc.*, 948 F.2d 536 (9th Cir. 1991).

*Byars v. Bluff City News Co.*, 609 F.2d 843 (6th Cir. 1980).


*City of Vernon v. Southern California Edison Co.*, 955 F.2d 1361 (9th Cir. 1992).


*MCI Communications Corp. v. AT&T Co.*, 708 F.2d 1081 (7th Cir. 1983).


United States v. AT&T Co., Decrees and Judgments in Federal Antitrust Cases, July 2, 1890-January 1, 1918, at 483 (D. Or. Mar. 26, 1914)).


