University of Connecticut

Department of Economics Working Paper Series

Technological Standards, Innovation, and Essential Facilities Toward a Schumpeterian Post-Chicago Approach

Richard N. Langlois
University of Connecticut

Working Paper 1999-07

December 1999

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

In this essay, I attempt to take seriously Schumpeter's perspective on competition as fundamentally about innovation. Drawing on literatures that concern themselves centrally with the patterns and processes of technological change, I focus on a set of issues very much on the present-day agenda: antitrust policy toward network industries in which technological standards are important. As both scholars and legal cases have suggested, one might logically view a set of standards as an ”essential facility” - a technological bottleneck - for those who wish to connect to the network.. I attempt to define the limits of the standard price-theoretic account for understanding the problem of essential facilities and offer instead a perspective drawing on the theory property rights in a regime of innovation. Contrary to what is suggested by traditional economic analysis, I argue that, as a logical matter, refusals to deal by essential-facility monopolists are not always equivalent to the exercise of existing monopoly power through price, and there are good theoretical reasons for an essential facility doctrine to concern itself with refusals to deal even when it fails to touch other exercises of market power by a legally acquired monopoly. I introduce the concept of the scope of an essential facility, understood in analogy with a similar concept in the economics of patents, and suggest that the degree to which antitrust policy should concern itself with the ownership or control of a technical standard ought to be proportional to the scope of the standard. At the same time, however, a Schumpeterian perspective reminds us that, in a world of dynamic technological competition, even possession of a standard with wide scope may afford only temporary protection, and the winds of Schumpeterian creative destruction may be a better bulwark against monopoly than the cumbersome and interest-laden processes of antitrust law and policy. Nonetheless, the notion of the scope of a standard may prove useful in many cases, including those involving regulated (or formerly regulated) industries or involving tradeoffs in intellectual property rights.

> Political controversy to one side, new technology should not be allowed to obscure an old truth. The basic problem is a rerun of the issues for rails and telecommunications: can outsiders connect to the network?
>
> — Richard Epstein[1]

Introduction.

Economics is a conservative discipline. Even its revolutions do not stray far from basic accepted principles, and sometimes they merely reassert basic accepted principles. The Chicago School of antitrust analysis scored victory after victory over the Harvard School by an assiduous application of static neoclassical price theory.[2] A new challenger, the so-called Post-Chicago approach, accepts the same basic principles, even as it dresses them up in the elegant if improbable clothes of mathematical game theory.[3] What unites these approaches is a devotion to a concept of competition oriented fundamentally around price and quantity. Other dimensions of competition may sneak in, such as competition for "quality" in the abstract. But price theory deals poorly with what is arguably the central impulse of actually existing competition: technological change. "In capitalist reality as distinguished from the textbook picture," wrote Joseph Schumpeter, the most important form of competition arises not from competition "within a rigid pattern of invariant conditions, methods of production and forms of industrial organization," but rather from

---

1 R. Epstein, Principles for a Free Society: Reconciling Individual Liberty with the Common Good 306 (1998).

2 See *infra*, note 27, and surrounding text.

3 See *infra*, note 33, and surrounding text.

“the new commodity, the new technology, the new source of supply, the new type of organization ... — competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of existing firms but at their foundations and their very lives.”[4]

In this essay, I attempt to take Schumpeter’s perspective seriously. Drawing on literatures that concern themselves centrally with the patterns and processes of technological change, I focus on a set of issues very much on the present-day agenda: antitrust policy toward network industries in which technological standards are important. As both scholars and legal cases have suggested, one might logically view a set of standards as an “essential facility” — a technological bottleneck — for those who wish to connect to the network..

Scholars have for the most part displayed skepticism of if not hostility to the essential facility doctrine on both legal and economic grounds.[5] The doctrine adds nothing new to the legal arsenal, critics have claimed, nor does the idea draw on any economic insights not gained from a traditional analysis of monopoly behavior. It is a central contention of this paper that there is a distinctive logic to a doctrine of essential facilities. But that logic becomes clear only by moving beyond standard price-theoretic accounts of the behavior of

---

4 J. Schumpeter, Capitalism, Socialism, and Democracy 84 (3d. ed. 1950).

5 See for example P. Areeda, Essential Facilities: An Epithet in Need of Limiting Principles, 58 Antitrust L. J. 841, 841 (1990); J. Ratner, Should There Be an Essential Facilities Doctrine? 21 U. C. Davis L. Rev. 327, 342 (1988); D. Reiffen and A. Kleit, Terminal Railroad Revisited: Foreclosure of an Essential Facility or Simple Horizontal Monopoly? 33 J. Law & Econ. 419 (1990); and D. Gerber, Rethinking the Monopolist's Duty to Deal: A Legal and Economic Critique of the Doctrine of “Essential Facilities,” 74 Virg. L. Rev. 1069 (1988).

firms to consider head-on the phenomenon of innovation. The pace and direction of innovation may not be the central issue in all "bottleneck" cases. But when technological standards are involved, we can presume that the dimension technological change will typically be at least as important as those of price and quantity.

In what follows I attempt to define the limits of the standard price-theoretic account for understanding the problem of essential facilities and offer instead a perspective drawing on the theory property rights in a regime of innovation. Contrary to what is suggested by traditional economic analysis, I argue that, as a logical matter, refusals to deal by essential-facility monopolists are not always equivalent to the exercise of existing monopoly power through price, and there are good theoretical reasons for an essential facility doctrine to concern itself with refusals to deal even when it fails to touch other exercises of market power by a legally acquired monopoly. I introduce the concept of the *scope* of an essential facility, understood in analogy with a similar concept in the economics of patents, and suggest that the degree to which antitrust policy should concern itself with the ownership or control of a technical standard ought to be proportional to the scope of the standard.

At the same time, however, a Schumpeterian perspective reminds us that, in a world of dynamic technological competition, even possession of a standard with wide scope may afford only temporary protection, and the winds of Schumpeterian creative destruction may be a better bulwark against monopoly than the cumbersome and interest-laden processes of antitrust law and policy. Nonetheless, the notion of the scope of a standard may prove useful in many cases, including those involving regulated (or formerly regulated) industries or involving tradeoffs in intellectual property rights.

Background.

Since the early 1980s, a doctrine of "essential facilities" has gained popularity in legal argument and judicial decision-making. Broadly speaking, the doctrine is this: "an owner of a crucial input cannot deny access if a firm seeking access cannot practicably obtain the input elsewhere."[6] Most accounts trace the idea to the 1912 case of *U. S. v. Terminal Railroad Association,* in which a consortium that included financier Jay Gould controlled railroad facilities for crossing the Mississippi at St. Louis.[7] The doctrine has also surfaced in cases involving such "bottleneck" inputs as a professional sports stadium, warehouse space in Providence, and a joint marketing arrangement for Colorado ski slopes.[8]

Recently, however, a number of cases have emerged in which the facility claimed essential is in the nature of technological knowledge and in which the

6 Ratner, *supra* note 5, at 330. Epstein, *supra* note 1 at 302, cites a more expansive formulation: "The essential facilities doctrine requires a firm with monopoly power in one market to deal equitably with competing firms operating in adjacent markets that depend on it for essential inputs." M. K. Kellogg, J. Thorne, and P. W. Huber, Federal Communications Law 12 (1992). The distinction between merely granting access and "dealing equitably" is an important one to which I return.

7 224 U. S. 383 (1912). For an analysis of this case, see Reiffen & Kleit, *supra* note 5.

8 *Hecht v. Pro-Football, Inc*., 570 F.2d 982 (D.C. Cir 1977), cert. denied, 436 U.S. 956 (1978); *Gamco, Inc. v. Providence Fruit and Produce Bldg., Inc*., 194 F.2d 484 (1st Cir.), cert. denied, 344 U.S. 817 (1952); and *Aspen Skiing Co. v. Aspen Highlands Skiing Corp*., 472 U.S. 585 (1985).

access desired is in the nature of connection to a network.[9] This category includes a number of cases involving physical networks like electricity or telecommunications, where there are clear elements of natural monopoly and the presence of explicit regulation.[10] More interestingly, perhaps, there have also been cases in which plaintiffs have desired access to what is in effect a "virtual" network, that is, a network in which participants are linked together by their economic complementarity and adherence to common technological standards rather than by physical interconnection. Two examples have involved the Eastman Kodak Company. In one case, a competitor in the production of film sued for access to the technical specifications for a new camera Kodak was developing, arguing that Kodak's dominance in film conferred on it the ability to set a universal standard for film and that conformance to such a standard was essential for any competitor.[11] More recently, the Supreme Court considered a case involving Kodak's refusal to sell replacement parts for its photocopiers to customers who purchased repair services from firms other than Kodak itself.[12] Since parts are standardized to a particular brand of copiers, competing repair

---

9 See *BellSouth Advertising v. Donnelley Information Publishing*, 719 F. Sup. 1551, 1566 (S.D. Fla. 1988). ("Although the doctrine of essential facilities has been applied predominantly to tangible assets, there is no reason why it could not apply . . . to information wrongfully withheld. The effect in both situations is the same: a party is prevented from sharing in something essential to compete.")

10 *Otter Tail Power Co. v. United States*, 410 U.S. 366 (1973); *MCI Communications Corp. v. American Tel. & Tel. Co.*, 708 F.2d 1081 (7th Cir. 1983), cert. denied, 464 U.S. 891 (1983)..

11 *Berkey Photo Inc. v. Eastman Kodak Co.*, 603 F.2d 263 (2d Cir. 1979), cert. denied, 444 U.S. 1093 (1980)..

12 *Image Technical Serv., Inc. v. Eastman Kodak Co.*, 903 F.2d 612, 616 n.3 (9th Cir. 1990), aff'd, 112 S. Ct. 2072 (1992). A District Court had issued a summary judgment in favor of Kodak, which was overturned on appeal. The Supreme Court upheld the reversal.

organizations are "locked in" to Kodak's standards for parts, plaintiffs argued, which standards then become an essential facility.

There is reason to think that cases involving technical standards as essential facilities are likely to rise in prominence with the continued advance of microelectronics and the Internet. One can clearly read many aspects of the ongoing cases against Microsoft as involving implicit (and sometimes explicit) claims that the technical standards of a computer operating system constitute an essential facility. And a federal district court recently granted a preliminary injunction to a maker of computer workstations that had claimed (among other things) that Intel's refusal to provide timely information on the specifications of its newest microprocessors amounted to foreclosure of an essential facility.[13]

Richard Epstein has recently given attention to the doctrine of essential facilities within a broader discussion of the law and economics of common carriers.[14] In general, he argues, rights of private property and freedom of contract constitute an institutional structure superior to that of government regulation, since (among other reasons) competition can usually be counted on to limit the arbitrary or exploitive exercise of private property rights. But in situations of impacted monopoly, notably those in which one contracting party controls facilities with the character of a public good, an exception to the general

---

[13] *Intergraph Corp. v. Intel Corp.*, CV 97-N-3023-NE (N. D. Ala.) 1998. The Federal Circuit vacated the injunction on November 5, 1999. (*Intergraph Corp. v. Intel Corp.*, 98-1308, U.S. Fed. Cir.).

[14] *Supra* note 1.

rule arises. Epstein argues that Anglo-American common law has always been sensitive to this exception, and that one of its important responses has been the law of common carriers. Broadly speaking, that body of law requires owners of bottleneck facilities to grant access on reasonable and non-discriminatory terms.[15] Historic examples included coaches and inns; modern examples include railroads and oil pipelines.

The doctrine of essential facilities thus constitutes a kind of special case of the logic of common carriers. As such, it has an intellectual heritage that stretches back before *Terminal Railroad.* Epstein points to the 1810 English case of *Allnut v. Inglis*.[16] Here the plaintiff, a wine importer, refused to pay the fee asked by a London customs warehouse that had the exclusive right to store wine from overseas. Finding for the plaintiff, Lord Ellenborough held that, although in general "every man may fix what price he pleases upon his own property or the use of it," the possessor of a monopoly, "if he will take the benefit of the monopoly, … must as an equivalent perform the duty attached to it on reasonable terms."[17]

Such a duty is especially clear in the case of a statutory monopoly like the customs warehouse in *Allnut*. Here the duty of a common carrier arises in essence out of contract: as part of the acceptance of a monopoly from the

15 Reinterpreting this in economic terms, a common carrier cannot refuse to deal, cannot price discriminate, and (depending on one's interpretation) cannot charge a full monopoly price.

16 12 East 525, 104 Eng. Rep. 206 (K. B. 1810).

17 Ibid. at 538, 104 Eng. Rep. at 210-211, cited in Epstein, *supra* note 1, at 283.

government, one agrees to serve all comers on reasonable and non-discriminatory terms.[18] Nonetheless, the law following *Allnut* viewed *de facto* as well as *de jure* monopoly as potentially falling under the principles of common carriers, even though the *de facto* case raises difficult issues both of practice and of concept. In the words of a seventeenth-century treatise cited in *Allnut*, even private monopoly can sometimes be "affected with the public interest."[19]

The problem of *de facto* monopoly arises also in present-day application of an essential facility doctrine. Although the doctrine has featured in cases with a clear regulatory contract,[20] in most cases the facility claimed essential has been under strictly private control. From the perspective of antitrust law, it is not so much the private character of the monopoly that is the issue as it is the extent to which the bottleneck monopoly was legally acquired. It is a well-established principle, encapsulated in a famous dictum of Learned Hand in the *Alcoa* case, that antitrust liability should not fall on a firm that gained market dominance through its own "superior skill, foresight, and industry." In such a case, wrote Hand, "a strong argument can be made that, although the result may expose the public to the evils of monopoly, the [Sherman] Act does not mean to condemn

[18] The idea of a "regulatory contract" has lately come into policy discourse in the context of the widespread deregulation of industries once thought of (and regulated) as natural monopolies. To the extent that such deregulation involves a repudiation by the government of its end of the deal (the statutory monopoly), deregulation may involve a taking under the meaning of the constitution. See for example J. Sidak and D. Spulber, Deregulatory Takings and Breach of the Regulatory Contract, 71 N. Y U. L. Rev. 851 (1996).

[19] Lord Matthew Hale (1609-1676), De Portibus Mari (posthumously published in the 1780s), cited in Epstein, *supra* note 1, at 282-283.

the resultant of those very forces which is its prime objective to foster: finis opus coronat. The successful competitor, having been urged to compete, must not be turned upon when he wins."[21]

The problem of private *de facto* monopoly affects the essential facility quite directly. The doctrine envisages *refusals to deal* by a monopolist, not the charging of a monopoly price. To put it differently, the issue is primarily access, not the terms of access.[22] Thus the doctrine does not in principle attempt to regulate the pricing behavior of a legally acquired monopoly. But why then insist on open access? The answer at some level must be that denying access is somehow worse than merely exploiting an existing monopoly through price. Exclusion must somehow extend or go beyond the legitimate exploitation of the legally acquired monopoly. This proposition is a fulcrum on which the doctrine of essential facilities turns.

The economics of vertical restraints.

The issues here fall into the broad area of vertical restraints, on which there is a large literature. The central question has long been this: can a firm with monopoly power in one market "leverage" that power into another market by

---

[20] See cases cited *supra* note 10. Moreover, the essential facility in *Terminal Railroad* was part of a system under federal regulation.

[21] *United States v. Aluminum Co. of America*, 148 F.2d 416, 430 (2d Cir. 1945).

[22] Of course, charging an exorbitant price can be a *means* of denying access, and courts have sometimes seen high pricing as a refusal to deal. L. Sullivan & A. Jones, Monopoly Conduct, Especially Leveraging Power from One Product or Market to Another, in Antitrust, Innovation, and Competitiveness 165, 177 (T. Jorde and D. Teece eds., 1992), citing *U.S. v. Western Electric*, 846 F.2d 1422, 1428 (D.C. Cir 1988).

some kind of vertical relationship (including the refusal to engage in a vertical relationship)?[23] It originally seemed obvious to many that this was so, a view given intellectual veneer (if not necessarily substance) by the writings of Joe Bain and the so-called Harvard School in the 1950s.[24] Indeed, Williamson describes the coalescence of these intuitions into an "inhospitability tradition" in antitrust policy: whenever a (vertical) contract appears non-standard — that is, whenever it involves anything other than the selling of goods at a linear spot price[25] — one must presume the contract has an anticompetitive motivation.[26]

In what is certainly the central drama in the recent intellectual history of antitrust, the package of intuitions and traditions represented by the Harvard School came under successful attack by a Chicago School wielding the bright sword of neoclassical price theory.[27] Chicago's logic here may be distilled into

---

23 By "vertical" one means a relationship between entities in successive locations in the chain of production, for example the relationship of a manufacturer with an input supplier or of a wholesaler with a retailer. This is in contrast to "horizontal" relationships involving competitors at the same level in the chain.

24 Meese refers to this as the "populist" view. A. Meese, Price Theory and Vertical Restraints: A Misunderstood Relation, 45 UCLA L. Rev. 143 (1997).

25 "Linear" pricing is when the buyer's expenditure on a good is exactly proportional to the amount purchased, with the constant of proportionality being the price. In far less opaque terms, it is the simple per-unit pricing we are all familiar with. This stands in contrast to multi-part tariffs and other "non-linear" pricing schemes, which I touch on below. On this terminology, see J. Tirole, The Theory of Industrial Organization 170 (1988).

26 O. E. Williamson, Assessing Vertical Market Restrictions: Antitrust Ramifications of the Transaction Cost Approach, 127 U. Penn. L. Rev. 953, 959 (1979).

27 R. Posner, The Chicago School of Antitrust Analysis, 127 U. Penn. L. Rev. 925 (1979). In fact, however, one could argue that the Chicago School was actually going beyond traditional price theory into what is now called the New Institutional Economics. Indeed, it is arguably the inhospitability tradition that, in seeing the non-standard as anticompetitive, drew on neoclassical price theory. On this point see R. Langlois, Contract, Competition, and Efficiency, 55 Brooklyn L. Rev. 831, 836 (1989); and Meese, *supra* note 24.

what is called the “chain-link” or “fixed-sum” theory of monopoly.[28] A firm that has monopoly power at one level in a chain of production can at best transmit that monopoly to other levels; it cannot create more monopoly power than it already has.[29]

Consider tying arrangements, a classic form of vertical restraint. Suppose that a firm with monopoly power ties the purchase of the good over which it has power to the purchase of another (usually, but not necessarily, complementary) good. For concreteness, consider the tying of the purchase of a copying machine to the purchase of toner and paper. The monopolist can extract the maximum amount of surplus by charging the optimal monopoly price for copiers. But the monopolist cannot increase that surplus by raising the price of the tied paper and toner over competitive levels, since to do so would be to raise above the optimal level the effective price to consumers of the copying machine. Why then engage in tying? The answer may involve either a desire to “meter” output in order to price discriminate or perhaps to protect brand-name assets by preventing the user from employing inferior complements that would degrade the quality of service.[30] Both of these increase economic efficiency, at least so

---

28 For the chain-link metaphor, see Gerber, *supra* note 5, at 1085, citing H. Hovenkamp, Economics and Federal Antitrust Law 191-202 (1985). For the fixed-sum terminology, see for example L. Kaplow, Extension of Monopoly Power Through Leveraging, 85 Colum. L. Rev. 515 (1985).

29 See generally R. Bork, The Antitrust Paradox: A Policy at War with Itself 225-245 (1978); Posner, *supra* note 27 at 926-927; F. Scherer & D. Ross, Industrial Market Structure and Economic Performance 522 (3d. ed. 1990); and Tirole, *supra* note 25, at 170.

30 Price discrimination as an explanation of tying was first offered by A. Director and E. Levi, Law and the Future: Trade Regulation, 51 Nw. U. L. Rev. 281 (1956) and W. S. Bowman,

long as the complementary inputs are used in fixed proportions.[31] The Chicago result generalizes to other kinds of vertical arrangements, including vertical integration. These cannot (under Chicago assumptions) increase existing monopoly power; and most have plausible efficiency explanations, such as the resolution of efficiency-impeding free-rider externalities.[32]

The Chicago School's victory quickly invited attack from so-called Post-Chicago critics, some of whom claim to have established a post-Chicago synthesis.[33] These critics largely accept that vertical arrangements can have efficiency rationales and that there is a fundamental truth to the fixed-sum-monopoly logic. Yet, they claim, there exist circumstances under which, and perspectives from which, vertical arrangements can have anticompetitive effects as well.[34] The principal criticism of the Chicago approach has been that, in

---

Tying Arrangements and the Leverage Problem, 67 Yale L. J. 19 (1957). On quality control and other possible efficiency justifications for tying, see W. Shughart, The Organization of Industry 478-479 (2d. ed. 1997).

31 Scherer & Ross, *supra* note 29, at 566. When substitution is possible among inputs, vertical restraints can sometimes lead to allocative inefficiency, but the models are complex and the results ambiguous. See ibid. at 522-526, 566-567; and D. L. Kaserman and J. W. Mayo, Government and Business: the Economics of Antitrust and Regulation 307-310 (1995). See also D. Reiffen and M. Vita, Comment: Is There New Thinking on Vertical Mergers? 63 Antitrust L. J. 917, 924 (1995). ("Although these models show that some profitable vertical mergers may result in price increases, they also show that such mergers may generate price decreases.")

32 L. G. Telser, Why Should Manufacturers Want Fair Trade? 3 J. Law & Econ 86 (1960); H. P. Marvel, Exclusive Dealing, 25 J. Law & Econ. 1 (1982).

33 H. Hovenkamp, Antitrust Policy after Chicago, 84 Mich. L. Rev. 213 (1985); J. Baker, Recent Developments in Economics that Challenge Chicago School Views, 58 Antitrust L. J.645 (1988); M. Riordan and S. Salop, Evaluating Vertical Mergers: A Post-Chicago Approach, 63 Antitrust L. J. 513 (1995).

34 That is, the post-Chicago position is only that vertical restraints should be treated under a rule of reason rather than as preemptively or per se legal. There is essentially no support

relying on conventional price theory, Chicago has taken an excessively static view. Price theory assumes a particular structure of the economic problem the monopoly faces. Within such a given structure, the monopoly can do no more than make the most of its existing monopoly. But often, as Louis Kaplow suggests, the "firm's motivation is to *change the structural conditions it faces* in the future in order that it may receive greater profits in the future."[35] And by changing those underlying structural conditions, the monopoly may well be able to "leverage" itself into a position even more powerful than the one from which it started.

To the extent, however, that the post-Chicago critics have added a dynamic element to the static analysis of Chicago, they have done so along one very particular and narrow dimension, namely, that of strategic analysis.[36] And, just as Chicago prevailed with the superior fire power of price theory, the critics — who have indeed established themselves now as the dominant mainstream — forged an even more high-tech weapon: mathematical game theory. Rather than confine attention to a single period, such strategic analysis looks at the behavior of firms in multiple periods and asks whether behavior in the short run might be explained not by advantages to be gained immediately but by advantages to be

---

among economists for the position that vertical arrangements of any kind should be per se illegal — as some still are under the law.

35 Kaplow, *supra* note 28, at 524, emphasis added.

36 For the characterization of the limits of the Chicago position as a failure to include strategic considerations (rather than as a more general inattention to dynamics and change), see O. Williamson, Antitrust Enforcement: Where It's Been, Where It's Going, 27 St. Louis U. L.J. 289 (1983) as well as references cited *supra* note 33.

gained in a future period. These models tend to be highly stylized, and they normally portray firms as possessing cognitive and maximizing abilities even more formidable than those upon which the Chicago School insisted. Nonetheless, strategic behavior (even in the narrow sense) is surely important to some extent in business practice, even if that extent may be much less than is suggested by the superabundance of game-theoretic models in the economics literature.

At the risk of only a little oversimplification, let me characterize the essence of the strategic criticism of Chicago in a way that will prove useful for understanding the issues of essential facilities and will also help make clear the limits of a strategic perspective — or at least the limits of a strategic perspective blinkered in by game theory. The core of the fixed-sum-monopoly argument is that it construes the firm's problem as that of short-run profit maximization. Vertical arrangements without independent efficiency rationales only impede that goal, and therefore are not "rational" in the short run. But if the firm is looking to a longer-run position, it may be rational to earn less than full monopoly profits in the short run in hope of greater future gain. But this reduces to a claim that the strategic firm is engaging in *predatory behavior* in order to disadvantage its rivals. Such behavior would include not only predatory pricing but also non-price actions that might (at some short-run cost) serve to eliminate or simply disadvantage rivals in the long run.[37]

The logic of strategic vertical restraints thus mimics that of predatory pricing. Pre-Chicago analysis took it for granted that firms with deep pockets would want to lower prices below short-run costs to drive rivals out of business, which would then allow such firms to recoup their short-run losses and more. Indeed, charges of predatory pricing go back to some of the earliest antitrust cases, including *Standard Oil*.[38] The Chicago school has been critical of the possibility of predatory pricing, which they view as typically an irrational (that is, non-optimal) strategy.[39] Post-Chicago critics would cast this as another instance of Chicago's myopic attention to the short run (in which predation is costly) to the neglect of the long-run (in which predation might pay off).[40] But I think it is possible to paint the Chicago criticism in a slightly different light, one that raises some serious issues for game theory: it is not so much that the critics of predatory pricing focus on that strategy's short-run costs but rather that they highlight the often ephemeral character of the long-run benefits. In the game-theoretic approach to strategic behavior, the future structural conditions (to use Kaplow's term) that firms try to influence today appear every bit as clearly and

---

37 Kaplow, *supra* note 28; W. Comanor & H. Frech, Strategic Behavior and Antitrust Analysis, 74 Am. Econ. Rev. 372 (1984); J. Ordover and R. Willig, An Economic Definition of Predation: Pricing and Product Innovation, 91 Yale L. J. 8 (1981).

38 *Standard Oil Co.* v. U. S., 221 U. S. 1 (1911), on which see J. S. McGee, Predatory Price Cutting: The Standard Oil (N.J.) Case, 1 J. Law & Econ. 137 (1958).

39 L. G. Telser, Cutthroat Competition and the Long Purse, 9 J. Law & Econ. 259 (1966).

sharply as do the conditions today.[41] To those who actually make the decisions, however, these possible future conditions appear far more hazy — they are but guesses. And firms do not try to execute closed-form optimal strategies but rather try to position themselves in ways they hope will prove favorable. The more uncertain the world, the harder to tell predation from dynamic competition.

Another question that arises in analyzing predatory behavior is to what extent there are structural barriers to prevent (re)entry once a rival has been eliminated or severely disadvantaged.[42] In many cases, for example, causing a rival to exit does not destroy the physical assets the rival once used to compete, leaving those assets available to new competitors. As Areeda and Hovenkamp put it, "[a]ny price is lawful once it appears that the prerequisites for successful predation — especially the ability to maintain monopoly prices after rivals are destroyed — are absent. In that event, predation is not likely to be present and, even if it were, there would be no 'dangerous probability' that monopoly would result."[43] The question of barriers to entry returns us very near the problem of

---

40 Kaplow, *supra* note 28, at 527-528.

41 Indeed, far from being highly uncertain, predators in most game-theoretic models have information superior to that of other players. See J. Ordover and G. Saloner, Predation, Monopolization, and Antitrust, in Handbook of Industrial Organization 537 (R. Schmalensee & R. Willig eds. 1989).

42 This is the basis for the "structural" test for predation advocated by P. Joskow and A. Klevorick, A Framework for Analyzing Predatory Pricing Policy, 89 Yale L. J. 213 (1979). See also K. Elzinga & D. Mills, Trumping the Areeda-Turner Test: the Recoupment Standard in Brooke Group, 62 Antitrust L. J. 559 (1994).

43 P. Areeda and H. Hovenkamp, Antitrust Law: An Analysis of Antitrust Principles and their Application, 1992 Supplement 631 (1992).

essential facilities. For, as Harold Demsetz has argued in an important paper, barriers to entry always boil down to property rights, whether *de jure* or *de facto*.[44] And the ownership of an asset essential to production may well constitute such a barrier.

The economics of essential facilities.

As a particular kind of vertical restraint, the denial of access to an essential facility raises both static and dynamic issues. From the static (or short-run) point of view, the issues revolve around whether *denying* access is rational behavior for a monopolist that could have extracted maximum surplus simply by *charging a monopoly price* for access.[45]

The logic here is the same as for tying. Suppose (to use the paradigm example) that a private monopolist controls the only possible access across a river, and that the monopolist secured its position legally. Suppose in addition that the monopolist also owns one of many competitors that must cross the river; for example, suppose the monopoly owner of a railroad bridge also owns one of several railroads that wish to cross the bridge The monopolist does not need to exclude competing railroads in order to maximize its rents; it only need charge all railroads a monopoly price to cross. Charging the competitors high prices will certainly advantage the railroad owned by the monopolist, but that will not increase the monopolist's total take. Even if the monopolist excluded all

44 H. Demsetz, Barriers to Entry, 72 Am. Econ. Rev. 47 (1982).

competing railroads, it would still be limited to the total rents available because of its ownership of the bottleneck. Indeed, the monopolist would want to charge its (now monopoly) railroad division an internal transfer price equal to what would have been the competitive price to cross the river, thus avoiding the problem of "double marginalization" that occurs if successive stages of production are both separately monopolized.[46]

Why then would an essential-facility owner ever want to refuse to deal with rivals? One answer is that a simple linear price to cross the river might not fully capture all the rents to be had from the bridge monopoly. Indeed, consumer surplus in the train example arises not because of anyone values the passage of the trains themselves but rather because consumers ultimately value the passage of the goods (widgets, let us say) those trains carry. To the extent that a per-train fee appears to shippers as a fixed cost, the fee won't enter into the railroad's marginal cost per widget and therefore (under competitive conditions) will not figure into the per-widget price the railroad charges widget-shippers. This may mean that the railroads operate at a loss, which may in turn imply that the bridge owner will have to reduce its fee to retain customers. But it certainly means that the bridge owner will not be able to capture all the surplus of the widget-shippers. In such a case, the bottleneck monopolist may

---

45 See for example Reiffen and Kleit, *supra* note 5, and Gerber, *supra* note 5.

46 J. Spengler, Vertical Integration and Antitrust Policy, 58 J. Political Econ. 347 (1950).

prefer to refuse to deal with competitors of its railroad so that it can earn all its profits from its (now monopoly) railroad rather than from the bridge.[47]

Before saying that refusals to deal can therefore be anticompetitive, one has to deal with two very different issues. The first is that there may well be pricing alternatives that could extract the full amount of surplus. One can avoid the problem described above by using a "nonlinear" price, that is, a two-part (or in general a multi-part) tariff. Instead of asking a fee per train, the bridge owner could ask a crossing fee (which doesn't depend on the number of widgets that cross) plus a per-widget "royalty."[48] Such a two-part tariff can approximate perfect price discrimination, and is normally efficiency enhancing.[49] Herein may lie the difficulty, however, as price discrimination is illegal under the Robinson-Patman Act.[50]

---

47 Notice that this strategy requires that there be barriers to reentry, which arise by hypothesis in this case from the "essential" — not easily duplicated — character of the bottleneck stage. The possession of the bottleneck asset is thus the exclusionary property right alluded to above. In a case in which there is no such barrier, the anticompetitive character of the vertical integration is less clear. At worst, competitors are disadvantaged because they would have to enter simultaneously at both stages. But if we assume that there are capabilities for entry already available at both stages, then what makes entry more difficult in the integrated case is only the cost of coordinating those existing capabilities into a unified entry by vertical partners, through a joint venture, for example.

48 I use this term deliberately to highlight that pricing problems of this sort abound in knowledge-based industries. For example, a book author who receives an advance plus royalties on copies sold is being compensated by a two-part tariff.

49 Tirole, *supra* note 25 at 146. The two-part tariff is also preferable to a per-widget charge alone, since "the monopolist can reduce the marginal price below the monopoly price and recoup lost profits through the fixed fee. The fixed fee thus induces the monopolist to lower prices, which is good for welfare." (Id.)

50 Gerber, *supra* note 5, at 1084.

The more difficult issue is this. If a refusal to deal is a method of extracting monopoly rent in situations in which appropriate pricing is either impossible or illegal, why should such refusals be treated any differently from the setting of a monopoly price? If the monopoly was acquired through "superior skill, foresight, and industry," then why should one method of exploiting the monopoly be any more anticompetitive than another? Indeed, a doctrine that permits the charging of a monopoly price but penalizes refusals to deal (or other non-price mechanisms of extracting monopoly rents) may encourage firms to evade sanction by inefficiently transforming facilities into forms that allow low-cost metering and pricing.[51]

There are two possible directions of response. One would be to reopen the question of whether an essential facility doctrine should consider price as well as access. This direction would return the essential facility doctrine to its common-carrier roots and transform it into an alternative mechanism for the regulation of privately owned public goods. But taking this path immediately leads to two barriers. The first is the matter of practicability: it may be far easier for antitrust policy to affect access than to regulate price, since the latter implies both a far greater information requirement and some kind of ongoing surveillance. Traditionally, price regulation in such cases has taken the form of administrative oversight, a procedure that, although not without costs of its own,

---

[51] Gerber, *supra* note 5, at 1093.

may be better suited to ongoing surveillance than are judicial processes.[52] The second barrier is the one implied in the Learned Hand dictum, namely the issue of incentives.[53] Using a logic similar to that lying behind the standard economic defense of patents, one could argue that to prohibit a firm from exploiting the benefits of a legally acquired monopoly is to discourage the very sort of briskly competitive behavior that is fundamental to economic efficiency.[54] (I return to this point later in the context of innovation.)

A second direction of response would be to limit the essential facility doctrine to situations in which the firm's behavior serves not to exploit a given monopoly position but somehow to extend that position — to create "more" monopoly power than was envisaged in the original (legally acquired) monopoly.[55] And, as Kaplow suggests, this can occur only when the firm's behavior somehow allows it to affect the structural conditions it faces rather than merely allowing it to maximize a fixed pie of profit.[56] In this situation, however,

---

52 Victor Goldberg has argued that one should view administrative regulation of this sort as a mechanism for carrying out an ongoing "administered contract" between the owner of the essential facility and its customers. See Goldberg, Regulation and Administered Contracts, 7 Bell J. Econ. 426 (1976).

53 *Supra,* note 21.

54 One may want to encourage the private creation of essential facilities because, even though the charging of a monopoly price is inefficient relative to the charging of a competitive price, a situation in which the facility exists (even if monopolized) is preferable the situation in which no facility exists. As in the patent case, the deadweight loss of monopoly is the price for whatever remaining social surplus the facility's existence generates.

55 In economic terms, this would presumably mean reducing welfare below what would have been implied in the exploitation of the legally acquired monopoly. It is logically possible that extending a monopoly could generate an *increase* in social surplus, for the same reason that the creation of the original monopoly might have done so. (Cf. *supra* note 54.)

56 Kaplow, *supra* note 28 at 524.

game theory has not provided a satisfactory strategic account of a situation in which the owner of a genuinely essential facility can benefit from attempting to monopolize a downstream market, since the owner of such a facility always benefits from greater competition in the downstream market through its monopoly upstream. It is only when the owner of an input is not fully a monopolist that vertical restraints can be anticompetitive, since those restraints can help increase the input owner's market power.[57] The owner of an input that is truly essential already has all the market power possible.[58]

Essential facilities and innovation.

Fortunately, the strategic account from game theory does not circumscribe the possible ways in which a firm might attempt to change the structural conditions it faces today in the hope of future gain. For example, a firm may try to change structural conditions by controlling the rate and direction of innovation in the system of which the essential facility is a part. Although such behavior is clearly "strategic" in some sense, it does not always take place in a world in which the possibilities of action fall neatly into the form of a game whose detailed contours and potential stakes are easily knowable in advance.[59]

---

57 Ordover & Willig, *supra* note 37; M. Whinston, Tying, Foreclosure, and Exclusion, 80 Am. Econ. Rev. 837, 840 (1990).

58 Of course, in many real-world essential facility cases, the facility claimed essential may not be literally the only alternative. For example, an operating system with more than 90 per cent of the market is an essential facility for most users, but it still has competitors. I return to this point below.

59 For a relevant attempt to cast innovation in such a form, however, see Ordover & Willig, *supra* note 37.

As a useful entry point to this argument, consider the analogy with the law and economics of patents suggested earlier. The dominant account of the economics of patents stresses the role of patents in inducing innovation. By creating a property right and thus a barrier to entry, a patent holds up to a would-be inventor the carrot of economic rents. According to this account, such a lure is necessary because of the peculiar nature of inventive knowledge: although its creation entails a fixed investment, the marginal cost of transmitting the knowledge is zero. Knowledge is also non-rivalrous, in that one person's consumption of it does not diminish the ability of others to consume it; and knowledge may also be difficult to exclude others from or to charge for.[60] Invention, in short, is a public good much like a bridge. Without the right to exclude people (and thereby receive economic rents), no one would expend the resources necessary to create the invention (or bridge) in the first place, since marginal revenues would never cover costs.[61] The deadweight loss of monopoly during the life of the patent is thus the price of having the invention at all, just as the costs of a monopoly gained through "superior skill, foresight, and industry" may be the price of having some essential facility at all.[62]

---

60 The canonical source of this description of inventive knowledge is K. Arrow, Economic Welfare and the Allocation of Resources to Invention, in The Rate and Direction of Inventive Activity: Economic and Social Factors (R. Nelson ed. 1962).

61 A more technical way to state the proposition is that the social return to the invention would exceed the private return, leading to an underinvestment in invention relative to the hypothetical social optimum. Cf. Arrow, *supra* note 60.

62 Gerber, *supra* note 5 at 1103-1110, has suggested using the logic of the patent law to determine whether the returns from an essential facility constitute a reward for the generation of net economic value. Citing *Hecht* (supra note 8), Sullivan and Jones, *supra*

However, recent scholarship has cast doubt on — or at least greatly reduced the reach of — this traditional model. For one thing, empirical studies have suggested that the unadorned model fits few industries.[63] In pharmaceuticals, for example, the model fits well enough: firms must expend large sums up front in the search for new molecules with therapeutic value. Once discovered, such molecules can be quickly described to other scientists and easily copied by other firms. Without patents, there would be little incentive for the necessary up-front research and development. But in many — perhaps most — other industries, often called "cumulative technology" industries, the model fits poorly. In these industries, inventions are not discrete and self contained; instead, technology advances through a succession of incremental improvements. Moreover, innovations in these industries reflect knowledge that is not easy to codify and transmit but is instead sticky and "tacit."[64] As a result, invention in cumulative-technology industries is less of a public good— and thus less in need of the lure of monopoly rents — than the traditional model suggests.

One implication is that the costs of patenting may outweigh the benefits in cumulative-technology industries. The traditional literature would suggest that

---

note 22 at 183, detect "a shared perception in many opinions that for many essential facilities, the defendant was less an innovator and more the lucky beneficiary of having entered a thin market first."

[63] R. Levin, A. Klevorick, R. Nelson, & S. Winter, Appropriating the Returns from Industrial R&D, 3 Brookings Papers on Econ. Activity 783 (1987).

[64] M. Polanyi, Personal Knowledge (1958).

such an imbalance could always be redressed by shortening the life of patents. The most interesting recent scholarship, however, has focused in instead on the problem not of optimal patent length but of patent *scope*.[65] Scope (or breadth) refers to the generality of the claims a patent asserts. For example, the infamous Selden patent in the early American automobile industry claimed the very idea of a gasoline-powered internal-combustion car.[66] The difficulty with such broad claims in a cumulative-technology industry is not merely that they may create deadweight losses in excess of benefits but rather that they foreclose avenues of future improvement and innovation.

Most cumulative technologies are in the nature of *system products*, that is, products that permit or require the simultaneous functioning of a number of complementary components. The next section takes up the theory of such products in more detail. For the moment, two observations from the theory of innovation are in order. First, improvement in system products is likely to take place differentially among the components, often focusing on those components that are bottlenecks in the system.[67] Second, innovation normally proceeds fastest when a large number of distinct participants are trying multiple

---

65 R. Merges and R. Nelson, On the Complex Economics of Patent Scope, 90 Colum. L. Rev. 839 (1990); R. Mazzoleni and R. Nelson, The Benefits and Costs of Strong Patent Protection: a Contribution to the Current Debate, 27 Res. Policy 273 (1998).

66 U. S. Patent No. 549,160, issued Nov. 5, 1895. The patent covered the carriage, the drive mechanism, and the engine.

67 N. Rosenberg, Perspectives on Technology 125 (1976). Thomas Hughes calls such bottlenecks "reverse salients." T. Hughes, The Dynamics of Technological Change: Salients, Critical Problems, and Industrial Revolutions, in Technology and Enterprise in a Historical Perspective 97 (G. Dosi, R. Giannetti, and P. A. Toninelli eds. 1992).

approaches simultaneously. Because of the complexity that system products normally exhibit, and because of the qualitative uncertainty inherent in the process of innovation,[68] multiple approaches and numerous participants provide greater genetic variety than would a single innovator (or small number of innovators), which leads to more-rapid trial-and-error learning.[69] The evolution of improvement in a system product typically results in the emergence of new bottlenecks, in surprising and often major changes in the relative importance of existing components, and in the introduction of wholly new components.[70]

A broad patent is one that locates control of multiple components in the hands of a single party. In so doing, a broad patent limits the number of approaches to and participants in the improvement of the system. Moreover, the patent holder can use the (often vague) scope of the grant to block or delay through litigation innovations elsewhere in the system that threaten the long-run value (and therefore the rent stream) of the package of components the patent holder controls.[71] Notice that blocking by patent is a form of exclusion whose

---

[68] On the idea that, especially in the context of innovation, uncertainty is of a more fundamental or "structural" sort than is usually represented in economic models, see R. Langlois, Internal Organization in a Dynamic Context: Some Theoretical Considerations, in Communication and Information Economics: New Perspectives 23 (M. Jussawalla and H. Ebenfield eds. 1984) and R. Langlois and M. Cosgel, Frank Knight on Risk, Uncertainty, and the Firm: A New Interpretation, 31 Econ. Inquiry 456 (1993).

[69] R. Nelson & S. Winter, In Search of More Useful Theory of Innovation, 5 Res. Policy 36 (1977).

[70] See R. Langlois and P. Robertson, Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries, 21 Res. Policy 297 (1992), hereinafter *Networks and Innovation*, as well as references cited *supra* note 67.

[71] R. Merges, Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents, 62 Tenn. L. Rev. 75 (1994); J Lerner,. Patenting in the Shadow of Competitors, 38 J. Law & Econ.463 (1995).

effects are not an indirect way of extracting an optimal rent from an existing monopoly. The intent and effect of blocking is to shape the future contours of the system product and its ownership, which can mean either (1) an attempt to maintain a monopoly position that might otherwise evaporate with the evolution of improvement in the system or perhaps (2) an effort to enhance an existing monopoly position by nudging innovation in directions that magnify the value of the components under the patent holder's control.[72]

Clearly, there is an analogy here with essential facilities. Since essential facility cases involve vertical relationships, the facility owner necessarily controls one component of what is at least minimally a "system." In many cases, of course, innovation may not be the central issue: *Terminal Railroad* was not about technological change and the future pattern of rents in the national rail system. But if we restrict our attention to cases in which the essential facility consists in intellectual property (of which technical standards are an important special case), then the analogy with patents is almost perfect. As the empirical literature on intellectual property suggests, there are ways other than patents in which a firm can protect its distinctive technological knowledge.[73] The firm may closely

---

72 Others have noticed that protecting an existing monopoly (rather than "leveraging" that monopoly) can be a motive for exclusion. See for example K. Baseman, F. Warren-Boulton, and G. Woroch, Microsoft Plays Hardball: The Use of Exclusionary Pricing and Technical Incompatibility to Maintain Monopoly Power in Markets for Operating System Software, 40 Antitrust Bull. 265 (1995). Also, the literature on "predatory product innovation" may be seen as a narrow variant of the possibility that an existing monopolist can attempt to enhance its position by influencing the direction of technical change. See Ordover and Willig, *supra* note 37. I return to these issues in more detail *infra*.

73 *Supra* note 63.

guard its trade secrets instead of patenting, or it may simply possess knowledge that is tacit and inherently costly to acquire. Moreover, as I will suggest presently, owners of technical standards (or the products that define and embody technical standards) can experience patent-like protection because of the phenomenon of network effects. An essential facility consisting of non-patented (but, by definition, protected) intellectual property can thus be analyzed along the lines marked out above.

Indeed, there may be a close analogy between the scope of a patent and the *scope of an essential facility*. A facility whose control by a single party encompasses several components of a system product would be broader than one whose control implicates fewer components. And the scope of the essential facility may well be a key consideration in antitrust policy.

Innovation in system products.

In order to proceed further, we need to look more carefully at the concept of a system product. In general, a system product is one in which several components must work together simultaneously to produce the desired output. The components thus must be compatible with one another or, to put it another way, must be held together by a common architecture. Figure 1 depicts such a system. Here the four components all connect to a common "compatibility module," which specifies the system's architecture.[74] If one party were the only

[74] Many important systems do in fact take the form of components plugged into a central unit like an automobile chassis, a computer bus, or a stereo amplifier. Nonetheless, it is possible

feasible supplier of, say, component 1, then component 1 would be a bottleneck, and the means to produce it an essential facility. In this case, the scope of the bottleneck would clearly be narrower than if the compatibility module had only one feasible supplier, since to control the compatibility module (the architecture) implicates all components in the system.

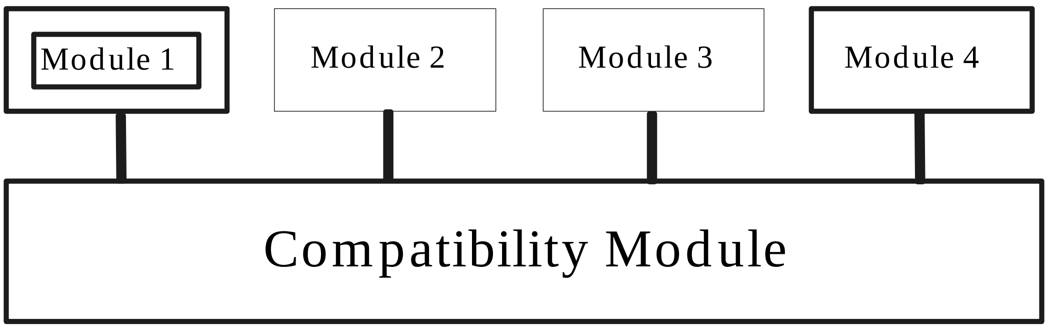

Figure 1
A modular systems product.

In general, systems are hierarchical: each of the components may itself be a system (or, if you prefer, a subsystem). For example, automobiles are system products, but they are also parts of a larger transportation system that includes roads, gas stations, repair shops, parking facilities, traffic courts, etc. Within any

in principle for the components to connect to one another directly without first connecting to a common module. In the language of computers, we can think of such a case as reflecting a peer-to-peer architecture. Even in the peer-to-peer case, however, there is still a "compatibility module," namely, the "virtual" architecture that connects the components together. Such a virtual architecture would then be in the nature of pure intellectual property not instantiated in a physical device. It would be the abstract set of standards on which the architecture is based. I return to the issue of standards presently.

system, then, there are potential bottlenecks of varying scope. Notice also that what is essential is relative to a particular system (or subsystem), a point to which I return.

Systems vary in the degree to which they are modular.[75] Modularity refers both to the structure of the system's design and to its standards, that is, it refers to the degree to which the system is in fact decomposable into modules and to the degree to which the interfaces among the modules are fixed and invariant.[76] If a system possesses at least some degree of modularity — as most do — we can talk about whether the system is *open* or *closed*. An open system is one in which information about the "interfaces" or connections among the components is publicly shared and available. A closed system is one in which the information is tightly guarded. Open systems are generally non-proprietary, in that the standard (the information about connections) is either unowned or owned by a collective body or by a non-profit organization like a trade association. A closed system is necessarily proprietary, at least *de facto*, and proprietary systems are ultimately closed at some level: although the owner of a proprietary system can choose to divulge relevant technical specifications to others, in the end that owner can always change the underlying standards

---

75 Langlois & Robertson, *Networks and Innovation*.

76 On modularity as a design principle, see H. Simon, The Architecture of Complexity, in *idem*, The Sciences of the Artificial (2d ed. 1981); C. Alexander, Notes on the Synthesis of Form (1964); and K. Clark, The Interaction of Design Hierarchies and Market Concepts in Technological Evolution, 14 Res. Policy 235 (1985).

unilaterally. In many real-world cases, systems will have a mix of proprietary and open elements.

Of course, ownership, and the structure of competition among owners, is at the heart of a concern with essential facilities. In the case of system products, competition can take two form: *inter*-system competition and *intra*-system competition. The former is in many ways the more familiar case. Here competition takes place at the level of the systems as a whole. For example, each automobile is a system of complementary components. But the choice we find in the market is among complete pre-packaged systems, and most components from one automobile do not fit on a different model.[77]

By contrast, intra-system competition refers to competition at the level of components within a particular system. Such competition requires at least some degree of openness and modularity. Similar modules compete as substitutes while remaining complementary to the other modules in the system. For example, one can choose among many different brands of computer monitors or modems that are compatible with a personal computer. As we will see, modularity requires a (relatively) fixed architecture, so that compatibility

---

77 Langlois & Robertson, *Networks and Innovation*, refer to such prepackaged systems as "entities." Paul David talks about "system variants." P. David, Some New Standards for the Economics of Standardization in the Information Age, in Economic Policy and Technological Performance 219 (P. Dasgupta and P. Stoneman eds., 1987).

modules typically do not compete against one another — or, rather, competition among compatibility modules is necessarily inter-system competition.[78]

It is possible, indeed even typical, for intra-system competition and inter-system competition to go on simultaneously. For example, two relatively open modular systems could compete in a classic "battle of the systems." Historical examples would include AC versus DC power before the turn of the century,[79] the 33 1/3-RPM versus the 45-RPM phonograph record,[80] and the VHS system versus the Beta system for videocassette recorders.[81] It is also possible to have inter-system competition among systems with different degrees of intra-system competition. In the limit, one could have an open modular standard (with intense intra-system competition) pitted against one or more closed proprietary systems. For example, certain relatively closed and proprietary systems of machinery for fabricating semiconductor chips supplied by Applied Material,

78 In the case of a system with open public standards, there can, of course, be competition among *physical* compatibility modules. But there remains only one "real" source of compatibility, namely, the standard. For example, the amplifier (technically, the preamplifier) is the principal physical compatibility module in a stereo system; but, because the connection among components is an open public standard, there are many competing producers of amplifiers. To coin a phrase, we might say that the open and non-proprietary character of the standard has "commoditized" the physical compatibility module. On stereos as modular systems, see Langlois & Robertson, *Networks and Innovation*; for a more detailed account see P. Robertson & R. Langlois, Modularity, Innovation, and the Firm: the Case of Audio Components, in Entrepreneurship, Technological Innovation, and Economic Growth: Studies in the Schumpeterian Tradition (F. M. Scherer and M. Perlman eds., 1992).

79 P. David, The Hero and the Herd in Technological History: Reflections on Thomas Edison and the Battle of the Systems, in FAVORITES OF FORTUNE: TECHNOLOGY, GROWTH, AND DEVELOPMENT IN THE INDUSTRIAL REVOLUTION (P. Higonnet, D. Landes, and H. Rosovsky eds., 1991); P. A. David and J. Bunn, Gateway Technologies and Network Industries, in EVOLVING TECHNOLOGY AND MARKET STRUCTURE 121 (A. Heertje and M. Perlman eds., 1990).

80 Robertson & Langlois, *supra* note 78.

Inc., compete with systems from multiple vendors adhering to an open public standard.[82] The relatively more closed architecture of the Apple Macintosh computer continues to compete with a personal computer architecture (once called the IBM architecture and now called the "Wintel" architecture) that is relatively more open and benefits from relatively more intra-system competition.[83] And the recent litigation between the Discover Card and the Visa credit-card network turned importantly on the relative merits of intra-system versus inter-system competition.[84]

The mix of inter-system and intra-system competition in an industry is influenced — but, as I will argue, not completely determined — by several

---

81 M. Cusumano, Y. Mylonadis, and R. Rosenbloom, Strategic Maneuvering and Mass-Market Dynamics: the Triumph of VHS over Beta, 66 Bus. Hist. Rev. 51 (1992).

82 R. Langlois, Capabilities and Vertical Disintegration in Process Technology: The Case of Semiconductor Fabrication Equipment, in Resources, Technology, and Strategy (N. Foss and P. Robertson eds., 2000).

83 R. Langlois, External Economies and Economic Progress: The Case of the Microcomputer Industry, 66 Bus. Hist. Rev. 1 (1992), hereinafter *External Economies*.

84 *SCFC ILC, Inc. v. Visa U.S.A., Inc.*, 819 F. Supp. 956, 963 & n.2 (D. Utah 1993), aff'd in part and rev'd in part, 36 F.3d 958 (10th Cir. 1994). Visa is an open system in which a consortium of financial institutions jointly manage the functions of clearing transactions, establishing brand image, and conducting research and development. Within this framework, the member institutions issue cards that compete with one another as well as with cards not in the Visa system. Like American Express and a few others, Discover is a closed proprietary system that manages all necessary functions internally. Dean Witter, then a division of Sears, Roebuck and the owner of the Discover Card, had purchased a Visa member bank from the Resolution Trust Corporation and sought to issue Visa cards. When Visa prohibited the issue, and then passed a by-law against the issuance of Visa cards by owners of cards that compete with Visa, Dean Witter sued, charging a violation of §1 of the Sherman Act. The court found in favor of Visa. For contrasting discussions of the relative merits of inter-system and intra-system competition in this case — by the economists who served as expert witnesses for the litigants — see D. Evan & R. Schmalensee, Economic Aspects of Payment Card Systems and Antitrust Policy Toward Joint Ventures, 63 Antitrust L. J. 861 (1995) and D. Carlton & A. Frankel, The Antitrust Economics of Credit Card Networks, 63 Antitrust L. J. 228 (1995). See also D. Balto, The Murky World of Network Mergers: Searching for the Opportunities for Network Competition, 42 Antitrust Bull. 793 (1997).

economic factors: (1) the extent of economies of scale relative to the extent of the market; (2) the diversity of consumer tastes; and (3) the possibilities for innovation. In general, innovation can come three ways: from changes in the components, from changes in the way the components are interconnected, or both. A system lends itself to *architectural* innovation when the parts are standardized but the connections among the parts are not, thus encouraging innovative recombinations of standardized parts.[85] By contrast, a system lends itself to *modular* innovation when the interfaces among the parts are standardized, thus encouraging improvement in the performance of the parts without changing the way in which they are hooked together.[86] And a system lend itself to *systemic* innovation when both the parts and the architecture of connection can easily change simultaneously.[87]

Consider automobiles as an example. There are significant economies of scale in the assembly of automobiles; but, relative to the size of the market for cars, those economies are exhausted early enough that many different firms can profitably offer consumers many different models, each a distinct system

---

[85] R. Henderson & K. Clark,. Architectural Innovation: the Reconfiguration of Existing Product Technologies and the Failure of Established Firms, 35 Admin. Sci. Q. 9 (1990).

[86] Langlois & Robertson, *Networks and Innovation.*

[87] D. Teece, Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy, 15 Res. Policy 285 (1986). The thrust of this term in its original context was somewhat different from that of architectural and modular innovation. Teece defines systemic innovation as innovation that requires simultaneous change across several stages of production, which stresses not design categories but organizational and institutional categories. Systemic innovation stands in contrast to *autonomous* innovation, in which change takes place strictly within stages of production and does not cross task boundaries. Clearly, autonomous innovation and modular innovation are closely related concepts.

product. Although there may be some competition among parts suppliers for the custom of the assembler, the primary mode of competition is inter-system. Why? Since economies of scale in assembly are exhausted early, there are no great benefits to consumers in having a single system — one kind of car.

Now, we could imagine a world in which manufacturers were all required to produce identical generic "people's cars" the parts for which were all standardized and interchangeable. This would create a regime of intense intra-system competition among assemblers and parts suppliers, leading to modular innovation in the parts. But the benefits of this competition would likely be outweighed by the costs. Consumers have a wide variety of different tastes for system packages, and many people thus might prefer a more-expensive vehicle tailored to specific tastes over a less-expensive generic car.[88] Of course, saying that parts are interchangeable doesn't mean that they are all identical, as there can be competition — including competition for quality and for innovation — within the constraint that the parts connect to standardized interfaces. Thus cars could be differentiated by the parts they use rather than by the overall design. To take a trivial example, one could plug in a high-quality car stereo system as easily as a low-quality one. But, in the case of cars, this enforced modular

88 Of course, if manufacturers could lower the price sufficiently, more and more consumers would be willing to forego differentiation. See K. Lancaster, Consumer Demand: A New Approach (1971). This was arguably what drove sales of Henry Ford's generic Model T — any color you want as long as it's black — in the early auto industry. By standardizing the design, he could take advantage of economies of scale and learning economies to reduce price so dramatically that it offset for many people the disbenefits of the car's generic style. This strategy began to fail, however, as the market for inexpensive but more-differentiated

system is not likely to be optimal despite the benefits of modular innovation it would produce. Architectural and systemic innovations can also benefit consumers, especially when tastes are diverse. And economies of scale in this case are insufficiently large to outweigh those benefits.

In the automobile example, the benefits of standardization come entirely from the economies of scale in production they enable. But, in other cases, standardization can have additional benefits that come not from the supply side but from the demand side. These benefits arise from the much-discussed phenomenon of *network effects*: the benefits to any individual of a would-be standard depend on how many other individuals already adhere (or are likely to adhere) to that alternative.[89] As more and more users commit to a standard, that standard becomes increasingly attractive to others; the commitment of those others makes the standard even more attractive — and so on in a cumulative fashion that is often described as "positive feedback."[90] These are called network effects because, in the first instance, they arise in the case of physical connection networks like telephone systems. The value to me of a phone system increases with the number of other people who are on the system. But the

---

used cars thickened by the 1920s. See R. Langlois & P. Robertson, Explaining Vertical Integration: Lessons from the American Automobile Industry, 49 J. Econ. Hist. 361 (1989).

89 The works most often cited on the subject are probably P. David, *Clio and the Economics of QWERTY*, 75 AM. ECON. REV. 332 (1985); M. Katz & C. Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424 (1985), and J. Farrell & G. Saloner, *Standardization, Compatibility, and Innovation*, 16 RAND J. ECON. 70 (1985). For recent surveys see P. David & S. Greenstein, *The Economics of Compatibility Standards: An Introduction to Recent Research*, 1 ECON. INNOVATION AND NEW TECH. 3 (1990) and N. Economides, *The Economics of Networks* 17 INT. J. IND. ORG. 673 (1996).

concept has been applied to "virtual networks" in which the connections are not physical but rather in the nature of economic complementarity.[91] For example, the benefit to consumers of a new digital television standard is proportional to the amount of programming they expect to be available on that standards, which is in turn dependent of the number of people who adopt the standard. Technological standards can generate economies of scale on both the demand side and the supply side simultaneously. For example, users of a computer operating system may benefit from a large installed base that generates lots of software and other complementary goods and services, while that same installed base allows software developers to reduce costs by focusing their efforts on a single standard platform.

As in the case of production economies of scale, network effects may be exhausted at scales smaller than the entire market. When this is the case, competition can take place between rival systems much as in the automobile case. Sometimes, however, the benefits of compatibility are so great that the network effects are not exhausted at any scale smaller than the entire market. In such a case an entire market can "tip" when a candidate standard has gathered a critical mass of adherent, leaving that candidate as *the* standard. It is this

---

90 W. B. Arthur, Positive Feedbacks in the Economy, 262 Scientific Am. 92 (1990).

91 Katz & Shapiro, *supra* note 89; *Arthur*, supra note 90.

possibility that Paul David popularized in his now-famous account of the QWERTY typewriter keyboard.[92]

Technological standards as essential facilities.

As we saw, the "essential" character of an input is always relative to the system of which it forms a part. To put it another way, "essentialness" is always an issue that speaks to *intra*-system competition. As a result, the analysis of an essential facility will depend crucially on the degree of *inter*-system competition in the industry. An input that is crucial to a system that is one of many alternatives has a status different from that of an equally crucial input to a system that is the only alternative.

The case of *Image Technical Services v. Kodak* raises the multi-system issues in a clear way.[93] Here there are many brands of copier, each brand, along with its complementary "aftermarket" for parts and service, representing a distinct system product. To put it another way, each maker of copiers supplies what is in effect a distinct compatibility module on which its aftermarket inputs depend. In refusing others access to the aftermarket (by tying the purchase of copiers to the use of Kodak's own parts and service), Kodak has foreclosed an essential facility. I do not wish to analyze the case in great detail. But the economic issues turn on the extent to which competition among copier makers disciplines

---

92 David, *supra* note 89.

93 *Supra* note 12.

Kodak's ability to engage in "installed-base opportunism."[94] If consumers have perfect information, or if reputation effects work well, consumers ought to take into account the effect of aftermarket lock-in when they make their initial purchasing decisions; and if there are many alternatives available, no copier maker should be able to engage in such behavior unless there are independent efficiency reasons for it.[95]

From the point of view of innovation, the issues appear somewhat different. Now our concern turns to the ways in which the pattern of ownership of the supply of modules influences the pace and direction of technical change. The facts of *Kodak* may not illustrate these points well, so consider instead the personal computer industry.[96]

When personal computers first emerged in the late 1970s, there were a number of incompatible platforms in active competition, many of them proprietary (like the Apple II) but many grouped around a mostly open standard (Intel-compatible microprocessor, the S-100 bus, and the CP/M operating system).[97] The nature of the market began to change in 1981 when

---

94 A firm engages in installed-base opportunism when it raises the price of complementary products after buyers have been "locked-in" to a standard by the purchase of the compatibility module the firm sells.

95 C. Shapiro and D. Teece, Systems Competition and Aftermarkets: an Economic Analysis of *Kodak*, 39 Antitrust Bull. 135 (1994); C. Shapiro, Aftermarkets and Consumer Welfare: Making Sense of *Kodak*, 63 Antitrust L. J. 482 (1995).

96 The following paragraphs draw on Langlois, *External Economies.*

97 A "bus" is a system of interconnection among components in the computer. The S-100 bus was an open public standard. The microprocessor architecture was a proprietary standard, but in fact the Intel 8080 faced competition from the Zilog Z80, a "clone" chip. CP/M was the property of Digital Research, Inc., but its technical details were widely available.

IBM introduced its personal computer. Because IBM wanted to develop its personal computer quickly, it relied almost exclusively on outside suppliers, including Intel for the microprocessor and Microsoft for the operating system. Because of this outsourcing strategy and the standards it necessitated, others could easily imitate the IBM hardware platform, which effectively became an open system at the component level, that is, any would-be maker of computers could obtain industry-standard components and compete with IBM. However, crucial parts of the overall standard (namely, the microprocessor and the operating system) remained largely under proprietary control. Because of the strength of the IBM name in generating network effects — principally because it created the expectation among users that the key vendor would continue to provide services long into the future and that a wide array of complementary devices and software would rapidly become available — the IBM standard became the dominant hardware platform, largely driving out competing alternatives during the decade of the 1980s.

In the case of the personal computer, the rise of single dominant — but largely open and non-proprietary — standard focused innovation in modular directions. And it is the ensuing rapid improvement in components, including not only the chips but various peripheral devices like hard disks and modems, as well as the proliferation of applications software, that has led to the rapid fall in the quality-adjusted price of the total personal computer system.[98]

Was this result inevitable? As I suggested earlier, such economic factors as economies of scale and diversity of tastes influence the pattern of competition among systems. But they do not fully determine it. The evolution of any industry is determined not only by economic forces existing at any time but also by a historical legacy of technological possibilities and of organizational and institutional structures.[99] By the time IBM developed its PC, the "dominant design"[100] of the computer as an open modular system had already been established and a network of capabilities already existed around that design. This legacy influenced IBM's design decisions. But consider the case of Japan, where the idea of the PC fell on very different ground. There the computer was taken up by a handful of vertically integrated systems houses more accustomed to architectural than to modular innovation. Each firm designed its own proprietary system incompatible with the systems of others even to the level of applications programs. The result was a vibrant inter-system rivalry, but, in the

98 Langlois, *External Economies*.

99 For an elaboration of this argument, see R. Langlois & P. Robertson, Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions (1995), hereinafter *Economic Change*.

100 On the concept of a dominant design see M. Tushman & J. P. Murmann, Dominant Designs, Technology Cycles, and Organizational Outcomes, 20 Res. in Org. Behavior 231 (1998).

absence of a dominant standard, a much-retarded development of the industry in comparison with that of the United States.[101]

One way to couch the problem of optimal inter-system versus intra-system rivalry is in terms of the tradeoff between the benefits of differentiated products and the benefits of the economies of scale (whether supply side or demand side) that come from standardization. This point has not been lost on the formal literature of standard-setting.[102] What seems not to have been noticed, however, is that the nature of this tradeoff is a function of whether variety must be provided through distinct systems with competing standards (which the literature assumes) or whether variety can be provided largely through recombinations of compatible modules. This in turn depends on the economies of scale in packaging the entire system. Where there are significant economies of scale —as in automobiles and probably copiers — we would expect that variety must be provided by competing systems. But where there are attenuated economies of scale in packaging the entire system — as in personal computers — it becomes cheaper to provide variety through the modules. As a rough approximation, we might say that there is less of a tradeoff between

---

[101] T. Cottrell, Fragmented Standards and the Development of Japan's Microcomputer Software Industry, 23 Res. Policy 143 (1994).

[102] J. Farrell and G. Saloner, Standardization and Variety, 20 Econ. Letters 71 (1986).

variety and standardization the greater the extent to which variety is a matter of "software" rather than "hardware."[103]

The implications of this analysis for the doctrine of essential facilities are far from clear. For one thing, antitrust is not an institution well suited to intervening in the evolution of industries at this level. Indeed, in view of the tremendous information requirements involved, it is not clear that any sort of industrial policy is suited to the task. Moreover, there may be reason to think that, although structures may be path-dependent in the short run, there exist long-run forces that may help to correct a non-optimal trade-off between inter-system and intra-system competition. Suppose that a system-product industry were organized as a rivalry among competing incompatible systems when a single standard would have been superior (in that it would have unleashed greater innovation and improved welfare). The existing players may be earning rents and may thus have an incentive to keep the structure intact. But if a move to a single standard would improve welfare, then there are unexploited gains

---

[103] In part, this is a matter of our perspective within the hierarchy of systems. Even if there are economies of scale in assembling major components of the system, there may not be economies of scale in offering variety in the system as a whole. For example, there are economies of scale in assembling televisions and videocassette recorders. But there are no such economies to hooking those components together and playing one of a wide variety of available tapes. Similarly, there are economies of scale in packaging variety in cars, but lower economies in packing variety in the wider transportation system, since travel to different destinations using complementary modular assets like roads is a source of variety. Notice that, in both cases, it is the software rather than hardware — the tapes in one case and the destinations in the other — that creates much of the variety in the larger system. Of course, cars and (perhaps a lesser extent) video hardware nonetheless continue to provide some element of the variety, since different hardware can interact with the software to produce slightly different experiences. Driving through the Berkshires in a Porsche is not the same as driving there in a Chevy; watching Star Wars on a wide-screen

from trade to be had by an entrant (or existing player) who champions such a standard.[104] Something very much along these lines occurred in the market for technical workstations. Here Sun Microsystems successfully championed an open-system strategy in the face of competition from systems that were far more closed and proprietary.[105]

Another way to see this issue is to note that, when there is vibrant inter-system competition, there are more possible entry points for innovation. Multiple competing systems provide a way not only of providing variety but also of experimenting with organizational and design alternatives.[106] The interesting question, then, is when and to what extent property rights can create barriers to experimentation. And this suggest that we pick up the other end of the stick and begin with the case of pure intra-system competition.

In many technological systems, extreme modularity is an important design alternative, making possible the provision of variety at the component

---

TV is not the same as watching it on a 19-inch screen. In the case of personal computers, however, even the hardware can be varied by recombining modules.

104 A similar point is made in a slightly different context by S. J. Liebowitz & S. Margolis, Network Externality: An Uncommon Tragedy, 8 J. Econ. Perspectives 133, 141 (1994).

105 R. Garud & A. Kumaraswamy, *Changing Competitive Dynamics in Network Industries: An Exploration of Sun Microsystems' Open Systems Strategy,* 14 STRATEGIC MANAGEMENT J. 351 (1993); J. Khazam & D. Mowery, *Tails That Wag Dogs: The Influence of Software-based "Network Externalities" on the Creation of Dominant Designs in RISC Technologies*, in THE INTERNATIONAL COMPUTER SOFTWARE INDUSTRY: A COMPARATIVE STUDY OF INDUSTRIAL EVOLUTION AND STRUCTURE 86 (D. Mowery ed., 1996); and C. Baldwin & K. Clark, *Sun Wars: Competition within a Modular Cluster, 1985-1990,* in COMPETING IN THE AGE OF DIGITAL CONVERGENCE 123 (D. Yoffie ed., 1997).

106 Indeed, the process of providing variety (searching the product space, in the lingo of economists) and the process of experimenting with organizational and technological

level. In such systems, the scale benefits of compatibility may not be exhausted by the extent of the market, leading to the eventual dominance of a single set of compatibility standards. The structure of property rights in such cases is crucial to the evolution of the product and of competition. In analogy with the law and economics of patents, we can think about how the structure of ownership may channel future innovative efforts.

In the first instance, proprietary ownership of a standard can be a property right that creates a barrier to entry. The logic is as follows. Products compete with one another along many (price-adjusted) dimensions of merit. Consumers choose those products that are superior along some or all dimensions. When there are many competing packages (inter-system competition), consumer choice typically fragments among differentiated products, and competing standards can coexist. In other cases, however, network effects may force all consumers into the same standard. But standards do still compete. If a contending standard offers a great enough improvement along enough dimensions of merit, a critical mass of adherents may switch to the alternative, thus establishing it as the new dominant standard.

Apart from the punctuated character of the switching among alternatives, standards competition does not differ fundamentally from competition among other sorts of products, except in the following respect. In standards

---

alternatives are really the same process. On this point see L. M. Lachmann, The Market as an Economic Process 15-16 (1986).

competition, one crucial dimension is the fact of universal compatibility itself. In thinking about the choice between vinyl records and compact disks when the latter first appeared, one would have considered (on a price-adjusted basis) the sound quality and other technical and aesthetic aspects of the two alternatives. But one would also have had to consider any existing sunk investment in vinyl disks and the fact that relatively few titles were then available on CD. Compatibility with an existing stock of complementary assets is one dimension of merit. But it is not, in a sense, a dimension of merit intrinsic to the product in the same way that sound quality is. Rather, it is a source of merit conferred on the standard by history.[107]

Standards are barriers to entry in the precise sense that competing products cannot gain advantage without *significant* superiority along non-compatibility dimensions of merit –- as was in fact the case with CDs.[108] Ownership of this barrier to entry is very much akin to the possession of a patent. The nature of such ownership will vary. In some cases, as when the standard is imbedded in the production of physical compatibility modules, the

---

[107] This is of course the source of the claim, hotly debated in the literature, that the existence of network effects can easily lead to the emergence of technically inferior and economically inefficient standards. Compare for example W. B. Arthur, Competing Technologies, Increasing Returns, and Lock-In by Historical Events, 99 Econ. J. 116 (1989) and S. J. Liebowitz & S. Margolis, Path-dependence, Lock-in, and History, 11 J. L. Econ. & Org. 205 (1995).

[108] D. Teece and M. Coleman, The Meaning of Monopoly: Antitrust Analysis in High-Technology Industries 63 ANTITRUST BULL. 801 (1998). ("The firm can earn a rent equal to its advantage in switching costs. The size of this rent is constrained by the size of the switching costs and the extent to which other suppliers can provide products that ease the transition of complementary goods to new platforms; or such an increase in performance to justify investment in new complementary goods.")

property right is protected in effect by trade secrecy or the inherent difficulty of copying. In other, cases, the standard may actually be protected by copyright.

In a well-known paper, Edmund Kitch argued that patents serve a "prospect function."[109] When patents are sufficiently broad, he claimed, they offer their holder a secure opportunity to orchestrate in an orderly fashion the subsequent development of the original idea. One could tell a somewhat analogous story in the case of proprietary rights to standards. Since standards are complex, there may be benefits when the development of the standard is under control of a single firm, since (1) that firm can think through the standardization issues synoptically without the tug of competing interests and (2) that firm has the necessary incentive to champion a superior standard even to the extent of subsidizing its early adoption.

It is certainly true that, by reducing the transaction costs of redeploying assets and coordinating among contracting parties, centralized ownership and control can have advantages in situations of truly systemic innovation.[110] And perhaps one could argue that wide proprietary control of a newly emerging standard might be such a situation. But the countervailing considerations weigh even more heavily. Precisely to the extent that a standard is complex and reflects an underlying cumulative technology, centralized control may actually limit the development of a standard. To work properly, complex standards

---

109 E. Kitch, The Nature and Function of the Patent System, 20 J.L. & Econ. 265 (1977). But compare references cited *supra* note 65.

require collaboration with users and with suppliers of the various components of the system. Indeed, as Hayek has argued, such complex standard sets as human languages or the Common Law could only have evolved as "spontaneous orders."[111] Even the proprietary developer of a standard needs to access the knowledge of a wide variety of collaborators, and even a proprietary standard is often something of a spontaneous order. Moreover, wide scope is not obviously necessary to motivate a potential "champion" for a standard. Any party with a long position in assets likely to appreciate if the standard is adopted has the incentive to expend resources to try to bring about that outcome.[112] For example, Sun Microsystems's incentive to champion a largely open system for technical workstations lay in its possession of distinctive knowledge and innovative ability complementary to the standard.[113]

Another important consideration, of course, is the extent to which the owner of a standard can manipulate that standard in ways that convey private advantage at the expense of more rapid development of the technology. For example, an owner can alter a standard strategically or can simply make the

---

110 Langlois & Robertson, *Economic Change*.

111 See, e.g., F. A. Hayek, 1 Law, Legislation, and Liberty (1973). The idea of spontaneous order as an approach to standard setting has gained considerable attention in the field of software. See E. Raymond, The Cathedral and the Bazaar (1999).

112 It has long been noticed that the ownership of complementary assets can be a substitute for patent protection. See, for example, J. Hirshleifer, The Private and Social Value of Information and the Reward to Inventive Activity, 61 Am. Econ. Rev. 561 (1971); M. Casson, The Entrepreneur: An Economic Theory 206-8 (1982); and Teece, *supra* note 87.

113 *Supra* note 105. Sun evidently believed that the rents to these distinctive capabilities would be greater with the large market an open standard offered than they would in the smaller market of inter-system rivalry among closed proprietary systems.

system less standard by deliberately reducing the degree of modularity in the components under its control.[114] Why might a standards owner do this? The literature offers two contrasting theories.

One set of theoretical stories falls under the rubric of "predatory product innovation."[115] The idea here is that a firm with market power that provides more than one complementary module for a system product may try to leverage its power by creating a new "generation" of the system that is incompatible with the old system. Even if this new system is superior in the eyes of consumers, economic welfare can suffer because the incompatibility freezes out independent suppliers of the previous generation of components, thus raising the market power of the innovating firm. The innovating firm raises its costs in the short run (in spending resources on research and development) in order to gain advantage in the future. What makes the behavior predatory in this account is that the future gain is not (all) the result of creating new value but comes in part from increased market power.

There are several problems with the idea of predatory innovation. Apart from the possibility that the bundling of components by a single manufacturer may have efficiency benefits in addition to the benefits of the innovation, the

---

114 The latter is of course related to the phenomenon of technological bundling, although demodularization is a stronger form of this phenomenon than is simply packaging modular components together. I do not wish to summarize the large literature on bundling here, except to suggest that there may be efficiency motives, having to do with systemic innovation or with the saving of transaction costs, that can be relevant in some cases.

115 Ordover & Willig, *supra* note 37.

theory raises disturbing practical issues, since it asks courts to distinguish "good" innovation from "bad" innovation, a decision with which unsuccessful competitors will be all too happy to assist.[116] Even in the clean and neat world of game theory, predatory product innovation is a logical possibility of almost Scholastic quality. It is yet more worrisome an idea in the uncertain world of innovation, where even the innovators cannot always reliably predict the effects of their innovation.[117]

The other brand of story one finds in the literature appears to take a diametrically opposite tack. Here the problem is not one of too much innovation but of too little. This possibility has not received the same theoretical (that is, mathematical) attention as predatory innovation, probably because it is a story that lends itself less easily to game theory. But the underlying theory seems to be the following.[118] As we saw, a standard may be displaced by a competitor offering significantly higher functionality. To the extent that the existing standard is open, this hurdle is lower, since a competitor can offer improvements in functionality while maintaining backward compatibility. The owner of a dominant standard may thus want to manipulate the standard in ways that close

[116] This first concern, and to some extent the second as well, are raised in J. Sidak, Debunking Predatory Innovation, 83 Colum. L. Rev. 1121 (1983). For a response, see J Ordover, A. Sykes, and R. Willig, Predatory Systems Rivalry: A Reply, 83 Colum. L. Rev. 1150 (1983).

[117] The practical problems of the Ordover-Willig analysis are compounded in my view by being tied to a test for predatory conduct that looks to the "front-end" of predation, namely the incurring of short-run costs and predatory intent. As I have argued, a test that looks to the "back end" — the property rights that may allow recoupment — makes far more sense in the uncertain world of innovation.

[118] See especially Baseman, Warren-Boulton, and Woroch, *supra* note 72.

off the possibilities for a competitor to achieve compatibility. This has a tendency to retard the generational advance of the system.

Here the idea of the "scope" of the standard becomes important. The owner of a standard that controls the compatibility of a large fraction of the components of a system is in a much better position to close off avenues of innovation that threaten the rent-earning potential of the standard. The owner of a standard with relatively smaller scope is always in danger of being "invented around" or made obsolete if it closes off access or otherwise exercises its market power unduly.

Technological standards antitrust policy.

What, then, are the policy implications of this analysis? At the *prima facie* level, they are these:

- Antitrust policy should err on the side of openness of standards, all other things equal.
- The attention that policy pays to a case ought to be proportional to the *scope* of the standard over which owner of the essential facility has *de facto* or *de jure* control.

Arrayed against these proposals are two sets of objections or qualifications. The first set arises from the analogy with patent policy itself, while the second arises from dynamic or Schumpeterian considerations more broadly.

A policy that requires a party to open to others the standard it controls is effectively a taking of intellectual property rights. Viewed strictly from the standpoint of economic efficiency, such a taking is desirable only when the benefits outweigh the costs. As I've suggested, and as the case of personal-computer hardware illustrates, there can be considerable benefits to an open system. The analogy with patent policy suggests two possible costs, however. The first is the potential benefit to inducing future innovation that comes from guaranteeing the innovator's prospective monopoly returns. But, in the case of a broad patent — or a broad standard — the remuneration that monopoly rights confer far outstrip the risk-discounted ex ante costs of innovation. Moreover, in the case of a broad patent or standard, the ability of the patent holder to block future innovation will do more to diminish the incentive for technological than will any weakening of intellectual property rights.

The second cost would arise if, as Kitch suggests in the case of patents, unified ownership would allow for a speedier and more coherent development of the standard.[119] As I suggested above, however, this benefit would be more likely in the very early stages of the development of a broad technical standard. Once a standard becomes mature, development of the standard increasingly takes on the character of a spontaneous order, even when the standard is under proprietary control. Moreover, there are mechanisms other than unified ownership — including not only formal standard-setting bodies but also informal mechanisms involving lead users and dominant players — to provide a more synoptic view and to effect relatively more systemic changes.

The larger dynamic or Schumpeterian issues are more difficult to deal with. At one level, I have argued that attention to the shapes and patterns of innovation does tend to give credence to the Post-Chicago view that a competitor on control of a standard can "leverage" its current position by manipulating the standard in ways that channel future innovation in directions beneficial to the standard owner but not necessarily as beneficial to society as possible. But what Schumpeter giveth to activist policy he may just as quickly taketh away.

I have suggested that ownership of a technological standard is a property right that creates a barrier to entry and confers Ricardian scarcity rents on its possessor. In Schumpeter's famous image, however, competition is a "perennial gale of creative destruction."[120] And, as David Teece and others have insisted, the rents we observe in this windswept kind of competition are at best quasirents (that is, temporary scarcity rents) and are more likely "Schumpeterian" rents (that is, temporary returns to entrepreneurship and innovation).[121] Clearly, the narrower the scope of a technological standard, the more temporary — the more "Schumpeterian" — the rents are likely to be. For example, major personal-computer applications programs like word processors

[119] Kitch, *supra* note 109.

[120] Schumpeter, *supra* note 4.

[121] See, for example, Teece & Coleman, *supra* note 108.

and spreadsheets involve technical standards, and competition among such programs involves network effects. This has led to dominant applications in the various program categories, and the owners of those dominant programs have presumably enjoyed rents during the period of dominance. But those periods have historically been relatively brief, as new dominant programs (embodying a new standard) cam to displace their predecessors in a process of "serial monopoly."[122]

Clearly, there can be standards — like, perhaps those embodied in an operating system — with greater scope than those of major applications programs. One would expect the duration of the "serial monopoly" in such cases to be longer, and the rents earned to approach something more like the Ricardian type. Nonetheless, it may well be that such rents are an inevitable part of the competitive process, and the gale of creative destruction can better deal with even such cases than can a ham-handed and politically motivated government: surgical antitrust policy is an illusion.[123]

If this is true, then the approach suggested here — like most other approaches to antitrust policy — is of little relevance, especially in cases of ordinary private monopoly under Sherman 2 and related statutes. In many cases involving standards as essential facilities, however, the issues revolve either around regulated (or formerly regulated) enterprises with a government-granted

---

122 S. Liebowitz and S, Margolis, Winners, Losers, and Microsoft (1999).

123 Teece & Coleman, *supra* note 108.

franchise or around tradeoffs in government-granted intellectual-property protection. In such cases, the winds of competition blow through what are already government-created canyons. Here at least the concept of the scope of a standard, and the analysis presented here more generally, may be of not-inconsequential value.