Systems Competition and Network Effects

Michael L. Katz, Carl Shapiro

Systems Competition and Network Effects

Michael L. Katz and Carl Shapiro

Many products have little or no value in isolation, but generate value when combined with others. Examples include: nuts and bolts, which together provide fastening services; home audio or video components and programming, which together provide entertainment services; automobiles, repair parts and service, which together provide transportation services; facsimile machines and their associated communications protocols, which together provide fax services; automatic teller machines and ATM cards, which together provide transaction services; camera bodies and lenses, which together provide photographic services. These are all examples of products that are strongly complementary, although they need not be consumed in fixed proportions. We describe them as forming systems, which refers to collections of two or more components together with an interface that allows the components to work together.

This paper and the others in this symposium explore the economics of such systems. Market competition between systems, as opposed to market competition between individual products, highlights at least three important issues: expectations, coordination, and compatibility. A recent wave of research has focused on the behavior and performance of the variety of private and public institutions that arise in systems markets to influence expectations, facilitate coordination, and achieve compatibility.

In many cases, the components purchased for a single system are spread over time, which means that rational buyers must form expectations about

Michael L. Katz is Professor of Business and Economics, University of California, Berkeley, California, and Chief Economist, Federal Communications Commission, Washington, D.C. The views expressed here do not reflect the opinions of the Commission. Carl Shapiro is Professor of Business and Economics, University of California, Berkeley, California.
availability, price, and quality of the components that they will be buying in the future. The use of a particular type of computer, for instance, may lead to large investments in human capital and software that work only with that type of computer. Once a certain system is chosen, switching suppliers is costly because new relation-specific investments have to be made. In such a situation, systems that are expected to be popular—and thus have widely available components—will be more popular for that very reason. The resulting positive-feedback effects have proven troublesome to economic theory, both technically (equilibrium may not exist, or multiple equilibria may exist) and in terms of market performance (the fundamental theorems of welfare economics may not apply).

Systems markets pose challenges for coordination among firms—and sometimes consumers as well. A firm contemplating whether to develop and release a new architecture of microprocessor, for example, must know whether software will be provided to work on the new microprocessor. Likewise, a firm may gain little by introducing a new audio format, such as digital compact cassette, unless programming will be available to play on that format. Naturally, issues of coordinating investment arise in any market; for example, firms in an industry have to reach the right level of capacity. However, the sort of coordination required by systems competition is often more extensive and explicit, employing tools including common ownership of various components suppliers, long-term contracts, and industry-wide standard-setting bodies.

Consumers can face two generic situations in which coordination can be valuable. One involves a communications network, such as the public telephone system, where various end users join a system that allows them to exchange messages with one another. Joining such a network is valuable precisely because many other households and businesses obtain components of the overall system (for an early analysis, see Rohlfs, 1974). Because the value of membership to one user is positively affected when another user joins and enlarges the network, such markets are said to exhibit “network effects,” or “network externalities.”

Another situation in which consumer coordination is vital arises when consumers must choose durable hardware, as when they purchase a device to play a new format of prerecorded music. In making such a choice, each consumer will have to form expectations about the availability of software (in this example, the availability of recordings in that format). In the presence of economies of scale in the production of software, the availability of software will depend on what other consumers do, which gives rise to positive-feedback effects.

This “hardware/software” paradigm applies to many markets: computer hardware and software (Katz and Shapiro, 1985; Church and Gandal, 1992); credit-card networks (the card is the hardware, merchant acceptance the software); durable equipment and repair services (the equipment is the hardware, the repair the software); and the typewriter keyboard (the typewriter is
the hardware, experience on that keyboard the software). These hardware/software systems can fruitfully be thought of as forming "virtual networks" that give rise to feedback effects similar to those associated with physical networks (Katz and Shapiro, 1985; Arthur, 1989, 1990).1

A third issue is compatibility: can a component designed to work in one system also work in another system? Classic examples of failures to attain compatibility include fire hoses that did not fit into fire hydrants, railroad cars that did not match railroad tracks, people who speak different languages, and computers that use different programming languages. It is tempting, but misleading, to view incompatibility as just another coordination failure. Although compatibility has obvious benefits, obtaining and maintaining compatibility often involves a sacrifice in terms of product variety or restraints on innovation. Thus, important questions revolve around how, and if, markets determine the right degree of compatibility.

The literature on systems competition examines how expectations, coordination, and compatibility affect three basic clusters of decisions. First, it examines technology adoption decisions: how many consumers purchase a given system, and what institutions or market mechanisms arise to internalize the network externalities associated with adoption? Second, it explores product selection decisions: what forces determine consumers' choices among rival incompatible systems? What variety of products and systems is available in equilibrium? Are (physical or virtual) network markets systematically biased against new technologies because no consumer will want to be a "guinea pig" testing a new system for others? Finally, the literature examines compatibility decisions: Which firms will seek compatibility, and which will not? How do intellectual property rights influence compatibility choices? How do the private and social incentives to produce compatible systems compare? What institutions arise to set product standards and achieve compatibility?

Our discussion of systems markets is organized around these three main questions. We then conclude with some thoughts on the policy implications of these results.

**External Adoption Effects: A Single System**

While the bulk of the recent literature has focused on competition between systems, it is useful to start by examining a market in which there is a single system. This structure is a straightforward one in which to examine some of the basic forces at work in the markets for systems, particularly the issue of whether network effects are in fact network externalities. We first examine

---

communications networks with their direct network effects; then we shall discuss the hardware/software paradigm, which involves indirect network effects.

Communications Networks

In a communications network, such as a network of electronic-mail users or a network of people who exchange WordPerfect files, each user desires to link directly to other users. Consequently, as has long been recognized, the demand for a network good is a function of both its price, and the expected size of the network. Owners of fax machines, for example, found those machines more valuable as others bought (compatible) fax machines.

The presence of these adoption effects can profoundly affect market behavior and performance. Although many of the issues are inherently dynamic, a simple static model usefully illustrates the basic forces at play. Suppose that there is a single period at the start of which firms sets their prices and during which consumers make all of their purchases. All firms supply access to a single network, say the network of fax machines. Under perfect competition, fax machines are available at cost. A user purchases a fax machine only if that user’s private benefits exceed the cost of the machine. However, the social benefits of one more user joining the network include benefits that accrue to others on the network. Since social marginal benefits exceed private marginal benefits—that is, since there are adoption externalities—the equilibrium network size is smaller than the socially optimal network size, and the perfectly competitive equilibrium is not efficient. Moreover, because of the positive-feedback nature of networks, even adoption externalities that are small at the individual level can lead to large social welfare losses.

The precise nature of the competitive equilibrium depends on how consumers form expectations about networks. One commonly proposed restriction to place on expectations is that they be “rational” in the sense that consumers’ expectations at any point in time correctly incorporate all information available at that time. For a simple model with no uncertainty, such rational expectations imply that consumers should be able to predict the market outcome correctly, so that their beliefs are confirmed in equilibrium. Of course, when there is uncertainty (perhaps regarding the pace of technological progress) or consumers are imperfectly informed about the market, then the most that one can hope for is that consumers use what information they have to form the best possible predictions.

\footnote{Economides and White (1993) emphasize the distinction between one-way and two-way communications networks. The examples in the text all are of two-way communications networks. One-way communications systems, like television and radio broadcast systems, do not exhibit the same direct network effects as two-way systems. Rather, these systems fall under the heading of hardware/software systems, where one user’s adoption has no direct impact on the utility of other users, but may have lagged, indirect effects through the provision of software (programming). See below for more on hardware/software networks.}
Unfortunately, even restricting attention to rational expectations equilibrium—which may strike many people as already placing unjustified faith in the computational ability of consumers—still allows multiple equilibria to occur.\(^3\) Sticking with the example of fax machines, clearly no consumer would value owning the only fax machine in existence. If each consumer supposes that no other consumer purchases a fax machine, then no one will purchase it, and there is a fulfilled expectations equilibrium with no sales.\(^4\) Suppose, however, that if all potential consumers were on the network, then each consumer would derive consumption benefits greater than the marginal costs of production. In addition, suppose that each consumer believes that a large number of other consumers are going to purchase fax machines. Then many would purchase fax machines, and this outcome is a second fulfilled expectations equilibrium. Under the hypothesized conditions, there exist (at least) two fulfilled expectations equilibria in this market.

Clearly, the two equilibria are rather different, and one would like to have a theory that includes the factors that lead to one outcome or the other. At an intuitive level, one can ask the question this way: What can consumers and firms do to influence the market outcome? Before tackling this question, we will draw the parallels between physical networks, which we have been calling communications networks, and virtual networks.

**The Hardware/Software Paradigm**

Now consider a system in which each consumer must buy two components to generate benefits, like computer hardware and software, or a television receiver and video programming. In a hardware/software market, one consumer's adoption decision (to buy the system or not) has no impact on other consumers, given the prices and varieties of software available. Therefore, any inefficiencies in market equilibrium in *static* hardware/software models are attributable to traditional market power, not network effects.

Network effects in hardware/software systems competition arise in the common case in which users make their purchases over time, either because consumers enter the market at different times, or because a given consumer may spread purchases over time as repairs are needed, preferences change, or updated and new components become available. In these markets, adoption

---

\(^3\)Rohls (1974) provides the earliest treatment of the multiple equilibria problem of which we are aware. See Leibenstein (1950) for an early treatment of fulfilled expectation-equilibria in the context of bandwagon effects.

\(^4\)We are presenting an extreme case to highlight the network effects. In fact, early fax machines were purchased by multi-location organizations that communicated with themselves, and thus could unilaterally break out of a "zero-output" trap. Often, large users who can internalize network effects take a lead in adopting a new technology subject to network effects. Once the use of fax machines spread beyond their use to communicate among different locations within a single news or other organization, owners of fax machines found those machines more valuable as others bought (compatible) fax machines.
externalities come about indirectly, through the impact of one consumer's adoption decision on the future variety or prices of components.

To illustrate, suppose that there are two components, being purchased over two periods. In the first period, a consumer chooses a piece of hardware. Since the consumer anticipates that this choice leads to being “locked-in” to the corresponding hardware/software system, the consumer must anticipate what will happen in the second period. As a result, demand in the first period will depend on the expectations that are formed about the second period.

Unlike communications networks, a perfectly competitive equilibrium is efficient in hardware/software markets, if cost conditions are in fact consistent with the existence of competitive equilibrium. Suppose that the components are competitively supplied by nonintegrated firms with textbook U-shaped average costs. With the components priced at marginal cost, the market is simply a competitive equilibrium with complementary goods, and the first-best is realized as the market equilibrium.

At the other extreme from perfect competition, consider a market in which a single firm is the sole source for two components that form a system. Two sets of issues arise from the fact that the two products form a system: multiproduct pricing and intertemporal pricing.

Set aside the issue of intertemporal pricing for a moment by supposing that the firm can fully commit to the prices of both components, either because it can make a contractually binding price announcement, or because of the overall importance to the firm of maintaining its reputation to make good on its promises. With a fixed proportions technology (that is, each consumer requires exactly \( k \) units of component \( B \) for each unit of component \( A \)), there is nothing different from the textbook case of a single-product monopolist. In this case, all that consumers care about is the overall price of the system, and the monopolist simply sets that price to maximize profits in the usual way.

With a variable proportions technology, however, issues of bundling and price discrimination arise. The firm would like to extract as much surplus as possible from the buyer, subject to limitations in the firm’s knowledge of characteristics of demand. In contrast with setting a single systems price, charging for each component separately provides the firm with two pricing instruments instead of one for meeting this goal. As surveyed by Katz (1989), a variety of schemes for extracting consumer surplus are possible. But again, any divergence between the market equilibrium and the social optimum is attributable to monopoly power, rather than externalities or a consumer coordination failure.

Now consider the dynamic effects that arise in a systems market when the supplier cannot commit to prices in advance. The key new issue that arises is

---

5As Burstein (1960) first pointed out, the firm faces what is essentially a taxation problem: the greater the number of commodities that the firm can subject to taxation, the more efficiently revenues can be raised and the greater are its profits. As Burstein noted, this relationship can hold even when the demands for the goods are completely unrelated.
what the firm can do to influence buyer expectations. A monopolist would like
to convince consumers that components will be available at low prices in the
future. Then, after consumers are locked in, this same monopolist may be
tempted to raise price to the monopoly level. Thus, the basis on which
consumers form their expectations of the software price becomes a critical issue
for market performance.

The literature on network effects has focused on situations in which the
quantity of hardware sold in the present serves as a signal of the future price of
software. In particular, if software is produced subject to declining marginal
cost, due to traditional economies of scale or to learning by doing, then a larger
base of hardware owners will lead to greater software sales, a lower marginal
cost of software, and a lower price. Similarly, the larger base may lead to a
greater variety of software or software of higher quality. With this common
pattern, the firm will have an incentive to lower the price of hardware to create
a larger network and thus a software “aftermarket” that is more favorable to
consumers. In other words, the elasticity of demand for hardware will be
increased by the signaling effect of the hardware price. In this common case, indirect network effects lead consumers to place a higher value on a more
popular system.

These indirect network effects are perhaps easiest to see when many firms
offer differentiated software. A glance back at some of the earlier examples of
software—television programs, computer programs, automobile service, and
credit-card acceptance—reveals that software is often supplied by many firms,
subject to some scale economies. When differentiated software is supplied by
many firms with low entry barriers, theoretical models of monopolistic competi-
tion (Salop, 1979; Dixit and Stiglitz, 1977; Spence, 1976) indicate that the
variety of software may be greater, and the price of software less, the larger is
the total demand for software (Church and Gandal, 1992; Chou and Shy,
1990). As with the integrated monopolist, this linkage increases the elasticity of

6See Farrell and Shapiro (1988), Klemperer (1992), and Beggs and Klemperer (1992) for more on
consumer lock-in, and Shapiro and Teece (1994) for a discussion of firms’ incentives to exploit
locked-in customers.
7If hardware sales occur gradually over time, economies of scale and learning by doing in the
production of hardware also can be used to send signals to consumers. By selling a large volume of
hardware today, the firm makes sure that future hardware costs and hence prices will be lower,
leading ultimately to a more popular system with more or cheaper software. Of course, in this
situation the firm must guard against buyers simply waiting and purchasing hardware in the future
when it is cheaper.
8Specifically, suppose that the ex post profit-maximizing price of software, \( p^* \), is negatively related
to the size of the installed base of hardware users, \( n \). Rational consumers will expect this price to be
charged: \( p^* = g^*(n) \). Consumers will expect greater consumer surplus when \( n \) is large. This
linkage has been captured by us and others using a reduced-form network benefit function, capturing
the fact that a hardware buyer will enjoy larger benefits, the more popular is the system chosen by
that consumer. As a consequence of these effects, the demand for hardware will be given by
\( D(p, g^*(n)) \), which is increasing in \( n \). As with communications networks, problems of positive
feedback effects and multiple fulfilled expectations equilibrium arise.
demand for hardware, inducing a hardware monopolist to price closer to marginal cost than otherwise.

The linkage between the size of the installed base of hardware and the price and variety of software has welfare implications. To illustrate, let us return to the case in which hardware is supplied competitively at marginal cost. Total surplus (profits plus consumer surplus) can in some circumstances be increased by offering hardware buyers a (small) subsidy. How can it be optimal to subsidize competitively-supplied hardware? After all, there are no obvious or direct externalities in this system. The answer is that there are indirect externalities that have welfare implications similar to those of the direct network externalities described above for communications networks. These indirect externalities arise because software is not priced at marginal cost. Indeed, if all goods were priced at marginal cost, these network externalities would merely be pecuniary externalities, and market equilibrium in hardware/software markets would be efficient.

In many software markets, however, economies of scale give rise to imperfect competition. It is a standard undergraduate exercise to show that providing a (small) subsidy to a monopolist increases welfare. It is not much harder to see that providing a small subsidy for a good that is complementary to the monopolist's product can easily increase welfare. This result carries over to a small subsidy applied to hardware that is complementary to software that is sold in a monopolistically competitive market at a price above marginal cost.\footnote{Of course, a subsidy to software production might also be welfare-improving. In some circumstances, software subsidies may be more difficult to administer than hardware subsidies and related policies.}

A subsidy for marginal hardware buyers will increase the installed base of hardware and thus may lead to greater variety and lower prices in the software market, thereby increasing the consumer surplus of all hardware buyers. There is no effect on the profits earned by software suppliers, since with free entry into software they earn zero profits in any event. Likewise, the competitive hardware suppliers earn zero profits with or without the hardware subsidy (for simplicity, suppose there is a perfectly elastic supply of hardware). Finally, as long as the subsidy is small, the usual deadweight loss associated with a subsidy will be second-order small, and thus dominated by the welfare gains to consumers.

**Market Responses to the Problems of Network Effects**

Since hardware/software networks and communications networks both exhibit positive adoption externalities, or network externalities, these networks are susceptible to under-utilization. We now ask what market institutions and/or firm strategies can effectively induce more users to join the network to
avoid such an inefficiency. We view this as an investment problem: who will invest in expanding the network? In turn, this leads us to ask who stands to benefit from a larger network.

Institutions: Ownership and Integration

Property rights may help solve externality and investment problems. Specifically, when there is a single owner of the network, that firm may be willing to sponsor the network by making investments in its growth that competitive hardware suppliers would not. The video game industry offers a good example of sponsorship: Nintendo, Sega, and Atari all sell proprietary hardware and complementary software. It appears that these firms take lower profit margins on hardware than software, recognizing that hardware sales contribute to a large proprietary network and thus stimulate future software sales.

Consider then what the owner of a network, who can price access to the network, would do to promote the network.\(^\text{10}\) We know that the total surplus generated by the network is maximized when the marginal benefits associated with a new user, *including the benefits flowing to other users*, just equal the marginal cost of serving the new user. Might a monopoly network owner come closer to this ideal than would competitive suppliers of network access?

To begin, suppose that the network owner simply sets a price \(p\) for network access, with no usage charges. This corresponds to the typical pricing of residential telephone service in the United States, where households are charged a fixed fee per month to have a telephone line but are not charged per local call. What does this imply for the pricing of network services and the internalization of external adoption effects? In the simple access pricing example just given, network ownership will be unlikely to solve the problem of adoption externalities. The network owner will price access above marginal cost, further reducing network size in comparison with marginal-cost pricing. Pricing above marginal cost leads to the usual monopoly deadweight loss; it also exacerbates the problem of adoption externalities in comparison with marginal-cost pricing of network access.

What about more general pricing strategies, whereby the network owner engages in price discrimination, perhaps by setting discriminatory access fees and also by charging on the basis of usage? These pricing instruments may well allow the network owner to internalize the adoption externalities: for example, by setting access fees at or below cost and earning profits on usage fees.\(^\text{11}\) As with all price-discrimination problems, the profit-maximizing pricing strategy

\(^{10}\) Of course, if there are large fixed costs of creating the network, ownership is crucial to the very emergence of the network, and competitive supply of access is not a realistic market structure in any event.

\(^{11}\) An interesting version of this is fads. A firm trying to start a fad may well subsidize key “trend-setters” to attract others. Here the network externalities are bandwagon effects that appear directly in users’ utility functions.
of the network monopolist depends upon the information available to the network owner about user benefits. The key question is what fraction of the benefits associated with a larger network can be appropriated by the network owner. In essence, offering users access to a larger network is like offering them a better product, so the analysis is similar to a monopolist choosing product quality.\textsuperscript{12}

Can network ownership overcome the "chicken and egg" problem in launching a new communications network? In theory, if there are multiple equilibria and network ownership favorably alters expectations by causing users to "have faith" that the system will be popular, ownership can improve performance. While advertising and other marketing activities undertaken by the network owner might improve consumer expectations, it is not obvious why network ownership should necessarily have this effect. In fact, network ownership may cause consumers to shy away, out of fear of being locked-in to a proprietary network.

However, the network owner can do far more than simply market its system to promote consumer confidence and adoption; the network owner can employ sophisticated pricing schemes to get the bandwagon rolling. For example, Dybvig and Spatt's (1985) analysis suggests that a monopolist could break any inefficient equilibrium by insuring potential buyers against the possibility of a small, low-value network. This would be accomplished by making the price paid by any one consumer contingent on the overall size of the network.\textsuperscript{13}

Network ownership is most effective in overcoming network externalities if the network sponsor captures some of the benefits derived from a larger network. This can occur if the hardware supplier has a stake in the supply of software as well as hardware, either through vertical integration, a joint venture, or contract. Alternatively, the hardware supplier may itself be a large end user, or may deal directly with end users. For example, an equipment vendor may sign a contract with a large buyer, such as the federal government, and promise to build a large network for which it charges the buyer. Whether the hardware vendor deals with software suppliers or end users, the network sponsor may be able to capture some of the social benefits of a larger hardware network, and thereby partially internalize network externalities through integration or contract.

In fact, large buyers are natural candidates to be the network sponsor. Perhaps one large user will begin its own network and encourage others to join; this has occurred in e-mail. Or a group of users may agree jointly to adopt a new communications system, thereby assuring each other that they will be able to communicate with other "founding" members. Notice that an open communications network shares many features with a public good; small users

\textsuperscript{12} See Spence (1975) on a monopolist's incentive to improve product quality.

\textsuperscript{13} In the next section of this paper, we explain how short-term leasing of equipment can have much the same effect, by giving consumers an easy out if the equipment is not supported in the future.
may free-ride on large users who may bear the costs necessary to create and market the network.

**Strategies to Attract Users to Networks**

The sponsor of a hardware/software network has various strategies available to expand the network by convincing consumers that software will be inexpensive in the future. If the network sponsor can make binding commitments, a promise to keep future prices low or to provide a variety of software will suffice.

If the network sponsor cannot make such a firm pricing commitment, then there is a credibility problem. This case is important in practice. After all, consumers buying Nintendo hardware lack any assurance as to the future price of Nintendo-compatible software, automobile dealers rarely make specific promises about the future prices they will charge for service or spare parts, and it would have been very difficult for Apple computer company to make price, quality and variety commitments for third-party software that runs on Apple’s machines. What strategies or signals might be used in this case?

One possibility is that the network sponsor can indirectly commit itself to a price path involving “competitive” second-period prices by opening the market to independent software suppliers. This is a form of second-sourcing, whereby a firm establishes an alternative source of supply to assure customers they will not be exploited (Farrell and Gallini, 1988). In practice, this may entail establishing an “open” system, so that third parties are permitted to supply components for the sponsor’s system on a royalty-free basis.14 Perhaps the most well-known example of this type of strategy is IBM’s decision to encourage independent software developers to write IBM-compatible software when IBM introduced its PC. Similarly, Nintendo had a very active program of third-party software suppliers for its video game system (but charged them royalties).

Another way for the network sponsor to assure customers that they will not be held up in the second period is by renting rather than selling the hardware. By renting, any capital loss associated with hardware due to a high price for software will be borne by the hardware vendor, not by the end user.15

Vertical integration can also serve as a commitment to supply both hardware and software. Generally, however, full integration is not required to establish a commitment. Witness the myriad of alliances and joint ventures recently in the market for multimedia systems, linking cable television companies, television studios, telephone companies, and consumer electronics firms. However, coordination in systems markets does typically require more than a

---

14 There is a close link between a firm’s strategy in a systems market and the zeal with which it protects its intellectual property rights. Indeed, some sponsors intentionally relinquish some of their intellectual property rights to promote their systems.

15 This analysis is much like the durable-goods monopoly problem posed by Coase. In that problem, a monopolist can avoid giving itself incentives to add to the supply of the durable good by renting the good and thus bearing the capital loss associated with increased supply.
simple spot-market transaction or short-term supply contract, if it is to constitute a credible commitment and thus a signal to buyers or suppliers of complementary goods.

A more direct approach is for the network sponsor to make sunk investments that commit it to the supply of software, and to communicate this to consumers. By lowering the marginal cost of software, the optimal price will fall as well. Nintendo adopted this strategy when it introduced its video game system into the United States.

Penetration pricing is yet another strategy to signal low software prices: by selling hardware below cost early on, the network sponsor is stimulating the demand for software, which may lead to a lower price of software if software is produced according to economies of scale or if the elasticity of demand for software is higher for marginal hardware consumers than for the average hardware consumer.

Still another mechanism for investment to serve as a commitment to the success of the product is to have an asset important to the firm serve as a hostage. The asset that we have in mind here is a firm’s reputation. In markets where network effects are present, a firm may benefit from having a reputation for selling “successful” products. Casual observation suggests that one reason that the IBM PC was so successful is that consumers expected the product to succeed since it was backed by IBM.

When reputation is a valuable asset, firms will find it profitable to invest in it. One form that this investment may take is to promote a product (through marketing or low prices) to a greater extent than would be profitable if the effects on reputation were ignored. IBM clearly was concerned with the effects that its abandoning the PCjr would have on the firm’s reputation, and may have delayed abandoning it for this reason. Moreover, to preserve its reputation for not stranding consumers, IBM may have a greater commitment to maintaining a software base and a parts inventory for these machines than it might otherwise have had, just as Xerox promises to service its copiers for years after it sells them.

Similarly, a firm may well refrain from exploiting its installed base for fear of losing future hardware sales due to damage to its reputation. Likewise, a firm that offers a broad product line, or is otherwise seen as taking a long-run perspective, may refrain from exploiting its installed base for fear of losing future sales in either the systems market itself or the other markets in which the firm is active. In summary, consumers’ knowing that a firm will act to preserve its reputation will raise the consumers’ expectations about the future network size and the availability of software.

In addition to serving as a commitment to high future sales, some sunk investments may serve as signals of the firm’s private information, perhaps about its costs of production or the level of market demand for its product. A simple announcement by a firm—say that “demand will be heavy”—may not be very credible. However, if a firm invests in a large, long-lived plant for the
production of hardware that can be profitable only if the firm expects sales to continue at a high level, then consumers would know that the firm actually expected to have high sales in the future. Thus, consumers would form more favorable expectations about future sales, and the present demand curve would shift outward.

While much of the discussion above has been couched in terms of hardware/software networks, parallel effects arise in communications networks when some consumers make their membership decisions now, while others wait until later. As with hardware/software, today’s potential customers are more likely to join the network, the more attractive terms they expect will be offered in the future to attract more members at that time.

**Competition Between Incompatible Systems**

So far, we have analyzed one system in isolation. We turn now to competition between “incompatible” systems. Two communications networks are incompatible if subscribers on one network cannot communicate with those on the other network. Two hardware/software systems are incompatible if the components of one system do not work with components of the other system.\(^\text{16}\)

Examples of incompatible rival systems are in the newspaper almost every week: VHS vs. Beta in videocassette recorders; phonographs vs. cassettes vs. compact discs vs. digital compact cassettes in audio equipment; analog vs. digital protocols for cellular telephone systems; Nintendo vs. Sega vs. Atari in home video game systems; 5 \(\frac{1}{4}\)" vs. 3 \(\frac{1}{2}\)" floppy disks and disk drives; e-mail vs. fax machines in instant written communications; conventional color television signals vs. high-definition signals in color television; and Visa vs. American Express vs. Discover in credit cards. The list can go on. Incompatible systems also can represent different generations of a single core technology: the Nintendo Entertainment System and the Super Nintendo Entertainment System accept different game cartridges.

Suppose there are multiple competing systems; each consumer can buy none, one, or several. How many systems will survive in the marketplace, and what factors determine the outcome? What strategies can a system owner employ to ensure an outcome favorable to it? And how does the market perform from a welfare perspective, regarding both consumers’ choices between the rival systems and the timing of those choices?

In markets with network effects, there is natural tendency toward de facto standardization, which means everyone using the same system. Because of the strong positive-feedback elements, systems markets are especially prone to

\(^{16}\text{Incompatibility can be one-way or two-way. Two-way incompatibility exists when components from one system simply don’t work in the other. One-way compatibility happens when a component from one system works in the other, but the reverse is not true. For example, WordPerfect 5.1 can read files from WordPerfect 5.0, but not vice versa.}\)
“tipping,” which is the tendency of one system to pull away from its rivals in popularity once it has gained an initial edge. Tipping has been observed in many situations, including AM stereo radio (Besen and Johnson, 1986); FM vs. AM radio (Besen, 1992); color vs. black and white television (Farrell and Shapiro, 1992); VHS vs. Beta in videocassette recorders (Cusumano et al., 1990); and typewriter keyboards (David, 1985). Tipping is reflected in static models in the form of multiple equilibria, very often multiple corner equilibria in which a single system dominates (Katz and Shapiro, 1985). In dynamic models, tipping is reflected in equilibria where new placements of the losing system simply dry up once a rival system is introduced or accepted in the marketplace (Farrell and Saloner, 1986a; Katz and Shapiro, 1992).  

Consumer heterogeneity and product differentiation tend to limit tipping and sustain multiple networks. If the rival systems have distinct features sought by certain consumers, two or more systems may be able to survive by catering to consumers who care more about product attributes than network size. Here, market equilibrium with multiple incompatible products reflects the social value of variety. In some cases—Apple vs. IBM computers, perhaps—important variety benefits might be lost through standardization. In other cases, such as VHS vs. Beta in videocassette recorders, any loss of variety seems a minor price to pay to achieve compatibility. Farrell and Saloner (1986b) discuss this tradeoff between standardization and variety.

The uncertainties of technological progress suggest another benefit to variety: standardizing on a single system can be very costly if the system selected turns out to be inferior to another system. With network effects, it can be very difficult to switch horses in midstream to a system that later proves superior. For example, the Japanese HDTV system is now widely regarded as inferior to the system being developed for use in the United States; NHK and other Japanese suppliers did not expect a workable all-digital system to be feasible before the turn of the century, so they focused their efforts on an analog system. Because the Japanese were promoting a single standardized system, they were not well placed to offer a digital system when such systems were recognized as feasible.

We suspect that in the long run the greatest difference between systems markets and other markets arises because firms’ innovation incentives are altered by network considerations. Rather little theoretical work has been done on R&D and technology choice in the presence of network effects and uncertain technological progress. But there is little reason to believe that, in the presence of network externalities, the marginal private and social returns to keeping one more technology in the portfolio of those under development are likely to be well-aligned. We see innovation in systems markets as a promising area for future research.

---

17 For very similar reasons, tipping also can occur in the presence of learning by doing, which is a close cousin of network effects and another type of dynamic increasing returns to scale.
Competitive Strategies in Systems Markets

Because a firm with a small, initial advantage in a network market may be able to parlay its advantage into a larger, lasting one, competition in network industries can be especially intense—at least until a clear winner emerges. One would expect the promotional strategies identified above for a monopoly network to be used quite aggressively in systems competition. For example, dramatic penetration pricing may emerge as the equilibrium outcome, as each firm seeks to establish an installed base and achieve leadership in a systems market.\textsuperscript{18} If the ultimate outcome is going to be one of tipping to a single system, the firms are effectively bidding for future monopoly profits. At the same time, it is important to recognize that merely observing a firm with a position of market dominance does not imply that the firm is earning supernormal profits: the firm’s quasi-rents may merely reflect costs incurred earlier to obtain the position of market leadership.

Early and visible sunk expenditures on software may signal to consumers that a hardware supplier is committed to the development of software, as noted above. In the context of systems rivalry, vertical integration of hardware manufacturers into software, or exclusive contracts for use of software, not only allow the hardware supplier to gain assured access to software, but can also serve to deny access to that software to rival hardware manufacturers. Nintendo adopted this strategy by signing exclusive contracts for games developed by third parties, making those games unavailable to Nintendo’s rivals, Atari and Sega.

More generally, a firm in a systems market has strong incentives to build up consumer beliefs about its own system, and to tear down consumer beliefs about rival systems, in trying to tip the market in their favor. Two recent examples come to mind. The WordPerfect corporation has filed a court complaint to block the Microsoft corporation from claiming that its word processing software was the most popular in the world.\textsuperscript{19} And Visa has had a long-running advertising campaign telling consumers that Visa cards are accepted “everywhere you want to be,” whereas merchants “don’t take American Express.”

In systems markets, even more so than in other markets, firms with established reputations, well-known brand names, and ready visible access to capital have competitive advantages. These are the firms that are less likely to choose an open-system strategy.

\textsuperscript{18}A milder version of penetration pricing is to charge a lower markup early in the lifetime of a product’s life cycle. Note that penetration pricing might easily be confused with predatory pricing: both are below cost, and both involve establishing a strong market position later to recoup losses incurred by selling below cost.

Do Network Effects Retard Innovation?

Popular opinion suggests that systems markets may tend to get locked-in to obsolete standards or technologies. Some theoretical models do indeed exhibit excess inertia; that is, users tend to stick with an established technology even when total surplus would be greater were they to adopt a new but incompatible technology. Today's consumers may be reluctant to adopt a new technology if they must bear the cost of the transition from one technology to the next, and if most of the benefits of switching will accrue to future users (Farrell and Saloner, 1986a). In terms of the Coase theorem, it is very difficult to design a contract where, say, the (potential) future users of HDTV agree to subsidize today's buyers of television sets to stop buying NTSC sets and start buying HDTV sets, thereby stimulating the supply of HDTV programming.

Although it seems plausible that the inertia associated with network effects has somehow deprived us of valuable new technologies, it is abundantly clear that many new, incompatible technologies are in fact successfully introduced. In fact, there is no general theoretical result implying excess inertia in market equilibria. Indeed, given the possibilities of multiple equilibria, markets may also exhibit the opposite of excess inertia, which we call "insufficient friction." In other words, the market may be biased in favor of a new, superior, but incompatible technology (Katz and Shapiro, 1986a; 1992). The reason is "stranding": today's buyers may ignore the costs they impose on yesterday's buyers by adopting a new and incompatible technology. Those who previously bought the old technology are stranded, and their capital investments may begin to depreciate because of a reduced flow of complementary software. Recognizing this problem, the Federal Communications Commission has taken actions to insure that those with NTSC television sets are not stranded as HDTV programming becomes available. Stranding effects will be small, however, if the established network is large and will remain so for some time; in this case, users of the old technology will be able to take advantage of most of the available economies of scale even without more new users.

One key factor in determining whether a given market exhibits excess inertia or insufficient friction involves possible asymmetries in sponsorship between old and new technologies. Specifically, the older technology is more likely to be competitively supplied, perhaps because certain patents have expired, whereas the new technology is (at least potentially) proprietary and thus may have one or a few sponsors. In Katz and Shapiro (1986a), we showed how the market could be biased in favor of a new sponsored technology over an old unsponsored one, because the sponsor can engage in penetration pricing and other forms of investment, whereas no competitive firm will sell the old system below cost since it knows competition from other suppliers of the same system would prevent recoupment. Competition between an older, competitively-supplied technology and a newer, sponsored one can lead to insufficient friction.
Compatibility Choice

To this point, we have taken the compatibility of systems as exogenous. In many markets, however, the degree of systems compatibility is at least partially subject to choice, and it may have a major impact on market performance. In this section we examine the private and social incentives to achieve "horizontal" compatibility (between two roughly comparable rival systems) and "vertical" compatibility (between successive generations of similar technology). The central task is to describe the costs and benefits of compatibility, and then describe the conditions under which compatibility is most likely to be socially desirable. To focus the discussion, we will consider the specific situation of two firms that are choosing whether to make their competing systems compatible.\textsuperscript{20}

Social Benefits and Costs of Compatibility

For communications networks, compatibility expands the size of each network to the total membership of both. This raises the gross consumption benefits enjoyed by a consumer who subscribes to only one firm's network, and avoids the cost of having to hold duplicate equipment to participate in two different networks to reach everyone. In hardware/software systems, the benefits of compatibility ultimately are due to lower production costs: when the components of different systems are interchangeable, there may be greater opportunities to take advantage of economies of scale, learning effects, and technological spillovers in the development and production of specific components.\textsuperscript{21}

As Matutes and Regibeau (1988) point out, compatibility also enhances variety by allowing consumers to mix and match (differentiated) components from various systems. This is a variant on the theme that economies of scale can be enjoyed when systems are compatible. Home audio components provide a classic example. Even with homogenous components, compatibility allows use of the cheapest one, component by component. Compatibility also allows consumers to exploit economies of scope (scale) in home production, as when a single audio amplifier is used with the television set, compact disk player, and tuner. Finally, with compatibility, consumers need not fear that the technology they have picked will end up being a loser, leaving them stranded or forcing them to purchase replacement equipment or reinvest in human capital (Berg, 1984).\textsuperscript{22}

\textsuperscript{20} In practice, partial compatibility is also a possibility and can manifest itself in at least two ways. First, there may be attenuated benefits, as when rival word processing programs can exchange text but not formatting codes, or when rival database programs can exchange raw data but not programs written to analyze the data. Second, the compatibility may extend to some components but not others, as when automobiles can use the same gas but not the same spare parts.

\textsuperscript{21} Thus, the social benefits of compatibility are lower if compatibility is attempted after two incompatible systems have already been developed.

\textsuperscript{22} Similar effects arise if consumers expect the future introduction of adapters that will allow previous incompatible systems to work together.
The potential costs of compatibility depend upon the mechanism by which compatibility is achieved. Broadly speaking, there are two mechanisms: standardization, whereby systems are designed to have interchangeable components; and adapters, which attach to a component of one system to allow it to interface with another system. With adapters, the principal cost is that of the adapters themselves, plus the fact that adapters may work imperfectly. By contrast, the primary cost of standardization is a loss of variety: consumers have fewer differentiated products to pick from, especially if standardization prevents the development of promising but unique and incompatible new systems.23

**Does Compatibility Intensify Price Competition?**

Most compatibility decisions are made by private firms and individuals. The issue naturally arises as to whether private firms are somehow biased for or against compatibility, by virtue of their focus on profits rather than total surplus. For example, if compatibility reduces competition among firms and allows them to appropriate more of the benefits that would otherwise accrue to consumers, firms should be biased toward standardization. A key question thus becomes how compatibility affects the degree of competition between systems suppliers.

For systems that are compatible, the locus of competition shifts from the overall package (including the network size) to the specific cost and performance characteristics of each component individually (Matutes and Regibeau, 1988; Economides, 1988). This general principle implies that if one firm has a distinctly superior overall package, including its product offering, its installed base, and its reputation, that firm is likely to prefer incompatibility and may in fact spend resources to block compatibility. However, if each firm has a distinctly superior component, both firms may prefer compatibility and may spend resources to achieve it.24

In Katz and Shapiro (1986b), we explore how compatibility affects pricing competition when the two systems compete over time. We found that

---

23 While in the text we focus on the costs and benefits of initially achieving compatibility, similar costs and benefits may arise from maintaining compatibility. In the case of the UNIX operating system, compatibility was lost as innovation took place and different versions of the program were developed, and there are now attempts to reestablish compatibility and to avoid further splintering of the UNIX standard.

24 To illustrate, consider two firms, each of which sells a two-component system. Suppose the firms’ systems are functionally equivalent, and the firms compete on prices (as Bertrand duopolists). Let \( c_i \) denote firm \( i \)'s constant unit cost of supplying component \( X \), and let \( d_i \) denote firm \( i \)'s constant unit cost of supplying component \( Y \). With compatibility, the firm with a lower-cost system earns profits equal to the difference in the systems costs: \( |(c_1 + d_1) - (c_2 + d_2)| \). With incompatibility, competition takes place component-by-component, and industry profits are proportional to \( |c_1 - c_2| + |d_1 - d_2| \). Whenever one firm has a lower cost of component \( X \), while the other has a lower cost of component \( Y \), the firms prefer compatibility. With incompatibility, a firm dissipates its advantage in one component by trying to compensate for its disadvantage in the other. See Farrell, Monroe, and Saloner (1994) for a more complete analysis along these lines, showing that the analysis is much richer with more than two firms.
compatibility relaxes competition early in the product life-cycle, because the threat of tipping is reduced. However, because compatibility prevents one firm from gaining control of the market, it tends to intensify competition later in the product life-cycle.

A general theme emerges from this analysis: since systems competition is prone to tipping, there are likely to be strong winners and strong losers under incompatibility. Therefore, if a firm is confident it will be the winner, that firm will tend to oppose compatibility. A firm might expect to become the dominant supplier for several reasons. First, the firm might expect that it will have lower costs in the future than will its rival. In particular, the sponsor of an emerging technology, whose costs are falling relative to those of its rival, may oppose compatibility for this reason (Katz and Shapiro, 1986b). A better reputation might also encourage a firm to oppose compatibility. When the competing technologies are compatible, firm reputation is less important to consumers since there is no danger of choosing the wrong technology and being stuck on the smaller network.

Product differentiation may have similar effects on compatibility incentives. When a set of consumers have a preference for one firm’s components, that firm has an advantage over its rivals in establishing an installed base. The firm may then prefer incompatibility, knowing that this brand preference by a subset of consumers can translate into an overall advantage to the firm—even in selling to those consumers with no brand preference—because of the positive feedback associated with network effects.

Asymmetries involving reputation, product differentiation, and installed base are especially likely when one of the firms is an entrant and the other an incumbent. Under incompatibility, the entrant will suffer an installed base disadvantage and may well suffer a reputational disadvantage as well. Incompatibility also discourages entry by requiring that entry must happen at a minimum size to be viable, which involves putting a sunk investment at risk. Of course, as discussed earlier, an entrant who has a superior technology may be the one that opposes compatibility (Katz and Shapiro, 1986b).

Institutions for Achieving Compatibility

Given that firms often disagree on the desirability of standardization, the market outcome may be strongly influenced by the process by which products are made compatible. If side payments are feasible, then it is more likely that the firms will be able to harmonize their interests and adopt the compatibility regime that maximizes industry profits. In this case, the change in consumer surplus is the only remaining wedge between the private and social compatibility incentives.

When side payments are infeasible, one must distinguish between markets in which a firm can unilaterally impose compatibility (say, by building an adapter) and markets in which a firm can unilaterally impose incompatibility (say, by using a proprietary interface). Suppose that one firm favors
compatibility, while another opposes it. When the first can unilaterally attain compatibility, that will be the outcome. But when the other firm can unilaterally block it, incompatibility will be the result. For example, Nintendo has successfully employed a “lock-out chip” to prevent unauthorized game cartridges from being played on Nintendo hardware, but Gillette has found it difficult to prevent rivals from making blades that fit into Gillette razors.

There are many different routes to reaching a standard. Farrell and Saloner (1988) compare three of them: explicit communication and negotiation before making any commitments (standards committees); unilateral action with the danger of making simultaneous incompatible commitments (market leadership); and a hybrid with both communications and unilateral commitment. Farrell and Saloner (1992) examine partial compatibility between two differentiated technologies that is achieved through imperfect two-way converters. Farrell (1993) examines two incompatible systems which vary in quality, but each system’s quality is the supplier’s private information. No general conclusions are available in this area, although most of the studies find a possibility that at least some explicit coordination may be mutually beneficial. In practice, industry voluntary standards bodies arise to try to solve these coordination problems. The study of these organizations involves a fascinating mixture of collective choice and industrial organization.

Conclusions

There are several reasons to expect equilibrium in systems markets to diverge from the social optimum: (1) due to economies of scale and product differentiation, these markets are often characterized by oligopoly or monopolistic competition, not perfect competition; (2) due to the importance of R&D and innovation, together with the high chance of tipping, these markets are often characterized by (temporary) monopolies; and (3) the network effects discussed above may indeed be network externalities, not internalized in any market transaction.

Since market outcomes may be inefficient, it is theoretically possible for government intervention to improve market performance. But there are several issues that must be addressed before concluding that government intervention is warranted in practice.

First, the extent of the market inefficiency is unclear, once recognition is given to the many private institutions that arise to achieve coordination and internalize externalities. As discussed throughout this paper, there are many possible responses of systems markets to these problems that involve no government intervention whatsoever.

Second, there is the question of whether the government would have incentives to improve matters. One plausible hypothesis is that the government will act to serve the current generation of producers and users, while acting to
block or impose inefficiently high costs on an emerging technology. Some believe this happened when the Federal Communications Commission (FCC) required that high-definition television signals fit into the 6 MHz bandwidth that traditionally has been used for broadcast television.

Third, even if policy-makers try to maximize total surplus, they may lack the information needed to do so. In the case of choosing a standard at the start of the product's life, it may be very difficult to determine which standard is the "correct" one. Moreover, the government may have a significant informational disadvantage relative to private parties when emerging technologies are involved. Many commentators feel that the FCC made a poor choice for color television in the 1950s, and that the European Community is making a losing choice in HDTV today.

In short, we are far from having a general theory of when government intervention is preferable to the unregulated market outcome.

A standard way to end a review of theoretical work, and to draw out policy implications, is to call for more empirical testing. While empirical study is surely worthwhile (we have not attempted a systematic report here of the empirical literature on systems markets), we believe that this is an area where additional theoretical research can still have a very high payoff in organizing thinking about which facts and relationships are most important.

Three theoretical questions seem especially pressing to us. One is the analysis of hardware/software linkages. Such linkages are of practical significance; in recent years, several of the largest mergers have been between hardware and software producers, like Sony's acquisition of Columbia Pictures and Matsushita's acquisition of MCA. And currently there is a frenzy of activity establishing alliances among hardware and software firms in the entertainment and telecommunications industries. Such linkages and mergers may be one of the central market responses to the potential inefficiencies of systems markets. Our second recommended area of research is the exploration of the dynamics of standards adoption, particularly on the difficult problem of understanding how coalitions form in this area and how voluntary standards bodies behave. Finally, we need to develop a more sophisticated understanding of the unique innovation incentives that arise in systems markets, particularly in the face of uncertain technological progress.

* Both authors thank Joseph Farrell for comments on an earlier draft, and the National Science Foundation for financial support. Shapiro also thanks the Sloan Foundation for financial support.
References


Farrell, Joseph and Garth Saloner, “Converters, Compatibility, and the Control of In-


