Network Issues and Payment Systems

*James J. McAndrews**

Networks play an integral part in the production and consumption of certain goods and services, including transportation, communications, and payment systems. A network good or service has two main characteristics: the value a person gets from the product increases as more people consume it and the technique a firm chooses to produce the product will depend on techniques chosen by other firms. For example, consider a telephone system. The greater the number of people connected by telephone lines, the greater the number of people any member of the system can call and the more he or she will enjoy belonging to that telephone network. Similarly, firms that offer phone service will produce switches and lines compatible with those of other firms that offer phone service, so that they can offer their customers the valuable service of connecting to all other parties.

It is helpful to think of network components as nodes connected by links.[1] Perhaps the most transparent example is a railroad system, a

*When this article was written, James McAndrews was a senior economist and research advisor in the Research Department of the Philadelphia Fed. He is now in the Research Department of the Federal Reserve Bank of New York.

[1]See the article by Nicholas Economides for a good survey of network economics and an example of an approach using nodes and links as the basic network components.

physical network composed of lines (the links) that connect destinations (the nodes). A railroad to one destination is of some value, but a railroad system that connects a traveler to many destinations potentially has great value. To create an extensive railroad system, regional rail lines must use compatible gauges. This complementarity between the components of a network leads consumers to place a higher value on larger networks and leads firms to take into account the production decisions of their rivals.

Other examples of physical networks include highways, oil and natural gas pipelines, water systems, and computerized airline reservation systems. Certain information services also have network characteristics. The Internet, for example, can be thought of as a network in which the computers are the nodes, and the software and the telephone lines to which the computers are connected form the links that allow files to be exchanged and seamlessly read by different machines.

Payment systems, such as credit cards, ATMs, currency, and checks, are also examples of network goods. Here, the nodes might be merchants, consumers, and banks, which are linked by the exchanges of information among them. In some cases, such as in an ATM network or a point-of-sale (POS) debit system, the links may also consist of telephone lines. In others, such as in the checking system, the links consist of methods of delivery of the check from the merchant to its bank, and from that bank, through a clearinghouse (similar to a telephone switching system), to the consumer's bank. In a credit card system, the complementarity between the components is obvious: as more people use credit cards, more merchants are induced to add terminals, since allowing customers a convenient means of payment will potentially increase their sales, and as more merchants permit credit card payment, the value to the customer of having a credit card increases, too.

Economists have recently renewed their interest in many of the unique issues that arise in network-dependent industries. Below, we'll discuss some of these issues, including compatibility and standard-setting among service providers, the role of an installed base of network facilities, and access to network facilities.[2] In addition, the more common economic issues of pricing policies, the tendency toward monopoly, and the introduction and adoption of alternative technologies take on new dimensions in network industries. Network economics is increasingly relevant in today's economy because of the growth of the communications industry and the computer hardware and software industries and the introduction of new forms of payment systems such as electronic money. An understanding of the economics of networks and the unique features of network goods gives insight into the organization of markets for these goods and provides the basis for formulating good business and public policy concerning these goods.

Below, we'll also analyze some payment-system issues from the perspective of network economics and show that formulating appropriate public policy would be difficult without a knowledge of the economics of payment networks.

NETWORK ISSUES

Not all goods have network characteristics. For non-network goods, firms compete to be the main producer, and the techniques one firm uses in producing the goods need not be related to the techniques used by other firms. Typically, the firm that is the most efficient producer will gain market share, and other firms will lose

[2]Among the many papers that explicitly analyze network issues are those by Jeffrey Rohlfs, Joseph Farrell and Garth Saloner, and Michael Katz and Carl Shapiro.

market share or be driven from the market entirely. Moreover, the pleasure one person receives from purchasing the good would be the same no matter how many other people purchase it. Think of ice cream: different firms compete to be the most popular brand, each using a technique it believes produces the tastiest product, and one person's pleasure from eating a cone doesn't depend on how many others buy ice cream cones.

But the situation is different for network goods. Consider a communications system: if one person uses Morse code and another uses semaphores, they could not communicate. For communication to flourish, a coordinated system of signals that can be mutually understood is necessary. So firms that want to provide some of these services must consider what other firms are providing. Rather than competing, these firms' decisions complement one another.[3] Furthermore, as more people adopt the communication system, its value increases, since it provides access to more people; this encourages larger networks.

Networks combine complementary components of a technology that makes possible the creation of goods and services.

Not only do the benefits increase as the network expands, but the per unit cost of production falls. One reason for the economies of scale is that networks are often set up with centralized switching facilities to route delivery of service. For example, in a local telephone network, rather than stringing wires from each house to all other houses, one line is strung from each house through a series of trunk lines to a central switch. As telephone traffic increases, the cost per call declines, since the fixed cost of the switch can be spread over more calls. This decline in average cost encourages larger networks.

In the 1970s and 1980s, economists began to recognize networks as distinct features of certain industries and subsequently outlined various economic issues unique to these industries.

Compatibility. One key to extending the size of a network is the compatibility of network components. Networks combine complementary components of a technology that makes possible the creation of goods and services. But the components' compatibility makes possible their complementarity. For many products, compatibility can be achieved only by adherence to technical standards.

Take the case of railroad gauges. U.S. railroads employed different gauges of track—the distance between the rails—for decades, necessitating the use of costly devices (including laying third rails in some cases, and having railroad cars with adjustable axle widths) to transport goods across different rail lines. In the 1830s, at the beginning of intercity rail service in the United States, three gauges emerged as the most popular. The three—4 feet 8.5 inches, 4 feet 10 inches, and 5 feet—varied only slightly from one another but were sufficiently different to prevent the interchange of rolling stock (railroad cars and engines). Hence, goods typically had to be unloaded and reloaded as they were shipped from one region's lines to another's. As long

[3]This aspect of networks is made clear in the paper by Philip Dybvig and Chester Spatt.

as the volume of shipments between regions was not too great, the different gauges could survive.

From the 1840s through the 1860s, additional gauges were introduced and survived, resulting in a balkanized railroad system. In a 1991 article, Douglas Puffert identified nine regions of the country that, in the 1860s, used common gauges within the region. By the 1890s, though, U.S. railroads had fixed on 4 feet 8.5 inches (called the standard gauge) as the measurement for rails across the country.

Puffert explained the evolution of railroad gauges in the United States in this way: In the early years of railroad development, purely local considerations were paramount. Railroads were built to gain access to ports, rivers, canals, and large regional markets and were not interlinked with other railroads. However, as the industrialization of the United States proceeded in the post-Civil War period, the higher cost of shipping goods across lines of different gauges became more apparent. In the late 1880s, those lines that did not use the standard gauge decided voluntarily to move their rails to adhere to the standard gauge across the United States. After that, an extensive system of car interchange developed among the railroads.[4]

Underproduction. Another issue is the possible underproduction of network services. Economic models suggest that market production of network services may often be inefficiently low because using a network imposes an external effect on other users of that network, an effect these other users typically disregard in making their own production decisions.[5]

Since expanding a network requires additional facilities, the new facilities create the possibility for new products and services. For example, suppose a business installs a fax machine, adding one machine to an existing network of 100 machines owned by other businesses. This installation allows the business to send messages to 100 other businesses, which, in turn, can send messages to the business with the new fax machine. The existing fax machine owners generally place a positive value on the extra machine, but typically do not subsidize its installation. And businesses deciding whether to install a new fax machine would not take into account the positive effect on other businesses. Thus, an externality exists in the purchase and use of network goods.[6] Because the prices for network goods and services do not typically reflect this externality, the consumption of network goods and services is expected to be inefficiently low in a competitive market: A business might decide it is too expensive to install the new fax machine, even though the value to the 100 other firms exceeds the cost of the machine.

Standards. The process of setting standards for network components is vital to achieving the compatibility that makes network complementarity fully possible. Setting standards can be done, as in the U.S. railroad case, by the marketplace, through cooperation (industry forums on setting standards), or by the government. Although the U.S. railroad indus-

[4]See the publications by Douglas Puffert; John Stover; and George Taylor and Irene Neu for excellent analyses of the history of U.S. railroads.

[5]This applies to models of markets with a competitive or a monopolistic structure.

[6]An externality exists when the decisions or activities of one entity affect, positively or negatively, the environment (excluding prices) of another. In the example in the text, the firm's decision to install a fax machine imposed a positive externality on the 100 other firms because it increased the ability of all the firms to communicate (and hence do business) more efficiently and more quickly with one another. The externality imposed by increasing network traffic need not always be positive, however. Network facilities, like many other economic facilities, can become congested: A negative externality is imposed when one party increases network traffic when the network is already operating at capacity.

try developed a standard gauge through market forces at work over half a century, one analyst of the issue called for legislation to lead the way in adopting one of the early gauges as a standard. Puffert quotes an unsigned commentary from an 1832 issue of the *American Railroad Journal:*

> It is a matter of regret with many of the friends of railroad improvements that no measure has been taken to insure a uniform width of track. The advantages of such uniformity must be perfectly obvious...we are forced to conclude that this discrepancy in the width of tracks will ultimately produce an infinitude of vexation, transfers and delays which might easily have been avoided.The establishment of a particular width, by statute, in two or three of the principal States, would probably have influence sufficient to produce the desired uniformity in most cases throughout the United States.

This commentator suggests the advantage of a mandatory, or legislated, standard. The "infinitude of vexation," occasioned by differing gauges, that persisted for decades could have been avoided. The disadvantage of the mandatory approach is that the legislature may decide on an inferior gauge.

Today, many industries cooperate in setting technical standards for products. For example, checks, smart cards, ATM cards, credit cards, and other components of the payment system are all carefully designed to maintain compatibility among different network components and providers of network services. The placement of information on the magnetic stripe on cards, the encryption devices and codes, and other technical standards must be common among the parties to a card-based payment for the system to operate. A cooperative industry-sponsored approach to standardization can achieve rapid adoption of standards while allowing those with the greatest interests and technical expertise to participate in setting the standards.

This type of cooperation among firms that are essentially supplying various components of an integrated product must be distinguished from collusion among competitors, which leads to price fixing and other anticompetitive outcomes. The practical difficulty lies in correctly identifying which type of cooperation is at work.

Installed Base. Sometimes, a technology used by an early leader in a network industry can establish a dominance that gives it an advantage over alternative technologies in the race for the industry standard. For example, the gauge eventually set for American rails was the early leader in number of miles of rail and the one most often used in the more industrialized Northeast and Midwest. This example reflects the fact that, in network industries, a large installed base of network facilities has an inherent advantage over new technologies that might otherwise satisfy consumer demand.[7] A technology that wins an early lead can serve as a template for other competing technologies: Those compatible with it have an advantage, and those not compatible have a disadvantage.

Furthermore, a large installed base of network facilities can increase demand for a particular system. Consider the competition among early telephone networks, which were not interconnected. For example, in a city with two telephone companies, the larger company could offer its subscribers wider service, making it unnecessary for those subscribers to purchase both companies' services. Hence, a large group of existing users enhances a system's chances for success.

Color television provides another example of the effect an installed base has on the pat-

[7]Indeed, Puffert points out that railroad engineers are still undecided on the technically preferred gauge.

tern by which people adopt a system. Early on, color TV was available in competing formats.[8] The first system approved by the Federal Communications Commission (FCC) in 1953 was incompatible with black-and-white receivers, and it never gained widespread acceptance. Also in 1953, the FCC approved an alternative system that was compatible with the existing black-and-white television system, so that consumers could receive programs transmitted in either format. Because local stations had to purchase expensive equipment to carry color programming, all programs weren't broadcast in color until 1970. Had consumers been forced to choose between keeping their old sets or buying new color sets—or owning both—and receiving limited programming on each system, the transition to color television would most likely have been delayed. This faster transition to color TV was accomplished at the cost of what some consider to be a lower quality system than others that were possible.

A well-established network with a large base of users can extend a firm's dominance into new products.

The influence that a large installed base has on the success of a network product points to the importance of three other common features in network industries: low introductory pricing, the role of expectations, and leveraging a firm's dominance in one product to dominance in another. All these potential business strategies reflect the explicitly dynamic (time-dependent) nature of network economics. The current size of the installed base influences the current demand for a service, but is itself the result of past decisions of those who supply and demand the service.

Low introductory pricing is a technique used to build a large base of users quickly, and it's a common one among many firms in network industries. If a firm succeeds in establishing its brand quickly among a sizable base of users, it has a good chance of charging higher prices to later users, and the higher prices will cover its costs. The competition between Netscape's Navigator and Microsoft's Explorer, in which both firms initially gave away their Internet browser, can be understood in this light.

The second feature related to an installed base is the role of expectations. When introducing new products and marketing existing ones, it's important to create expectations among current and potential consumers that the product will have a large installed base, even if it doesn't at present. For example, advertising for credit cards touts their ubiquity. Many people are establishing e-mail accounts, since e-mail is expected to become a permanent means of communication. For competitive firms, using false advertising and falsely undermining consumers' expectations about a rival's products are techniques that can illegitimately affect market expectations.

The success of compact disc players shows the importance of expectations. The widespread adoption of CD technology happened quickly, even though compact disc players were incompatible with the existing record technology and in spite of the installed base of record players. A key difference between the case of color TV

[8]The article by Neil Gandal and Rafael Rob presents the history of the adoption of color TV and compact disc players.

and that of CDs is that the firms that developed the compact disc player, Philips and Sony, jointly operated the first two plants for producing compact discs, ensuring a steady supply of discs for the players to use. Unlike the case with TV, recording studios did not have to purchase costly equipment to produce music for compact discs. Hence, potential purchasers could expect that virtually all recorded music soon would be available on compact discs, whereas some TV programs were produced and broadcast only in black and white for almost 20 years after the introduction of color TV.

A well-established network with a large base of users can extend a firm's dominance into new products. A firm that controls a network product can, in some cases, control the standard for a complementary product by incorporating the second product into its offering of the first. For example, the maker of a dominant computer operating system may incorporate more and more "applications" software in each succeeding generation of the operating system software. This tactic essentially leverages the ability of the provider of network services to influence the market's choice of new technologies. In some cases, of course, the dominant firm does so by offering the best possible second product, but this need not be the case. Indeed, by limiting the ability of competitors to introduce complementary (or next generation) products that are compatible with current technology, the dominant network's provider can gain an advantage for its own complementary products (which may be inferior to those sponsored by competitors).

A more worrisome abuse of the power held by a dominant network's provider is tying the sale of network services to ancillary products that consumers would otherwise be able to buy more cheaply from competing suppliers. In this case, the dominant provider enjoys an advantage because of the large base of users. Alternative providers of network services cannot compete, since their small size makes their product less valuable. Therefore, the dominant provider can overcharge for ancillary services because users will not defect to the alternative (smaller) network.

In an anti-trust suit, the U.S. Department of Justice accused Electronic Payment Services (EPS), the operator of the MAC ATM network, of tying the sale of ATM processing (an ancillary product) to the sale of ATM network access. In the 1994 consent decree, EPS agreed to allow other processors to compete for that ancillary service.

Access. Once established, a network that has a large base of users must determine which firms will have access to its facilities, that is, whether the network standard will be "open" or "closed." With an open standard, many firms can design and sell products compatible with the standard; a closed system limits the number of firms that can use the standard to sell products. In the late 1970s and the early 1980s, bank customers began to have access to their deposit accounts through ATMs. Most of these systems were proprietary and therefore closed. In the mid 1980s, many banks struck agreements to share access to ATMs, thereby creating shared ATM networks—an open standard.

Successful networks can create a type of monopoly called bottlenecks or essential facilities. By restricting access to such facilities, their owners place competing producers of a service at a significant competitive disadvantage. For example, the Telecommunications Act of 1996 directed the FCC to establish the detailed conditions under which competitors to the "baby Bells" could gain access to the local telephone network's lines and switches to provide telephone service. Without mandated access, a local phone network has little incentive to give competing providers access to its facilities (even at a cost). And without access to the local telephone network, the alternative provider would have to build a large network facility to attract a critical mass of users. By denying access to its competitors, the local network enjoys a considerable advantage over entrants into the market. Such bottleneck monopolies have been successfully challenged under the antitrust laws of the United States.

APPLICATION TO PAYMENT SYSTEMS

Recognition of the network characteristics of payment systems can yield insight into important public policy issues. To be successful, payment systems, which are technologies for the exchange of value among participants, must have wide acceptability. A card-based payment product, whether a debit or credit card, requires that consumers have the cards and that merchants have authorization terminals. These two pieces of equipment are complementary, and the more terminals that retailers deploy, the more potential transactions are available to a cardholder. Currency and coin, too, require a network of facilities for reading, counting, and sorting so that bills and coins can be accepted at vending machines.

The failure of the Susan B. Anthony $1 coin can be better understood once we recognize the dominance of the network effects that support the dollar bill. John Caskey and Simon St. Laurent argue that the popular explanation for the failure of the $1 coin—that it was poorly designed because it was hard to distinguish it from the quarter—is suspect. Although the Susan B. Anthony coin is similar to the quarter in terms of its color, reeded edge, and thickness, it weighs 43 percent more than a quarter, has the same size relationship to the quarter as the quarter has to the nickel, and has distinctly different engraving from the quarter. Instead of design, Caskey and St. Laurent focused on network effects, primarily those involving vending machines. The $1 coin can make higher value transactions easier in vending machines, but only if vending machine owners spend the resources necessary to convert their machines to accept the coins. They will do so if they expect the public to use the coins. Likewise, the public will use the coins if they expect them to be widely accepted. Neither of these expectations was met with the Susan B. Anthony because the $1 note remained in circulation.

In Canada, the adoption of a $1 coin (with some alternative design attributes—in particular, a gold color) was similarly met with a disappointing level of adoption, even though Canada's marketing campaign was much more extensive than the one the United States used for the Susan B. Anthony. However, the Bank of Canada began withdrawing the $1 note, an action that led vending machine companies to rapidly convert their machines to accept the coins. Today the $1 coin is the only circulating dollar in Canada. This experience focuses attention on the installed base of note users (and the machines and system by which the notes are handled). A large installed base that favors notes makes it difficult to influence expectations that the coin will gain general acceptance. And wide acceptance is needed for a coin to displace a successful (although more costly) note.

ATM networks yield numerous examples of the importance of network effects. Dennis Carlton and Alan Frankel offer one example of how compatibility can increase network output and convenience. They examine the output effects of the merger of the two ATM networks in Chicago—Cash Station and Money— in 1986. They point out that such a merger can lead to greater convenience and service because of the complementarity of the network components: the bank cards and ATMs. Carlton and Frankel state, "As the number of participants and terminals on the network increases, consumers might still be better off as a result of the increased network size and geographic density. As the number of participants and terminals on the network increases, consumers can rely more on the network. The *full* cost of using ATM services, including search costs and the risk of being unable to find an operating terminal, might have fallen even if some fees increased." Carlton and Frankel also show that the number of machines on the network and the number of transactions conducted by the network's members increased at more than double the national rate in the first full year following the merger. In addition, when the seven years following the merger are looked at as a whole, the number of machines and number of transactions increased faster than the national rate as well.

The merger of those two ATM networks made the cards of almost any Chicago bank compatible with any machine in the Chicago area, and it resulted in a significant increase in output, an outcome consistent with the theory of network effects.[9]

The recent growth of the "off-line" debit cards of Visa and MasterCard presents us with an example of the importance of an installed base of network facilities. The Visa and MasterCard off-line debit cards, also called check cards, can be used at the point-of-sale (POS) to electronically debit the cardholder's deposit account. Other POS card systems, known as "on-line" debit, are offered by the regional ATM networks, such as MAC, Honor, Star, and Pulse. The primary technical difference is that the on-line systems use a personal identification number (PIN), and the transaction is routed to the cardholder's bank for authorization; the off-line systems use a signature rather than a PIN and are routed to Visa and MasterCard for authorization. The off-line transactions typically are settled with the cardholder's bank within a few days after the transaction, while on-line systems typically settle the same day as the transaction or the following day.

The off-line systems have piggybacked on the extensive system of credit card authorization devices in retail operations around the world. Visa and MasterCard have insisted that retailers accept their check cards as long as they accept their credit cards. Off-line systems thus have a huge network with which their cards are compatible, resulting in great convenience for consumers.

In contrast, on-line systems have had to sell their product retailer by retailer. The retailer (who may already have a credit card authorization terminal) typically must purchase a PIN pad. Furthermore, while the off-line systems are accepted and have cardholders across the nation, each on-line ATM / POS system is accepted and has cardholders from only a particular region of the country. This more limited acceptance of the cards reduces their desirability for some retailers. By leveraging the widespread acceptability of the credit card authorization systems, the off-line cards quickly became the more heavily used of the two systems.

Some retailers have challenged Visa and MasterCard over their requirement that retailers must accept their check cards if they accept their credit cards. In an antitrust lawsuit, they allege that the credit card associations are guilty of an illegal tying arrangement, using their dominance in credit card acceptance to acquire dominance in the debit card marketplace. The check card transactions of Visa and MasterCard typically carry a higher fee for the retailer than do the on-line POS card transactions of the regional ATM / POS networks.

Another instance of the insight network economics can provide is in the continued dominance of the check for consumer bill payments and business-to-business payments. Checks are often derided as an inefficient means of payment compared with electronic alternatives. Kirstin Wells estimates that the total cost to society of a check is roughly double that of an automated clearinghouse payment.

Why haven't individual businesses and banks done more to convince check writers (possibly through sharing the potential cost savings) to move to electronic payment? The

[9]My 1995 article points out similar effects after the partial merger, the so-called duality agreement, between the two largest national networks, Plus and Cirrus, in 1991.

answer, in part, has to do with the large installed base of network facilities and business practices that support the check. The paper invoice that usually accompanies a check payment is universally accepted and understood. The current electronic alternative, electronic data interchange, is used only by a relatively small group of banks and firms. Until an electronic alternative to the paper invoice is widely available and gains dominance, payment by check will remain relatively convenient. And its very convenience reduces the incentives for firms to adopt an alternative system.

CONCLUSION

Networks have characteristics that create distinct business-policy issues for the providers and consumers of network services, including compatibility, access to network facilities, and the creation and exploitation of dominance in the provision of network facilities.

The economics of networks is an important advance in the economics of industrial organization, lending insight into important industries, including the payment system. In payment systems, the need for compatible facilities for the exchange of value gives rise to fundamental complementarities among system facilities, which is the hallmark of network economics.

Our understanding of the failure of the Susan B. Anthony dollar coin, the success of the off-line debit cards of Visa and MasterCard, the superior convenience of merged ATM systems, and the difficulty of replacing the check as a dominant means of payment are all enhanced by an understanding of network economics.

REFERENCES

Carlton, Dennis W., and Alan S. Frankel. "Antitrust and Payment Technologies," Federal Reserve Bank of St. Louis *Review*, November / December 1995, pp. 41-54.

Caskey, John P., and Simon St. Laurent. "The Susan B. Anthony Dollar and the Theory of Coin / Note Substitutions, Part 1" *Journal of Money, Credit and Banking*, August 1994, pp. 495-510.

Dybvig, Philip H., and Chester H. Spatt. "Adoption Externalities as Public Goods," *Journal of Public Economics*, Vol. 20, 1983, pp. 231-47.

Economides, Nicholas. "The Economics of Networks," *International Journal of Industrial Organization*, Vol. 14, No. 6, October 1996, pp. 673-99.

Farrell, Joseph, and Garth Saloner. "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *American Economic Review*, Vol. 76, 1986, pp. 940-55.

Gandal, Neil, and Rafael Rob. "The Dynamics of Technological Adoption in Hardware / Software Systems: The Chicken and Egg Problem," mimeo, The University of Pennsylvania, September 1996.

Katz, Michael, and Carl Shapiro. "Network Externalities, Competition and Compatibility," *American Economic Review*, Vol 75 (3), 1985, pp. 424-44.

McAndrews, James J. "Commentary," Federal Reserve Bank of St. Louis *Review*, November / December 1995, pp. 55-59.

Puffert, Douglas J. *The Economics of Spatial Network Externalities and the Dynamics of Railway Gauge Standardization*. Ph.D. Dissertation, Stanford University, March 1991, UMI Dissertation Information Service.

Rohlfs, Jeffrey. "A Theory of Interdependent Demand for a Communication Service," *Bell Journal of Economics*, 5, 1974, pp. 16-37.

Stover, John F. *American Railroads*. Chicago, 1961.

Taylor, George R., and Irene Neu. *The American Railroad Network 1861-1890*. Cambridge, Massachusetts, 1956.

Wells, Kirstin E. "Are Checks Overused?" Federal Reserve Bank of Minneapolis *Quarterly Review*, Fall 1996, pp. 2-12.