Are Bank Runs Contagious?

*Ted Temzelides**

Banks are a vital part of the economy because they provide an important channel through which many businesses get their financing. However, as we know from the history of the United States and other countries, banks can be subject to runs and panics. A panic that encompasses a large part of the banking system can seriously disrupt economic activity.

During a run, a bank experiences much heavier demand for deposit withdrawals than it can easily meet. If the run is severe enough, the bank will not be able to meet the demands of all depositors trying to withdraw money and, consequently, will have to suspend payments. During a panic, runs occur on a large number of banks.

Panics may occur because of regional or economywide problems, such as a real estate bust, during which the portfolios of many banks lose value. If depositors have not completely lost confidence in the banking system, they will transfer their deposits from failing banks to solvent banks. But panics may also occur because runs on a few banks cause depositors at other banks to lose confidence and, therefore, to withdraw *indiscriminately* from both solvent and insolvent banks. These types of panics, which involve runs on a few banks spreading to otherwise solvent banks, are said to involve contagion.

*When this article was written, Ted Temzelides was an economist in the Research Department of the Philadelphia Fed. He is now in the Department of Economics at the University of Iowa.

This general definition of contagion does not specify the precise reasons that bank runs might spread. Later, we will discuss several recent studies that test for different factors that can lead to contagion.

In the late 1800s and early 1900s, there were several episodes of widespread runs on banks. Even the 1980s saw a number of well-publicized runs, including those involving S&Ls in Maryland and Ohio, Penn Square Bank, and the Seattle First National Bank. Despite these more recent occurrences, the number of runs has fallen dramatically since deposit insurance was established in 1933.

Of course, deposit insurance is not without its own problems. For example, it has recently been criticized for distorting banks' incentives for taking excessive risk—the so-called moral hazard problem. This can lead to problems for institutions that are poorly capitalized. In fact, excessive risk-taking, fueled by distortions caused by deposit insurance, has been implicated in the S&L debacle of the 1980s. While a bank enjoys higher profits if risky projects pay off, it does not always have to pay for taking the additional risks. Rather, the FDIC bears the cost of paying off depositors when the bank cannot. Also, insured depositors, knowing the FDIC will repay them if their banks cannot, lose the incentive to try to assess the riskiness of their banks and, therefore, do not require higher interest rates from banks with riskier portfolios.

Although the FDIC Improvement Act of 1991 introduced risk-based deposit insurance premiums to try to mitigate excessive risk-taking, today's premiums do not vary much with the riskiness of a bank's portfolio. According to FDIC data, 94.4 percent of banks, which hold 96.8 percent of total deposits, pay the same premium (which is currently zero) for FDIC insurance. These banks are not identical in terms of the riskiness of their portfolios. Thus, the current risk-based premiums are unlikely to have a large effect on banks' risk-taking behavior.[1]

Because of the potential distortions caused by deposit insurance, some economists advocate scaling it back or even eliminating it. But before such steps are taken, analysis is needed to weigh the costs from the problems created by deposit insurance against its benefits in preventing panics. One necessary step in this analysis is to assess how serious the problem of runs would be if there were no deposit insurance. An insolvent bank should be allowed to close; however, its failure mustn't spread to healthy banks. If runs are contagious, deposit insurance, regulation, and the ability to borrow at the Federal Reserve's discount window may all play an important role in preventing runs from spreading. In this article, we will review the basic theory and present some recent evidence on contagious bank runs.

ILLIQUID ASSETS AND LACK OF INFORMATION

One of the major roles of banks is to provide liquidity in the economy by allowing depositors to withdraw money from their bank accounts whenever they want to. But while banks have liquid liabilities, they invest a large part of their portfolios in long-term illiquid assets, for example, real estate or business loans.[2] In normal circumstances, the bank's loan portfolio has some returns from loans that borrowers

[1]Part of the reason most banks currently pay the lowest premiums is that the Deposit Insurance Funds Act of 1996 restricts the FDIC from collecting excessive reserves. Section 7 of the Federal Deposit Insurance Act sets the target ratio of reserves to estimated insured deposits at 1.25. Since the banking industry has been performing well, failures and, therefore, insurance payouts, have been low, allowing the insurance premiums banks pay to fall. Nevertheless, uniform premiums, regardless of whether they are high or low, can distort banks' incentives for taking on more risk than is best for society.

[2]A typical bank holds a higher fraction of loans than securities in its portfolio; however, these fractions do vary over the business cycle. On average, over the last nine years, loans as a fraction of banks' total assets were 60.5 percent, while securities, which tend to be more liquid, were 19.6 percent of total assets.

are paying back. Also, the bank holds enough liquid assets, such as Treasury securities, to meet the usual demand for withdrawals.

However, if too many depositors want to withdraw their money, the bank will have to begin liquidating some of its long-term assets, for example, by selling them in a secondary market, before they mature. Typically, this early liquidation means the assets will not pay off as much as they would have, had the bank been able to hold them to maturity: the bank may have to sell the assets at "fire sale" prices. In other words, deposits are liquid—depositors can withdraw their money from the bank at any time. But loans are illiquid—it can be very costly to recall them and difficult for the bank to find a suitable buyer for them. While innovations in financial markets have permitted bank portfolios to become increasingly liquid—for example, through the securitization of mortgages and consumer loans—other bank assets, such as corporate loans, remain illiquid.

Since all banks keep only a fraction of their deposits as cash, any level of illiquidity makes them vulnerable if demand for withdrawals is high enough. This problem can become so severe that it can lead to insolvency. In a world without deposit insurance, if a depositor believes, for whatever reason, that her bank is about to become insolvent, she has an incentive to be the first to get her money out before the bank runs out of cash. If enough depositors panic and demand to withdraw their deposits, a run is created. Even healthy banks, whose assets would pay off in full if held to maturity, could fail if faced with a sufficiently large and unexpected amount of withdrawals. And the run might spread if depositors at other banks become worried as well.

When depositors at one bank start a run, why do depositors at other banks often follow suit? Banks' ability to handle unusually large withdrawals depends on what proportion of their assets is liquid and the quality of their illiquid assets. If a depositor believes that other depositors at her bank plan to withdraw their funds, she may start worrying about her own money. She knows that if withdrawals are large enough, the bank could fail. In this case, an amount less than the initial deposit will be left for her if she waits too long, so she may decide to withdraw her deposits immediately. If all depositors share her beliefs, a run could start and that bank could fail regardless of the condition of its assets. A run on one bank may lead depositors at other banks to form similar beliefs about the behavior of other depositors and to start a run on their banks. In this case, failures could spread among both solvent and insolvent banks because runs on a large number of banks could lead depositors to lose confidence in the banking system as a whole.

Alternatively, depositors might have some information about the quality of their bank's assets. If the assets turn sour—for example, during a period of unfavorable economic conditions—these depositors might start a run on the bank. Subsequently, depositors at other banks may start runs if they think their banks have assets similar to those of the first bank. Thus, panics can be triggered when depositors, in the light of new information, revise their beliefs about the quality of their banks' assets.[3]

In this case we might expect informed depositors to start runs mainly on troubled banks. Then, as they got more information about which banks were solvent, we would expect them to move their money from failing banks to healthy ones. Therefore, this type of run appears to be less costly for society. On the whole, it could even be beneficial, since monitoring bank performance helps to distinguish between good and bad banks. However, accurate monitoring relies on depositors' having perfect information about their banks' condition—but information about the economic condition of banks is almost never perfect. In times of financial distress, de-

[3]See, for example, Calomiris and Gorton (1991).

positors are particularly sensitive to any kind of news and may start runs on some liquidity-constrained but otherwise healthy banks, thereby causing them to fail.

Usually, economists view these as opposing theories of why runs occur, but real world episodes probably contain features of both. This article will discuss the evidence on whether certain historical episodes of bank failures have involved contagion and why or why not.

ONE VIEW: NO CONTAGION EFFECTS

During the National Banking Era (1863-1914), there were five major banking panics: 1873, 1884, 1890, 1893, and 1907, roughly one a decade. All these panics occurred before either the Federal Reserve System or deposit insurance was created.[4] In most of these episodes, large numbers of banks temporarily suspended the convertibility of de-

[4]See the box *Selected Bank Runs During the National Banking Era* for brief descriptions of the proximate causes and the main sequence of events during selected panics from this period.

Selected Bank Runs During the National Banking Era*

1873
Economic Environment: Railroad boom; four years of rapid economic growth.
Proximate Cause: Excessive loan expansion to railroad companies and decline of the railroad business.
Major Events:
Sept. 8: Suspension of the N.Y. Warehouse and Security Company, which was financially involved with the Missouri, Kansas, and Texas Railroad.
Sept. 13: Failure of Messrs. Kenyon, Cox & Co., which endorsed Canada Southern Railway paper.
Sept. 18: Major bank runs started.
Sept. 20: Closing of N.Y. Stock Exchange.
Sept. 24: Suspension of currency payments by N.Y. banks.

1893
Economic Environment: Monetary disturbance caused by Sherman Silver Purchase Act and international gold flows.
Proximate Cause: Stock market collapse in May.
Major Events:
Feb. 26: Failure of Philadelphia and Reading Railroad.
May 4: Failure of the National Cordage Company (trust company), which caused the stock market collapse.
July: Numerous bank failures throughout the nation. Reserve outflows from N.Y. banks.
Aug 5: Suspension of payments by N.Y. banks.

1907
Economic Environment: Steady rise in the price level and economic boom in the 1900s.
Proximate Cause: Failure of an attempt to corner copper stocks.
Major Events:
Oct. 17: Runs on the Mercantile National Bank, which attempted to corner the stock of the United Copper Company.
Oct. 21: Runs on Knickerbocker Trust Company, which was financially involved with the Mercantile National Bank. Major bank runs started.
Oct. 22: Suspension of Knickerbocker Trust Company.
Oct. 23: Major runs on trust companies.

*Taken from Sangkyun Park, "A Triggering Mechanism of Economywide Bank Runs," in Allin F. Cottrell, Michael S. Lawlor, and John H. Wood, eds., *The Causes and Costs of Depository Institution Failures,* Kluwer, Boston, 1995. Used with permission.

posits into cash. In other words, depositors could not withdraw their money from the bank. A suspension would typically start at banks in the East, commonly New York, and subsequently spread westward. Panics tended to occur during the fall when the demand for liquidity was higher, mostly because of the seasonal increase in demand for currency to cover needs related to agriculture and the seasonal movement of crops. Panics were also associated with recessions, during which nonbank businesses also experienced difficulties. But the nature of the economy is not the only determinant of banking system stability.

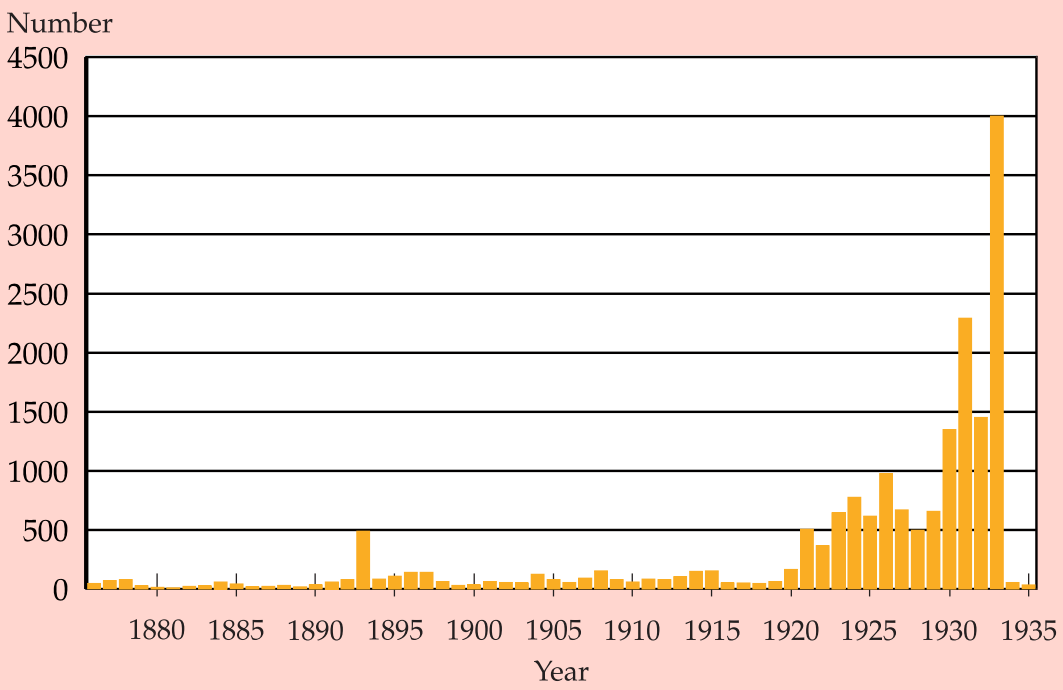
While the economies of the United States and Canada were quite similar in the late 1800s and early 1900s, the experiences of their banking systems were very different. During 1930-33, more than 9000 banks failed in the United States, but none failed in Canada. (Figures 1 and 2 show the number and percentage of bank failures in the United States from 1876-1935.) And unlike in the United States, panics were not widespread in Canada.

Stephen Williamson has argued that this difference in failures and panics was partly due to the structure of the banking systems in the two countries. Because of branching restrictions, the

FIGURE 1

Number of Bank Failures

1876 - 1935



Data Source: Table 2, Chapter 2, in George J. Benston, Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman, *Perspectives on Safe and Sound Banking: Past, Present, and Future.* Cambridge, MA: MIT Press, 1986.

U.S. system consisted of a large number of relatively small banks. In fact, in 1890, there were more than 7000 banks in the United States.[5] At the same time, Canada had a branch banking system without geographic restrictions—about 40 chartered banks with about 400 branches. Williamson argues that the ability of bigger banks to diversify in Canada was one of the factors that prevented widespread crises there in the early 1930s.[6]

Researchers are not in complete agreement about whether certain historical episodes in the United States were severe enough to constitute banking panics or were merely less significant episodes in which a small number of banks failed. Moreover, even when researchers agree that a certain episode was a panic, they do not always agree about whether contagion effects played an important role in the panic's development.

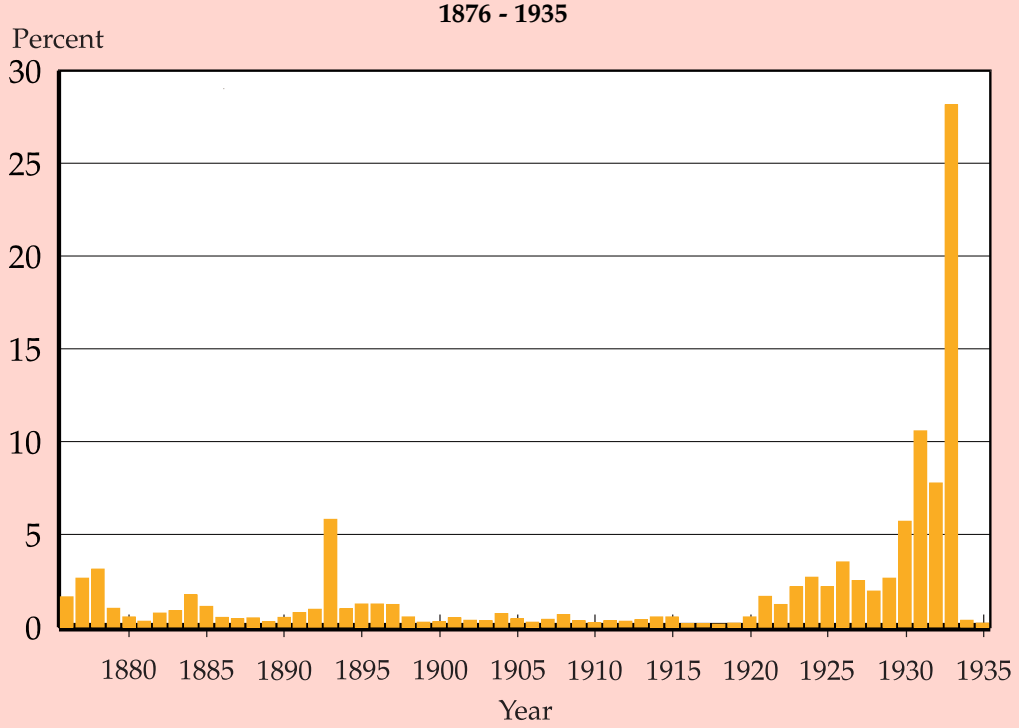
[5]By 1920, the number of banks in the United States had grown to more than 30,000.

[6]Nationwide branching was permitted in the United States as of June 1997.

FIGURE 2

Bank Failure Rate*

1876 - 1935



*Failures during the year as a percent of the total number of banks at year-end.

Data Source: Table 2, Chapter 2, in George J. Benston, Robert A. Eisenbeis, Paul M. Horvitz, Edward J. Kane, and George G. Kaufman, *Perspectives on Safe and Sound Banking: Past, Present, and Future.* Cambridge, MA: MIT Press, 1986.

Several researchers have questioned the widespread existence of contagion effects.[7] Instead, they argue that panics are the result of bad economic times that cause weak banks to become insolvent. They think it unlikely that depositors' loss of confidence in banks or the banking system can, by itself, cause a financial crisis. They argue that depositors who withdraw funds generally transfer them to another bank that is considered safe, in which case the total deposits of the banking system are not affected.

In addition, banks' own actions helped stop contagion. During some panics in the National Banking Era, bankers lent money among themselves to help one another meet the high demand for withdrawals. In several episodes, a coalition of banks, such as the New York Clearing House, acted collectively by issuing clearing house loan certificates to help banks that needed temporary liquidity. Because these certificates were joint liabilities of the entire group, this action helped the coalition's member banks that were in good standing but that were nevertheless facing liquidity demands they could not meet. While such cooperative arrangements did not succeed in eliminating panics altogether, there is evidence that they were successful in reducing the frequency and severity of panics. The "no contagion" view, then, maintains that, on the whole, contagious bank panics were rare events.[8]

A recent study by Charles Calomiris and Joseph Mason investigates whether, in the absence of deposit insurance, differences in information across depositors induced runs on solvent banks and involved contagion effects in the June 1932 banking panic in Chicago. Although the number of bank failures at the national or state level was not particularly high that month, there was a very strong concentration of bank failures in Chicago during the week of June 20. Calomiris and Mason report that 40 bank failures took place in Chicago in June, 26 of them during that week. In addition, the pattern of deposits shows that Chicago banks experienced a large decline in deposits during late June, and this pattern was not observed in other areas of the country.

By focusing on a particular location, Calomiris and Mason ensured that the banks they studied faced similar economic environments, e.g., the mix of industries of potential borrowers, and the incomes of potential depositors. Using a variety of measures, they investigated whether banks that failed during this episode were weaker and thus—panic aside—more vulnerable to declines in the prices of assets than banks that survived. The authors looked at market-to-book value of equity, interest rates paid on borrowings from other banks, and other accounting measures that gave them information about the probability of a bank's failure. They compared these factors at banks that failed during the panic with those at banks that survived. They found that the banks that went on to fail began with lower market-to-book values of equity, higher estimated probabilities of failure, and higher borrowing rates. In other words, they were weaker banks, and they shared characteristics with banks that failed outside the panic period.

Calomiris and Mason concluded that bank failures in Chicago in June 1932 were due to declines in asset values at the failed banks and not to panic-induced withdrawals because of depositors' confusion about the status of the banks or contagion.[9] Furthermore, they argued that contagion to solvent banks was avoided because the banks knew each other's status bet-

[7]See the articles by George Benston and coauthors; George Kaufman; and Charles Calomiris and Gary Gorton.

[8]Even this camp considers the events of late 1932 through early 1933, which brought the entire banking system to a halt, to be consistent with contagion effects. The disagreement between researchers seems to be about how frequently we would expect to observe such events in an environment without government deposit insurance.

ter than depositors did. Therefore, they helped each other by making loans backed by very high-quality assets. Hence, solvent banks that could post collateral avoided failure during the crisis. Some banks with sufficient collateral also borrowed from the Fed's discount window. But, in general, banks preferred borrowing from one another, since this action was less public: they feared that borrowing from the Fed might be misinterpreted by depositors as a sign of weakness.

Although it is hard to generalize on the basis of one event, the authors' findings are consistent with many studies that suggest that private interbank cooperation may be sufficient to reduce, although not necessarily prevent, contagious panics.

ANOTHER VIEW: CONTAGION EFFECTS

While it is difficult to distinguish between runs that occur at the same time in many banks and contagious ones, some researchers have been able to identify contagious runs.[10] Milton Friedman and Anna Schwartz have interpreted the panic of 1930 as a purely autonomous disturbance largely unrelated to the Depression and, thus, a candidate for a contagious panic.[11] Elmus Wicker studied the same event and argued that the crisis was precipitated by the collapse of the Caldwell financial empire. Caldwell and Company, located in Nashville, Tennessee, controlled the largest chain of banks in the South. Wicker attributes this failure to Caldwell's "weak and precarious financial state on the eve of the Depression." Caldwell's collapse caused depositors to revise their expectations about future deposit losses and affected more banks in later months. Wicker thus disputes the view that the panic of 1930 was a wholly autonomous event. His view does not rule out contagion, but makes it less likely that contagion was present.

Soon after Caldwell's closing, the Bank of United States also failed. Friedman and Schwartz maintained that the bank's name led to confusion about its official status, constituting a serious blow to depositor confidence.[12] They concluded that the banking panic of November-December 1930 was the result of a contagion of fear that spread among depositors, accelerating the bank failure rate, reducing the money stock, and worsening the economic downturn. (See *The Banking Panic of 1930*.)

Anthony Saunders and Berry Wilson found evidence for significant contagion effects during the period 1930-32 but no evidence of contagion during the panics of 1929 or 1933. Using regression techniques, Saunders and Wilson investigated the determinants of deposit withdrawal rates during these periods. Their analysis compares the deposit withdrawal rates at failing banks in the three years prior to the year of a bank's failure with the withdrawal rates at a matched sample of surviving banks. Each bank in the sample of banks that survived was

[9]But they did acknowledge that some confusion among depositors was present during this episode.

[10]George Kaufman, who maintains that the importance of contagion effects has been exaggerated, nonetheless cites examples of runs on neighboring banks that occurred after the announcement of negative news about the solvency of one institution. Kaufman also points out four periods in which the level of deposits in the banking system declined (1878, 1893, 1908, and 1930-33), a condition consistent with contagion.

[11]In addition to being a candidate for a panic in which contagion effects were present, the panic of 1930 provides an example of a crisis that developed after the Federal Reserve System was created but before deposit insurance was established.

[12]Confusion may have arisen because of the similarity in names between the Bank of United States and the first Bank of the United States and the second Bank of the United States, both of which were early attempts at establishing a central bank. However, the Bank of United States was a commercial bank with no special ties to the government.

The Banking Panic of 1930*

In several of the panics that occurred before the Federal Reserve System was established, banks temporarily suspended convertibility of deposits into currency. Such suspension of payments, often coordinated by banks in the New York Clearing House, successfully prevented panics from spreading when banks were suffering mainly from temporary liquidity problems. Restriction of payments by banks during the early signs of a panic protected the banking system by giving time for depositors' fears to wear off and for banks to regain liquidity. Once the danger of widespread runs had passed, banks resumed converting deposits into currency. During these suspensions, banks typically carried on with their usual operations: making loans, transferring deposits by check, and, in certain cases, converting limited amounts of deposits into cash, for example, so that firms could meet their payrolls. Such limited suspension was not without costs, but the costs were far smaller than those of the panics of the 1930s.

During the panic of 1930, early suspension of convertibility did not occur. Milton Friedman and Anna Schwartz maintain that, as a result, the panic of 1930 became the first of a series of crises that ended only after the banking holiday of March 1933.

Friedman and Schwartz emphasize the importance of the failure of the Bank of United States on December 11, 1930. This bank was the largest that had failed in the United States up to that time, and its failure provides an example of how the methods for stemming incipient panics had changed after the Fed came into existence. Despite various plans sponsored by the Federal Reserve Bank of New York and others to save the bank, the member banks of the New York Clearing House withdrew support and did not provide the new capital funds that would have helped in reorganizing the bank. Personal appeals by the state superintendent of banks and the lieutenant governor of New York were unsuccessful at changing the position of the clearing house. Instead, the president of the New York Clearing House suggested that the effects of closing the bank would be only local.

The bank, a member of the Federal Reserve System, borrowed from the Fed, but this borrowing was not sufficient to save it. It is not certain whether the bank could have raised collateral of high enough quality to back more substantial lending. In general, banks avoided borrowing from the Fed during periods in which fearful depositors were looking for signs of weakness, trying to predict which banks were likely to fail. Thus, the Bank of United States failed, and many others followed.

Friedman and Schwartz argue that under the pre-Federal Reserve banking system, banks would probably have restricted payments to depositors during the final months of 1930, which might have bought time for the panic to subside. They claim that the existence of the Federal Reserve prevented suspension by reducing the concerns of stronger private banks, which in the past had taken the lead in such a move. Furthermore, many people, assuming the Federal Reserve would deal with such crises, believed such a move was unnecessary. Had suspension of convertibility taken place during this episode, the Bank of United States might have been able to reopen, since this bank eventually paid off 83.5 percent of its liabilities at its closing, despite the fact that it had to liquidate a large fraction of its assets under unfavorable conditions.

*Some parts are based on Milton Friedman and Anna Schwartz, *A Monetary History of the United States.*

matched to a failed bank in terms of deposit size and the city in which it was located.[13] The authors reasoned that if a bank run was not due to a contagious panic, depositors should withdraw their money from bad banks and re-deposit it with good banks in the same locality. If, on the other hand, bank runs were developing into contagious panics, the authors would observe increased withdrawal rates at both good and bad banks as the time of failure approached.

Saunders and Wilson found that, for banks that failed in the period 1930-32, deposit withdrawals at good banks increased over the three years leading up to the failure of their matched bad banks.[14] Withdrawals at banks that failed and those that survived differed little in the years prior to failure. In the year of failure, failing banks lost, on average, a higher fraction of their deposits than banks that survived. For example, banks that failed in 1932 experienced, on average, withdrawals of 51.8 percent, while the matched control banks suffered withdrawals of 19.3 percent in that year.

The authors interpreted these observations as evidence consistent with contagion. In addition, regression analysis showed that in the period 1930-32, the rate of withdrawals at the matched surviving banks was significantly higher if the rate of bank failures in the surviving bank's state was higher or if the deposit outflows at their matched failing banks were higher. We would expect neither of these factors to be positively related to deposit withdrawals at the surviving banks if contagion were not present because, in the absence of contagion, we would expect to see funds moving from failing banks to healthy ones instead of withdrawals from both types.[15]

Joseph Aharony and Itzhak Swary studied the behavior of bank stockholders in more recent bank failures. Since these stockholders are not insured, how they acted might illuminate how depositors might behave in the absence of deposit insurance. The authors empirically tested whether stockholders draw inferences about the health of a bank by observing similar banks. Such observations may be one mechanism through which contagion arises.

The authors focused on the failure of five large banks in the southwestern United States during the mid-1980s.[16] They concentrated on a set of bank characteristics as a measure for the information on which depositors base their assessments of banks' riskiness. For example, the distance of a nonfailing bank's headquarters from a failed bank's headquarters may be particularly important: it's a good indication that both banks have similar loan portfolios, and hence face similar risks, because banks in the same location are subject to similar economic conditions and have similar types of borrowers. Size may also be important, since banks of different sizes may engage in different types of activities. For example, large banks tend to be more involved in wholesale activities, such as offering credit to large firms. So the failure of a large bank, if there is contagion, will likely have more of an impact on other large banks.

The authors' results indicated that the closer a large solvent bank is to a large failing bank, the stronger is the negative impact of the fail-

[13]If no such matched bank existed within the city, they chose the matching bank from a city of similar size.

[14]This was true even when differences in local economic conditions that could affect deposit withdrawals, for example, differences in personal income, were taken into account.

[15]Saunders and Wilson also show that contagion effects grew worse over the period 1930-32.

[16]The failure of the First National Bank of Midland, Texas, for example, in October 1983 resulted from a run by large depositors. This was the second biggest commercial bank failure in the United States.

ure on the solvent bank's stock return.[17] Joseph Aharony and Itzhak Swary also found that the larger the solvent bank is, the larger the decline of its stock return after the failure of another large bank. The results are consistent with information-based runs, in which the market assumes that similar banks are likely to have similar problems. The authors interpret their results as suggesting that the failure of a bank in one region should make regulatory authorities concerned about the possibility of contagion and, therefore, of consecutive runs on similar, but otherwise healthy, banks in the same region.

CONCLUSION

The study by Saunders and Wilson suggests that contagion effects were present in some cases and not in others. But additional research is needed before the *magnitude* of contagion effects during different episodes is documented with certainty. Research similar to the Calomiris and Mason study, which used bank-level data on local panics that occurred before federal deposit insurance existed, could prove useful in this endeavor as could more formal statistical tests.

Historically, the possibility of panics, and therefore of contagious panics, appears to depend on the structure of the banking system and the specific private arrangements designed to prevent such events. But private arrangements have limitations; historically, they did not eliminate banking crises altogether. Current evidence suggests that contagion effects have to be taken seriously in the debate about deposit insurance. Before drastic changes are made to the safety net, we should remember what banks experienced when they had to work without a net.

[17]Since data on stock returns are easily accessible, both depositors and stockholders may use them to evaluate the condition of their bank.

SELECTED BIBLIOGRAPHY

Aharony, J., and I. Swary. "Additional Evidence on the Information-Based Contagion Effects of Bank Failures," *Journal of Banking and Finance,* 20, 1996.

Bhattacharya, S., and A.V. Thakor. "Contemporary Banking Theory," *Journal of Financial Intermediation,* 3, 1993.

Benston, G.J., and others. *Perspectives on Safe and Sound Banking. Past, Present, and Future.* Cambridge: MIT Press, 1986.

Calomiris, C., and G. Gorton. "The Origins of Bank Panics: Models, Facts, and Bank Regulation," in R. Glenn Hubbard, ed., *Financial Markets and Financial Crises.* Chicago: University of Chicago Press, 1991.

Calomiris, C.W., and J.R. Mason. "Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic," manuscript (November 1995).

Friedman, M., and A.J. Schwartz. *A Monetary History of the United States 1867-1960.* Princeton: Princeton University Press, N.J., 1963.

SELECTED BIBLIOGRAPHY (continued)

Kaufman, G.G. "Bank Contagion: Theory and Evidence," manuscript (June 1992).

Lucia, J.L. "The Failure of the Bank of United States: A Reappraisal," *Explorations in Economic History,* 22, 1985.

Mester, L. "Curing Our Ailing Deposit-Insurance System," Federal Reserve Bank of Philadelphia, *Business Review* September/October 1990.

Park, S. "A Triggering Mechanism of Economywide Bank Runs," in Allin Cottrel, Michael Lawlor, and John Wood, eds., *The Causes and Costs of Depository Institution Failures.* Kluwer Academic Press, 1995.

Saunders, A., and B. Wilson. "Contagious Bank Runs: Evidence from the 1929-1933 Period," *Journal of Financial Intermediation* 5, 1996.

Wicker, E. "A Reconsideration of the Causes of the Banking Panic of 1930," *Journal of Economic History,* 6 (September 1980).

Williamson, S. "Bank Failures, Financial Restrictions, and Aggregate Fluctuations: Canada and the United States, 1870-1913," Federal Reserve Bank of Minneapolis, *Quarterly Review,* Summer 1989.

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 consid-

erable 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 num-

ber 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.