

2009-PB-01
January 2009

Two-Sided Markets, Bank Card Payment Networks, and Public Policy

David VanHoose

Abstract: This policy brief reviews recent studies that have sought to apply analysis of network externalities and the theory of two-sided markets to card payment networks, and it evaluates the public policy implications of this body of work. Three general conclusions emerge from the bulk of the research to date. First, this industrial-organization focus on the roles of network externalities and market two-sidedness in shaping the card payment industry's pricing structure is likely to prove much more fruitful for analysis of relevant public policy issues than the alternative approach to payment economics offered in the modern monetary economics literature that emphasizes dynamic general equilibrium modeling of payment systems. Second, in light of the complex balancing act entailed in managing competing interests of card issuers, payment acquirers, cardholders, and retailers in the presence of network externalities, governmental regulation of card payment networks' fee structures is unlikely to yield welfare improvements and in fact could yield significant inefficiencies. Third, welfare-improving competition among card payment networks is best promoted by permitting card issuers, payment acquirers, cardholders, and retailers to participate in multiple card payment networks, suggesting that recent antitrust challenges of restrictive rules imposed by some card payment networks have been appropriate.

About the Author: **David VanHoose** is a Senior Fellow at Networks Financial Institute. VanHoose earned his Ph.D. at the University of North Carolina at Chapel Hill and is currently the Herman W. Lay Professor of Private Enterprise and Professor of Economics at the Hankamer School of Business at Baylor University. He has published articles in such professional journals as the *Quarterly Journal of Economics*, the *International Economic Review*, the *Southern Economic Journal*, the *Journal of Money, Credit, and Banking*, *Economic Inquiry*, and the *Scandinavian Journal of Economics*.

Keywords: Bankcard payment networks, two-sided markets.

JEL Codes: G28, L41, L51

The views expressed are those of the individual author and do not necessarily reflect official positions of Networks Financial Institute. Please address questions regarding content to David VanHoose at David_VanHoose@baylor.edu. Any errors or omissions are the responsibility of the author.

NFI working papers and other publications are available on NFI's website (www.networksfinancialinstitute.org). Click "Research" and then "Publications/Papers."

Two-Sided Markets, Bank Card Payment Networks, and Public Policy

David VanHoose

1. Introduction

As noted by Kahn and Roberts (2009), the provision of payment-clearing services is a big business, with about \$75 in U.S. payments processed each year per \$1 of U.S. gross domestic product. More than 90 billion payments clear annually in the United States. These payments amount to more than \$1,000 trillion per year.

Although banking institutions are most commonly referred to as financial intermediaries in light of their roles in addressing asymmetric information and various risks related to asset transformation, they also function as *payment intermediaries*. Based on an examination of income statements of the 25 largest banking companies, Radecki (1999) has developed estimates of revenues generated by the banks' payments-related activities. Based on these estimates, he concludes that revenues generated by fund transfers on behalf of depositors accounted for 28 percent of the banks' total operating revenues. Credit card payment processing generated 7 percent of their revenues, and securities payment processing contributed an additional 3 percent. All told, therefore, Radecki concludes that transactions relating to payments processing generated about 38 percent of the banks' revenues.

In light of banks' traditionally significant role as payment intermediaries, Lacker and Weinberg (2003) review several studies related to banks' role in payment systems and suggest that more work should be done on "payment economics." Most of the work that Lacker and Weinberg surveyed examines large-value, or wholesale, payment systems—such as

the U.S. Clearing House Interbank Payment System or the Fedwire system— from central-planning perspectives of monetary economists. Some studies relating to such payment issues that matter most acutely to central banks, such as VanHoose (2000) and Holthausen and Rochet (2006), utilize an industrial organization point of view.

Nevertheless, and perhaps unsurprisingly, most payment-economics work related to the monetary economics literature focuses on funds-market and various liquidity-, credit-, and systemic-risk issues germane to monetary theory and policy. This orientation perhaps helps to explain why monetary economists commonly apply overlapping-generations models to the study of payment flows in overnight and intraday funds markets [see, for instance, Kahn and Roberds (2001), Martin (2004), Bhattacharya et al. (2007), and Monnet and Roberds (2009)]. Application of monetary theory to analysis of payments has provided interesting insights into issues relating to the role of fiat money in clearing payments, netting versus gross settlement, and monetary-theoretical issues such as the exact nature of the fiat-money equilibrium that results under different assumptions about payment sequencing or the potential optimality of the Friedman rule of optimal money growth and inflation.

Nevertheless, the more recent, industrial-organization-oriented research on payment economics is more relevant to regulation and antitrust issues confronted in retail payment systems, such as bank card payment networks, which over the years have been subjected to a variety of lawsuits, legal regulations, and antitrust actions. On the legal and regulatory front, most of the attention has been placed on interchange fees. In 1979, National Bankcard Association unsuccessfully challenged in U.S. courts the existence of interchange fees, which a payment network charges an issuing bank on

behalf of an acquiring bank. A few years ago, the Reserve Bank of Australia banned no-surcharge rules and began regulating interchange fees. In Denmark, interchange fees on domestic card payments are prohibited, in Sweden and the Netherlands, no-surcharge rules requiring retailers accepting a payment network's cards not to provide discounts for cash use are illegal, and in the European Union the European Commission recently intervened to bring down interchange fees. On the antitrust front, U.S. payment card networks have faced antitrust suits by the Department of Justice challenging restrictions on issuers and blocking a proposed merger between acquirers. Payment card networks have also confronted private lawsuits challenging the legality of "Honor All Cards" rules requiring retailers that honor a network's credit cards to accept its debit cards as well.

This policy brief reviews recent studies that have sought to apply analysis of network externalities and the theory of two-sided markets to card payment networks, and it evaluates the public policy implications of this body of work. In light of the early status of much of this work, these implications are tentative. Nevertheless, these provisional inferences challenge the efficacy of regulating the fee structure of card payment networks while simultaneously providing support for antitrust challenges of certain restrictive rules that some card payment networks have attempted to impose on participating banks and retailers.

2. Two-Sided Markets and Network Externalities

Chakravorti (2003) provides a helpful review of analyses of theories of the economic functioning of credit card networks. Most of the theories that he reviews, and virtually all subsequent work, are based on the concept of two-sided markets. So let us begin our discussion of regulatory and

antitrust issues relating to payment networks by explaining how researchers apply theories of two-sided markets to networks that clear payments initiated using payment cards such as those commonly issued by banks.

To consider the essential economics of a two-sided card payment market, consider Figure 1. In a two-sided market, one or more *platforms* facilitate dealings between *end users*. In the case of a card-payment network, the network itself—for instance, Visa, MasterCard, Discover, or American Express—constitutes the platform. Owners of the networks, which may or may not include banks, serve as *platform providers*. The end users are retailers and individual cardholders.

Figure 1: A Two-Sided Card Payment Network

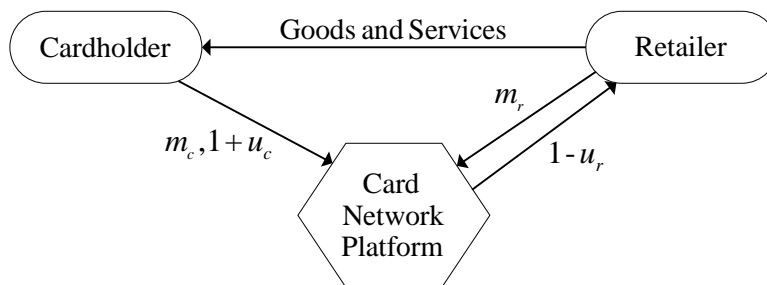


Figure 1 depicts an environment in which a card network potentially charges end users both usage fees (u_c and u_r) and membership fees (m_c and m_r). Let's ignore the latter for the time being—see discussion of membership fees below—by supposing that membership fees are equal to zero, so that the card network collects only usage fees. The network platform does so by collecting from a cardholder an additional per-dollar fee u_c for each dollar of payments that it transmits to a retailer on the cardholder's behalf. Out of each dollar of payments transmitted to retailers, however, the card network deducts an amount equal to u_r and thereby charges this usage fee to retailers. Thus, the total per-unit usage fee charged

to end users is $u = u_c + u_r$.

As discussed by Rochet and Tirole (2006a), if the aggregate volume of transactions processed by the card network platform were to depend solely on this total price—that is, if a change in u_c with u kept unchanged leaves aggregate card transactions unaffected—then the market would only be one-sided. Only the overall price would matter across all end users, implying that the network platform providers can vary that price without inducing asymmetrical responses on the part of cardholders and retailers. If the aggregate volume of transactions varies with u_c while u remains unchanged, however, then the market is two-sided. A change in the relative price structure alters total activity on the network. Consequently, the network platform providers must perform a delicate balancing act in developing a pricing structure that takes into account an asymmetry of price effects on the two sides of the market—analogously to the problem faced in other two-sided markets, such as markets for computer games and online dating services.

In the bank payment card industry, the values that end users assign to access to a given payment network depends on the number of other end users. For instance, an individual who owns a card entitling her to access payment-clearing services of a network benefits when others opt to use the same card, because this increases the likelihood that retailers will choose to accept the network's card. Thus, economists say that products such as payment services provided via card payment networks are subject to *network externalities*, or third-party spillover effects onto some network participants resulting from decisions by others with whom they engage in no direct transactions.

Why do retailers accept payment cards? Undoubtedly, the primary

explanation is transactional benefits—in the form of sales that otherwise would not take place—accruing to the retailer. Hayashi (2006), for instance, motivates retailers' acceptance of credit and debit cards by the upward shifts in product demands confronting individual merchants, which thereby yield additional sales. In an analysis of retailers' motivations in accepting credit card payments, Chakravorti and To (2007) propose an additional incentive. They emphasize the revolving function of credit cards and proposed that competing merchants accept credit cards because doing so permits them to close current sales with illiquid consumers rather than confronting uncertainty regarding future sales. Through recognition of actual and anticipated sales expansions, Rochet and Tirole (2006b) suggest, retailers essentially internalize the benefits experienced by customers, generating benefits to themselves, as well, and hence opt to accept a payment network's cards as long as these and other potential benefits exceed the network's fees. Of course, the cardholder benefits that retailers seek to internalize are subject to network externalities, so merchant end users also confront externalities in card payment systems.

The two-sidedness of card payment networks' pricing structures hinges on the existence of these external spillovers. As discussed by Rochet and Tirole (2006a), the Coase theorem indicates that in the presence of clearly defined property rights and in absence of transaction costs and asymmetric information, private parties should be able to establish contracts that internalize usage externalities and attain a social optimum. Two-sidedness of the price structure of card payment networks, they suggest, results primarily from transaction costs and constraints established by the networks or policymakers that prevent implementation of such welfare-improving contractual agreements.

The act of providing a card payment network entails fixed expenses unrelated to variable usage of the network. Retailers incur fixed costs to establish card-acceptance systems, and acquirers experience significant fixed expenses in developing systems for clearing payments. Issuers also incur considerable fixed costs related to marketing and advertising. All of these fixed costs influence decisions regarding whether or not to participate in a card payment network, thereby determining the magnitudes of external benefits experienced by end users.

Why are there also often membership fees to participate in networks? It is well known from the industrial organization literature regarding price discriminating two-part tariffs and Ramsey pricing that including access fees can efficiently enable the recouping of fixed costs. This fact undoubtedly helps to explain why membership fees such as those depicted in Figure 1 are often observed within the overall pricing structure of card payment networks. Rochet and Tirole (2006a) also point out that if a desired pricing structure includes a role for cross-subsidization, access prices such as membership fees can be employed to capture portions of end-user surpluses for such purposes. Armstrong (2006) provides a recent model in which flat membership charges emerge as important tools for balancing the externalities faced by end users in two-sided markets.

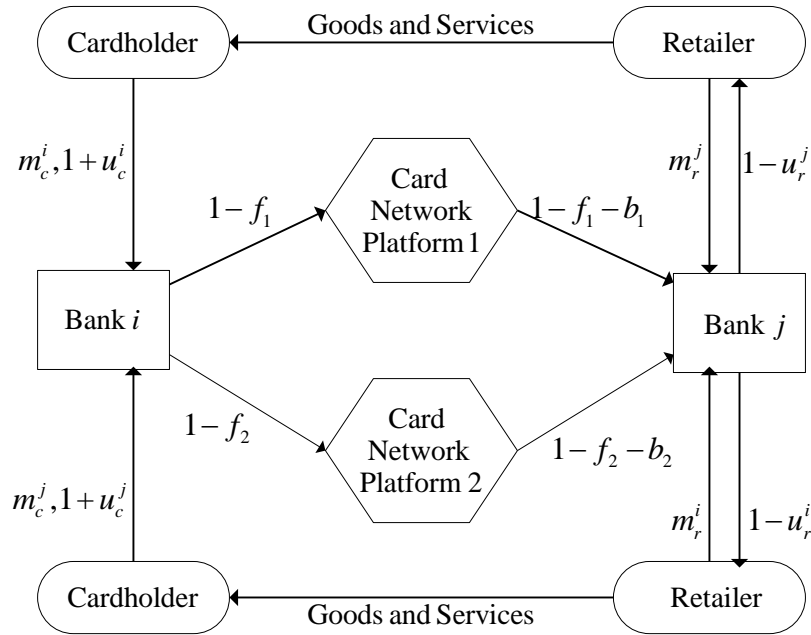
3. Multiple Levels of Participation and Pricing Structures in Card Payment Networks

As explained by Schmalensee (2002), the manner in which usage fee structures of a card payment network are established hinges on whether the network is *proprietary* — that is, owned and operated as an individual profit-maximizing “closed network,” as in the case of the American Express card

network— or *cooperative*— that is, owned and operated by an association of coordinating institutions, such as the Visa bank card “open network.” The basic two-sided structure depicted in Figure 1 is most applicable to a *unitary* proprietary network that engages in all issuing and acquiring activity on its own behalf. As Schmalensee notes, an alternative proprietary system is a *non-unitary* system in which the network contracts with other parties, such as a selected set of institutions, to do some of the issuing and/or acquiring. In contrast, in a cooperative card payment network, independent institutions establish a contractual arrangement for coordinating card payments via the network.

Figure 2 provides a stylized depiction of co-existing card payment networks. The figure displays two institutions, Bank i and Bank j , that are members of two networks, Card Network Platform 1 and Card Network Platform 2. For the sake of simplicity, the figure shows Bank i as an issuer with respect to both card network platforms and displays Bank j as a payment acquirer with respect to the two platforms. In fact, some banks perform both roles in cooperative networks, although increasingly banks that operate as acquirers contract with third parties to process payments.

Figure 2: Competing Card Payment Networks



Note that pricing-structure complexity is greater in Figure 2. The banks continue to charge membership and usage fees to cardholders and retailers. In addition, the networks charge two types of fees to member banks. The fees b_1 and b_2 denote membership and/or usage fees that the network charged the banks, and the fees f_1 and f_2 denote interbank fees—the interchange fees that the network collects from an issuing bank on behalf of an acquiring bank. Baxter (1983) provided the first formal analysis of interchange fees within a bank payment-clearing association, which he showed must take into account the two-sided nature of the payment transactions. In Baxter’s model, with perfectly competitive net-issuing and net-acquiring banks, the association itself is indifferent regarding alternative levels of the interchange fee—a result sometimes called “interchange-fee neutrality” in the literature on card payment networks.

As discussed by Schmalensee and by Rochet and Tirole (2002), allowing for imperfect competition results in considerable scope for

variation in pricing structures across card payment networks. Based on the analysis of card payment systems provided by Evans and Schmalensee (1999), most studies of networks' two-sided pricing structures presume low search costs and little product differentiation on the acquiring side of card payment networks, which researchers often offer as support for assuming perfectly competitive behavior among acquiring institutions. (As Evans and Schmalensee document, however, concentration among acquirers has increased noticeably.) Search costs and the scope for product differentiation are presumed to be higher on the issuing side, resulting in some market power for issuers.

To minimize complexities, some studies, such as those of Rochet and Tirole (2002), Bolt and Tieman (2006), and Wright (2003, 2004) assume that rival card payment networks process a volume of transactions that is either arbitrarily fixed or essentially predetermined via an exogenously specified distribution of consumer transaction choices. Rochet and Tirole (2002) further examine a setting in which issuers have no incentive to operate as acquirers, which naturally further simplifies their framework of analysis. In the context of their model, Rochet and Tirole find that in the absence of other pricing complications such as surcharges for purchases with cash instead of cards, the interchange fee that maximizes overall social welfare is never higher than the profit-maximizing level. Indeed, without surcharges the profit- and output-maximizing interchange fees coincide, implying that, if anything, outcomes can arise in which there is a socially inefficient overprovision of payment cards. When card surcharges are permitted within the Rochet-Tirole framework, there is an underprovision of payment cards, and welfare implications are ambiguous.

Wright (2003) allows for heterogeneity in retailers' benefits from

network participation not considered by Rochet and Tirole and concludes that attaining the welfare-maximizing interchange fee must balance differing externalities faced by heterogeneous cardholders and retailers. As a consequence, the profit-maximizing aggregate usage level typically lies below the socially optimal level. Wright (2004) concludes that the welfare effects of imposing a “one-price policy” in which surcharges for card purchases are prohibited depends on the amount of market power possessed by retailers. If retailers possess market power, then the ability to impose surcharges would enable retailers to set prices to extract surplus from cardholders, thereby reducing revenues accruing to issuers. In the absence of retailer market power, however, a one-price policy results in a bifurcation of the retailing market into cash-only and card-only segments.

Schmalensee (2002) also studies the functioning of a single card payment platform in the context of a basic usage-pricing framework. He finds that a unitary proprietary platform that simply maximizes its own private value establishes a set of fees that generates the smallest transaction volume. Essentially, a unitary proprietary platform acts as a profit-maximizing monopoly and restrains output via a structure of sufficiently high fees that restrains production of transaction services to maximize its profits. In contrast, in Schmalensee’s model, a cooperative system that aims to attain an output objective while covering operating costs balances benefits from spreading payment volumes across separate net-issuing and net-acquiring groups utilizing the payment network.

Within this single-network framework that abstracts from banks’ interactions with cardholders and presumes fixed weights in network decision-making by banks that are either issuers or acquirers, Schmalensee derives a decomposition of the optimal interchange fee into two parts, with

one part depending on the difference between demand elasticities across cardholders and retailers and the other part depending on the difference in costs faced by issuers and acquirers. He concludes that a cooperative platform typically will set a pricing structure that yields, relative to unitary and non-unitary proprietary platforms, the highest volume of payment transactions. A double-marginalization (or double-monopoly-markup) problem confronts multiple participants in both a cooperative network and a non-unitary network, with the latter producing a volume of payment transactions that is intermediate between the volumes forthcoming from cooperative and unitary proprietary networks.

Eisenmann et al. (2008) contemplate governance issues faced by two-sided platforms that include both platform providers responsible for serving as users' primary point of contact with a platform and platform sponsors that are involved in determining who may participate in a platform-intermediated network and for developing its technological workings—much as banks do in card payment networks such as Visa and MasterCard. Eisenmann et al. argue that, over time, forces tend to push two-sided networks toward constructions that are hybrids of proprietary and cooperative structures. In the long run, they conclude, such networks naturally gravitate toward centralized control over platform technology via closed sponsorship of a platform and shared responsibility for serving end users via openness of platform providers. This could help explain why the American Express and Discover networks have recently sought to encourage participation by banks.

Bolt and Tieman (2006) study a model of a monopoly two-sided platform such as the one depicted in Figure 1 and hence ignore the issue of interchange fees. In addition, they abstract from membership fees. In this

simplified setting, they demonstrate that when the socially optimal pricing structure is implemented, the monopoly platform fails to cover its operating costs. Thus, they conclude, network externalities faced by the card payment platform create a second-best pricing problem analogous to that faced by a natural monopoly, in which first-best, allocatively efficient pricing generates negative economic profits. A possible consequence, Bolt and Tieman suggest, is a card payment network pricing structure that includes cross-subsidization schemes, higher interchange fees, and no-surcharge rules preventing retailer participants from giving discounts to customers who pay with cash.

When card network platform providers compete for end users, such as the case of a card-network duopoly illustrated in Figure 2, the range of complexities faced in analyzing pricing structures broadens considerably. Nearly all of the work analyzing such competitive-network settings within the context of two-sided-market frameworks has appeared very recently. In these settings, end users can “multihome”: cardholders can participate in more than one network by utilizing more than one payment card, or retailers can accept cards from more than one network from their customers.

Rochet and Tirole (2002) briefly touch on the impacts of competition between payment systems via multihoming by cardholders. They contend that if one network seeks to undercut another by setting a slightly lower interchange fee, then retailers have an incentive to respond by accepting the former system’s card and rejecting the card of the other system, potentially reducing welfare relative to the monopoly case.

In a more general two-sided-market framework, Rochet and Tirole (2003) examine a situation in which “buyers” (cardholders in the payment-network context) utilize the services of differentiated platforms that either

are profit-maximizing firms or non-profit associations. Their model applies to an extension of the setting in Figure 1 to the case of more than one platform competing for the same end users. Rochet and Tirole include buyers and sellers of different types, with a network platform competing to obtain “marquee buyers” —cardholders who generate a particularly high surplus to retailers—from another platform and taking into account the existence of “captive buyers” that remain loyal to their initial platform choice. Other things being equal, the presence of marquee buyers naturally raises the seller price, but the presence of captive buyers skews the pricing structure to the benefit of sellers. Naturally, an increase in predisposition to multihoming on the part of buyers also is beneficial to sellers. These results could be sensitive, however, to Rochet and Tirole’s assumption of no fixed membership-type fees, which Armstrong (2006) suggests could complicate joint determination of pricing structures in competing two-sided markets. Rochet and Tirole find that under some circumstances interchange fees could be lower and total market surplus greater in a monopoly setting than when two networks compete.

Cabral (2005) suggests, however, that Rochet and Tirole’s ambiguous welfare results regarding monopoly versus competition between two-sided networks does not necessarily rule out the existence of other more-than-offsetting welfare losses that might arise from existence of a monopoly network. Furthermore, a recent analysis by Chakravorti and Roson (2006) studies a variety of interactions between payment networks, including duopoly and cartel behavior with or without symmetry of the networks’ payment instruments (such as credit cards offered by one platform and debit cards on another). In the context of their framework, which includes some narrowing assumptions such as monopolistic retailers, Chakravorti

and Roson conclude that competition unambiguously raises the welfare of end users, although not generally uniformly, even in the case of symmetric competition.

Guthrie and Wright (2007) attempt to provide a classification of the wide array of equilibrium pricing-structure outcomes that can result from competition between two card identical payment platforms depending on the natures of potentially multihoming cardholders and retailers and on their behavior. In general, Guthrie and Wright find that either both networks set the same structure of fees with at least one side of end users (cardholders or retailers) multihoming, or only one network survives and attracts the exclusive participation of all end users. Guthrie and Wright conclude that in a setting with homogeneous retailers, competition between networks cannot boost fees charged retailers. They argue that if, in contrast, retailers are more realistically heterogeneous, platform competition can result in higher fees being assessed on retailers, thus leading to higher interchange fees. The fee structure can be further biased against retailers if consumers, knowing that retailers will accept more than one card in equilibrium, choose a preferred card to hold and thereby benefit from platform competition that focuses on attracting cardholders rather than retailers. In contrast, if retailers know that cardholders always hold multiple cards in equilibrium, retailers can “steer” cardholders to their preferred network, resulting in platform competition focusing on attracting retailers, to the benefit of the latter group of end users.

4. Regulatory and Antitrust Issues in Card Payment Networks

Is it appropriate for interchange fees and no-surcharge rules to be regulated, as is a common practice in Europe? Should card-payment-

network rules be subjected to antitrust scrutiny? There has been some initial work evaluating these regulatory and antitrust issues. In Schmalensee (2002), the levels of interchange fees under either cooperative or non-unitary proprietary systems are driven primarily by different weights within the network on the part of net-issuing versus net-acquiring banks. In the context of his analysis of a single card network platform, Schmalensee finds that the proprietary, private-value maximizing interchange fee theoretically may be higher or lower than an output-maximizing one that cooperative network platform providers might set. Furthermore, his analysis suggests that a policy action that might raise the interchange fee conceivably could boost network transaction output and that a policy action that reduces the interchange fee potentially could depress network transactions output. Policy intrusions into the fee-setting process, he concludes, could place cooperative networks at a competitive disadvantage with respect to proprietary networks, leading institutions to abandon higher-output cooperative networks in favor of lower-output proprietary networks. Schmalensee's conclusion, therefore, is that there is no clear economic argument in favor of antitrust policy that interferes in the setting of interchange fees.

As noted by Rochet and Tirole (2008), regulation of interchange fees is often motivated as a means of promoting purely cost-based fees. As they point out, however, applying this logic would yield the conclusion that the same types of governmental fee regulations should be applied to the television industry, newspapers, videogame platform providers, and other two-sided industries. Furthermore, Rochet and Tirole (2006) question whether retailers actually experience harm as a result of interchange fees and suggest that some perceived harm may result from the fact that

heterogeneous retailers internalize customer benefits at different levels.

Schwartz and Vincent (2006) utilize a card-network model with variable transaction volumes to examine the implications of no-surcharge rules. They argue that such rules imbalance the fee structure between end users, resulting in harm to retailers, with networks responding by reducing cardholder fees and granting rebates to card users if possible, with cardholder welfare potentially declining if rebates are not feasible. Network profits increase, and overall welfare increases only if there a sufficiently large number of cash users.

Rochet and Tirole (2006c) examine the effects of an honor-all-cards rule. In their benchmark model based on Rochet and Tirole (2002), they find that in fact such a rule benefits not only the multi-card network platform providers that impose it but also allows the platform providers to optimally re-balance externalities across end users. The result is greater volumes of both credit and debit payments processed by the network, resulting in greater social welfare. Generalizing their framework to allow for heterogeneous retailers, differentiated platforms, and varying substitutability between credit and debit transactions yields the same re-balancing effect on the multi-card network's pricing structure. The implications for welfare, however, tend to become ambiguous depending on parameter values.

Emch and Thompson (2006) contemplate the application of the Department of Justice merger guidelines to card payment networks [see VanHoose (2009) for a detailed discussion of these guidelines]. Emch and Thompson show that it is possible that monopolization of a two-sided market considerably raises the prices charged to cardholders but not retailers, or *vice versa*. They derive price markup formulas for fees charged

to end users. Based on these relationships, they propose applying the guidelines' standard test for a "small but significant and non-transitory increase in price" that might result from a proposed merger to the sum of the prices the network charges to both groups of end users. White (2006) notes, however, that the pricing relationships derived by Emch and Thompson hinge on their assumption of fixed proportions. He also concludes that their results likely depend on the simple two-sided framework they utilize, which more readily apply to proprietary networks than to open card associations.

Sun and Tse (2007) apply a differential game analysis to both monopolistic and competing payment networks in an effort to examine the impacts of multihoming. They conclude that a greater tendency for end users to engage in multihoming makes it more likely that networks can co-exist but that the steady-state market shares of competing networks can diverge considerably depending on participants' propensities to engage in multihoming. They also argue that network distributors, such as card issuers, play a crucial role in maintaining and expanding a network. Sun and Tse's analysis provides some measure of support for the Department of Justice's actions charging Visa and MasterCard with antitrust violations for their rules forbidding banks from becoming issuers of American Express and Discover cards.

5. Conclusion

Whereas monetary-theoretic frameworks of analysis emphasize how the nature of payments can complicate the nature of a monetary equilibrium, the industrial organization approach takes as given the existence of a monetary system in which payment networks are imbedded.

Furthermore, in contrast to monetary-theoretic models of payment systems, the modern industrial organization approach focuses on the two-sided nature of card payment network platforms with which banks interact. This approach highlights the importance of two-sided pricing structures constructed with an aim to take into account network externalities that differentially affect cardholders and retailers.

So far, this industrial organization approach has yielded a number of tentative conclusions and policy implications. First, the privately and socially optimal pricing structures of two-sided card payment networks result from a complicated balancing of the incentives faced by end users that confront network externalities.

Second, platform providers are likely to charge cardholders and retailers unbalanced fees, and the optimal setting of interchange fees typically will reflect this asymmetrical pricing structure. Thus, regulation of interchange fees will not necessarily improve overall social welfare, nor will legal restrictions on card payment networks' rules governing surcharges. Indeed, in many cases, regulation of fees and rules could well be harmful.

Third, some recent work indicates that in light of the fact that the pricing structure of card payment networks can be decomposed and analyzed using standard industrial organization techniques, it may be possible to develop methods of applying current antitrust policy guidelines to card payment networks. Nevertheless, this conclusion must be regarded as tentative given that this initial work is predicated on a number of simplifying assumptions.

Fourth, a tentative conclusion is that rules established by platform providers requiring retailers to honor debit cards if they accept the network's credit cards are not necessarily anti-competitive. Such rules may

be consistent with an aim to effectively balance interests of all end users within a coherent pricing structure. More work must be done to establish conditions under which honor-all-cards rules are welfare-improving rather than welfare-reducing.

Finally, the application of the theory of two-sided markets to card payment networks suggests that rules restricting participation of banks in competing networks may be anti-competitive. Efforts to steer banks away from a competing platform naturally help restrain the growth of issuers within the competing network and thereby reduce the potential for the competing network to establish a critical mass of distributors. As a consequence, rules restricting bank participation in competing networks reduce the likelihood of a multihoming outcome involving multiple rival card payment networks, thereby potentially reducing social welfare.

In their recent survey of payment economics, Kahn and Roberds (2009, p. 17) reach the following conclusions:

“While [the industrial organization, I/O] literature has provided numerous theoretical insights, some notable challenges remain. The first is a closer integration of the I/O literature with the monetary literature surveyed above. The second is a need for more consensus building in economists’ assessment of current pricing structures in the industry, a process that has been hindered by the lack of available data.

It is difficult to argue with the second conclusion. As Kahn and Roberds note, however, a significant difficulty faced by researchers in the payment economics area is a paucity of available data. Certainly, most data from retail payment systems such as card payment networks are proprietary and not within easy reach of economists.

The first conclusion is more debatable. Many modern monetary

economists have a preference for theories driven by specifications of underlying preferences. A common complaint by proponents of the monetary-theoretic approach is that the industrial-organization methodology is often too “static” and lacks development from “first principles” utilized to generate dynamic-equilibrium outcomes. Only in such theoretical settings, it is often claimed, are the most general welfare implications to be obtained.

It is arguable, however, that a contrary conclusion is appropriate, namely that it is the monetary-theoretic approach to payments economics that should be integrated with the industrial-organization, two-sided-markets methodology. The former tends to abstract from issues, such as network externalities and imperfect competition, that are highly relevant to public policy issues involving regulation and antitrust to which the industrial-organization methodology naturally lends itself. The monetary-theoretic approach’s insistence on imbedding a single equilibrium type of monetary configuration within a particular specification of payments technology also fails to reflect the reality of widespread diversity and continuing evolution of payment-network technologies within prevailing monetary arrangements. That is, monetary-theoretic models typically examine a static payment system in order to analyze dynamics of an economy’s monetary structure, whereas in the world we actually inhabit, payments-system developments tend to be dynamic within a relatively static monetary structure.

In the near term, therefore, the industrial-organization methodology is much more likely to yield results of greatest use to policymakers. Researchers new to the payment economics area would be well advised to begin their study with the industrial organization literature on two-sided

markets. Furthermore, monetary economists who desire to evaluate deeper implications of payments systems for the optimal quantity of money, interest rates, and inflation would be well advised to consider how to imbed such questions into models of two-sided payment-network platforms.

REFERENCES

- Baxter, William, 1983, Bank interchanges of transactional paper: Legal and economic perspectives, *Journal of Law and Economics* 26, 541-588.
- Bhattacharya, Joydeep, Joseph Haslag, and Antoine Martin, 2007, Federal Reserve Bank of New York Staff Report No. 281, April.
- Bolt, Wilko, and Alexander Tieman, 2006, Social welfare and cost recovery in two-sided markets, *Review of Network Economics* 5, 103-117.
- Cabral, Luis, 2005, Market power and efficiency in card payment systems: A comment, *Review of Network Economics* 5, 15-25.
- Chakravorti, Sujit, 2003, Theory of credit card networks: A survey of the literature, *Review of Network Economics* 2, 50-68.
- Chakravorti, Sujit, and Roberto Roson, 2006, Platform competition in two-sided markets: The case of payment networks, *Review of Network Economics* 5, 118-140.
- Chakravorti, Sujit, and Ted To, 2007, A theory of credit cards, *International Journal of Industrial Organization* 25, 583-595.
- Eisenmann, Thomas, Geoffrey Parker, and Marshall Van Alstyne, 2009, Opening platforms: How, when, and why? In Annabelle Gawer, ed., *Platforms, Markets, and Innovation*, Cheltenham, UK: Edward Elgar, Forthcoming.
- Emch, Eric, and T. Scott Thompson, 2006, Market definition and market power in payment card networks, *Review of Network Economics* 5, 45-60.
- Evans, David, and Richard Schmalensee, 1999, *Paying with Plastic: The Digital Revolution in Buying and Borrowing*, Cambridge, Massachusetts: MIT Press.
- Guthrie, Graeme, and Julian Wright, 2007, Competing payment schemes, *Journal of Industrial Economics* 55, 37-67.

- Holthausen, Cornelia, and Jean-Charles Rochet, 2006, Efficient pricing of large value interbank payment systems, *Journal of Money, Credit, and Banking* 38, 1797-1818.
- Kahn, Charles, and William Roberds, 2009, Why pay? An introduction to payments economics, *Journal of Financial Intermediation* 18, 1-23.
- Kahy, Charles, and William Roberds, 2001, Real-time gross settlement and the costs of immediacy, *Journal of Monetary Economics* 47, 299-319.
- Lacker, Jeffrey, and John Weinberg, 2003, Payment economics: Studying the economics of exchange, *Journal of Monetary Economics* 50 (2), 381-387.
- Martin, Antoine, 2004, Optimal pricing of intraday liquidity, *Journal of Monetary Economics* 51, 401-424.
- Radecki, Lawrence, 1999, Banks' payment-driven revenues, Federal Reserve Bank of New York *Economic Policy Review* 5, 53-70.
- Rochet, Jean-Charles, 1992, Capital requirements and the behavior of commercial banks, *European Economic Review* 36, 1137-1178.
- Rochet, Jean-Charles, and Jean Tirole, 2008, Competition policy in two-sided markets, with a special emphasis on payment cards, in Paulo Buccirossi, ed., *Handbook of Antitrust Economics*, Cambridge, Massachusetts: MIT Press, pp. 543-582.
- Rochet, Jean-Charles, and Jean Tirole, 2006a, Two-sided markets: A progress report, *Rand Journal of Economics* 37, 645-667.
- Rochet, Jean-Charles, and Jean Tirole, 2006b, Externalities and regulation in card payment systems, *Review of Network Economics* 5, 1-25.
- Rochet, Jean-Charles, and Jean Tirole, 2006c, Tying in two-sided markets and the honor all cards rule, Working Paper, Toulouse University, February 9.
- Rochet, Jean-Charles, and Jean Tirole, 2003, Platform competition in two-sided markets, *Journal of the European Economic Association* 1, 990-1029.

- Rochet, Jean-Charles, and Jean Tirole, 2002, Cooperation among competitors: Some economics of payment card associations, *Rand Journal of Economics* 33, 549-570.
- Schmalensee, Richard, 2002, Payment systems and interchange fees, *Journal of Industrial Economics* 50, 103-122.
- Schwartz, Marius, and Daniel Vincent, 2006, The no surcharge rule and card user rebates: Vertical control by a payment network, *Review of Network Economics* 5, 72-102.
- Sun, Mingchun, and Edison Tse, 2007, When does the winner take all in two-sided markets? *Review of Network Economics* 6, 16-40.
- VanHoose, David, 2008, Policy Implications of Endogenous Sunk Fixed Costs in Banking: Has U.S. Antitrust Policy Been on the Wrong Track? Networks Financial Institute *Policy Brief* 2008-PB-06.
- VanHoose, David, 2000, Central bank policy making in competing payment systems, *Atlantic Economic Journal* 28, 117-139.
- White, Lawrence, 2006, Market definition and market power in payment card networks: Some comments and considerations, *Review of Network Economics* 5, 61-71.
- Wright, Julian, 2004, The determinants of optimal interchange fees in payment systems, *Journal of Industrial Economics* 52, 1-26.
- Wright, Julian, 2003, Optimal card payment systems, *European Economic Review* 47, 587-612.