

Economics of payment cards: A status report

Wilko Bolt and Sujit Chakravorti

Introduction and summary

The proliferation of payment cards—that is, debit, credit, and prepaid cards—has dramatically changed the way we shop and merchants sell goods and services. Today, payment cards are indispensable in most advanced economies. According to a recent U.S. survey, the percentage of payment cards used for in-store purchases increased from 43 percent in 1999 to 56 percent in 2005 (American Bankers Association and Dove Consulting, 2005). For Europe, Bolt and Humphrey (2007) report that the number of card payments increased by 140 percent across 11 European countries during the period 1987–2004.¹ Amromin and Chakravorti (2009) find that greater usage of debit cards has resulted in lower demand for small-denomination bank notes and coins that are used to make change.² Furthermore, without payment cards, Internet sales growth would have been substantially slower.

Debit, credit, and prepaid cards are three forms of payment cards. Debit cards allow consumers to access funds at their banks³ to pay merchants; these are sometimes referred to as “pay now” cards because funds are generally debited from the cardholder’s account within a day or two of a purchase.⁴ Credit cards allow consumers to access lines of credit at their banks when making payments and can be thought of as “pay later” cards because consumers pay the balance at a future date. Prepaid cards can be referred to as “pay before” cards because they allow users to pay merchants with funds transferred in advance to a prepaid account.⁵ (We ignore prepaid cards in our discussion.)

Recently, some merchants have started to accept only card payments for safety and convenience reasons. For example, a cafe in Washington, DC, stopped accepting cash for purchases primarily because the cost of safekeeping cash was too expensive (Rafsanjani, 2006). Also, many quick service restaurants and coffee shops now accept payment cards to capture greater sales and increase transaction speed. Wider acceptance and

usage of payment cards suggest that a growing number of consumers and merchants prefer payment cards to cash and checks.

As more consumers and merchants adopt payment cards, providers of these products may benefit from economies of scale and scope. In the United States, being able to operate on a national level allowed some issuers (banks that issue cards to consumers), acquirers (banks that convert payment card receipts into bank deposits for merchants), and payment processors to benefit from economies of scale and scope. Some European payment providers might enjoy these benefits in the future as greater cross-border harmonization occurs with the introduction of the Single Euro Payments Area (SEPA).⁶ The primary focus of SEPA is to create a uniform framework not only for card payments, but also for electronic credit transfers and direct debits, so that these retail payments can be completed in the euro area without intermediation by other banks. The potential advantages of SEPA are increased competition among a greater number of payment providers and the realization of scale economies and more-efficient payment instruments.⁷

The increased usage of cards has increased the value of payment networks, such as Visa Inc., MasterCard Worldwide, Discover Financial Services, and others. (We describe how payment networks operate in more detail later.) Earlier this year, Visa Inc. had the largest initial public offering (IPO) of equity in U.S. history, valued at close to \$18 billion (Benner, 2008). The sheer

Wilko Bolt is a senior economist in the Research Department at De Nederlandsche Bank (the Dutch central bank). Sujit Chakravorti is a senior economist in the Financial Markets Group at the Federal Reserve Bank of Chicago. The authors thank Gene Amromin and Anna Paulson for comments. The views expressed are those of the authors and do not represent the views of De Nederlandsche Bank and the European System of Central Banks.

magnitude of the IPO suggests that financial market participants value Visa's current and future profitability as a payment network. One potential reason for Visa to change its corporate structure from a card association to a publicly traded company is to reduce antitrust scrutiny by regulators and to lower the threat of lawsuits filed by certain payment system participants.⁸ In 2006, MasterCard Worldwide became a publicly traded company. Also, in 2007, Discover Financial Services was spun off by Morgan Stanley.

Some industry observers have suggested that the high profitability of payment card providers has increased scrutiny by public authorities in many jurisdictions.⁹ Several U.S. merchants have filed lawsuits against MasterCard and Visa regarding the setting of interchange fees. Interchange fees are generally paid by the merchant's bank to the cardholder's bank. These fees are set by the network—and not bilaterally negotiated among the banks in the network. In addition, the U.S. Congress is considering legislation about the level and determination of merchant fees. Recently, the European Commission (EC) ruled that the (multilateral) interchange fees applied by MasterCard in Europe violated Council Regulation (European Commission) No. 1/2003: The EC said that MasterCard's fee structure restricted competition among acquiring banks and inflated the cost of card acceptance by retailers without leading to proven efficiencies.¹⁰

To date, there is still little consensus—either among policymakers or economic theorists—on what constitutes an efficient fee structure for card-based payments. In this article, we discuss several theoretical economic models that analyze whether intervention by public authorities might improve the welfare of payment system participants. These models consider the costs and benefits of payment card usage versus other types of payments—for example, cash—and the underlying pricing of payment services under various types of market structures for payment providers and merchants. We address the following questions in this article:

- What is the optimal structure of payment fees between consumers and merchants?
- Will competition among payment providers, networks, or instruments improve consumer and merchant welfare?
- What guidelines should policymakers follow when regulating fees for payment services?

The rest of our article is organized as follows. We first explain how a payment network operates. Having established the payment network framework, we discuss the costs and benefits of providing and using payment cards relative to various other types of payment

instruments. Next, we review the key contributions to the theoretical payment card literature. We consider papers with models that focus on interchange fees, price differentiation at the point of sale, network competition, the role of credit, and the pricing of payment services when a bank provides competing payment instruments. We also discuss the impact of these factors on social welfare.

Payment flows in a card network

A payment network must convince both a buyer and a seller to use its payment service before a transaction can take place. For the purpose of this article, we define a payment service narrowly as the provision of a payment instrument that is used by consumers to pay merchants for goods and services. The consumption of a payment service requires participation of two distinct end-users—consumers and merchants.

The two-sided market literature has been used to analyze the structure of prices paid by consumers and merchants.¹¹ The price structure or balance is the share that each type of end-user pays of the total price of the payment service.¹² This literature combines the multiproduct firm literature, which studies how firms set prices on more than one product, with the network economics literature, which studies how consumers benefit from increased participation in networks by other consumers.¹³ Rochet and Tirole (2006) define a two-sided market as a market where end-users are unable to negotiate prices based on costs to participate on a platform and the price structure affects the total volume of transactions. In the payments context, consumers and merchants generally do not negotiate prices of goods and services based on the payment instrument used to make a purchase.¹⁴ For example, the prices for goods and services are the same regardless of whether the consumer pays in cash or with a payment card. However, when merchant fees increase, some merchants might refuse to accept payment cards, resulting in fewer potential card transactions. Similarly, raising consumer fees may reduce consumer participation. In other words, if there exists a ratio of consumer fees to merchant fees where more transactions occur than another ratio of fees, that market is said to be two-sided.

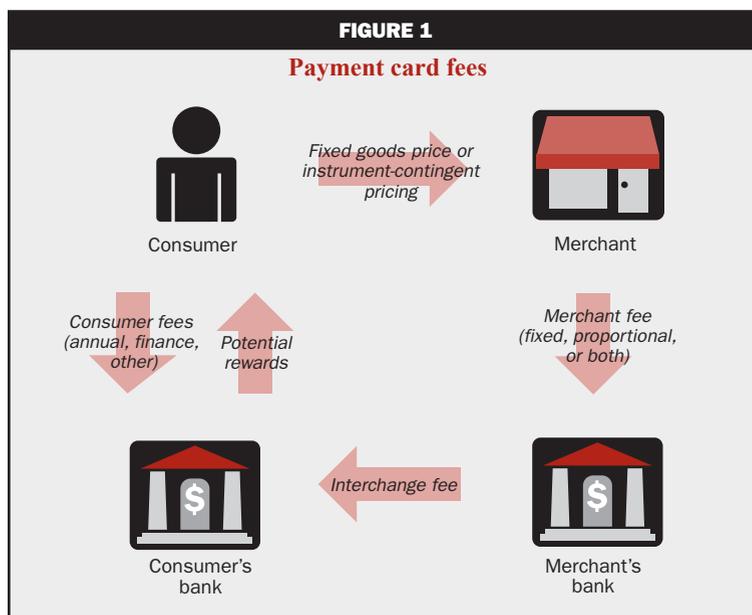
Most payment card transactions occur in three- or four-party networks.¹⁵ These networks comprise consumers and their banks (known as issuers), as well as merchants and their banks (known as acquirers). Issuers and acquirers are part of a network that sets the rules and procedures for clearing and settling payment card receipts among its members.

In figure 1, we diagram the four participants and their interactions with one another. First, a consumer establishes a relationship with an issuer and receives a payment card. Consumers generally do not pay per transaction fees but often pay annual membership fees to the banks that issue the payment cards. In addition, many payment card issuers give their customers per transaction rewards, such as cash back or other frequent-use rewards. Second, a consumer makes a purchase from a merchant. Generally, the merchant charges the same price regardless of the type of payment instrument used to make the purchase. (In some instances, merchants may set different prices for the same good or service based on the type of payment instrument used.) Third, if a merchant has established a relationship with an acquirer, it is able to accept payment card transactions. The merchant pays either a fixed per transaction fee (more common for debit cards) or a proportion of the total purchase amount, known as the merchant discount fee (more common for credit cards), to its acquirer.¹⁶ For credit cards, the merchant discount can range from 1 percent to 5 percent, depending on the type of transaction, type of merchant, and type of card, as well as whether the card is present or not, among other factors. Fourth, the acquirer pays an interchange fee to the issuer.

If acquiring markets are competitive or if merchants have significant bargaining power, the merchant discount approaches the interchange fee. The interchange fee is set by the network rather than by each issuer and acquirer bilaterally. The interchange fee has drawn antitrust scrutiny in various jurisdictions on the grounds that centrally setting interchange fees may reduce competition and harm consumers and merchants.

Costs and benefits of different payment methods

Here, we explore the costs and benefits of providing and using various payment instruments. Studying the costs to banks to provide payment services is difficult, given the proprietary nature of the cost data. However, there are some European studies that attempt to quantify the real resource costs of several payment services. In these studies, social cost refers to the total cost for society net any monetary transfers between



participants, and reflects the real cost of resources used in the production and usage of payment services. For the Netherlands in 2002, Brits and Winder (2005) report that the social cost of all point-of-sale (POS) payments (cash, debit cards, credit cards, and prepaid cards) amounted to 0.65 percent of gross domestic product (GDP). The social cost of payment services for Belgium in 2003 was 0.75 percent of GDP (Quaden, 2005). Bergman, Guibourg, and Segendorff (2007) find that the social cost of providing cash, debit card, and credit card payments was approximately 0.4 percent of GDP in Sweden for 2002. For Norway, Humphrey, Kim, and Vale (2001) estimate the cost savings from switching from a fully paper-based system (checks and paper “giro,” or a payment in which a payor initiates a transfer from her bank to a payee’s bank) to a fully electronic system (debit cards and electronic giro) at the bank level at 0.6 percent of the nation’s GDP.¹⁷ Based on a panel of 12 European countries during the period 1987–99, Humphrey et al. (2006) conclude that a complete switch from paper-based payments to electronic payments could generate a total cost benefit close to 1 percent of the 12 nations’ aggregate GDP.¹⁸

These numbers confirm the widespread agreement that the ongoing shift from paper-based payments to electronic payments may result in large economic gains. Compared with cash, electronic payments also offer benefits in terms of greater security, faster transactions, and better recordkeeping; in addition, electronic payments offer possible access to credit lines.¹⁹ Merchants may also benefit from increased sales or cost savings by accepting an array of electronic payment instruments.

However, these benefits to consumers and merchants are often difficult to quantify.

Using U.S. retail payments data, Garcia-Swartz, Hahn, and Layne-Farrar (2006) attempt to quantify both the costs and benefits of POS payment instruments.²⁰ They find that shifting payments from cash and checks to payment cards results in net benefits for society as a whole, but they also conclude that merchants may be paying a disproportionate share of the cost. Much of the payment card literature focuses on the proportion of the total price paid by merchants and consumers. In other words, economists are trying to answer the question: Do the sum of prices to end-users for card payments and their peculiar asymmetric structure reflect the exercise of market power by card providers, or do they reflect the nature of the service provided? In the next section, we consider how the economics literature has attempted to answer this question.

Economic models of payment cards

In this section, we review some important contributions to the theoretical payment card literature. The results from various economic models in this literature may differ because of differences in their underlying assumptions. The early models of payment cards ignored strategic interactions of consumers and merchants, and focused on the aggregate demand of each type of end-user and the level of the interchange fee. These models were extended to include explicit consumer and merchant interactions but assumed inelastic consumer demand for goods, no price differentiation by merchants based on the payment instrument used by consumers, and exogenous benefits from card usage. Several models extended this literature by considering merchants' ability to separate consumers by charging different prices. Another set of models expanded the literature by considering the ability of payment cards to increase sales because the cards provided greater security and eased consumers' liquidity and credit constraints. More recently, models of payment cards have considered network competition, as well as competition among different types of payment instruments.

Models focusing on interchange fees

Here, we discuss the academic literature on interchange fees.²¹ Baxter (1983) considers an environment where consumers are homogeneous, merchants are perfectly competitive, and the market for issuing and acquiring payment cards is competitive. He argues that the equilibrium quantity of payment card transactions occurs when the total transactional demand for credit card services, which are determined

by consumer and merchant demands jointly, is equal to the total transactional cost for credit card services, including both issuer and acquirer costs, or:

$$f + m = c_I + c_A,$$

where f is the willingness to pay for a consumer, m is the willingness to pay for a merchant when demand for payment services equals the supply of payment services, and c_I and c_A are the issuer's marginal cost and the acquirer's marginal cost, respectively. A consumer's willingness to pay is based on her net benefits received, b_B , and is greater than or equal to the fee in equilibrium.²² Similarly, the merchant's fee, m , is less than or equal to the net benefits it receives, b_S . Note that this equality does not mean that simultaneously $f = c_I$ and $m = c_A$. Hence, pricing each side of the market based on marginal cost—as would be suggested by economic theory for one-sided competitive markets—need not yield the socially optimal allocation. To arrive at the socially optimal equilibrium, a side payment may be required between the issuer and acquirer. To achieve the socially optimal prices, the side that receives more than its cost pays the one that earns less than its cost via the interchange fee.

Unfortunately, Baxter's framework does not allow us to study the optimal setting of interchange fees by banks, since their profits are zero regardless of the level of the interchange fee. Extensions of Baxter's model relax the assumption of perfectly competitive markets for payment services and consumption goods. We explore these contributions in turn and focus on their implications to the welfare of market participants.

Schmalensee (2002) extends Baxter's analysis by allowing issuers and acquirers to exercise market power but still assumes that merchants operate in competitive markets. His results support Baxter's conclusions that the interchange fee balances the demand for payment services by each end-user type and the cost to banks to provide them. Furthermore, the socially optimal interchange fee is not likely to be zero, even when issuers and acquirers have market power.²³ Schmalensee finds that the profit-maximizing interchange fee of issuers and acquirers may also be socially optimal for a conventional measure of social welfare with a strong set of assumptions.²⁴

Unlike Baxter (1983) and Schmalensee (2002), Rochet and Tirole (2002) consider strategic interactions of consumers and merchants. In their model, issuers have market power, but acquirers operate in competitive markets. Thus, any increases in interchange fees are passed on to merchants completely. They consider two identical Hotelling merchants²⁵

in terms of their net benefits of accepting a payment card for sales (that is, they have the same b_s) and the goods that they sell. Consumers face the same fixed fee, f , but are heterogeneous in terms of the net benefits, b_B , they derive from using the payment card. Only consumers with $b_B \geq f$ will adopt a payment card and use it for all purchases. Furthermore, Rochet and Tirole (2002) assume that the total number of transactions is fixed and changes in the prices of payment services do not affect the demand for consumption goods.

Rochet and Tirole (2002) have three main results. The first result is that their socially optimal interchange fee is higher than the socially optimal Baxter (1983) interchange fee, since issuers exert their market power and capture merchants' surplus. Their second result is that the interchange fee that maximized profit for the issuers may be more than or equal to the socially optimal interchange fee, depending on the issuers' margins and the cardholders' surplus. Third, merchants are willing to pay more than their net benefit if they can steal customers from their competitors or retain their customers by accepting cards.²⁶ However, overall social welfare does not improve when merchants steal customers from their competitors by accepting payment cards.

Wright (2004) extends Rochet and Tirole (2002) by considering a continuum of industries where merchants in different industries receive different benefits from accepting cards. His model is better able to capture the trade-off between consumer benefits and merchant acceptance when the interchange fee is increased because some merchants will not accept cards. This is in stark contrast to the knife-edge decision to accept or reject cards by all merchants obtained by Rochet and Tirole (2002). In Wright's (2004) environment, both consumer and merchant fees are per transaction. Each consumer buys goods from each industry. Issuers and acquirers operate in markets with imperfect competition. Wright (2004) assumes that consumers face the same price regardless of which instrument they use to make the purchase.

Wright (2004) concludes that the interchange fee that maximizes overall social welfare may be higher or lower than the interchange fee that maximizes the number of transactions. In particular, restricting the total number of transactions by setting higher interchange fees raises total welfare if the gain in surplus of the marginal card user who now *starts* using his card, along with all those merchants who accept his card, exceeds the loss in surplus of the inframarginal merchant who now *stops* accepting cards, along with all those card users who can no longer use their cards for purchases at her store. To be socially optimal, the fee

structure should reflect this asymmetry in inframarginal benefits, something that a card provider's private choice of interchange fee may not take into account.

Models with price differentiation at the point of sale

The models discussed so far have largely ignored the ability of merchants to pass on part or all of their payment cost to consumers—whether in the form of higher prices to their card-based consumers or as a higher uniform price to all consumers.²⁷ In some cases, merchants are not allowed to add a surcharge for payment card transactions because of legal or contractual restrictions.²⁸ However, in jurisdictions where merchants are free to set higher prices for purchases made with payment cards, they usually do not.²⁹ Even if differential pricing based on the payment instrument used is not common, the possibility to do so may enhance the merchants' bargaining power in negotiating their fees. If merchants charged different prices, cash-paying consumers would either not subsidize or only partially subsidize merchant fees for processing card transactions.³⁰

Wright (2003) extends Rochet and Tirole (2002) to consider the effects of no-surcharge rules when merchants are monopolists or Bertrand competitors.³¹ Wright (2003) assumes that each consumer demands a unit of each good that makes up a basket of consumption goods. Similar to the consumers and merchants in Rochet and Tirole (2002), consumers are heterogeneous in terms of the benefits they receive from using cards, and merchants are homogenous in terms of the benefits they receive from card acceptance. Another key assumption is that consumers and merchants make their decisions to participate in payment networks prior to knowing goods prices.

Wright (2003) finds that no-surcharge rules generate higher welfare than when monopolist merchants are allowed to set prices based on the payment instrument used.³² He argues that merchants are able to extract consumers' surplus *ex post* from payment card users, while cash users are unaffected. This result is driven by cash users paying the same price regardless of whether there is one price or multiple prices for the same good. Thus, cash users are made no worse off by differentiated prices. Because the monopolist merchant has already fully extracted surpluses from consumers who use cash when setting one price, an increase in the cash price would result in these consumers not making any purchases because of their inelastic demand for each good. Wright (2003) only considers equilibriums where merchants will continue to sell to cash users. When merchants are allowed to surcharge, they extract "too much" surplus *ex post* from customers who use payment cards because merchants set higher prices

for card purchases. This results in lower welfare for the remaining consumers making card purchases compared with the scenario in which merchants set one price regardless of payment instrument used. In addition, some card customers in a uniform price environment convert to cash in an environment that differentiates prices depending on which payment instrument is used; this conversion results in a complete loss of these consumers' surplus.

Schwartz and Vincent (2006) study the distributional effects among cash and card users with and without no-surcharge rules.³³ They relax the assumption that the demand for the consumption good is fixed. However, they assume that consumers are exogenously divided into cash and card users and cannot switch into the other group. In this environment, if merchants charge one price regardless of the instrument used, the price rises for cash users because merchants have to support more expensive card purchases. Unlike previous models, their model considers the issuer and acquirer as one entity. While they are unable to explicitly model interchange fees, they are able to study the price structure of payment services. If either the issuing or acquiring market is competitive, results obtained in this setting would be identical to those derived in a four-party network.

Schwartz and Vincent find that the absence of pricing based on the payment instrument used increases network profit and harms cash users and merchants. The payment network prefers to limit the merchant's ability to separate card and cash users by forcing merchants to charge a uniform price to all of its customers. When feasible, the payment network prefers rebates (negative per transaction fees) given to card users.³⁴ Granting such rebates boosts card user demand while simultaneously forcing merchants to absorb part of the corresponding rise in the merchant fee, since any resulting increase in the uniform good's price must apply equally to cash users. In this way, the network uses rebates to indirectly extract surplus from cash-paying customers in the form of higher prices. If rebates are feasible, card users are always better off. Overall welfare rises if the ratio of cash users to card users is sufficiently large and merchants' net benefits from card acceptance are sufficiently high.

Gans and King (2003) argue that, as long as there is a form of "payment separation," the interchange fee is neutral regardless of the market power of merchants, issuers, and acquirers. The interchange fee is said to be neutral if a change in the interchange fee does not change the quantity of consumer purchases and the profit level of merchants and banks. When surcharging is costless, merchants will implement

pricing based on the payment instrument used, taking away the potential for cross-subsidization across payment instruments and removing the interchange fee's role in balancing the demands of consumers and merchants. In effect, the cost pass-through is such that lower consumer card fees (due to higher interchange fees) are exactly offset by higher goods prices from merchants. Payment separation can occur if one of the following is satisfied: There are competitive merchants, and they separate into cash-accepting or card-accepting categories, in which each merchant only serves one type of customer and is prevented from charging different prices; or merchants are able to fully separate customers who use cash from those who use cards by charging different prices. Therefore, Gans and King argue that policymakers should remove any merchant pricing restrictions, such as no-surcharge rules.

Models with competition between networks

We have not yet considered models where competition among payment networks is explored. Economic theory suggests that competition generally reduces prices, increases output, and improves welfare. However, with two-sided markets, competition may yield an inefficient price structure. A key aspect of network competition is the ability of end-users to participate in more than one network. When end-users participate in more than one network, they are said to be "multi-homing." If they connect only to one network, they are said to be "singlehoming." As a general finding, competing networks try to attract end-users who tend to singlehome, since attracting them determines which network has the greater volume of business. Accordingly, the price structure is tilted in favor of end-users who singlehome.

Rochet and Tirole (2003) extend their previous work by considering network competition. Their primary focus is on the price structure or balance between consumers and merchants in a three-party network. They do not explicitly model the interchange fee but study the impact of competition on the structure of prices. Under a set of plausible assumptions they find that the price structures for a monopoly network and competing platforms are the same, and if the sellers' demand is linear, this price structure in the two environments generates the highest welfare under a balanced budget condition.

Guthrie and Wright (2007) extend Rochet and Tirole (2003) by assuming that consumers are able to hold one or both payment cards and that merchants are motivated by "business stealing" in deciding whether to accept payment cards in a four-party network. They only consider networks that provide identical payment services, and they find that network competition

results in higher interchange fees than those that would be socially optimal. Competition results in both networks charging the same interchange fee or all transactions occurring on one network. This result is not surprising given that both networks offer identical payment products.

Chakravorti and Roson (2006) extend Rochet and Tirole (2003) by considering the effects of network competition on total price and on price structure where networks offer differentiated products. They only allow consumers to participate in one card network, whereas merchants may choose to participate in more than one network. They compare welfare properties when these two networks operate as competitors and as a cartel where each network retains demand for its products from end-users. Like Rochet and Tirole (2003) and Guthrie and Wright (2007), they find that competition does not necessarily improve or worsen the balance of consumer and merchant fees from the socially optimal one. There are other fee structures for a given sum of consumer and merchant fees that would improve consumer and merchant welfare. However, they find that the welfare gain from the drop in the sum of the fees from competition is generally larger than the potential decrease in welfare from less efficient fee structures.

Models accounting for the role of credit

So far, we have considered models that ignore the extension of credit as a benefit to consumers and merchants.³⁵ Given the high level of antitrust scrutiny targeted toward credit card networks, we find this omission in most of the academic literature surprising. In the long run, aggregate consumption over consumers' lives may not differ because of access to credit, but such access may enable consumption smoothing that increases consumers' utility. From a merchant's perspective, extension of credit may lead to intertemporal business stealing. In other words, merchants attract consumers who do not have funds today by accepting credit cards, resulting in merchants tomorrow being unable to make sales to consumers who bought today on credit. In addition to extracting surplus from all consumers and merchants, banks have an additional source of surplus—liquidity-constrained consumers. How much surplus can be extracted depends on how much liquidity-constrained consumers discount tomorrow's consumption.

We define liquidity-constrained consumers as those who do not have funds at the time of purchase. These models consider a positive probability that some consumers will be unable to meet their credit obligations. The cost of these consumer defaults may be passed on by banks to merchants, certain types of consumers,

or both. Both models discussed here consider a different extreme in terms of who ultimately pays for the credit card services, including the cost of credit default.

Chakravorti and Emmons (2003) consider the costs and benefits of consumer credit in a four-party network where consumers are subject to income shocks after making their credit card purchases and some are unable to pay their credit card debt. To our knowledge, they are the first to link the insurance aspect of credit cards to their payment component. All markets for goods and payment services are assumed to be competitive. Observing that over 75 percent of U.S. card issuer revenue is derived from cash-constrained consumers, they consider the viability of the credit card system if it were completely funded by these types of consumers.³⁶ Note that their model does not assume zero interchange fees, but assumes that merchants fully pass on their payment costs to consumers. They impose an incentive constraint on individuals without liquidity constraints such that they will only use cards if they are guaranteed the same level of consumption were they to use cash. Such a constraint allows them to study the convenience use of credit cards—that is, the usage by those who do not need credit to make purchases.

Chakravorti and Emmons derive three main results. First, if consumers sufficiently discount future consumption, liquidity-constrained consumers who do not default would be willing to pay all credit card network costs *ex ante*, resulting in all consumers being better off. The key assumption is that at least a certain number of consumers face binding liquidity constraints and do not default. Second, if merchants charge a single price for a good regardless of how consumers pay, and if there are no side payments made by issuers to convenience users, card-accepting merchants who charge a single price for all purchases attract only liquidity-constrained consumers because some merchants charge a lower price and only serve cash-paying customers. Note, only those consumers who are liquidity-constrained use credit cards, and there is no convenience use. Third, if card issuers extend rebates to convenience users, a merchant can, under certain conditions, attract all types of consumers—including consumers who carry a credit card balance month to month and those who do not—when a single price is charged.

Chakravorti and To (2007) consider a scenario with monopolist merchants and a monopolist bank that serves both consumers and merchants where the merchants absorb all credit and payment costs in a two-period dynamic model.³⁷ Consumer demand for consumption goods is inelastic. They focus on credit extended to consumers who face income uncertainty.

They depart from the payment card literature in the following ways. First, similar to Chakravorti and Emmons (2003), rather than taking a reduced form approach where the costs and benefits of payment cards are exogenously assigned functional forms, a model that endogenously yields costs and benefits to consumers, merchants, and banks from payment card use—that is, consumer credit—is specified. Second, the model uses a dynamic setting where there are intertemporal trade-offs for all of the parties involved. Third, they consider consumption and income uncertainty.

Their model yields the following results. First, if merchants earn a sufficiently high profit margin and the cost of funds is sufficiently low, the economy is able to support credit cards. In other words, the benefits to consumers and merchants must be greater than the cost to support the credit card network. Second, the discount fee that merchants are willing to pay their banks increases as the number of credit-constrained consumers increases. Third, a prisoner's dilemma situation may arise: Each merchant chooses to accept credit cards, but by doing so, each merchant's discounted two-period profit is lower. In other words, there exists intertemporal business stealing among merchants across different industries, potentially resulting in all merchants being worse off.

Models with competition among payment instruments

Most of the literature ignores competition between payment instruments, with one payment provider offering multiple payment options to its customers and setting prices to maximize profits. Moreover, most economic models of payment cards generally do not consider price incentives offered by merchants to steer consumers to certain types of card payments. Furthermore, much of the payment literature focuses on the extensive margin—how prices influence membership—instead of the intensive margin—how prices affect usage.

In Bolt and Chakravorti (2008), we study the ability of banks and merchants to influence the consumer's payment instrument choice when they have access to three payment forms—cash, debit card, and credit card. To our knowledge, this model is the first to analyze payment network competition by combining elements of models that stress price balance with those that consider consumers' liquidity constraints and security concerns. In addition, we consider how banks set prices when they participate in multiple payment networks.

In our model, consumers participate in payment card networks to insure themselves from three types of shocks—uncertain income flows, theft, and the merchant they are matched to. Consumers only derive utility from consuming goods from the merchant they are matched to. Merchants differ on the types of payment instruments that they accept and the types of consumption goods that they sell. Unlike most two-sided market models, where benefits are exogenous, we explicitly consider how consumers' utility and merchants' profits increase from additional sales resulting from greater security and access to credit lines, and we consider the optimal allocation of those costs between consumers and merchants.³⁸ Before the realization of the three shocks, consumers are homogeneous. To focus on the intensive margin, we consider equilibria where consumers have access to all three payment instruments. Each merchant chooses which instruments to accept based on its production costs, and each merchant is categorized as cash only, cash and debit card, or full acceptance (cash, debit card, and credit card). Merchant heterogeneity is based on differences in production costs. We consider the merchants' ability to pass on payment processing costs to consumers in the form of higher goods prices.

Our key results can be summarized as follows. The structure of prices is determined by the level of the bank's cost to provide payment services and the level of aggregate credit losses. We find that the optimal strategy is for the bank to first fully extract consumers' surplus and then to extract merchants' surplus. In addition, we identify equilibria where the bank finds it profitable to offer one or both payment cards. Finally, uniform price policies increase bank profits when the bank supplies both types of payment cards than when merchants adopt pricing based on the payment instrument used. However, consumers and merchants are worse off when consumers without liquidity constraints use credit cards because they do not receive the proper price incentives, resulting in use of a less efficient payment instrument.

Conclusion

In summarizing the payment card literature, we find that no single model is able to capture all the essential elements of the market for payment services. It is a complex market with many participants engaging in a series of interrelated bilateral transactions. Moreover, determining appropriate pricing arrangements for payment instruments is difficult, since payment networks are subject to large economies of scale and give rise to strong usage and network externalities.

Much of the debate over various payment card fees is concerned with the allocation of the surpluses from consumers, merchants, and banks, as well as who is able to extract surpluses from whom.

We are able to draw the following conclusions. First, a side payment between the issuer and the acquirer may be required to get both sides on board. There is no consensus among policymakers or economists on what constitutes an efficient fee structure for card payments. Second, while consumers generally react to price incentives at the point of sale, merchants are reluctant to charge higher prices to consumers who benefit from card use. Third, network competition may not improve the price structure but may significantly reduce the total price paid by consumers and merchants. Fourth, consumers and merchants both value credit extended by credit card issuers (along with other benefits such as security), and consumers and merchants are willing to pay for it.

Sound public policy regarding the allocation of payment fees is difficult. The central question is whether the specific circumstances of payment markets are such that intervention by public authorities can be expected to improve economic welfare. The efficiency of payment systems is measured not only by the costs of resources used, but also by the social benefits generated by them. While the theoretical literature on the economics of payment cards is growing, the empirical literature is too limited to provide much guidance to policymakers. Public competition authorities around the world are considering or have imposed interchange fee regulations, along with the removal of merchant pricing restrictions based on the type of payment instrument used. Eventually, the data from such cases where the authorities have intervened may provide a useful “natural experiment” to test and refine the various theories discussed here.

NOTES

¹These 11 countries are Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom.

²They study 13 countries—Austria, Belgium, Canada, Finland, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States.

³In this article, we define banks broadly as depository institutions.

⁴There are countries, for example, France, where the cardholder’s account is debited much later. These types of cards are referred to as delayed debit cards.

⁵For a summary of prepaid cards, see Chakravorti and Lubasi (2006).

⁶SEPA applies to all countries where the euro is used as the common currency. The implementation of SEPA started in January 2008 with the launching of the SEPA credit transfer scheme and should be completed when all national payment instruments are phased out; these instruments may not be entirely phased out until 2013.

⁷Beijnen and Bolt (2009) provide estimates of scale economies that quantify the potential benefits of SEPA arising from consolidation of electronic payment processing centers across the euro area. It is likely that SEPA will trigger substantial consolidation of payment infrastructures and processing operations across borders to allow banks and their customers—both consumers and merchants—to benefit from these cost efficiencies in the form of lower payment fees.

⁸According to some industry press reports, a major reason for Visa’s IPO is to shield itself from antitrust litigation (see Enrich, 2006). Berry and Breitkopf (2006) also note that “as it was for MasterCard, the desire to reduce exposure to antitrust claims is a major factor in Visa’s plan to go public.”

⁹For a summary of antitrust challenges in various jurisdictions, see Bradford and Hayashi (2008) and Weiner and Wright (2005).

¹⁰On December 16, 2002, the Council of the European Union adopted Council Regulation (EC) No. 1/2003 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty Establishing the European Community (that is, the 1997 consolidated version of the Treaty of Rome). The new regulation came into effect on May 1, 2004. For more discussion on the EC’s ruling on MasterCard, see Bolt (2008).

¹¹Payment networks are one type of two-sided market platform. Other types of two-sided market platforms include computer game platforms, newspapers, and online dating sites. These markets have platforms that provide goods and services to two distinct sets of end-users and must convince both sides—such as game developers and game users, newspaper readers and advertisers, and males and females—to participate.

¹²An important empirical observation of two-sided markets is that platforms, generally, and payment networks, specifically, tend to heavily skew the price structure to one side of the market to get both sides “on board,” using one side as a “profit center” and the other side as a “loss leader,” or at best financially neutral. See Bolt and Tieman (2008) for an explanation for this phenomenon based on curvature of the demand functions.

¹³For a more general treatment of two-sided markets, see Armstrong (2006), Caillaud and Jullien (2003), Jullien (2001), and Rochet and Tirole (2006).

¹⁴While not common, some merchants do post different prices based on the payment instrument used to make the purchase. For example, some U.S. gas stations have started to discount cash prices recently in response to shrinking profit margins. We discuss models that consider instrument-contingent pricing on pp. 19–20.

¹⁵There are two types of payment card networks—open (four-party) and proprietary (three-party) networks. Open networks allow many banks to provide payment services to consumers and merchants, whereas in proprietary networks, one institution provides services to both consumers and merchants. When the issuer is not also the

acquirer, the issuer receives an interchange fee from the acquirer. Open networks have interchange fees, whereas proprietary systems do not have explicit interchange fees because one institution serves both consumers and merchants using that network's payment services. However, proprietary networks still must set prices for each side of the market to ensure that both sides are on board.

¹⁶In some instances, merchants are charged a fixed fee *and* a proportional fee.

¹⁷Bolt, Humphrey, and Uittenbogaard (2008) show that Norway rapidly shifted from cash and paper-based instruments to electronic modes of payments by effectively applying differentiated bank payment pricing to consumers.

¹⁸These 12 countries are Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom.

¹⁹Some key benefits of using cash include privacy and anonymity, which payment cards do not provide.

²⁰Carbó Valverde et al. (2008) conduct a similar exercise for Spain, and find that when summing net costs and benefits across participants, debit cards are the least costly and checks are the most costly, with credit cards and cash ranking second and third, respectively.

²¹Chakravorti (2003), Evans and Schmalensee (2005), and Rochet (2003) provide excellent discussions of academic models of interchange fees.

²²Net benefits for consumers and merchants are defined by the difference in benefits from using a payment card and using an alternative payment instrument.

²³Schmalensee (2002) defines the socially optimal interchange fee as the one that maximizes the sum of consumer and merchant surplus. Such a measure is appropriate if card acceptance is not used as a strategic tool to steal customers from another merchant.

²⁴Schmalensee (2002) assumes that there are a single issuer and a single acquirer, demand curves are linear, and there are no fixed costs.

²⁵The Hotelling merchants are spatially separated from their consumers. A merchant can charge a lower price to capture the other merchant's customers, but these customers may face an additional transportation cost to arrive at the merchant's competitor.

²⁶Rochet and Tirole (2002) also consider some extensions regarding merchant heterogeneity, differentiated prices for cards and cash, and system competition.

²⁷Carlton and Frankel (1995) extend Baxter (1983) by considering when merchants are able to fully pass on payment processing costs via higher consumption goods prices. They find that an interchange fee is not necessary to internalize the externality if merchants set pricing for consumption goods based on the type of payment instrument used.

²⁸No-surcharge rules do not allow merchants to impose surcharges for payment card purchases. However, merchants may still be allowed to offer discounts for noncard payments. For more discussion about no-surcharge rules and cash discounts, see Barron, Staten, and Umbeck (1992); Chakravorti and Shah (2003); Kitch (1990); and Lobell and Gelb (1981).

²⁹For discussion about the reluctance of merchants to set different prices, see Frankel (1998). IMA Market Development AB (2000) and Bolt, Jonker, and Van Renselaar (2008) discuss the effect of removing no-surcharge rules in Sweden and the Netherlands, respectively.

³⁰The assumption here is that credit card transactions are more expensive in terms of explicit and implicit costs than other forms of payment. However, there are instances when card payments were discounted vis-à-vis cash payments. During the conversion to the euro, one German department store offered card discounts because of the high initial demand for euro notes and coins to make change for cash purchases (Benoit, 2002).

³¹In Bertrand competition, two firms compete on price, resulting in each firm setting the price that would exist in perfect competition.

³²For economic models that we discuss, no-surcharge rules are economically the same as restricting merchants to charge the same price regardless of the payment instrument used, although the latter is a more restrictive policy.

³³Carlton and Frankel (1995) discuss this cross-subsidy.

³⁴In this context, rebates is a general term that captures incentives for consumers to use their cards, such as cash back and other frequent-use rewards.

³⁵We limit our focus here to consumption credit. Payment credit—the credit that is extended by the receiver of payment or a third party until the payment instrument is converted into good funds—is ignored. For more discussion, see Chakravorti (2007).

³⁶For a breakdown of issuer revenue percentages, see Green (2008).

³⁷Because Chakravorti and To (2007) ignore revolving credit (when cardholders carry a balance month to month), they consider the extreme case of merchants covering all operating and credit default costs. Many consumers do not pay annual fees in the United States and about 40 percent do not carry debt, suggesting that their use is subsidized by others. In the United States, it is not unusual for merchants to extend subsidized credit for large-value purchases that are below market rates, suggesting that some merchants are willing to subsidize consumer credit.

³⁸McAndrews and Wang (2006) and Wang (2006) are notable exceptions. McAndrews and Wang (2006) find that payment card adoption costs, the distribution of consumer incomes, and firm size are key determinants of the structure of fees borne by each side. Their model predicts that large merchants accept both cash and cards, medium-sized merchants are specialized and only accept cash or cards (but not both), and small merchants only accept cash. In a similar model, Wang (2006) finds that card networks raise interchange fees to maximize issuers' profits as card payments become more efficient. In equilibrium, consumer rewards and card transaction volume also increase, while consumer surplus and merchant profits may not.

REFERENCES

- American Bankers Association and Dove Consulting**, 2005, *2005/2006 Study of Consumer Payment Preferences*, report, Washington, DC, October.
- Amromin, Gene, and Sujit Chakravorti**, 2009, “Whither loose change?: The diminishing demand for small-denomination currency,” *Journal of Money, Credit, and Banking*, forthcoming.
- Armstrong, Mark**, 2006, “Competition in two-sided markets,” *RAND Journal of Economics*, Vol. 37, No. 3, Autumn, pp. 668–691.
- Barron, John M., Michael E. Staten, and John Umbeck**, 1992, “Discounts for cash in retail gasoline marketing,” *Contemporary Economic Policy*, Vol. 10, No. 4, October, pp. 89–102.
- Baxter, William F.**, 1983, “Bank interchange of transactional paper: Legal and economic perspectives,” *Journal of Law and Economics*, Vol. 26, No. 3, October, pp. 541–588.
- Beijnen, Christine, and Wilko Bolt**, 2009, “Size matters: Economies of scale in European payments processing,” *Journal of Banking and Finance*, forthcoming.
- Benner, Katie**, 2008, “Visa’s Record IPO rings up 28 percent gain,” CNNMoney.com, March 19, available at http://money.cnn.com/2008/03/19/news/companies/visa_ipo_opens.fortune/index.htm.
- Benoit, Bertrand.**, 2002, “Defiant C&A reignites debate on German shopping laws,” *Financial Times*, January 9, p. 2.
- Bergman, Mats, Gabriella Guibourg, and Björn Segendorff**, 2007, “The costs of paying—Private and social costs of cash and card,” Sveriges Riksbank, working paper, No. 212.
- Berry, Kate, and David Breitkopf**, 2006, “Big step for Visa may prove bigger for industry; merchants gain more clout; association model fades further,” *American Banker*, Vol. 171, No. 196, October 12, p. 1.
- Bolt, Wilko**, 2008, “The European Commission’s ruling in MasterCard: A wise decision?,” GCP, April 1, available by subscription at www.globalcompetitionpolicy.org/index.php?id=981&action=907.
- Bolt, Wilko, and Sujit Chakravorti**, 2008, “Consumer choice and merchant acceptance of payment media,” De Nederlandsche Bank and Federal Reserve Bank of Chicago, mimeo, March 24.
- Bolt, Wilko, and David B. Humphrey**, 2007, “Payment network scale economies, SEPA, and cash replacement,” *Review of Network Economics*, Vol. 6, No. 4, December, pp. 453–473.
- Bolt, Wilko, David B. Humphrey, and Roland Uittenbogaard**, 2008, “Transaction pricing and the adoption of electronic payments: A cross-country comparison,” *International Journal of Central Banking*, Vol. 4, No. 1, March, pp. 89–123.
- Bolt, Wilko, Nicole Jonker, and Corry Van Renselaar**, 2008, “Incentives at the counter: An empirical analysis of surcharging card payments and payment behavior in the Netherlands,” De Nederlandsche Bank, mimeo, August 15.
- Bolt, Wilko, and Alexander F. Tieman**, 2008, “Heavily skewed pricing in two-sided markets,” *International Journal of Industrial Organization*, Vol. 26, No. 5, September, pp. 1250–1255.
- Bradford, Terri, and Fumiko Hayashi**, 2008, “Developments in interchange fees in the United States and abroad,” *Payments System Research Briefing*, Federal Reserve Bank of Kansas City, April.
- Brits, Hans, and Carlo Winder**, 2005, “Payments are no free lunch,” *Occasional Studies*, De Nederlandsche Bank, Vol. 3, No. 2.
- Caillaud, Bernard, and Bruno Jullien**, 2003, “Chicken and egg: Competition among intermediation service providers,” *RAND Journal of Economics*, Vol. 34, No. 2, Summer, pp. 309–328.
- Carbó Valverde, Santiago, David B. Humphrey, José M. Liñares Zegarra, and Francisco Rodríguez Fernández**, 2008, “A cost–benefit analysis of a two-sided card market,” Fundación de las Cajas de Ahorros (FUNCAS), working paper, No. 383.
- Carlton, Dennis W., and Alan S. Frankel**, 1995, “The antitrust economics of credit card networks,” *Antitrust Law Journal*, Vol. 63, No. 2, Winter, pp. 643–668.

- Chakravorti, Sujit**, 2007, "Linkages between consumer payments and credit," in *Household Credit Usage: Personal Debt and Mortgages*, Sumit Agarwal and Brent W. Ambrose (eds.), New York: Palgrave MacMillan, pp. 161–174.
- _____, 2003, "Theory of credit card networks: A survey of the literature," *Review of Network Economics*, Vol. 2, No. 2, June, pp. 50–68.
- Chakravorti, Sujit, and William R. Emmons**, 2003, "Who pays for credit cards?," *Journal of Consumer Affairs*, Vol. 37, No. 2, pp. 208–230.
- Chakravorti, Sujit, and Victor Lubasi**, 2006, "Payment instrument choice: The case of prepaid cards," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 30, No. 2, Second Quarter, pp. 29–43.
- Chakravorti, Sujit, and Roberto Roson**, 2006, "Platform competition in two-sided markets: The case of payment networks," *Review of Network Economics*, Vol. 5, No. 1, March, pp. 118–143.
- Chakravorti, Sujit, and Alpa Shah**, 2003, "Underlying incentives in credit card networks," *Antitrust Bulletin*, Vol. 48, No. 1, Spring, pp. 53–75.
- Chakravorti, Sujit, and Ted To**, 2007, "A theory of credit cards," *International Journal of Industrial Organization*, Vol. 25, No. 3, June, pp. 583–595.
- Enrich, David**, 2006, "Visa plans restructuring, sees IPO in 12–18 mos.," *Dow Jones Newswires*, October 11.
- Evans, David S., and Richard L. Schmalensee**, 2005, "The economics of interchange fees and their regulation: An overview," AEI–Brookings Joint Center for Regulatory Studies, related publication, No. 05-12.
- Frankel, Alan S.**, 1998, "Monopoly and competition in the supply and exchange of money," *Antitrust Law Journal*, Vol. 66, No. 2, pp. 313–361.
- Gans, Joshua S., and Stephen P. King**, 2003, "The neutrality of interchange fees in payment systems," *Topics in Economic Analysis and Policy*, Vol. 3, No. 1, article 1, available at www.bepress.com/bejeap/topics/vol3/iss1/art1.
- Garcia-Swartz, Daniel D., Robert W. Hahn, and Anne Layne-Farrar**, 2006, "A move toward a cashless society: A closer look at payment instrument economics," *Review of Network Economics*, Vol. 5, No. 2, pp. 175–198.
- Green, Jeffrey**, 2008, "Exclusive bankcard profitability study and annual report 2008," *Card and Payments*, May 1, pp. 36–38.
- Guthrie, Graeme, and Julian Wright**, 2007, "Competing payment schemes," *Journal of Industrial Economics*, Vol. 55, No. 1, pp. 37–67.
- Humphrey, David B., Moshe Kim, and Bent Vale**, 2001, "Realizing the gains from electronic payments: Costs, pricing, and payment choice," *Journal of Money, Credit, and Banking*, Vol. 33, No. 2, May, pp. 216–234.
- Humphrey, David B., Magnus Willeson, Göran Bergendahl, and Ted Lindblom**, 2006, "Benefits from a changing payment technology in European banking," *Journal of Banking and Finance*, Vol. 30, No. 6, June, pp. 1631–1652.
- IMA Market Development AB**, 2000, "Study regarding the effects of the abolition of the nondiscrimination rule in Sweden," report, Lerum, Sweden, February 29.
- Jullien, Bruno**, 2001, "Competing in network industries: Divide and conquer," IDEI (Industrial Economic Institute) and GREMAQ, University of Toulouse, mimeo, July.
- Kitch, Edmund W.**, 1990, "The framing hypothesis: Is it supported by credit card issuer opposition to a surcharge on a cash price?," *Journal of Law, Economics, and Organization*, Vol. 6, No. 1, Spring, pp. 217–233.
- Lobell, Carl D., and Joseph W. Gelb**, 1981, "The Cash Discount Act," *New York Law Journal*, December 31, pp. 1–4.
- McAndrews, James J., and Zhu Wang**, 2006, "Microfoundations of two-sided markets: The payment card example," Federal Reserve Bank of Kansas City, Payments System Research, working paper, October 24.
- Quaden, Guy (presenter)**, 2005, "Costs, advantages, and disadvantages of different payment methods," National Bank of Belgium, report, December.

Rafsanjani, Nazanin, 2006, “Plastic only: Cafe refuses to accept cash,” *Morning Edition*, National Public Radio, October 11, available at www.npr.org/templates/story/story.php?storyId=6246139, with transcript from Factiva.

Rochet, Jean-Charles, 2003, “The theory of interchange fees: A synthesis of recent contributions,” *Review of Network Economics*, Vol. 2, No. 2, June, pp. 97–124.

Rochet, Jean-Charles, and Jean Tirole, 2006, “Two-sided markets: A progress report,” *RAND Journal of Economics*, Vol. 37, No. 3, Autumn, pp. 645–667.

_____, 2003, “Platform competition in two-sided markets,” *Journal of the European Economic Association*, Vol. 1, No. 4, pp. 990–1029.

_____, 2002, “Cooperation among competitors: Some economics of payment card associations,” *RAND Journal of Economics*, Vol. 33, No. 4, Winter, pp. 549–570.

Schmalensee, Richard, 2002, “Payment systems and interchange fees,” *Journal of Industrial Economics*, Vol. 50, No. 2, June, pp. 103–122.

Schwartz, Marius, and Daniel R. Vincent, 2006, “The no-surcharge rule and card user rebates: Vertical control by a payment network,” *Review of Network Economics*, Vol. 5, No. 1, March, pp. 72–102.

Wang, Zhu, 2006, “Market structure and credit card pricing: What drives the interchange?,” Federal Reserve Bank of Kansas City, Payments System Research, working paper, No. 06-04, December 20.

Weiner, Stuart, and Julian Wright, 2005, “Interchange fees in various countries: Developments and determinants,” *Review of Network Economics*, Vol. 4, No. 4, December, pp. 290–323.

Wright, Julian, 2004, “The determinants of optimal interchange fees in payment systems,” *Journal of Industrial Economics*, Vol. 52, No. 1, pp. 1–26.

_____, 2003, “Optimal card payment systems,” *European Economic Review*, Vol. 47, No. 4, August, pp. 587–612.