Externalities in Payment Card Networks: Theory and Evidence

Sujit Chakravorti

November 18, 2009
Externalities in Payment Card Networks: Theory and Evidence

Sujit Chakravorti*
Federal Reserve Bank of Chicago
November 18, 2009

Abstract
Payment cards continue to replace cash and checks in advanced economies. Along with the growth of payment card transactions has come greater scrutiny by public authorities of certain payment network rules along with the level of certain fees. This article reviews the growing payment card literature and discusses the impact of several regulatory interventions on card adoption, usage, and social welfare.

*Sujit Chakravorti is a senior economist at the Federal Reserve Bank of Chicago and can be reached at sujit.chakravorti@chi.frb.org. I thank Wilko Bolt, Santiago Carbó Valverde, Bill Emmons, Emery Kobor, Francisco Rodriguez Fernandez, Roberto Rosen, and Ted To for shaping my views over the years regarding the economics of retail payment networks. I also thank participants at “The Changing Retail Payments Landscape: What Role for Central Banks” held at the Federal Reserve Bank of Kansas City for which this article was originally written for. I also thank Anna Lunn for excellent research assistance and Han Choi for suggestions to improve the article’s readability. The views expressed are my own and do not necessarily reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.
The proliferation of payment cards has dramatically changed the ways we shop and merchants sell goods and services. Today, payment cards are indispensable in most advanced economies. 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 in 13 advanced economies.¹ Recent payment surveys also indicate that consumers are using payment cards instead of checks.

Some merchants have started to accept only card payments for safety and convenience reasons. For example, American Airlines began accepting only payment cards for in-flight purchases on all its domestic routes since June 1, 2009. 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. In addition, payment cards may allow access to credit that can be used to attract consumers without funds.

Debit, credit, and prepaid cards are three forms of payment cards. Debit cards allow consumers to access funds at their banks (defined broadly as depository institutions) 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.² 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

¹ Amromin and Chakravorti study 13 countries—Austria, Belgium, Canada, Finland, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States.
² There 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.” Furthermore, many U.S. debit card issuers extend credit lines as well, primarily as overdraft protection. For more discussion, see Chakravorti (2007).
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.³

Greater usage of cards has increased the value of payment network operators, such as Visa, Inc., MasterCard Worldwide, Discover Financial Services, and others. In 2008, Visa had the largest initial public offering (IPO) of equity, valued at close to $18 billion, in U.S. history (Benner, 2008). The sheer 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 (Enrich, 2006). 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.⁴ Several U.S. merchants have filed lawsuits against MasterCard and Visa regarding the setting of interchange fees. These fees are paid by the merchant’s bank to the cardholder’s bank and are set by the network operator.⁵ In April 2009, MasterCard reached an interim understanding with the European Commission on interchange fees for cross-border consumer payments in the European Union. Effective July 1, 2009, MasterCard Europe

---
³ For a discussion of the economics of prepaid cards, see Chakravorti and Lubasi (2006).
⁴ For a summary of antitrust challenges in various jurisdictions, see Bradford and Hayashi (2008).
⁵ In Australia, the interchange fee for debit card transactions is paid by the card issuer (banks that issue cards to consumers) to the acquirer (banks that convert payment card receipts into bank deposits for merchants), but this is an exception.
established cross-border interchange fees for consumer card transactions that, on average, do not exceed 30 basis points for credit cards and 20 basis points for debit cards.

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, I discuss several types of externalities that are present in payment networks. The first and, perhaps, the most researched externalities are adoption and usage externalities. In addition to these externalities, underlying fee structures may affect the welfare of individuals or firms participating (or not participating) in the payment network. Finally, I will discuss the limited evidence that exists regarding the effectiveness of some policy interventions.

There are several conclusions that I draw from the academic models, recent interventions in payment card markets, and discussions about potential policy interventions. First, many economic models suggest that the socially optimal interchange fee structure may not be systematically lower than the network profit-maximizing fee. Second, removing merchant pricing restrictions generally improve market price signals. Third, merchant, card issuer, or network competition may result in lower social welfare contrary to generally accepted economic principles. Fourth, if warranted, fees set by the authorities should not only consider costs but also benefits received by consumers and merchants, such as convenience, security, and access to credit that may result in greater sales.

Finally, the motivation for why public authorities intervene differs across jurisdictions. The type of public institution that regulates payment cards also differs. The

---

6 Rochet and Tirole (2006a) provide an overview of some externalities in card systems that I cover in this article.
institution may be an antitrust authority, a central bank, or a court of law. Often public authorities intervene because the interchange fee is set by a group of competitors and the level of the fee is deemed to be excessive. In other cases, by mandating fee ceilings, authorities expect greater number of merchants to adopt payment cards instead of cash. Alternatively, some policymakers argue that lowering card issuers’ interchange revenue may reduce incentives to cardholders to use more costly payment cards (for example, credit cards instead of debit cards).

The rest of the article is structured as follows. In the next section, I discuss externalities in payment card markets in the context of theoretical models. I also explore two externalities that have been less researched. In the following section, I investigate market interventions, along with the motivation of the authorities for such interventions and whether they met their objectives. Finally, I offer some concluding remarks.

**Externalities**

Before discussing the externalities present in payment card networks, let us review the key participants and the monetary transfers among them. Payment networks comprise consumers (more generally, buyers) and their banks (known as issuers), as well as merchants (more generally, sellers) and their banks (known as acquirers), along with the network operator and other participants that facilitate these transactions. Payment card transactions involve a set of interrelated bilateral transactions. First, a consumer establishes a relationship with an issuer and receives a payment card. Second, a

---

7 In addition to cash handling and safekeeping costs, some public authorities may find the inability to trace cash transactions an unattractive feature of cash.

8 In the case of prepaid cards, the identity of cardholders may not be known to the issuer, but there still exists a relationship.
consumer makes a purchase from a merchant. Third, if a merchant has established a relationship with an acquirer, the merchant is able to accept payment card transactions. Fourth, the acquirer receives payment from the issuer. A network operator facilitates these bilateral relationships.

In figure 1, the four key participants and their monetary transfers are diagrammed. When the consumer establishes a relationship with a bank, she agrees to pay an annual fee if one is charged, finance charges if she borrows long term, and other fees. In addition, she may receive per transaction rewards to promote greater usage of the card. When the consumer uses her card to make a purchase, the merchant may impose an additional fee for card acceptance or pass on the cost to all consumers in the form of higher prices. To convert the payment card receipt into a bank deposit, the merchant pays a fee to its bank. In addition to per transaction fees that may be fixed or proportional to the amount of the purchase, the merchant may also pay fixed fees. The merchant’s bank pays interchange fees to the cardholder’s bank. In this section, I study the effect of a bilateral payment transfer on other bilateral relationships in the network and potential externalities that might arise.

Adoption and usage externalities

The two-sided market literature has been used to analyze the structure of fees paid by consumers and merchants. Payment networks are one type of two-sided market. Other types of two-sided market platforms include computer game platforms, newspapers, and online dating sites. These platforms provide goods and services to two

---

9 For a review of the academic literature on two-sided payment networks, see Bolt and Chakravorti (2008b).
or more distinct sets of end-users and must convince all sides to participate. The price structure or balance is the share that each type of end-user pays of the total price of the payment service.

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.\textsuperscript{10} Rochet and Tirole (2006b) 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.

A key externality examined in the payment card literature is the ability of the network to convince both consumers and merchants to participate in a network. Initially, the literature focused on per transaction fees and ignored fixed costs. In such an environment, there is no distinction between adoption and usage. Baxter (1983) argues that the equilibrium quantity of payment card transactions occurs when the total transactional demand for payment card services, which are determined by consumer and merchant demands jointly, is equal to the total transactional cost for payment card services, including both issuer and acquirer costs, or:\textsuperscript{11}

\[ f + m = c_I + c_A, \]

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

\textsuperscript{10} For a more general treatment of two-sided markets, see Armstrong (2006), Caillaud and Jullien (2003), Jullien (2001), Rochet and Tirole (2006b), Rysman (2009), and Weyl (2009).

\textsuperscript{11} Baxter (1983) considers an environment where consumers are homogeneous, merchants are perfectly competitive, and the market for issuing and acquiring payment cards are competitive.
consumer’s willingness to pay is based on her net benefits received, $b_s$. The consumer will participate when her net benefit is greater than or equal to the fee in equilibrium.\textsuperscript{12} Similarly, if the merchants’ fee, $m$, is less than or equal to the net benefits it receives, $b_s$, merchants will accept cards. 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.

Schmalensee (2002) extends Baxter’s (1983) analysis by considering issuers and acquirers that have market power, but still assumes that merchants operate in competitive markets.\textsuperscript{13} His results support Baxter’s conclusions that the interchange fee balances the demands for payment services by each end-user type and the cost to banks to provide them. Schmalensee finds that the profit-maximizing interchange fee of issuers and acquirers may also be socially optimal.\textsuperscript{14}

Given the simultaneous consumption of payment services by consumers and merchants, a side payment may be necessary to get both sides on board if there are asymmetries of demand between consumers and merchants and/or of costs to service consumers and merchants. This result is critically dependent on the inability of merchants to price discriminate between card users and those who do not use cards or among

\textsuperscript{12} Net benefits for consumers and merchants are defined by the difference in benefits from using a payment card and using an alternative payment instrument.

\textsuperscript{13} Schmalensee assumes a single issuer, single acquirer, linear demand curves, and no fixed costs.

\textsuperscript{14} Schmalensee defines the socially optimal interchange fee as the one that maximizes the sum of the 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.
different types of card users. While most economists and antitrust authorities agree that an interchange fee may be necessary, the level of the fee remains a subject of debate.

**Merchant competition**

A common reason given by merchants when asked why they do not reject cards instead of paying high fees to the card networks for accepting them is that they would lose business to their competitors. Some merchants argue that merchants as a whole would be better off by not accepting certain types of payment cards. Some economic models have predicted that merchant competition may increase the ability of networks to set higher interchange fees.

Unlike Baxter (1983) and Schmalensee (2002), Rochet and Tirole (2002) consider strategic interactions of consumers and merchants. They have two main results. First, the interchange fee that maximizes 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. Second, merchants are willing to pay more than the socially optimal fee if they can steal customers from their competitors. 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

---

15 Rochet and Tirole consider two identical Hotelling merchants in terms of their net benefits of accepting a payment card for sales 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. They assume that the total number of transactions is fixed and changes in payment fees do not affect the demand for consumption goods.
benefits and merchant acceptance when the interchange fee is increased because some merchants will not accept cards. Wright 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.

These models suggest that merchant competition may actually lead to a greater ability by network operators to extract surplus from them. Furthermore, there is no systematic bias in the social-welfare-maximizing and profit-maximizing interchange fee. In the next section, I explore the ability of merchants to steer consumers to the merchant’s preferred payment instrument by using price incentives.

Instrument-contingent pricing

The two-sided market literature assumes that end-users are not allowed to negotiate prices of platform services. In many jurisdictions, merchants are not allowed to add a surcharge for payment card transactions because of legal or contractual restrictions. If consumers and merchants were able to negotiate prices based on differences in costs that merchants face and the benefits that both consumers and merchants receive, the interchange fee would be neutral, assuming full pass-through. 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. Generally, the merchant charges the same price regardless of the type of payment instrument used to

---

16 In Wright’s environment, both consumer and merchant fees are per transaction fees. Each consumer buys goods from each industry. Issuers and acquirers operate in markets with imperfect competition. Wright assumes that consumers face the same price regardless of which instrument they use to make the purchase.

17 No-surcharge restrictions do not allow merchants to impose surcharges for payment card purchases. However, merchants may be allowed to offer discounts for noncard payments. For more discussion about no-surcharge rules and discounts, see Chakravorti and Shah (2003).
make the purchase. Frankel (1998) refers to merchants’ reluctance to set different prices even when they are allowed to do so as price cohesion.

Even if price differentiation based on the payment instrument used is not common, the possibility to do so may enhance the merchants’ bargaining power in negotiating their fees. Merchants can exert downward pressure on fees by having the possibility to set instrument-contingent pricing. Payment networks may prefer non-instrument-contingent pricing because some consumers may not choose payment cards if they had to explicitly pay for using them at the point of sale (POS).

Carlton 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. Furthermore, they argue that cash users are harmed when merchants set one price because they subsidize card usage.

Schwartz and Vincent (2006) study the distributional effects among cash and card users with and without no-surcharge restrictions. They find that the absence of pricing based on the payment instrument used increases network profit and harms cash users and merchants.\(^\text{18}\) 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

\(^{18}\) Schwartz and Vincent relax the common assumption made in the literature 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.
to card users.\(^\text{19}\) Granting such rebates to card users boosts their demand for cards while simultaneously forcing merchants to absorb part of the corresponding rise in the merchant fee, because 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.

Gans and King (2003) argue that, as long as there is “payment separation,” the interchange fee is neutral regardless of the market power of merchants, issuers, and acquirers. 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.

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. He argues that merchants are able to extract consumers’ surplus ex post from payment card users, while cash users are unaffected. Wright only considers equilibria where merchants will continue to sell the same quantity of goods to cash users at the same price. When

\(^{19}\) In this context, a rebate is an incentive for consumers to use their cards—for example, cash back and other frequent-use rewards.
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.

Economic theory generally suggests that if merchants were able to recover their payment costs, the impact of the interchange fee would be severely dampened. However, the potential for merchants to charge more than their processing costs exists and consumer welfare could be harmed by such practices. The most interesting puzzle may be why merchants choose not to price differentiate even when they are allowed to do so. Some observers suggest that merchant competition may prevent price differentiation.

*Network competition*

Economic theory suggests that competition generally reduces prices, increases output, and improves welfare. However, with two-sided markets, network 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 “multihoming.” 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.\(^{20}\) Even if consumers adopt more than one payment card, Rysman (2007) finds that consumers may have strong preferences to use only one of them.

\(^{20}\) For more discussion, see Evans (2003).
Some models of network competition assume that the sum of consumer and merchant fees is constant and focus on the price structure.\textsuperscript{21} Rochet and Tirole (2003) find that the price structures for a monopoly network and competing platforms may be 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” when deciding to accept payment cards. They find that network competition can result in higher interchange fees than those that would be socially optimal.

Chakravorti and Roson (2006) consider the effects of network competition on total price and on price structure where networks offer differentiated products.\textsuperscript{22} 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. 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.

Unlike one-sided markets, competition does not necessarily improve the balance of prices for two-sided markets. Furthermore, if competition for cardholders is more intense because consumers ultimately choose the payment instrument, issuers may provide greater incentives to attract them. If issuers have greater bargaining power to

\textsuperscript{21} The motivation behind this assumption was based on the earlier cooperative structure of the two large networks. However, the two largest networks changed their structure from associations to for-profit firms.\textsuperscript{22} Chakravorti and Roson only allow consumers to participate in one card network, whereas merchants may choose to participate in more than one network. However, unlike Guthrie and Wright (2007) and Rochet and Tirole (2003), Chakravorti and Roson consider fixed fees for consumers. Chakravorti and Roson compare welfare properties when the two networks operate as competitors and as a cartel, where each network retains demand for its products from end-users but the networks set fees jointly.
raise interchange fees, they can use this power to partially offset the cost of consumer incentives. I will discuss later the funding of rewards to entice more consumers in the context of the Reserve Bank of Australia’s interchange fee regulation.

**Surplus from revolvers**

So far, among the models that I have discussed, the benefits of consumer credit are not considered. Given the high level of antitrust scrutiny targeted toward credit card fees, including interchange fees, this omission in most of the academic literature is rather 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. In addition to extracting surplus from all consumers and merchants, banks may extract surplus from liquidity-constrained consumers. How much surplus can be extracted depends on how much liquidity-constrained consumers discount tomorrow’s consumption.

Chakravorti and Emmons (2003) consider the costs and benefits of consumer credit where consumers are subject to income shocks after making their credit card purchases and some are unable to pay their credit card debt. To my knowledge, they are the first to link the insurance aspect of credit cards to their payment component.

Observing that over 75 percent of U.S. card issuer revenue is derived from cash-

---

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

24 The empirical literature on credit cards has suggested interest rate stickiness along with above-market interest rates, although some have argued that the rate is low compared with alternatives such as pawn shops. For more discussion, see Ausubel (1991) and Brito and Hartley (1995).

25 All markets for goods and payment services are assumed by Chakravorti and Emmons to be competitive. Chakravorti and Emmons impose a participation constraint on individuals without liquidity constraints such that the individuals will only use cards if they are guaranteed the same level of consumption as when they use cash including the loss of consumption associated with higher prices for consumption goods.
constrained consumers, they consider the viability of the credit card system if it were completely funded by these types of consumers.\textsuperscript{26} They find that 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 than a world with no credit cards. However, they also find that the inability of merchants to impose instrument-contingent prices results in a lower level of social welfare because costly credit card infrastructure is used for transactions that do not require credit extensions.

Most of the payment card literature ignores consumer finance charges and other types of consumer fees, such as annual, over-the-limit, and cash advance fees. In the United States, the regulation of consumer fees on credit cards has increased and new restrictions have been implemented. Perhaps, with reduced revenue from these sources coupled with greater usage of debit cards, interchange fee revenue may become more critical. Of course, as mentioned previously, these fees continue to face regulatory pressure as well.

\textit{Merchant fees and consumer credit}

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.\textsuperscript{27} Their model yields the

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{26} For a breakdown of issuer revenue percentages, see Green (2008).
\item \textsuperscript{27} Chakravorti and To 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, they construct a model that endogenously yields costs and benefits to consumers, merchants, and banks from credit card use. Second, their model considers a dynamic setting where there are intertemporal tradeoffs for all participants. Third, they consider consumption and income uncertainty.
\end{itemize}
\end{footnotesize}
following results. First, the merchants’ willingness to pay bank fees increases as the number of credit-constrained consumers increases. Note that up to a point, merchants are willing to subsidize credit losses in exchange for additional sales. Second, 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. Unlike the merchants in the previous models, the merchants in this one do not sell the same type of goods and may enjoy significant market power. In other words, business stealing may occur across merchants that sell the same or similar goods or across consumption periods between merchants that sell completely different types of goods.

*Competition among payment instruments*

Most of the payment card literature ignores competition between payment instruments.\(^2^8\) Furthermore, much of the payment literature focuses on the intensive margin—how fees influence usage—instead of the extensive margin—how fees affect adoption—or does not distinguish the two.\(^2^9\) Much of the policy debate is about market forces behind consumer choice and merchant acceptance among multiple types of payment instruments.

If consumers carry multiple types of payment instruments, merchants may be able to steer them away from more costly payment instruments. Rochet and Tirole (2007) argue that merchants may choose to decline cards after they have agreed to accept them.

\(^{28}\) Farrell (2006) studies the impact of higher interchange fees on consumers who do not use cards. While the redistributive effects generally do not affect social welfare, he argues that the impact of pricing of a payment instrument in one network affecting the usage of other payment instruments should be of concern to policymakers.

\(^{29}\) Bedre and Calvano (2009), Bolt and Chakravorti (2008a), and Chakravorti and Roson (2006) are notable exceptions.
They define the “tourist test” as when the merchant accepts cards even when it can “effectively steer” the consumer to use another payment instrument. Rochet has often given the example of an experience that he had in southern Italy, where after having a meal, the restaurant claimed that its payment card terminal was broken and payment had to be made in cash. After visiting a nearby ATM, Rochet paid the bill with cash. In this example, the merchant did not pass the tourist test. The restaurant figured out that being a gentleman, Rochet would not leave the bill unpaid. However, if the consumer is unable to access cash or another form of payment, the merchant would lose the sale.

Merchants may steer consumers through price incentives, if allowed to do so. Bolt and Chakravorti (2008a) study the ability of banks and merchants to influence the consumers’ choice of payment instrument when they have access to three payment forms—cash, debit card, and credit card. Unlike most two-sided market models, where benefits are exogenous, they explicitly consider how consumers’ utility and merchants’ profits increase from additional sales resulting from greater security and access to credit.

Bolt and Chakravorti’s (2008a) key results can be summarized as follows. With sufficiently low processing costs relative to theft and default risk, the social planner sets the merchant fee to zero, completely internalizing the card acceptance externality. The

---

30 I have often had similar experiences at the end of cab rides when I try to pay with my credit card and the driver chooses not to accept it, even though there are multiple signs stating that credit cards are accepted.

31 In Bolt and Chakravorti’s model, consumers only derive utility from consuming goods from the merchant they are matched to. In addition, some consumers prefer to consume before their income arrives. Merchants differ on the types of payment instruments that they accept and type of consumption good they sell. 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. Bolt and Chakravorti consider the merchants’ ability to pass on payment processing costs to consumers in the form of higher uniform and differentiated goods prices.

32 While default rates and theft will differ across countries, Bolt and Chakravorti provide some estimates. For Italy, Alvarez and Lippi (2009) estimate the probability of being pickpocketed at around 2 percent in 2004. For the United States, Scholtes (2009) reported that credit card default rates hit a record of more than 10 percent in June 2009.
bank may also set the merchant fees to zero, but only if merchants are able to sufficiently pass on their payment fees to their consumers or if their payment fees are zero. If the real resource cost of payment cards is too high, the social planner sets a higher merchant fee than the bank does, resulting in lower card acceptance and higher cash usage. Bolt and Chakravorti (2008a) find that bank profit is higher when merchants are unable to pass on payment costs to consumers because the bank is better able to extract merchant surplus. The relative costs of providing debit and credit cards determine whether the bank will provide both or only one type of payment card.

**Payment fraud and liability**

An aspect of payment networks that has received little attention in the payment network literature is the incentive that each participant has in maintaining the integrity and safety of the system as a whole. An externality arises if one participant on account of negligence and lack of incentives allows a fraudster to gain access to information that may be used to make fraudulent purchases.\(^{33}\)

For example, consumers often face no liability for fraudulent transactions if proper procedures are followed for payment card transactions. While such a liability waiver encourages greater usage of cards vis-à-vis other payment instruments with less protection, it may also have the unintended consequence of consumers not maintaining appropriate antifraud precautions.\(^{34}\) Primarily because of this liability shift, the card networks have implemented various fraud prevention strategies, such as real-time verification, the ability to shut down accounts rapidly, and the tracking of spending

---

\(^{33}\) See Amromin and Porter (2009) and Braun et al. (2008).

\(^{34}\) See Douglass (2009).
patterns of cardholders over the last few decades. While U.S. issuers and networks limit consumer liability, consumers may bear losses associated with fraudulent transactions if they do not adopt risk-reducing procedures in other countries. For example, an Italian banker explained to me that most Italian banks shift the liability back to consumers if they do not use the recommended security procedures for Internet card payments.

Merchants also enjoy certain protections (though more limited than those for consumers) if they follow set guidelines when accepting payment cards.

Similarly, the lack of merchant and processor data security measures may pose negative externalities. For example, while the cost of not protecting payment information for an individual entity may be small, its impact on the system as a whole may be significant. Recently, the industry has been exploring various procedures to reduce this risk.

Market participants have expressed the view that better enforcement of current laws regarding payment fraud and greater adoption of existing industry-wide standards would greatly aid in reducing and containing fraud. Some observers have suggested that public authorities should establish standards, provide mechanisms for sharing information on data breaches, and formulate appropriate responses when wide-scale fraud occurs. Understandably, market participants may be reluctant to share or publicize breaches because of the potential loss in future business.

**Dynamic efficiency and innovation**

Dynamic efficiency and innovation have generally been ignored by economists and policymakers. Some market participants have argued that positive profits are

---

35 See Nocera (1994).
necessary for payment networks to innovate. In other words, regulatory solutions to correct “excessive” interchange fees by using a cost-based approach may stifle future innovation. When general-purpose payment cards were first introduced, issuers and networks faced significant losses and many left the industry to only return later, suggesting that investments in new products and processes may require significant time to recover.

Historically, the card networks have been more innovative than other payment networks, such as those that process checks. In the United States, a law had to be passed relatively recently to facilitate the widespread acceptance of substitute checks instead of the original physical check enabling rapid migration to the truncation of the physical check. In contrast to the networks processing checks, credit card networks were exchanging payment information electronically for more than two decades. In addition, the card networks established real-time authorization systems in the 1970s to combat payment fraud.\textsuperscript{36} Interestingly, fees charged by third parties to guarantee checks are pretty close to or higher than merchant fees for credit cards. When similar protections against payment default are included for checks, the cost of check acceptance with similar protections converges to the cost of payment card acceptance, suggesting that payment instruments may differ with respect to the benefits to merchants. Furthermore, some merchants may be willing to forgo certain benefits because of the type of customers that they serve.

\textsuperscript{36} For more discussion about innovations in the payment card market, see Chakravorti and Kobor (2005), Evans and Schmalensee (1999), and Nocera (1994).
Market Interventions

Policymakers in different jurisdictions are encouraging the replacement of cash and checks with electronic substitutes, such as payment cards at the point of sale. In some U.S. municipalities, acceptance of payment cards for cab rides has been mandated. A primary reason cited is the safety of passengers and cab drivers (who are often the targets of muggings). In Mexico, the government gave away terminals to merchants to increase the acceptance of payment cards versus cash (Castellanos et al., 2008). However, forced acceptance of payment cards and government-subsidized merchant terminals are not common. In this section, I explore several market interventions in various jurisdictions and study the impact of those interventions.

Removal of no-surcharge policies

There are several jurisdictions where merchants are able to impose surcharges. Some of the academic research cited previously suggests that if merchants are allowed to surcharge, the level of the interchange fee would be neutral. In this section, I discuss examples where merchants are able to post differentiated prices.

The Australian authorities were concerned about the substitution of credit cards by debit cards; they argued that consumers did not receive the proper price incentives to use debit cards, the less costly payment instrument. The Reserve Bank of Australia (RBA) reported that the average cost of the payment functionality of the credit card was

---

37 In the United States, some payment providers have introduced decoupled debit as a competitor to traditional payment cards. These types of payments use the automated clearinghouse (ACH) network to transfer funds from consumers to merchants for point of sale transactions.

38 Prager et al. (2009) review the U.S. payment card market and consider potential regulations.
AUS$0.35 higher than a debit card using a consistent AUS$50 transaction size.\textsuperscript{39} To encourage better price signals, the RBA removed no-surcharge restrictions in 2002.

While most Australian merchants do not impose surcharges for any type of payment card transaction today, the number of merchants who do are increasing. At the end of 2007, around 23 percent of large merchants and around 10 percent of small and very small merchants imposed surcharges. Large merchants surcharged around 15 percent of the time. The average surcharge for MasterCard and Visa transactions is around 1 percent, and that for American Express and Diners Club transactions is around 2 percent (Reserve Bank of Australia, 2008a).\textsuperscript{40} Using confidential data, the Reserve Bank of Australia (2008a) also finds that if one network’s card is surcharged more than other networks’ cards, consumers dramatically reduce their use of the card with the surcharge. After analyzing consumer surveys, the Reserve Bank of Australia (2008a) noted that nearly 40 percent of credit card convenience users (that is, credit card users who do not need credit to make purchases) did not use a debit card during the time of the survey; this suggests that using credit cards is still preferred by many of those who do not need to borrow.\textsuperscript{41}

Some economists have stressed that merchants may surcharge consumers more than their costs. A potential regulatory response is to cap the surcharge. In responding to the 2007/08 review of reforms by the Reserve Bank of Australia, some market participants suggested that merchants might be imposing higher surcharges than their cost

\textsuperscript{39} Reserve Bank of Australia (2008a), 17.
\textsuperscript{40} Note that in other jurisdictions, card networks may prevent merchants from imposing different surcharges on credit cards from different networks.
\textsuperscript{41} Of course, even those credit card users who pay off their balances every month may benefit from short-term loans because of timing asymmetries between their incomes and purchases.
to accept payment cards. The RBA has considered setting a limit for the surcharge amount but has not gone ahead with implementing one.

In the United States, merchants are allowed to offer cash discounts but may not be allowed to surcharge credit card transactions. In the 1980s, many U.S. gas stations explicitly posted cash and credit card prices. Barron, Staten, and Umbeck (1992) report that gas station operators imposed these policies when their credit card processing costs were high but later abandoned these policies when acceptance costs decreased because of new technologies such as electronic terminals at the point of sale. Recently, some gas stations brought back price differentiation based on payment instrument type, citing the rapid rise in gas prices and declining profit margins.

In the Netherlands, Bolt, Jonker, and van Renselaar (2009) study the impact of debit card surcharges. They report that a significant number of merchants are setting different prices, depending on whether cash or a debit card is used. Debit card surcharges are widely assessed when purchases are below 10 euro, suggesting that merchants are unwilling to pay the fixed transaction fee below this threshold. Bolt, Jonker, and van Renselaar find that merchants may surcharge up to four times their fee. In addition, when these surcharges are removed, they argue, consumers start using their debit cards for these small payments, suggesting that merchant price incentives do affect consumer payment choice. Interestingly, in an effort to promote a more efficient payment system, the Dutch central bank has supported a public campaign to encourage retailers to stop surcharging and for consumers to use their debit cards for small transactions.

There are instances when card payments were discounted vis-à-vis cash payments. During the conversion to the euro from national currencies, one German
department store offered discounts for using cards because of the high initial demand for euro notes and coins to make change for cash purchases (Benoit, 2002). It should be noted, however, that the retailer was in violation of German retailing laws for doing this. In a more permanent move, the Illinois Tollway charges motorists who use cash to pay tolls twice as much as those who use toll tags (called I-PASS), which may be loaded automatically with credit and debit cards when the level of remaining funds falls below a certain level.\(^{42}\) In addition to reducing cash handling costs, the widespread implementation of toll tags decreased not only congestions at toll booths but also pollution from idling vehicles waiting to pay tolls, since tolls could be collected as cars drove at highway speeds through certain points on the Illinois Tollway. In both of these cases, the benefits of using cards outweighed the costs for society in general. However, benefits from card acceptance vary considerably across merchants.

**Regulation of interchange fees**

There are several jurisdictions where interchange fees were directly regulated or significant pressure was exerted by the public authorities on networks to reduce their interchange fees. In this section, I will discuss the impact of interventions in three jurisdictions—Australia, Mexico, and Spain.

Concluding that surcharges alone would not put sufficient downward pressure on interchange fees, the Australian authorities imposed explicit interchange fee targets for the two large four-party payment networks—MasterCard and Visa—but did not impose

---

\(^{42}\) For more discussion, see Amromin, Jankowski, and Porter (2007).
any restrictions on three-party networks—American Express and Diners Club. In 2002, the RBA imposed weighted-average credit card interchange fee caps and later imposed per transaction targets for debit cards. As of April 2008, the weighted-average credit card interchange fees in the MasterCard and Visa networks must not exceed 0.50 percent of the value of transactions. The Visa debit weighted-average interchange fee cap must not exceed 12 cents (Australian) per transaction. The EFTPOS (electronic funds transfer at point of sale) interchange fees for transactions that do not include a cash-out component must be between 4 cents (Australian) and 5 cents (Australian) per transaction.

The Reserve Bank of Australia (2008a) reports that the interchange fee regulation, coupled with the removal of the no-surcharge rule, improved the price signals that consumers face when deciding which payment instruments to use. Specifically, annual fees for credit cards increased and the value of the rewards decreased. The Reserve Bank of Australia (2008a) calculates that for an AUS$100 transaction, the cost to consumers increased from –AUS$1.30 to –AUS$1.10 for consumers who pay off their balances in full every month. A negative per transaction cost results when card benefits such as rewards and interest-free loans are greater than payment card fees.

In its recent five-year review of their payment card policies, the Australian Payments System Board suggested that the explicit regulation of interchange fees be removed subject to certain conditions. This policy can be described as regulatory contestability. In other words, the authorities will remove restrictions if the payment

---

43 In four-party networks, the issuer and the acquirer need not be the same. In three-party networks, the issuer and acquirer are the same resulting in no explicit interchange fee between issuers and acquirers.

44 For more discussion about the effect of rewards on card use, see Carbó-Valverde and Liñares-Zegarra (2009) and Ching and Hayashi (2006).

45 The notion of contestability is a bit different than the normal usage in economics because the regulator threatens regulation but does not threaten to enter the market to put downward pressure on prices.
card networks do not raise their fees beyond some threshold. However, the actual threshold is not quantified.

Those who oppose the Australian interchange fee regulation argue that consumers have been harmed by reduced rewards and higher fees and have not shared in the cost savings—in terms of lower prices for goods and services. However, measuring price effects over time of interchange fee regulation is difficult.

Another interesting case where government authorities exerted pressure to decrease interchange fees occurred in Mexico. Similar to the RBA in Australia, the Bank of Mexico—the Mexican central bank—has the authority to regulate retail payment systems throughout the country. Unlike the RBA, the Bank of Mexico used moral suasion to reduce interchange fees. The motivation of the Mexican authorities to reduce interchange fees was to reduce merchant fees that were preventing greater adoption and usage of payment cards in Mexico.

Mexico’s Bank Association (ABM) set different interchange fees for debit and credit cards in August 2004; prior to this time, the fees were the same for both types of cards. Interchange fees were set based on a merchant’s monthly transaction volume. By August 2005, debit card interchange fee for the largest merchants fell from 2.00 percent to 0.75 percent while the credit card interchange fee fell from 2.00 percent to 1.80 percent. The category that applied to the smallest merchant was eliminated; as a consequence the interchange fee of this group fell from 3.50% to 1.95% and 3.50% to 2.70% for debit and credit cards, respectively. The ABM also proposed interchange fees based on a formula where the interchange fee balances out the issuing and acquiring

---

46 My discussions with Bank of Mexico staff, especially José Luis Negrín, were critical to my understanding of the Mexican payment card market.
banks’ profits (net of interchange), and where profits are normalized by revenue (net of interchange). A reference rate is obtained and specific interchange fee levels are calculated for a number of merchant categories using proxies of the demand elasticity for each category.

In 2008, ABM further reduced debit and credit card interchange fees. The new IF levels implied a reduction in the weighted average of 12.5% and 9% for credit and debit, respectively.47 As expected, merchant fees also decreased. In order to follow the evolution of merchant fees, Bank of Mexico gathered information from a sample of 1000 firms that accepted card payments. The results are that from 2005 to 2008, the average merchant discount rate has decreased 12.3% and 23.3% for credit and debit, respectively.48 As a result of these reductions, the number of POS terminals installed increased to 446,025 by the end of 2008 compared to 129,971 in 2002. POS transactions increased from 52 million in 2002 to 215 million by the end of 2008 of which 46% were credit card transactions.

The installation of POS terminals was subsidized through a private, nonprofit trust fund called FIMPE that was initially funded by the banks. The banks received a tax credit from the government for their investment. It is important to note that there may be significant fixed and variable costs. As a result of these interchange fee reductions and terminal subsidies, the number of POS terminals installed increased to 418,237 by the end of 2007 compared with 129,799 in 2002. POS transactions increased from a 135 million in 2002 to 698 million by the end of 2007, of which 48 percent were credit card transactions.

47 The weighted average interchange fee for credit cards decreased from 1.84% to 1.61% and for debit cards decreased from .78% to .71%.
48 From 2005 to 2008, the average merchant fee decreased from 2.85% to 2.50% and the average debit merchant fee declined from 2.53% to 1.94%.
transactions. The reduction in interchange fees resulted in lower per transaction costs, and the terminal subsidies reduced the fixed costs.

Unlike in Australia or Mexico, the antitrust authority, and not the central bank, intervened in payment card markets in Spain. Part of the motivation was based on directives by the European Commission regarding fees that were set by networks that had significant market power. Over the period 1997–2007, the number of debit cards increased by 40.9 percent and the number of credit cards has increased by 207.1 percent. During the same period, debit card transactions increased from 156 million to 863 million and credit card transactions increased from 138 million to 1.037 billion. Furthermore, the average number of POS transactions per card per year increased from 7.1 to 27.8 during the same period.

The first intervention occurred in May 1999, when the Spanish government mandated the three Spanish payment card networks to gradually reduce maximum interchange fees from its initial value of 3.5 percent to 2.75 percent by July 2002. These maximum fees varied significantly across merchant categories.

In April 2002, Spain’s antitrust authority requested the Spanish networks to provide information on how they determined their interchange fees. From 2003 until 2005, several attempts from the industry to maintain their “special authorization” for the setting of interchange fees were refused. Eventually, the networks were requested to set levels of interchange fees that only reflected operating costs and those due to fraud. In December 2005, the Ministry of Industry, Tourism, and Trade decided that the multilateral interchange fees should not exceed the costs to provide card services.
Finally, a new regulatory framework stated that from 2009 onward, each of the card networks would audit their operations and provide a cost-based analysis for debit and credit cards. From January 2006 to December 2008, the highest interchange fee levels had to be reduced in a stepwise manner. Furthermore, a distinction had to be made between debit card and credit card interchange fees, with the former being a fixed amount per transaction and the latter being a percentage amount per transaction. For merchants with an annual value of less than 100 million euro in POS card payment receipts, the credit card interchange fee was set to decrease from 1.40 percent per transaction in 2006 to 0.35 percent in 2009; for those same merchants, the debit card interchange fees (regardless of the purchase amount) were reduced from 0.53 euro per transaction in 2006 to 0.35 euro per transaction in 2009. These fees are the maximum allowable, and in some cases the actual fees are lower. Additionally, price differences between debit cards and credit cards, merchant sectors, and intrasystem and intersystem operations should also be progressively reduced.

Carbó Valverde, Chakravorti, and Rodriguez Fernandez (2009) study the effects of interchange fee regulation in Spain from 1997 to 2007. To my knowledge, they are the first to use bank-level data to study the impact of several episodes of interchange fee regulation for debit and credit cards. They find that intense issuer competition coupled with high interchange fees may have made consumers, merchants, and banks worse off. Clearly, merchants benefit from lower fees and consumers benefit when more merchants accept payment cards if the benefit of greater acceptance outweighs any additional cost to payment providers. Surprisingly, they find that revenues increase among the banks in their sample, even though interchange fees decreased. The effect of these regulations is
clear on banks’ revenues; however, their effect on banks’ profits could not be determined because of data limitations. Furthermore, there may be a critical interchange fee below which issuer revenue decreases. Unfortunately, their data does not allow them to find this critical interchange fee. Additionally, in the absence of adoption and usage externalities, the level of the interchange fee may not affect social welfare.

**Honor-all-cards rules**

A payment card network may require that merchants that accept one of its payment products to accept all of its products. There are different forms of the honor-all-cards rule. The honor-all-cards rule may extend to any payment card that is issued by a member of a network. In other words, if a merchant accepts a network’s credit card, it must accept debit and prepaid cards from issuers belonging to that network. Such a rule enables a card network to innovate by producing different products that when introduced will have a large base of merchants that accept them bypassing the chicken-and-egg problem. The introduction of payroll cards, a type of prepaid card, is an example of an innovation that leverages a card network’s existing infrastructure.

In the United States, around 5 million merchants sued the two major networks, MasterCard and Visa, over the required acceptance of the network’s signature-based debit card when accepting the same network’s credit card. The case was settled out of court. In addition to a monetary settlement, MasterCard and Visa agreed to decouple merchants’ acceptance of their debit and credit products. While few merchants have declined one type of card and accepted another type, the decoupling of debit and credit card acceptance may have increased bargaining power for merchants in negotiating fees.
As part of the payment system reforms in Australia, MasterCard and Visa were mandated to decouple merchants’ acceptance of their debit and credit cards as well. The Payments System Board (Reserve Bank of Australia, 2008b, 16) is unaware of any merchant that continues to accept debit cards but does not accept credit cards from the same network.

A subset of the honor-all-cards rule is the honor-all-issuers rule. In other words, if a merchant accepts a credit card from one issuer, it must also accept credit cards from another issuer within the same network. Such a policy levels the playing field between large and small issuers through a base product, which each issuer can customize. Otherwise, small issuers would not be able to compete with the large issuers. Larger issuers also benefit from the underlying network effects.

Another type of honor-all-cards rule could cover the acceptance of different credit or debit cards from the same issuer. For example, issuers may have a plain vanilla credit card and also have others that earn different types of rewards. While merchants may not care what types of rewards their customers receive from their banks, merchants may pay different fees based on the type of card used by their patrons. More recently, policymakers are considering allowing merchants to discriminate within a card classification, such as a credit card, based on differences in interchange fees.

Conclusion

In summarizing the payment card literature, I find that no one 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.
Much of the debate over various payment card fees is concerned with the allocation of surpluses from consumers, merchants, and banks, as well as the question of who is able to extract surpluses from whom.

I am 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. However, 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 may be reluctant to charge higher prices to consumers who benefit from card use. However, surcharging is increasing in jurisdictions where it is allowed. Third, network competition may not improve the price structure but may significantly reduce the total price paid by consumers and merchants. Fourth, both consumers and merchants value credit extended by credit card issuers (along with other benefits such as security), and consumers and merchants are willing to pay for it. Fifth, evidence from recent interventions suggests that market-based fees may not maximize social welfare.

Determining 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. Efficiency of payment systems is measured not only by the costs of resources used, but also by the social benefits generated by them. Clearly, further research is warranted to explore the complex market for payment services, and policy recommendations should be based on more in-depth research, especially empirical studies that focus on the effects of government intervention.
References


Bedre, Ozlem and Emilio Calvano (2009), “Pricing payment cards,” mimeo, Toulouse School of Economics and Harvard University.


Jullien, Bruno (2001), “Competing in network industries: Divide and conquer,” IDEI (Industrial Economic Institute) and GREMAQ, University of Toulouse, mimeo, July.


FIGURE 1

Payment card fees

Consumer

Fixed goods price or instrument-contingent pricing

Merchant

Consumer fees (annual, finance, other)

Potential rewards

Consumer’s bank

Interchange fee

Merchant’s bank

Merchant fee (fixed, proportional, or both)