

The Economics of Payment Cards*

Marc Rysman
Boston University

Julian Wright
National University of Singapore

Abstract

This paper surveys the economics literature on payment cards, focusing on the role of interchange fees, merchant internalization, surcharging and two-sided markets. The paper aims to integrate a wide range of theoretical papers with the relevant empirical research and important policy questions. We address not only the accomplishments of the research so far, but also the most important limitations in the existing theory and gaps where future research should be directed. Drawing on the insights developed, we provide a detailed evaluation of different policy options.

* *Acknowledgements: This research was supported by a grant from Visa Inc. The views expressed are strictly our own. The authors thank Özlem Bedre-Defolie, Emilio Calvano, the editor Martin Peitz and an anonymous referee for their helpful comments.*

1. Introduction

Payment cards, such as Visa and MasterCard debit and credit cards have become pervasive in modern economies. However, these debit and credit card systems have also generated heated controversy in numerous countries. Both merchants and public authorities have initiated regulatory and legal proceedings, with varying success (see, for example, Bradford and Hayashi, 2008, and Prager *et al.*, 2009). A variety of policy-makers, such as the Reserve Bank of Australia, the European Commission and the United States Government Accountability Office, and economists (e.g., Carlton and Frankel, 1995; Cabral, 2005; Vickers, 2005; Farrell, 2006) have concluded that payment systems rely too heavily on subsidies to cardholders, leading to excessive usage of cards, excessive merchant fees for card acceptance, and inflated retail prices. Further criticism falls on the no-surcharge rule, which prevents merchants setting a higher price for payment by cards. The literature on the economics of two-sided markets (e.g., Rochet and Tirole, 2003; Armstrong, 2006; Weyl 2010) studies these issues from a quite different perspective, emphasizing that skewed pricing arises naturally and can enhance welfare when there are externalities between consumers and merchants. Developing policy for payment cards requires a thorough understanding of the economics of these markets, and the skewed fee structures that often emerge.

This paper provides a snapshot of the state of the literature on payment cards. The paper aims to integrate a wide range of theoretical papers with the relevant empirical research and important policy questions. In so doing, we point to directions for new theoretical and empirical work to bridge the gap between economic theory and policy.

Much of our discussion evaluates interchange fees, the fee that the merchant's bank pays to the consumer's bank as part of a typical purchase transaction. The setting of interchange fees has come under much regulatory and antitrust scrutiny. In Section 2, we provide an explanation of why interchange fees exist and why they are valued by open payment card systems, such as Visa and Mastercard, which rely on issuers and acquirers to deal with consumers and merchants.¹ We also compare interchange fees to analogous decisions by closed systems, such as American Express and Discover, that deal directly with consumers and merchants. We discuss related issues, such as the use of surcharging by merchants to steer consumers towards one card type or another. In Section 3, we discuss whether interchange fees and related policies represent anticompetitive behavior from the perspective of antitrust policy. In Section 4, we discuss the computation of optimal interchange fees, from the perspective of social welfare. This discussion leads naturally to Section 5, which analyzes the sorts of market failures that may take the industry away from a socially optimal output, and to Section 6, which discusses potential justifications and prospects for regulation.

Throughout, we focus on several key concepts. One is the role of the interchange fee in *balancing* terms to consumers versus terms to merchants. With no payment between a

¹ The issuer, or issuing bank, issues payment cards to consumers and is responsible for collecting funds from the consumer when its consumers make purchases with their cards. The acquirer, or acquiring bank, provides the merchant with the ability to accept card payments from consumers and is responsible for remitting funds to merchants when it accepts payment by card.

merchant's bank and a cardholder's bank, the terms faced by consumers and merchants may not lead to the optimal use of the card system. Thus, an interchange fee can potentially help balance the terms to increase the usage of cards and possible welfare. A second is the concept of *merchant internalization*, under which merchants take into account the surplus they offer to cardholders in deciding how much they are willing to pay to accept cards since the additional surplus offered to cardholders allows them to increase their price or capture additional demand (possibly from rivals). Merchant internalization can affect the balancing act played by interchange fees, leading to more favorable terms to consumers and less favorable terms to merchants than would be socially efficient. A third concept is that of asymmetry across the two sides of the market, which can drive apart the socially and privately optimal interchange fees. Asymmetry can take a variety of forms, such as different margins or pass-through rates on the consumer and merchant sides, the differential ability of the platform to price discriminate across the two sides, the asymmetry in the choices facing consumers and merchants, and the differential willingness of consumers and merchants to sign up to multiple types of payment instruments.

Among the most contentious issues in payment cards has been the determination of interchange fees applicable to banks that participated in the card platform when it was structured as a joint venture. We argue that the collective setting of interchange fees should not be characterized as a form of collusion, or as a restraint to competition. Rather, it is better thought of as a balancing device that allows competing banks to provide a valuable service as part of an open system. If banks were to instead set interchange fees unilaterally or bilaterally, or open systems were broken up into competing closed systems, we explain why the market outcomes would likely be substantially worse. Having said that, it is possible that interchange fees as determined by card platforms are not at the socially optimal level. We review different theoretical reasons for why this is possible. The main avenues leading to market failure are somewhat subtle. Reflecting this, standard responses such as increased entry or improved information (e.g., making interchange fees more transparent to end-users) may do little or nothing to improve market outcomes. This provides a potential rationale for regulation, albeit a controversial rationale given the subtle nature of the potential market failures.

We discuss different approaches to regulating the interchange fee, noting that current regulatory practice is based on arbitrary measures that do not accord with any sound economic principles or analyses. This reflects, in part, the difficulty in identifying the socially optimal level of interchange fees, a reason we think more study is needed at this point. A promising way forward, as a starting point, is to try to calculate the Baxter (1983) interchange fee, or more recent related variations (i.e., the merchant indifference criterion or tourist-test interchange fee). These regulatory benchmarks are designed to get consumers to overcome a fundamental externality, that is, to take into account the effect their payment decisions have on the merchants they buy from. While far from straightforward, such approaches can be implemented using a careful and detailed study of merchants' avoided cost from accepting cards. We discuss some of the pitfalls of existing studies and how future studies may be better designed. We argue that more theoretical work on how to adjust the calculation of these regulatory benchmarks to handle realistic features of the market is also needed, pointing to specific limitations of the existing theory. In our view, regulatory initiatives should proceed with caution.

One feature of this literature is that there are already a number of overview articles and reviews in existence, including a recent survey by Verdier (2011). Our survey differs from these previous works in the extent to which we focus on uncovering the key economic insights from the literature rather than focusing on the different models that have been developed. Modeling is important to our discussion, but rather than review how each article contributes to the development of the modeling of payment card systems, our point of difference is to focus on the relevant insights that come from these models and what needs to be done to better connect them with the relevant policy issues. Empirical testing of various theories is an important element of this discussion. Thus, we address not only the accomplishments of the research so far, but also the most important limitations in the existing theory and gaps where future research should be directed, as well as what are the important empirical questions that can be potentially answered with the right data.

We are particularly interested in using the economics literature to inform public policy towards payment cards. The industry has been the focus of a number of different antitrust suits and regulatory initiatives, as we discuss further below. We hope to provide a framework for this discussion, and address the most basic questions about whether government intervention is warranted, either by antitrust or regulation. What market failures are present in this industry, and what are the implications for market performance? If there are problems, what are the conditions under which government intervention will improve these outcomes?

Despite the existence of a number of literature reviews, and although we cover a large number of papers ourselves, this literature is still growing and there many unanswered questions. Clear policy guidance is lacking on many issues, as we point out below.

2. A positive theory of interchange fees

This section focuses on how interchange fees are determined, absent regulation. It does not address whether the level that will be set is efficient or desirable, which is something addressed in subsequent sections.

A question of balance

Closed systems, or three-party systems, such as Discover and American Express typically *issue* cards to consumers and *acquire* merchants to accept the card. They set fees to both sides, which largely consist of an annual fee, an interest rate (for credit cards) and a rewards program for consumers, and a fee for merchants (termed the *merchant discount* in the industry). A closed card platform can choose any structure of prices that it so desires between cardholders and merchants. In particular, if low cardholder fees and high merchant fees are what generate the most card transactions or profit, then the closed card system is free to set this structure of prices.

In contrast, many of the largest systems separate the clearing-house services from the task of obtaining consumers and merchants. This is true for Visa and MasterCard, as well as for debit networks such as NYCE and Pulse. Thus, the direct customers of these systems are banks. The systems are *open* in the sense that any bank or equivalent financial institution

can join. In these systems, banks join and then seek to issue cards to consumers and acquire merchants to accept cards. When a consumer makes a purchase from a merchant, the payment is authorized and routed from the issuing bank through the payment system (i.e., Visa or NYCE) to the merchant's account with the acquiring bank; subject to liability rules governing fraud, payments are generally guaranteed to the merchant, and the issuer is responsible for collecting funds from the consumer. Since many banks are typically associated with each system, there can be substantial competition among banks to offer access to the system. The competition takes place over the terms mentioned above, fees and rewards, and other features such as consumer protection and customer service. Open systems are often referred to as *four-party systems*, referring to the merchant, the consumer and the two banks, although the network owner is really a fifth party.

Typically, the owner of the open system collects a fee, often called the *switch fee* for the service. These are fairly small and, to date, have been largely uncontroversial. In addition to the switch fee, open systems set an *interchange fee*. The interchange fee is an amount that the merchant's bank pays to the consumer's bank as part of a typical purchase transaction.

Figures 1 and 2 illustrate. They contrast the two types of card systems. Consider a consumer that purchases a \$100 good from a merchant using a payment card. In an open system, the issuing bank effectively transfers \$100 to the acquiring bank through the payment system. The acquiring bank deposits the \$100 in the merchant account, minus a merchant fee or *merchant discount*. When authorized, as mentioned, the issuing bank is responsible for the consumer's funding of the transaction. The payment network collects a switch fee from both banks. Most importantly from the perspective of this paper, the acquiring bank effectively pays the issuing bank the interchange fee. If the interchange fee is set at 1%; this means that the issuer effectively receives \$1 from the acquirer on a \$100 transaction (this occurs, operationally, by a reduction in the settlement amount paid by the issuer for that transaction).

While banks compete on terms to consumers and merchants, the default interchange rate is generally set by the network. Typically, the structure of the fee is fairly simple. Some systems use a simple percentage of transaction value, whereas some include a fixed fee per transaction, or both. Interchange fees may differ depending on the industry the merchant belongs to or the type of card used (e.g., platinum or regular). The interchange fee is an important element of the cost for the acquiring banks. Raising the interchange fee will typically lead acquiring banks to increase their fees. Similarly, the interchange fees provide revenues based on transaction values to the issuing bank, which will enable issuing banks to offer consumers better terms, and particularly encourage usage, such as by lower fees, enhanced consumer protections, better customer service, or rewards programs.

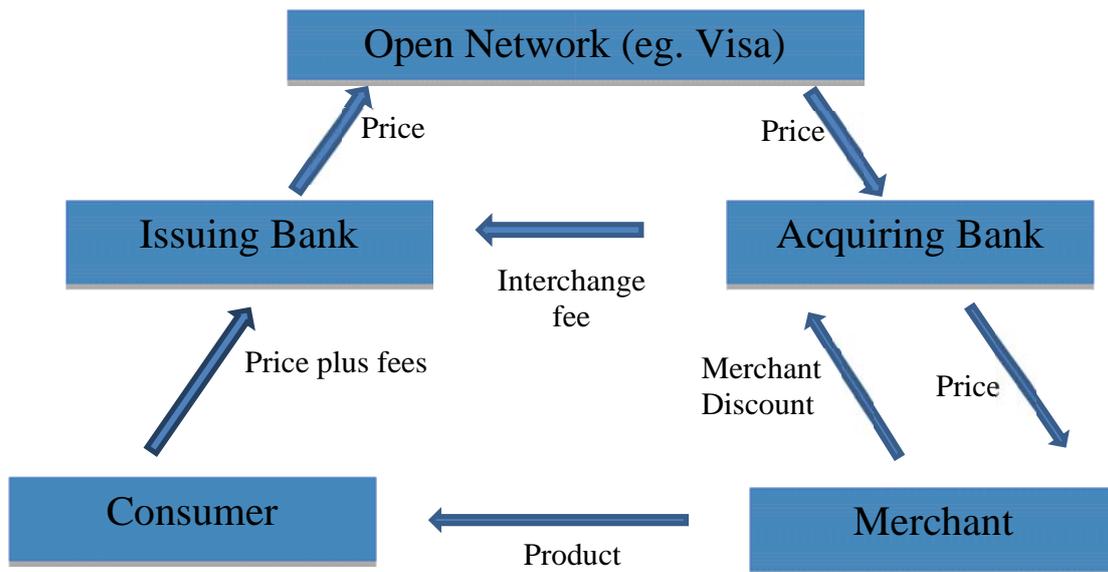


Figure 1: An Open Network System

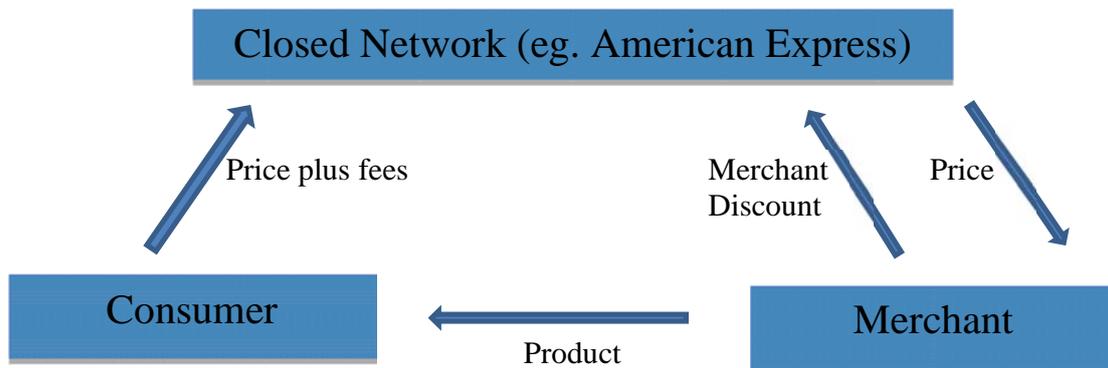


Figure 2: A Closed Network System

To understand the role of interchange, it is easiest to start by considering a closed system. As in Figure 2, the closed system sets fees directly to the consumer and merchant, and does not use an interchange fee. However, there is an implicit interchange fee, without which the closed system would be unable to balance the two sides. Consider what would happen if a closed card system was divided into a separate issuing department and a separate acquiring department, without any transfer pricing allowed between the two departments. Each department would set its own prices to maximize its own separate profit. In particular, the issuing department would not take into account that attracting additional cardholders and card usage generates additional revenues for the acquiring department by

increasing the number of card transactions. The issuing department would therefore tend to set its price to cardholders too high from the perspective of the system as a whole, resulting in too few card transactions. This externality could be internalized if the acquiring department was able to pay the issuing department a transfer fee for each card transaction it generated. It would be in the interests of the system as a whole to have such a transfer payment. As a result of such a transfer fee set by the system, the issuing department would set lower prices to cardholders and conversely, the acquiring department would set higher prices to merchants.

Of course, a similar externality also exists in the other direction, in that each additional merchant acquired will generate additional transactions and revenue for the issuing department. What is relevant for understanding interchange fees is whether, on net, the business can expand as a whole by transferring more revenues in one direction or the other.² In the case of the very first card platforms, which were closed, it can be inferred that this required a transfer from the acquiring departments to the issuing departments. This is consistent with the observations of Evans and Schmalensee (2005, pp. 80-81) that "*During the 1950s and into the early 1960s, based on available data, Diners Club earned about 70 percent of its revenues - and probably most of its gross margin - from the merchant side of the business.*" Similarly for American Express "... available data indicate that for the last four-plus decades, American Express has earned upward of 65 percent of its transaction-related revenues from merchant fees."

Similar issues arise for platforms in other industries where platforms have to bring together two types of customers that want to interact (see the discussion in Rochet and Tirole, 2003, for example). Consider a newspaper that was run as two separate departments, a circulation department that determined the price of the newspaper to readers and an advertising department that determined the price of ads to advertisers. If the two departments had to price independently and without any transfer fee between them, then such a newspaper would not be competitive with other newspapers that could coordinate their prices (for example, through the use of a transfer fee paid by the advertising department to the circulation department for each newspaper sold). As a result, readers would be charged too much from the perspective of the newspaper as a whole. With the wrong pricing structure, the newspaper would not attract as many readers, which would then make it less attractive to advertisers, resulting in lower overall demand and profit, and resulting in it being less competitive vis-à-vis other newspapers.

Of course, in practice, a closed card platform (such as a newspaper publisher) does not need to make use of an explicit transfer fee. It can directly coordinate the actions and prices of its issuing and acquiring departments. Diners Club did this when it started up in 1950, by initially charging merchants 7% of their bill and nothing to cardholders (it subsequently introduced an annual fee of \$3 and a slightly negative transaction fee in the form of float), even though it only had a very small customer base. Likewise, Bank of America, which initially started as a closed card platform started with a merchant service fee of 5 percent of each transaction.

² As Schmalensee (2002) points out, in such a case there would also be a double marginalization problem resulting in fees on both sides being too high. A closed card platform can deal with this directly by lowering its fees accordingly. An open-card platform deals with this by having competition among issuers and among acquirers.

This is in contrast to open card platforms such as those run by MasterCard and Visa, which rely on the decisions of many different independent issuers and acquirers who compete to attract cardholders and merchants, respectively. Each issuer/acquirer will seek to independently maximize its individual profit, in competition with other issuers and acquirers, and with other payment systems. Thus, if the interchange fee was fixed at zero, the open-card system would face the same type of problem faced by a closed system without any explicit or implicit transfer between its issuing and acquiring departments (with each department maximizing its own profits). Externalities across issuing and acquiring would not be internalized. If the net benefit of attracting additional cardholders or card usage exceeds that of attracting additional merchants to accept cards, cardholder fees would be too high and merchant fees too low from the perspective of the system as a whole, card transactions within the system would not be maximized, and overall system profit would be lower than what it otherwise would be.

By setting a positive interchange fee, a fee paid by acquirers to issuers per the value or volume of transactions between them, the card system can increase the cost of acquirers and provide revenue for issuers, which would tend to result in higher fees to merchants and lower fees (or, enhanced protections, increased rewards, etc.) to cardholders. An interchange fee can therefore be viewed as the analogue in an open system to the implicit transfer fee that exists between the acquiring department and the issuing department in a closed system.

A higher interchange fee, therefore, can contribute to lower card fees, enhanced consumer protections, services and card rewards, and so expands the use of cards at merchants that accept them. On the other hand, a higher interchange fee raises merchant fees, and may make some merchants unwilling to accept cards, or more aggressive in steering consumers to other forms of payment, thereby losing all or many card transactions that otherwise would have occurred at these merchants. This highlights the basic balancing role played by an open card system's setting of an interchange fee.

Existing models of interchange fees

The existing literature has modeled how interchange fees have been determined, based on the above conceptualization. In this literature, each card platform is assumed to set an interchange fee that maximizes the aggregate profit obtained from its card transactions by all its issuers and acquirers. This was motivated traditionally by the idea that MasterCard and Visa were associations representing the interests of their member issuers and acquirers, a situation which is still true for Visa Europe. In the present day structure in which MasterCard and Visa (outside of Europe) are investor-owned, MasterCard and Visa extract profit through switch fees (also called network fees) charged to issuers and acquirers. A justification for continuing to assume MasterCard and Visa set interchange fees to maximize the aggregate profit obtained from its card transactions by all its issuers and acquirers is that given the wide scope for price discrimination in setting switch fees, the profitability of the platform should be closely related to the profit these issuers and acquirers obtain from card transactions. However, this is at best a convenient approximation.

Alternatively, and possibly more realistically, some authors have assumed that each card platform seeks to maximize the value or volume of card transactions. If card platforms can set switch fees based only on the value or volume of transactions and not on the revenues obtained by issuers and acquirers arising from card usage, card platforms can be modeled in this way. Note that these two approaches (profit maximization and volume maximization) become equivalent if issuers and acquirers pass-through interchange fees into their respective prices equally. In this case, when the interchange fee is changed, issuers' and acquirers' aggregated profit will only change to the extent that the value or volume of card transactions will change. Therefore, maximizing the value or volume of card transactions will be equivalent to maximizing the profit of issuers and acquirers, thereby providing equivalence between these two approaches.

To date, no one has modelled open systems in which switch fees and interchange fees are set optimally, to understand how this affects the determination of interchange fees in case the above assumptions do not hold. The ability to set switch fees to issuers and acquirers would seem to make interchange fees redundant, since positive switch fees could be set to acquirers and negative switch fees (i.e. subsidies) could be set to issuers, to mimic any interchange fees. Yet, switch fees do not seem to have been used in this way. Interchange fee setting remains contentious in the presence of switch fees, suggesting switch fees have not made interchange fees redundant. Thus, understanding the different roles of switch fees and interchange fees remains an important area for future research.

As will become clear, the models in the existing literature have no problem explaining either the choice of positive interchange fees by open card platforms, or the idea that merchants feel they must take payment cards (as put forward by Vickers, 2005). Here, we first review the most important assumptions and differences across the models. We will discuss some limitations of these models and areas where more modeling is needed.

Most of the literature focuses on the case of a single card platform with some exogenously provided alternative, like cash. The platform generates some social benefit. For instance, consumers get convenience benefits of using cards rather than the alternative (i.e., cash) and merchants get convenience benefits of accepting cards compared to the alternative (i.e., cash).³ Only a handful of articles have considered the case where there are competing card systems (e.g., MasterCard and Visa). Merchants are assumed to set a single retail price for a good regardless of whether consumers pay with a card or cash. In particular, merchants are assumed not to be able to set a surcharge to consumers for using cards. One justification for making this assumption is the no-surcharge rules card platforms have traditionally applied to merchants. Another is the lack of widespread surcharging for cards (or discounting for cash) even in countries where card platforms have had to remove rules preventing merchants from adding surcharges for their cards. Moreover, even where discounting for alternatives to cards has long been permitted, the practice tends to be concentrated in a few merchant sectors. We discuss surcharging further below.

Existing theories that provide a positive theory of interchange fees fitting these assumptions include those introduced by Rochet and Tirole (2002), Schmalensee (2002), Wright (2004),

³ Rochet and Tirole (2011) equivalently refer to the convenience cost of using or accepting cash, where the convenience cost of using or accepting cards is normalized to zero.

Guthrie and Wright (2007), and Bedre-Defolie and Calvano (2013) among others. Key differences across these models are whether (i) they allow for merchants to be heterogenous, (ii) whether they allow for merchant internalization, (iii) whether cardholder pricing is restricted to usage pricing, fixed (or annual) fees or two-part tariffs that combine both, and (iv) whether they allow for competing platforms.

Rochet and Tirole (2002) focus on the case of a single card platform in which merchants all obtain the same convenience benefits of accepting cards (homogenous merchants) and consumers face fixed (e.g., annual) fees for holding a card. They are the first to introduce merchant internalization, which is the property that merchants that set a single price regardless of how consumers pay will tend to internalize the surplus they offer to cardholders in their decision about how much they are willing to pay to accept cards. In their model, this result comes from the fact that merchants take into account that card acceptance allows them to offer more surplus to their customers, who get a benefit from using cards, and so can set a higher price without losing market share.

Subsequent research (Farrell, 2006; Wright, 2010, 2012; Rochet and Tirole, 2011) has shown that merchant internalization is a much more general phenomenon, holding in some cases even when merchants are monopolistic, Cournot competitors or perfectly competitive. Simply put, when a merchant accepts cards it is improving the quality of the service it offers consumers, the option of using cards for payment, and it is only natural this allows it to charge a higher price. The more surplus it can offer consumers the more it is willing to incur a cost to do so. This phenomenon is no different from any other service a merchant may employ to attract customers for which it does not set a separate price.⁴

Rochet and Tirole (2002) focus on the fact that merchant internalization makes merchants less resistant to accepting cards, and so merchants are willing to pay fees above the convenience benefits they receive from accepting cards. Rochet and Tirole use this set-up to explain why merchants sometimes feel they must take cards (Vickers, 2005), even though at the point of sale they would rather not. Once the customer has decided to buy from the merchant, the merchant may want to steer the consumer to pay by cash because the merchant discount is larger than the convenience benefit of accepting the card, but from an *ex-ante* perspective, the merchant also takes into account the increase in store attractiveness brought about by the option of paying by card. Thus, the merchant is left with conflicting sentiments, the *ex-ante* belief in the value of accepting the card and the *ex-post* desire for the consumer to pay by cash (Rochet and Tirole, 2011, p. 467).

Schmalensee (2002), another early contribution to the literature, provides a contrasting treatment. He takes as given the existence of some "quasi-demands" by consumers and merchants for cards. The terminology "quasi-demands" refers to the fact that a customer may want to use cards but whether it can do so at a particular merchant also depends on whether the merchant accepts cards (and similarly for merchants with respect to a particular customer). In his model, the quasi-demand of each type of user can be derived assuming users only consider convenience benefits. Cards are used only if the convenience benefits

⁴ Chakravorti and To (2007) focus on the ability of a merchant to bring forward sales by accepting credit cards rather than risk the possibility of the sale going to its rival in the next period if it turns down credit cards. This creates an intertemporal business-stealing reason to accept cards that has a similar effect to merchant internalization.

from doing so exceed the respective usage fee. Thus, Schmalensee's theory ignores merchant internalization.⁵ On the other hand, by specifying a quasi-demand function for card acceptance by merchants, it allows for the fact that merchants get heterogeneous benefits of accepting cards.

In his formulation, with linear demands, only differences across the two sides affect the privately determined interchange fee. Thus, the privately determined interchange fee is linear in two terms, one reflecting the difference in costs across issuing and acquiring and the other reflecting the difference in demand elasticities across the two sides. Interchange fees will be positive if issuing costs are higher than acquiring costs, as is usually the case, and if cardholder demand is more price sensitive so that card transactions can be expanded by loading less of the costs on the cardholder side.

Wright (2004) combines the merchant internalization inherent in Rochet and Tirole (2002) and heterogeneous merchants from Schmalensee (2002). Additionally, he assumes card fees are usage based, so rewards arise per transaction and therefore will also be internalized by merchants under merchant internalization (unlike the case of Rochet and Tirole, 2002 in which issuers' fees are fixed fees). Merchant internalization makes the balancing role of privately determined interchange fees more subtle than in Schmalensee (2002). It not only makes merchants more willing to pay to accept cards (less resistant in the terminology of Rochet and Tirole, 2002), but also explains why their demand is less sensitive to interchange fees and so why positive interchange fees are to be expected.

To see this, consider a higher interchange fee that increases merchant fees and decreases card fees (e.g., increases rewards or features) by the same amount. This higher interchange fee would seem to have no impact on whether merchants will accept cards given that merchants internalize the higher rewards their customers get from using cards when deciding whether to accept them. By itself, this would make merchants totally insensitive to changes in the fee structure. The reason merchants are still somewhat sensitive to changes in the fee structure and less likely to accept cards when interchange fees are higher, is due to the following additional effect. Facing higher rewards, some consumers will start using cards that previously did not. These consumers therefore have a benefit of using cards that is lower than the average consumer. Thus, the mix of consumers shifts towards those with lower values of using cards. Given that merchants also internalize the benefit their customers get from using cards, merchants will be less likely to accept cards following an increase in interchange fees.

Taking the same linear demand model that Schmalensee (2002) considered but allowing for merchant internalization, with full pass-through of interchange fees on both sides (competitive issuing and acquiring), Wright (2004) finds that issuing and acquiring costs do not factor in the privately determined interchange fee at all. Even if the two sides are otherwise symmetric, the interchange fee is positive. With less than perfect pass-through,

⁵ This can be rationalized in very special cases. One special case is where the merchant is a monopolist facing unit demands with a mix of cash customers and card customers. If a monopolist increases its retail price above the maximum amount cash customers are willing to pay in order to try to extract some surplus from those paying by cards, the monopolist will lose all cash customers. Therefore, if there are enough cash customers, the monopolist cannot extract any surplus from card customers. Wright (2003a) and Bedre-Defolie and Calvano (2010) make use of this setup.

and less than full merchant internalization and/or non-linear quasi-demand, relative costs and demand elasticities will, in general, matter across the two sides.

Among articles focused on single-card platforms, Bedre-Defolie and Calvano (2013) provide a recent contribution that differs in that they model issuers setting two-part tariffs, including a fixed or annual fee as in Rochet and Tirole (2002) and a linear usage fee as in Wright (2004), with cardholders making both a decision about holding a card and then whether to use the card at the point of sale. While usage fees to cardholders decrease one-for-one with any increase in interchange fees in their baseline setting (full pass-through), the fixed fee set by issuers offsets some of this benefit to cardholders. In other words, some of the additional cardholder surplus to cardholders from lower usage fees or higher rewards will be captured by higher fixed card fees. In their model, the privately determined interchange fee takes into account the effect of lowering usage fees or increasing cardholder rewards on the option value to consumers of being able to use cards. This provides a reason to increase interchange fees above the level that would arise in a model without any fixed fee to consumers, since there is a double dividend of increasing cardholder surplus. Maximizing buyer surplus thus implies (high) positive interchange fees. However, while their analysis allows for heterogenous merchants it does not allow for merchant internalization.

As mentioned previously, all of the above models are based on a single-card platform that sets an interchange fee to maximize profits. What happens to the determination of interchange fees if there are competing card platforms, each setting its own interchange fees? Guthrie and Wright (2007) are the first to focus on this case. They extend the models of Rochet and Tirole (2002) with homogenous merchants, and Wright (2004) with heterogenous merchants, to the case with two identical card platforms, in both cases keeping the assumption of merchant internalization. Before discussing their results, note that the effect of inter-system competition on the level of interchange fees is far from obvious. It is similar to asking whether more competition between rival newspapers will result in publishers charging less to readers or less to advertisers. Greater inter-system competition should help reduce at least one of the prices, but in general it is theoretically possible for greater competition to increase the prices charged on one of the sides (Rochet and Tirole, 2003).

Guthrie and Wright (2007) show that the equilibrium determination of interchange fees with competing systems depends on the extent to which consumers and merchants join multiple competing platforms (multihome) versus join only one.⁶ In their model, in equilibrium it is possible for there to be multihoming on one side or other, reflecting that if users on one side join both platforms, then users on the other side do not need to. They characterize the range of possible outcomes, which varies between the case in which consumers' interests alone determine the equilibrium interchange fee and, at the other extreme, the case in which merchants' interests alone determine the equilibrium interchange fee. To illustrate, consider the extreme case in which all consumers choose to hold one card only. This leads to a competitive bottleneck equilibrium similar to that analyzed in Armstrong (2006). In the competitive bottleneck equilibrium, card systems compete to sign up consumers exclusively.

⁶ See also Rochet and Tirole (2011) for some related analysis. Chakravorti and Roson (2006) extend the research on platform competition by relaxing the assumption that the total price is constant. Rysman (2007) provides empirical evidence on this issue.

Merchants want to access these consumers, and each card system then sets an interchange fee that maximizes consumers' interests. In the case with merchant heterogeneity, at one extreme interchange fees will be set above those set by a single card platform, reflecting that maximizing buyer surplus calls for a particularly high interchange fee, and at the other extreme, interchange fees will be set below those set by a single card platform, reflecting that maximizing seller surplus calls for a particularly low interchange fee.

Whether inter-system competition drives up or down interchange fees is, therefore, an empirical question which remains wide open. Weiner and Wright (2005) take a very preliminary look at the issue, noting in a cross-section of twenty countries that there is no discernable relationship between interchange fees and the proportion of card transactions that Visa conducts, compared to Visa and MasterCard together. For instance, they found those countries where Visa and MasterCard have more equal market shares have no higher interchange fees (on average) than those where Visa attracts most of the business. In this respect it is notable that the U.S., which is perhaps subject to the most intense inter-system competition between card platforms, also has relatively high interchange fees compared to certain other OECD countries. These findings of Weiner and Wright are consistent with the fact that theoretical models do not predict any particular correlation between the level of interchange fees and network market shares.

Using time-series data (1990-2003) from the U.S., Weiner and Wright (2005) plot interchange fees and issuer concentration through time, finding some tentative evidence of a positive correlation between the two (although this depended on how issuer market concentration was measured). Noting that individual issuers prefer a card system that offers them a higher interchange fee⁷, they speculate that interchange fees increased due to the increasing ability of large issuers to play one platform off against another in the U.S., which was the result of a shift towards a more equal distribution of market shares among the top 5 issuers over the period and strong inter-system competition.⁸ Each of the episodes in which interchange fees jumped up was associated with a battle to keep issuers.⁹ Developing a theoretical model that captures these issues would be very useful. It would also be useful to try to understand how this inter-system competition relates to the setting of switch fees which have largely been ignored in the existing literature.

Surcharging

Until now we have assumed, as does each of the papers discussed above, that merchants cannot add a surcharge to consumers who pay by card. As mentioned earlier, this can reflect either that card platforms impose no-surcharge rules on merchants, or where no-surcharge rules have been banned by regulators or otherwise do not apply, that for some

⁷ This could be either because the card system maximizes the profit from issuing and acquiring, whereas issuers ignore the acquiring side, or because the card system maximizes profit from switch fees which leads it to care more about card volume than issuers do. Thus, without competition for issuers, card systems may set a lower interchange fee than the issuers would like.

⁸ Wang (2010) develops a very different theory of interchange fees in which the entry and exit of heterogeneous card issuers and falling issuing costs are the main driving forces of privately set interchange fees.

⁹ Weiner and Wright (2005) note "In 1990/91 there was a rise in large non-bank issuers that the networks sought to attract; in 1998/99 there was the realignment of Citibank's business to MasterCard; and from 2003/04 American Express has increasingly been courting existing MasterCard and Visa issuers by offering the equivalent of higher interchange fees."

other reason many merchants still do not add surcharges for payment by card. We turn to the issue of the extent of surcharging and theories to explain it below. Initially, we explain what the existing literature on interchange fee determination in the presence of surcharging has focused on.

The main result in the literature with regard to surcharging is known as neutrality. This says that the level of the interchange fee will be irrelevant for the decisions of cardholders and merchants when merchants can set a surcharge for consumers who pay by card. Gans and King (2003) give the most general neutrality result. As with the rationale for interchange fees, neutrality is easiest to explain in the case of a closed card platform like that of American Express. The merchant accepting American Express cares only about the total amount it pays when it accepts cards (the merchant fee less any surcharge it charges consumers) and the consumer that uses cards cares only about the total amount it pays when it uses cards (the card fee plus any surcharge it faces from the merchant). Since there are three prices set by American Express and the merchant (the card fee, merchant fee and the merchant's surcharge for cards), there is a redundant price. Thus, the structure of fees between cardholder and merchant will be neutral. This situation is analogous to real-estate agent fees, for which it does not matter, in theory, how much of the real-estate agent's commission is paid by the buyer and how much is paid by the seller. In principle, the sale price will just adjust accordingly.

As Gans and King (2003) establish, the same logic extends to an open card platform where issuers set card fees and acquirers set merchant fees. Suppose that an open system passes the interchange fee to consumers and merchants, and the system raises the interchange fee. Thus, fees go up for merchants and down for consumers by an offsetting amount. The merchant can raise the surcharge to make up the increased merchant fee, so that the merchant still collects the same amount per transaction, on net. Similarly, the consumer will pay lower card fees but a higher surcharge to the point of indifference between the two pricing structures. The change does not affect the quantity of transactions. With full pass-through of interchange fees to merchants, surcharges will reflect merchant fees, less any convenience benefit (or avoided cost) to the merchants of accepting cards.

If this theory held exactly, interchange fees would be irrelevant in countries in which surcharging is allowed, such as Australia, Czech Republic, Denmark, Ireland, Hungary, Netherlands, New Zealand and the U.K. The various parties should not care about interchange fees. In reality, the card platforms and the various other parties do seem to care a lot about interchange fees. Moreover, the limited evidence available suggests that surcharging is not widespread in any of these countries. Australia has seen the greatest use of surcharging following the removal of no-surcharge rules there. Even there, by 2011, the majority of merchants were still not surcharging and the vast majority of card transactions were still conducted without surcharges. Moreover, of those merchants that do surcharge, the *average* surcharge is around twice that of the merchant fee they face, and this average surcharge has increased even while merchant fees have decreased.¹⁰

¹⁰ This led to a government inquiry. See <http://www.rba.gov.au/publications/consultations/201106-review-card-surcharging/pdf> for data on the level of surcharging and complaints about excessive surcharging. In June 2012 the Reserve Bank announced its decision to allow card scheme rules to limit surcharges to the reasonable cost of card acceptance, effective January 1st, 2013. Similarly, see http://www.ofc.gov.uk/shared_ofc/super-

Surcharging seems to be used by some merchants as a form of add-on pricing, adding additional fees that appear only at the register, where it may be more difficult for consumers to opt out or make another payment selection (e.g., if they don't have sufficient cash on hand), and where doing so makes price comparisons between merchants more difficult. Thus, the neutrality theory and associated models that assume perfect cost-based surcharging do not seem to explain observed behavior well.

Wright (2003a) provides an analysis of surcharging in which merchants are local monopolies, showing how they will surcharge to extract the convenience benefits of their customers from using cards, which they are unable to do without surcharges in his model.¹¹ In his model, consumers first decide whether to hold a card and later whether to use it. By extracting the usage value of cards through surcharges, merchants make fewer consumers willing to hold cards in the first place. However, this externality is not internalized by merchants when deciding whether to surcharge since they take the number of cardholders as given (as each merchant is just one of the very many merchants that consumers have a chance of dealing with). This analysis provides an answer to why negotiations along the lines of Coase (1960), in which consumers and merchants negotiate prices that internalize any externalities that exist between them, will not take place when surcharging is allowed, and provides one reason why neutrality might break down. Each merchant knows that a customer will use the card across a wide range of merchants, so that an individual merchant would never be able to capture the benefits of its own subsidy, designed to get consumers to join a card network. Merchants free ride on the actions of other merchants to provide surplus for cardholders.

More generally, the theory of interchange fees when only some merchants surcharge, or would do so if interchange fees are increased beyond some threshold level, is yet to be fully worked out. To do so in a meaningful way will require first providing a convincing theory of why merchants sometimes do not surcharge, and other times seem to surcharge excessively, with little or no discounting for alternatives to payment cards. Existing theories of interchange fees either involve all merchants surcharging or all merchants not surcharging.

Perhaps the simplest way to proceed is to assume that merchants face some cost of surcharging. Then, depending on the convenience benefit of accepting cards, some merchants will reject cards, some will accept cards but surcharge and others will accept cards and not surcharge. An alternative theory that would account for the disparate behavior of merchants is that for some types of merchants, surcharges are particularly salient from consumers' point of view, perhaps more so than setting higher retail prices and so such merchants will try to avoid using surcharges. On the other hand for other types of merchants, surcharges may not be salient at all, so merchants can use a surcharge to add a hidden fee to their customers, similar to other forms of add-on pricing.¹² Neither type of situation will lead interchange fees to be neutral. Whether one develops one of these

[complaints/OFT1349resp.pdf](#) for complaints of excessive surcharging in the U.K. and the OFT's statement on November 9th, 2012, that "a ban on payment surcharges above the cost of processing a transaction should apply to all businesses."

¹¹ This is due to the existence of sufficient cash customers, who constrain the monopolist from increasing its retail price beyond the monopoly price for cash customers in the absence of surcharging given consumers are assumed to have unit demands.

¹² See Ellison (2005) for a discussion and theory of add-on pricing.

theories, or some other (e.g. based on the perceived fairness of surcharges), it remains for future research to explore the implications of surcharging for the setting of interchange fees, taking into account this type of heterogeneity in retail behavior.

This discussion also suggests that empirical research on surcharging behavior would be very valuable. While some research exists, there certainly could be more. What determines a firm's decision to surcharge, and how does the firm determine the surcharge level? Is it affected by the type of product, the type of consumer, or the level of competition? What is the relationship between surcharging and the cost of card acceptance, both directly in terms of the merchant fee and more generally in terms of the merchant's overall cost of accepting a card? To what extent, if at all, do merchants reduce prices for cash customers when they surcharge for card use? How does surcharging vary across different payment options that a consumer might use? An issue with this line of questioning is that surcharging practices are relatively recent and may still be evolving. But overall, we view new empirical evidence on surcharging as particularly valuable.

Section 3 Anti-competitive behavior?

In this section, we focus on the proposition that card systems engage in anti-competitive behavior that should be subject to punishment under antitrust laws. These concerns can be divided into two types. The first concern is that positive interchange fees are an anti-competitive device used by competing issuers (under the guise of a card platform) to enable them to collectively increase merchant fees. This concern centers around a charge of collusion, based on the claim that card platforms represent a collection of banks that otherwise compete with each other. We deal primarily with this concern, comparing outcomes with centrally set interchange fees to those that arise in counterfactuals in which centrally set interchange fees are ruled out. A second concern is that the contracts that acquirers sign with merchants involve rules imposed by card networks that have been claimed to exploit their market power to extract merchant surplus and limit competition between card systems. We discuss primarily the “no-surcharge” provision, although other contract items have been brought up in the past. This concern is not dependent on collective behavior by banks and instead centers around the actions of the card system as a whole.

Concerns about collusion?

We begin with a discussion of what we call the “cartel theory.” Historically, Visa and MasterCard were organized as associations of banks, so the interchange fee was set by banks that had ownership and governance rights within the associations. Thus, advocates of the cartel theory claimed there was a collective determination of the interchange fee that represented an example of collusive price-fixing. In this view, the card system uses high interchange fees to extract revenue from merchants and at the same time induces consumers to use cards so that merchants must accept the card. More recently, Visa (outside of Europe) and MasterCard have incorporated to become investor-owned companies that have banks as clients. Articles in the “cartel theory” vein include those by Carlton and Frankel (1995), Frankel (1998), Balto (2000), and Frankel and Shampine (2006).

Opposing points of view have been offered by Evans and Schmalensee (1995), Chang and Evans (2000), and Klein *et al.* (2006).

The cartel theory has never been formalized. However, many of the assumptions behind the cartel theory are not fundamentally different from the existing positive theory of interchange fees. What the cartel theory views as the exploitation of market power is viewed by the two-sided markets literature as optimal pricing in the response to the given elasticities of demand on each side of the market. Consider the following: In the cartel theory, higher interchange fees are used to increase merchant fees. This is a feature of all of the models of interchange fees discussed in Section 3. Indeed, in several of these models (including in Rochet and Tirole, 2002), acquiring is perfectly competitive and involves no margins. The pass-through of interchange fees to acquirers is 100%. In the cartel theory, this increase in interchange fees generates a higher profit margin on card transactions for issuers. Provided issuers do not rebate the entire higher margin into higher rewards and lower fees (or other activities like marketing and card innovations), this will also be true in the models of Section 3. The possibility of less than 100% pass-through of interchange fees to cardholders is allowed in most existing models. The fact that cardholders do get some positive rewards, which the cartel theory interprets as rent-seeking behavior by issuers, is of course consistent with some positive pass-through of interchange fees. That interchange fees are set to maximize the profit of issuers is also a feature in a number of existing models, like those of Rochet and Tirole (2002, 2011) given perfect acquiring competition.

The primary defense of the right of a card association to set interchange fees rests on the concept of a joint venture. Joint venture law specifies when collective action is legal. Briefly, an important tenet of antitrust law is that having more products available is preferable to having fewer, and so joint action is legal when it makes a product available that would not otherwise be so. For example, creating a sports league requires the collective action of individual teams to set rules, schedules and champions. Although this coordination among teams could be characterized as a form of horizontal restraint, sports leagues have long been recognized to be legal under joint venture law since professional sports as a product could not be delivered without some coordination between independent teams.

Similarly, the standard defense of the interchange fee set by a joint venture is that banks could not offer a payment card without it. No bank is big enough to deliver the ubiquity required to generate a successful payment card. In this interpretation, open card systems represent a standard-setting organization that allows disparate issuers and acquirers to offer a national and international payment product. Without some rule about how issuers and acquirers were to compensate each other, there would be no way to deliver this product. Under this argument, Visa successfully defended itself against a collusion charge in 1984. (*Nabanco vs. Visa, 1984*).

The cartel view, while sharing some of the same assumptions of existing models of interchange fees, is quite inconsistent with the interpretation of interchange fees given in Section 2 as a balancing device. In the cartel view, the reason why interchange fees are positive is not the result of some balancing considerations, but rather is because positive interchange fees artificially raise merchant fees beyond some competitive level. According to the cartel theory, interchange fees do not allow open card platforms to replicate (or

compete) with the pricing structure a closed card platform would set directly. Rather, they allow open card platforms to go beyond the merchant fees that would otherwise arise in some competitive alternative by allowing the competing issuers to collude.

In our view, the cartel theory is flawed. The principal problem with the cartel view is that the counterfactual (i.e., what would happen absent the alleged cartel behavior) is no more competitive than when interchange fees are set collectively. We review several possibilities here. First, the standard response of a competition authority to a cartel is to break up the cartel. For instance, that might mean forcing each bank to market its own individual card product, such as Discover or American Express. As we have argued above, it is not clear that an increase in the number of independent card systems would deliver lower interchange fees. The notion of breaking up Visa or MasterCard into thousands of individual card producers does not appear to have been seriously considered by even the most ardent critics of the card system. In our view, the fact that break-up is not desirable is a strong signal that collusion is not the primary problem.

A second alternative is to maintain the card system but not allow for a system-wide interchange fee. In this approach, each issuing and acquiring bank would have to negotiate a bilateral interchange fee. This would prove onerous for consumer and merchants. In this scenario, a consumer that wanted to use a Visa card in a store that accepted Visa would still first have to check if the consumer's bank and the merchant's bank had reached an agreement on interchange. If not, the card would not be usable. Such a scenario could undermine the value of the card altogether, which is what leads card systems to adopt the "honor all cards" rule – merchants that accept Visa cards must accept all Visa cards.¹³ However, it is difficult to combine the "honor all cards" rule with bilateral interchange negotiation because it leads to a hold-up problem, in which a single issuer might refuse to deal with acquirers unless they accept its demands for a particularly high interchange fee. Klein *et al.* (2006) have argued that a common interchange fee allows a four-party card network to maintain an honor-all-cards rule, some form of which is an important factor underlying the value of the network.

Furthermore, bilateral negotiation with an honor-all-cards rule could result in substantially higher interchange fees and substantially lower system output (a possibility shown in Small and Wright, 2002). If any particular issuer increases its interchange fee, it will capture the benefit of the higher fee, but the harm caused by that increase - lower merchant acceptance of cards - will be shared by all issuers. This last feature, free-riding by individual issuers, is in contrast to the interchange fees that would be set by issuers setting interchange fees collectively, even if acquirers' interests are not counted at all.

Another concern about bilateral negotiations is that they could make it hard for new entrants given the need to deal with each individual financial institution on the other side, some of which will likely be competing on the same side, to get access to the open platform.¹⁴

¹³ We note that in certain countries, including the U.S., Visa and MasterCard have "bifurcated" their respective "honor all cards" rules, permitting separate acceptance decisions by merchants of credit versus debit category products.

¹⁴ See RBA: "Reform of Debit Card Systems in Australia – A Consultation Document," December 2005. One of the ways it notes entry could be made more difficult under bilaterals is (pp.3-4) "for an existing participant to be unwilling to agree to an interchange fee that is similar to the fee paid to, or received from, existing participants." It goes on to state, "These concerns are longstanding and have been raised in discussions with the Bank over a number of years. It has taken some

Authorities such as the Reserve Bank of Australia (RBA) that have considered bilaterally-set interchange fees have tended to dismiss them. One interesting exception is the New Zealand Competition Commission, which in 2009, reached a settlement with banks and card platforms that involved each issuer posting its own interchange fees to acquirers.¹⁵ To protect against the holdup problem noted above, card platforms were allowed to set maximum interchange fees that individual issuers could adopt. MasterCard and Visa proceeded to set maximum interchange fees at similar levels to those already in place at the time, and individual issuers have not chosen to set significantly lower interchange fees.

In addition to break-up and bilateral negotiation, perhaps the most popular method for addressing collectively set interchange fees is to set them to zero. The European Commission successfully prosecuted MasterCard for collusion by arguing that collectively set interchange fees were not necessary to provide the card product, since MasterCard could provide the card product for an interchange fee of zero. Note the crucial difference in alternatives considered in the U.S. case (*Nabanco*) and the European case: the U.S. case used bilaterally negotiated interchange fees as the alternative whereas the European case used an interchange fee of zero. Thus, an interchange fee set by the network was found to be necessary to provide the product in the U.S., but in Europe, the Commission concluded that an interchange fee greater than zero was unnecessary.

We view the European approach as unusual in that antitrust law does not typically inquire into price levels once prices are determined to be set in a way that is consistent with antitrust principles. For instance, the European case suggests that if MasterCard had collectively set an interchange fee of zero, it would have been legal, a highly unusual twist for a collusion case. Furthermore, the decision of the European Commission recognized that a non-zero interchange fee may raise total welfare. In fact, the court did not ultimately impose a zero interchange fee. It has instead moved towards a fee that it determines to be socially desirable. Thus, the court has turned towards interchange fee regulation. While we further discuss the issues with regulation below, at this point we note that we know of no other collusion case that has ended in the collusive instrument being regulated rather than banned, which, in our view, undermines the original claim of collusion.

A fourth response to the collusion charge has actually already been undertaken by the card systems, which is to switch from an association of banks to an incorporated firm owned by investors that serves banks as clients. In this case, the decision-making power over the interchange fee lies with a publicly held corporation, rather than with a joint venture owned and governed by banks, which seemingly undercuts the claim of collusion. Thus, joint venture law no longer naturally applies. In this case, the card systems argue that there is no collective setting of prices, but instead, the interchange fee is a vertically imposed rule or condition which is pro-competitive since it is output expanding. Critics tend to dismiss the importance of the change in organization.¹⁶ Indeed, under the theory of balancing

direct connectors in the EFTPOS system many years to establish bilateral connections with some of the larger participants, who have been reluctant to give priority to establishing these connections. Some current direct connectors have not been able to establish complete sets of bilateral connections with all other direct connectors despite efforts over a number of years.”

¹⁵ In addition, no-surcharge rules were removed.

¹⁶ For example, see Point 3 in the Executive Summary of the decision of the European Commission in regards to MasterCard in 2007. (*Commission Decision of 19/XX/2007, COMP/34.579 MasterCard.*)

interpretation of interchange fees, the new form implies little change in the underlying incentives facing card systems (e.g. to maximize the usage of the card system).

Finally, we consider the cartel theory in light of the historical record. Prior to the introduction of collectively set interchange fees, one might expect there was some more competitive marketplace that was undermined by this collective action. Merchant fees ought to have been lower prior to the creation of this cartel. This would be consistent with the cartel being formed in order to raise these merchant fees so as to raise issuers' margins, reduce output, and lessen competition. As noted in Section 2, the historical experience does not support this view. Prior to the introduction of open card platforms and positive interchange fees, card platforms were closed platforms that involved much higher merchant fees. Moreover, the use of positive interchange fees has undoubtedly led to the expansion of card transactions, the presumption by both sides of the debate, and the opposite of the anti-competitive output restriction that would be consistent with the predictions of a cartel theory. Interchange fees have not decreased post the Visa and MasterCard IPOs. American Express, a single entity, still charges higher merchant fees (on average) than Visa or MasterCard.

A final point on this is that advocates of the cartel theory sometimes question the use of the theory of two-sided markets to discuss card systems. They would argue that since card systems such as Visa are close to ubiquitous, there is no longer a network effect or positive feedback loop between consumer usage and retail acceptance. That is, two-sided markets should not matter for a “mature” market, such as payment cards. This reasoning is flawed. Even if every consumer and merchant were a member of a card network, there is still the important question of usage of the card, which will respond to incentives provided by the interchange fee. Indeed, most of the models of interchange fees discussed in Section 3 focus exactly on the question of usage with network effects playing no role. Furthermore, the card market is in a greater state of flux than is commonly assumed by academics and policy makers. Europe has transitioned from a signature verification system to a “chip-and-PIN” system. Innovations such as contactless payments and mobile payments represent important possible future developments for the industry. Also, we argue below that decoupled debit, a debit card supplied by a firm other than the consumer’s bank, represents a possible innovation.

Concerns about Contracting?

There have been a number of other charges against the card systems. Whereas the cartel theory is clearly flawed, charges based on a variety of contracts that card platforms sign with merchants may warrant more detailed analysis. Note that these charges are not specific to open platforms. In practice, they have been applied to American Express as well. We briefly discuss the issue of surcharging, which has been subject to litigation and regulation.

The “no-surcharge” requirement captures the general restriction that merchants are not allowed to steer consumers towards the merchant’s preferred payment mechanism by adding a fee to the standard price when the cardholder chooses to use a particular payment

method.¹⁷ It is perhaps better termed a “no-discrimination” requirement, if merchants are not allowed either to surcharge the card or discount alternatives.¹⁸ The antitrust argument against such a restriction is that it plays an exclusionary role, in that it is difficult for a new card system that is predicated on low merchant fees to enter the market and effectively compete. Even if the merchant finds the new card system to be attractive, it has no way of steering consumers to use it, beyond refusing to accept alternative cards. Furthermore, the inability to surcharge (or discount) perhaps gives greater power to the card system to raise merchant fees. More generally, a wholesaler that conditions the merchant’s price of its product on the price of a competing product could raise antitrust concern. The Department of Justice prosecuted under these types of arguments and reached a settlement with Visa and MasterCard under which they have agreed to allow merchants to discount cards, and to apply different discounts to different cards.¹⁹

One counterargument is that only the largest firms in an industry should be able to exploit market power. American Express is smaller than Visa and MasterCard and still uses a no-surcharge rule, and the Discover card used one up until 2006. Indeed, no-surcharge rules have existed since the inception of the industry, long before card platforms could face any charge of market power. However, the theory of two-sided markets has a potential explanation. Card networks have monopoly power over access to those consumers to the extent that the consumers only want to use that form of payment (e.g. if consumers single home). In that case, card platforms can theoretically have a type of “market power” with regard to merchants even if the card platforms have relatively small market shares among consumers.²⁰

We discussed above that recent experience from surcharging suggests that merchants use surcharging not just to pass costs along but also to price discriminate against card users (e.g. by imposing add-on type pricing). While merchant mark-ups are a part of any market, perhaps it is reasonable for a firm to act to prevent price discrimination against its own customers. In fact, the Reserve Bank of Australia will allow card schemes to limit surcharges to the reasonable cost of card acceptance from 2013 in order to prevent abuse. Overall, further theoretical research on surcharging from the perspective of a form of resale price maintenance or a “most favored nation” type of covenant would be useful.

Several policy concerns are less central to antitrust in the sense that they do not deal with exclusionary behavior. For instance, as we have outlined above, under certain sets of assumptions, surcharging renders the interchange fee neutral, in that final consumer and

¹⁷ Under some versions of a no-surcharge rule, merchants are allowed to offer discounts for the use of cash, but may not offer discounts or surcharges that steer consumers among different cards. There exist instances in which such discounts are permitted.

¹⁸ In some countries, card systems allow discounting but not surcharging. Given that discounting is not widely used in such countries, it appears that in practice the no-surcharge rules have a similar effect to no-discrimination rules.

¹⁹ At the time of this writing, there is still no settlement with American Express. Thus, although Visa and MasterCard allow merchants to offer a discount for the fact it may be cheaper for them to accept Discover cards, American Express does not do so. This means that merchants that accept American Express cards cannot discount for the use of a Discover Card, Visa or MasterCard. Additionally, a merchant class action lawsuit against the card companies has generated a proposed settlement that would allow both surcharging and discounting, although there are several hurdles before it would be implemented.

²⁰ Note in this case, that market power exists despite possibly intense inter-platform competition. The result is known in the literature as a competitive bottleneck (see Armstrong, 2006).

merchants prices are not affected by the interchange fee, so that the interchange fee has no effect on the number of transactions. This can be undesirable, since the interchange fee in a no-surcharge setting can raise welfare. Also, the interchange fee has been an important element in moving the economy to a digital payments system that is more efficient than the one it replaced, so that undercutting the interchange fee is not necessarily desirable. In contrast, a problem with a no-surcharge rule is that the cost of card acceptance must be collected through regular prices, which are paid by both card and non-card users. Thus, some commentators have developed models under which card users impose an externality on non-card users through higher prices. According to this theory, as card users tend to be wealthier than non-card users, this externality raises potential questions about the effect of interchange fees on inequality (Schuh *et al.* 2010).

Section 4. Welfare maximizing interchange fees

In an article that was well ahead of its time, Baxter (1983) provided the first formal modeling of interchange fees. He provided a theory of welfare-maximizing interchange fees in which there is perfect competition of issuers and acquirers so both obtain zero profits.²¹ In a simple theoretical setting in which there are just two payment alternatives (say cards and cash) that can be used to complete a single transaction, he notes that efficient card transactions arise when consumers use cards if and only if the joint convenience benefits from doing so (i.e., theirs and the merchant's) exceed the joint costs to the corresponding issuer and acquirer from providing the service. However, with interchange fees set at zero and absent differential pricing by the merchant (or more generally, Coaseian negotiations between the cardholder and the merchant), the cardholder will face a fee equal to the issuer's cost and the merchant will face a fee based on the acquirer's cost. This will not, except by coincidence, lead to the efficient usage of cards. An externality will arise.

Suppose instead the interchange fee paid to issuers from acquirers is set equal to the merchant's convenience benefit of accepting cards, less the acquirers' cost of providing the acquiring service to the merchant. This is known as the Baxter interchange fee. Assuming perfect competition of issuers and acquirers, this interchange fee will be passed back to cardholders (e.g. through rebates or reduced fees). Consumers will therefore face all the costs of cards (issuer's and acquirer's) and all the benefits (their own and that of the merchant). Their decisions about whether to use cards will fully internalize the associated externality.

This simple logic underpins most of the subsequent analysis of welfare-maximizing interchange fees. It provides a simple yet powerful benchmark. If each merchant faces the Baxter interchange fee, then by construction, the associated merchant fee would be set at the level to leave the merchant indifferent *ex-post* (i.e., at the point of sale) between whether consumers use cards or the relevant alternative (e.g. cash). The latter principle is known as the "merchant indifference criterion" (Farrell, 2006) or the "tourist test" (Rochet and Tirole,

²¹ As Schmalensee (2002) notes, Baxter does not provide a positive theory of interchange fees. Given that issuers and acquirer are assumed to make no profit, the level of the interchange fees is not determined in his theory, which is why we did not discuss the Baxter theory until this point.

2011)).²² These alternative principles are equivalent to the Baxter interchange fee when the acquiring margin is zero, and in the benchmark case considered by Baxter, all correspond to the welfare maximizing interchange fee.

When acquirers obtain a positive margin, which they presumably must do if they are to cover their fixed costs, the merchant indifference criterion or equivalent tourist test imply an interchange fee benchmark that is set too low. In particular, they imply an interchange fee below the Baxter interchange fee, which remains equal to the welfare-maximizing interchange fee given that the fee cardholders face should still be set so as to induce them to internalize the merchants' convenience benefit of accepting cards and the associated costs of acquiring. On the other hand, with positive (constant) issuer margins but zero acquiring margins, the different approaches all imply an interchange fee benchmark which is too low.²³ That is, the Baxter interchange fee is now below the welfare-maximizing interchange fee (see Rochet and Tirole, 2011). The divergence arises because the welfare maximizing interchange fee needs to be set to offset the distortion arising from positive issuing margins, thereby ensuring that consumers face the joint costs and benefits of the two sides when deciding whether to use cards (but no more), whereas the Baxter interchange fee does not make any such adjustment.²⁴ Instead, the Baxter (or tourist test) interchange fee corresponds to the level which maximizes what Rochet and Tirole (2011) call total user surplus (i.e., the convenience benefit to each side net of the price they pay issuers and acquirers, respectively), which they show is equivalent to consumer surplus under some conditions. The arguments for using a consumer surplus or total welfare standard will be discussed later when discussing different regulatory approaches to interchange fees. In this section the focus is on what the literature has to say with respect to the usual objective in economic analysis: overall efficiency (i.e., maximizing total welfare).

Consider now what happens when merchants differ in their convenience benefits of accepting cards, so that some will accept cards and others will not, at the same interchange fee. Assuming these differences cannot be observed by the social planner, the objective in setting an interchange fee will no longer be just to get consumers to use cards when doing so generates social surplus. The objective will also be to induce merchants to accept cards when doing so generates social surplus. The welfare-maximizing interchange fee involves a trade-off between getting the right price signal for consumers and getting the right price signal for merchants. The right price for consumers is such that they use cards whenever the sum of their convenience benefits from using cards and the average convenience benefits from merchants they purchase from with cards exceeds joint costs. The right price for merchants is such that they accept cards whenever the sum of their convenience benefits from accepting cards and the average convenience benefits of their card-paying

²² The tourist test asks: does the interchange fee lead to a merchant fee that would induce the merchant to turn down the card for a tourist who is known to have the cash and is already at the store, assuming that the merchant has this discretion and will not see the tourist again? The highest interchange fee that does not lead the merchant to turn down the cards in such circumstances is the tourist test interchange fee.

²³ By constant margins, we mean that margins do not change with a change in the interchange fee, which means interchange fees are fully passed-through.

²⁴ When socially optimal interchange fees are above the level implied by the merchant indifference test, we are assuming cards will still be accepted by merchants, which will not be an issue if merchant internalization is present so that merchants accept cards, in part, to attract more business, as discussed in Section 2.

customers exceeds joint costs. Except for very special cases, these goals are conflicting and a single instrument (the interchange fee) cannot achieve both conditions.

One case where both goals can be met simultaneously is when issuing and acquiring margins are zero, and merchant internalization holds. With merchant internalization, merchants already internalize cardholder's surplus, so there is no need to correct for an externality on that side. Thus, the focus is as before, on getting consumers to face the full costs and benefits of their decision to use cards. Since the resulting externality applies to all merchants that a typical cardholder deals with, it is the average of these merchants' convenience benefits that should be used in the Baxter interchange fee (Wright, 2003b). This will ensure the theoretically most efficient price is set on each side.

Rochet and Tirole (2011) note that this "average" Baxter interchange fee implies that the average merchant (among those that accept cards) is *ex-post* indifferent as to the means of payment chosen by the consumer, so that some merchants will want to reject cards *ex-post* and some will strictly prefer to accept them. Efficiency cannot require that the tourist test be met by all participating merchants since cardholders must internalize the surplus of the average merchant they deal with. Moreover, the average Baxter interchange fee (or tourist test) retains relevance even if the social planner can identify different merchants by their different convenience benefits and set differential interchange fees accordingly. Provided consumers face card fees and rewards or other benefits which do not depend on a particular merchant's identity, then they cannot be induced to exactly internalize the welfare generated by their card usage for each merchant but rather can only do so for the average merchant that accepts cards.

Beyond such special cases, welfare maximization requires a trade-off between getting the price right on each side. Thus, for instance, in the presence of positive issuing and acquiring markups, the welfare maximizing interchange fee will be complicated by the need to try to offset these markups. Since the interchange fee acts as a transfer between the two sides, obviously reducing one markup will exacerbate the other. When issuing margins are higher than acquiring, as is sometimes proposed (see Katz, 2005, at p.132 and Tirole, 2011, at p.143), then issuer margins suggest that the welfare-maximizing interchange fee will be above the average Baxter level discussed above.

Calculating the socially optimal interchange fee will be further complicated by the fact that issuers face costs that may arise per cardholder as well as per transaction, and by the fact that consumers make joining as well as usage decisions, both of which will be affected by card fees (and so interchange fees). This introduces the potential for an additional trade-off in determining the welfare-maximizing interchange fee: between getting enough consumers to hold the payment card in the first place, and preventing excessive usage of cards by those holding them. We will return to this theme when discussing possible regulatory benchmarks for interchange fees in Section 6.

Like the positive theory of interchange fees, the normative theory of interchange fees has not been well developed for the case in which some, but not all, merchants add surcharges for card payments. Surcharging is an important consideration given in some jurisdictions where interchange fees are regulated (e.g. Australia), merchants are also able to surcharge. Interchange fees based on the tourist test should remove any incentive for the average

merchant to surcharge since they will imply merchants are, on average, indifferent about whether cards are used or not (from an *ex-post* point of view). However, merchants with particularly low convenience benefits of accepting cards may prefer to add a surcharge for payment by cards and those with particularly high convenience benefits of accepting cards may prefer to offer a discount for payment by cards. Whether allowing for such surcharges and discounts changes the calculus of welfare-maximizing interchange fees remains to be studied, especially in more realistic environments in which some merchants will not surcharge even if they can, and others take advantage of being able to surcharge to add on additional fees.

Section 5. Market failure?

In an open Board Meeting of the Federal Reserve on the regulation of interchange fees, Federal Reserve Chairman Ben Bernanke raised the following pertinent question: “*There’s a presumption that prices will be set by market competition, generally, but then, of course there are counter examples such as electric utilities, for example, where the government intervention can be justified ... for various reasons. Can you ... help us thin[k] about ... what are the arguments for and against allowing interchange fees to be determined in the market versus having a regulatory intervention when we think about the economics?*”²⁵ This section addresses one aspect of this question, which is whether there is any particular market failure arising from the private determination of interchange fees. Thus, we start by comparing privately and socially optimal interchange fees to see whether there is any systematic distortion arising between them. To be clear, the existence of a distortion does not imply that card systems are acting anticompetitively. These distortions are likely to arise whether the card platform is open or closed, or whether there is only one card platform in the market or many. Rather, any policy response should be under the rubric of regulation. Section 6 will focus on analyzing possible regulatory interventions.

The economics literature identifies a substantial set of distortions in the card market that might drive the market away from a socially optimal performance. However, these might be difficult to act upon in a regulatory context. In some contexts, privately set interchange fees will be too high, but under equally plausible conditions they will be too low. Indeed under some specific settings, noted below, the privately and socially optimal interchange fees exactly coincide. Recent research has highlighted some possible sources of distortions, which provide reasons for why privately set interchange fees may be set too high, but we still have little empirical research evaluating the relative importance of these different theories.

In explaining these claimed distortions in the card industry, it is worth keeping in mind that they depend on specific sets of simplifying assumptions. Even with these simplifying assumptions, the biases that emerge are subtle in nature compared to what one would normally consider sufficient grounds for regulatory intervention. Indeed, reflecting this, it is sometimes a challenge to articulate the rationale behind the biases that result in simple terms, as will become clear in what follows.

²⁵ See, Federal Reserve Board of Governors Holds an Open Meeting, CQ Financial Transcripts, December 16, 2010, pp. 8-9.

Privately determined interchange fees may be efficient

It is instructive to first consider three different sets of assumptions that result in privately and socially optimal interchange fees that *exactly* correspond. The list is not meant to be exhaustive. Other combinations of assumptions are likely to lead to the same conclusion. The list is however instructive, since it illustrates the point that a diverse range of reasonable assumptions can lead to privately set interchange fees to be set efficiently or without any systematic bias. This is quite a different situation from some other settings (e.g. the unregulated pricing of electricity distribution or transmission). While we do not believe that any set of these assumptions holds exactly in practice, these results make the point that whatever bias exists could, in general, be in either direction.

Schmalensee (2002) established that privately and socially optimal interchange fees coincide for an open card platform represented by a monopoly issuer and a monopoly acquirer when quasi-demands are linear and merchant internalization does not arise. In this case, the balancing exercise of the card platform is exactly the same as that faced by a social planner maximizing welfare, with only the differences in costs and demand elasticities across the two sides affecting the optimal level of the interchange fee. Wright (2004) derived the same equivalence between privately and socially optimal interchange fees under similar assumptions but allowing for competition between issuers and acquirers, provided their margins do not change with interchange fees.

In contrast, Rochet and Tirole (2002) obtained the equivalence of the privately and socially optimal interchange fees in a situation in which issuing is imperfectly competitive, issuers do not pass back interchange fees fully so margins increase in the interchange fee, and acquiring is perfectly competitive. In their setting, due to homogenous merchants, interchange fees are set to the maximum level at which merchants still accept cards. If this implies card fees that are still above the efficient level (i.e., due to a lack of issuing competition), a lower card fee will be required. Since a lower card fee requires a higher interchange fee, this is not possible without causing merchants to stop accepting cards. The card platform's preferred interchange fee can therefore coincide with the (constrained) socially optimal level.

The final setting we note in which the same equivalence result can be obtained is one in which there are competing platforms, and consumers all hold cards from both platforms. Even in the presence of merchant internalization, this will enable merchants to drop any card that charges above the level that leaves them indifferent to whether consumers pay with the payment card or the alternative instrument (e.g. cash), knowing consumers will hold the competing platform's card. When issuing and acquiring margins are competed down to zero, this corresponds to the welfare-maximizing level of interchange fees (Guthrie and Wright, 2007). Thus, the equilibrium (competitive) interchange fee exactly equals the fee that maximizes welfare (i.e., the Baxter interchange fee, or tourist test interchange fee, in this setting).

Despite the consensus that there is no clear or obvious distortion in the private determination of interchange fees, public authorities such as the Reserve Bank of Australia, the European Commission and the United States Government Accountability Office,

together with a number of economists (e.g. Carlton and Frankel, 1995; Katz, 2001; Cabral, 2005; Vickers, 2005; Farrell, 2006), have argued that the fee structure in debit and credit cards are likely to be distorted, with merchants paying too much to accept payment cards and cardholders paying too little compared to the efficient fee structure, resulting in excessive usage of payment cards by consumers, a cost that is ultimately passed on to consumers paying by cash. We turn now to explain three settings where such a bias arises.

Revenue-shifting hypothesis

Schmalensee (2002) first pointed out that if the interests of issuers and acquirers are not weighted equally in the determination of interchange fees, the privately determined interchange fee will be affected by the desire to shift profit from one side of the system to the other. He focuses on the case of a monopoly issuer and monopoly acquirer, but the issuer's profit is given more weight in determining the platform's interchange fee, thereby leading the platform to prefer a higher interchange fee than otherwise would be the case. All else equal, this profit-shifting motive leads to a bias towards excessive interchange fees compared to the socially optimal interchange fee.

In a similar vein, Wright (2004) considers the case in which a higher interchange fee can shift revenues to the side within a card system where there is less competition, which reduces the extent to which revenues are dissipated through competition. Thus, in his setting, the profit of each side is given equal weight but interchange fees are not passed through equally by issuers and acquirers. Katz (2005, p.132) writes²⁶ "*Industry wisdom is ... that acquirers generally pass through a higher percentage of fee changes to their customers than do issuers.*" If this is true, then a card platform that increases interchange fees will increase merchant fees more than it will reduce card fees (or increase rewards and other cardholder benefits). Assuming the higher interchange fee revenues are not competed away in other ways (e.g. higher investment on the issuing side, or greater marketing expenses), this will enable the card platform to increase the total margin obtained by issuers and acquirers collectively, albeit at the cost of reduced card transactions. Without this asymmetry in pass-through rates, the card platform would maximize the profit of issuers and acquirers by setting an interchange fee to maximize the volume of card transactions. Under some conditions (e.g. in the absence of merchant internalization), maximizing the volume of card transactions leads to the economically efficient outcome. Thus, in such a case, asymmetric pass-through rates result in excessive interchange fees.

Compared to the other biases that will be discussed, the revenue-shifting hypothesis is the closest to a traditional market power concern. Note that the hypothesis says that privately set interchange fees may reduce card transactions compared to the outcome if the interchange fee is set to maximize the number of transactions. But note privately determined interchange fees may still expand card transactions compared to other benchmarks (e.g. zero interchange fees, bilaterally determined interchange fees, or even socially optimal interchange fees in case this differs from the volume-maximizing level of interchange fees).

²⁶ One possible reason for this, despite the fact acquiring is generally more concentrated than issuing, is that acquiring services are relatively homogenous and merchants (especially large ones) negotiate contracts with acquirers known as "interchange fee plus contracts" in which interchange is fully passed on to the merchant.

The possibility of asymmetric pass-through is an empirically testable hypothesis. In an empirical study of issuing and acquiring in Europe, the European Commission (2006) indeed found a pass-through rate of 0.4 for acquiring versus -0.25 for issuing (based on their preferred fixed-effects specification). The rather low rate of pass-through for acquiring, among other unusual findings, casts doubt over the empirical analysis in their study. Studies of the regulatory experiment in Australia in 2003, in which interchange fees were reduced by 40 basis points in 2003, suggest a roughly 100% pass-through into lower merchant fees, but that cardholder fees (net of rewards and interest-free benefits) have increased by a more modest amount (Chang *et al.*, 2005). However, neither empirical exercise properly controls for other influences on fees. As such, neither provides a proper empirical test of the shifting-revenues hypothesis.

Future empirical research could usefully be directed towards such a study. Similar issues have been analyzed in the mobile-phone termination literature, in which the effect of regulated reductions in termination charges on retail prices for mobile subscribers has been studied (in that literature, the effect that lower mobile termination charges will tend to increase the prices faced by mobile subscribers is called the "waterbed effect"). Genakos and Valletti (2009) provide an empirical analysis of the waterbed effect.

Here we note some of the issues a proper study of the shifting-revenues hypothesis should account for. While it is interesting to know how pass-through rates differ across the two sides, this is not the only determinant of whether banks collectively can benefit from shifting revenues. Issuing margins may be competed away in ways other than lower card fees, such as improved service, innovation and increased marketing efforts. What really matters is the effect of a change in interchange fees on the profit margins of each side. While harder to measure, this would provide a more robust test of the shifting-revenues hypothesis. The period of time in which any change in margins is measured is also likely to be important. In Australia, some banks increased their annual card fees in the months prior to the widely anticipated RBA-imposed decrease in credit card interchange fees in 2003, possibly reflecting that consumers pay these fees in advance through one-year contracts. Other changes, such as changes to rewards, innovations and marketing are likely to be felt only gradually over time. Other important econometric issues that arise include how to address the endogenous determination of interchange fees, which, apart from exogenous regulatory changes, may be affected by the same factors causing bank fees and margins to change, and how to properly control for other factors affecting bank fees and margins.

According to the revenue-shifting hypothesis, a modest regulated reduction in interchange fees should lead to an increase in card transactions.²⁷ While consumers will use their cards less at each merchant they buy from, more merchants will accept cards, and the latter effect will dominate. This does not seem very plausible given merchant acceptance rates are already high, especially among large retailers, which is an argument against the revenue-shifting hypothesis. The revenue-shifting hypothesis would require there be enough merchants (possibly in new merchant categories) that are induced to accept cards by the reduction in interchange fees.

²⁷ Recall, under the hypothesis, that interchange fees are set above the level which maximizes card transactions (in order to raise the total margin on each card transaction).

One can empirically test the revenue-shifting hypothesis by looking at episodes where interchange fees have been reduced by regulation below their privately determined level. The evidence from Australia does not point to any substantial change in card transactions following the RBA's regulated reduction in interchange fees (Hayes, 2010). There are at least three problems with using such evidence to draw conclusions on the existence of a revenue-shifting hypothesis. First, we do not know what would have happened to card transactions in the counter-factual. Second, policies changed other than the reduction in interchange fees (e.g. surcharging was allowed, and some entry rules for banks were modified). Third, the regulated reduction in interchange fees was large, so it is possible that the interchange fee shifted from one that was above the volume-maximizing level to one that is now below the volume-maximizing level. Thus, the volume of card transactions (output) could remain roughly constant in the face of the regulated change even though at the unregulated level, interchange fees were set too high in order to shift revenues to the issuing side.

Merchant internalization

Rochet and Tirole (2002) were the first to show formally that merchant internalization could result in merchants paying too much to accept cards, lowering their resistance to paying high merchant fees, and resulting in overprovision of cards due to interchange fees being set too high. In their model, with homogenous merchants, this happens when issuing margins are not too high, so that there is no need for efficient interchange fees to be set very high to offset these. They write (p. 559) "*a low merchant resistance is the worst case scenario for the social optimality of an issuer-determined interchange fee.*"

Allowing for heterogeneous merchants, Rochet and Tirole (2011, p. 485) conclude in contrast that "*... when the merchant homogeneity assumption is relaxed, the price structure chosen by a monopoly platform, in the absence of a regulation, is no longer systematically biased in favor of cardholders.*" Intuitively, with constant margins, issuers want to maximize volume which requires appealing to the marginal cardholder and merchant (the surpluses of the average cardholder and merchant are ignored). Any bias induced by a private choice of interchange fees then depends on the relative surplus of the two groups of users (i.e., that of the average user versus the marginal user).²⁸ However, as Wright (2012) has recently shown, these results do not exploit the full implications of merchant internalization. Merchant internalization does still create a systematic bias towards excessive interchange fees when merchants are heterogeneous, although the mechanism is rather subtle.

The theory for why interchange fees might be excessive is easiest to understand in the case of a monopoly platform that is able to perfectly price discriminate across the two sides. Perfect price discrimination implies that, in order to capture maximum profit, the platform fully takes into account users' average willingness to pay on each side. Merchant internalization implies the amount each merchant is willing to pay to accept cards takes into account the average surplus its customers expect to get from using cards. Provided consumers face the same retail price for goods regardless of how they pay, this surplus

²⁸ Wright (2004) reached the same conclusion. Weyl (2010) focuses on the difference between the surplus of marginal and average users in explaining distortions in pricing, which he calls Spence distortions.

determines what consumers are willing to pay to hold the card in the first place. Thus, the consumers' surplus from card usage gets counted twice, once when the platform extracts surplus from each merchant that accepts cards, and once when it extracts surplus from the consumers who hold cards.

For a platform that cannot price discriminate on either side, the logic behind the bias is more subtle. The hypothetical case in which cardholders and merchants draw symmetric convenience benefits of cards provides the clearest intuition. The symmetric nature of demand from the two sides means welfare will be maximized by an interchange fee that equalizes the benefits of the marginal user on each side. Since merchants internalize the benefits of cardholders in their card acceptance decision, this requires merchants be charged more than cardholders. However, a monopoly platform will want to shift even more charges from cardholders to merchants, beyond the point where the benefits of the marginal users are equalized. This reflects that the merchants' participation on the card platform is also relatively insensitive to a change in the interchange fee, since a higher interchange fee raises the fees charged to merchants but it also increases the cardholders' rewards or other benefits, which the merchant takes into account due to merchant internalization. Wright (2012) shows the result holds more generally, allowing for an arbitrary asymmetry in the distribution of convenience benefits of using cards across cardholders and merchants. Thus, the bias towards excessive interchange fees does not depend on the particular distribution of cardholder and merchant preferences for using cards.

The result, however, does rely on some key assumptions, most importantly that merchant internalization is sufficiently strong and, for the general case (without price discrimination), that the resulting quasi-demand function for card acceptance by merchants is not too convex (namely, it is log-concave). These assumptions are not easy to verify empirically. However, future research might consider empirically establishing the shape of the quasi-demand function for card acceptance by merchants in order to test the log-concavity property, as well as to empirically validate the extent of merchant internalization.

Asymmetric choice between consumers and merchants

Bedre-Defolie and Calvano (2013) emphasize a different asymmetry between consumers and merchants in explaining a bias towards cardholders in the setting of interchange fees. They shut down the merchant internalization channel. Rather, they focus on the fact consumers make two decisions (whether to hold cards and then whether to use them) whereas merchants only make one (whether to accept them). Thus, like the previous papers, they assume that merchants cannot steer consumers towards using particular payment instruments (say through surcharging, discounting, promotions or otherwise) and that merchants must make a once and for all decision to accept all cards from a particular platform or not.²⁹ Their insight is that the fact consumers make two decisions means if the

²⁹ That decisions are delegated to consumers reflects that consumers have benefits of using cards that vary across their different transactions whereas a merchant's benefit of accepting cards is likely to be reasonably constant across its transactions.

card platform (or its issuers) can set fixed fees to consumers in addition to usage fees (or rewards), it will be able to extract more cardholder surplus (i.e. some of the option value of using cards), which shifts the tradeoff in favor of maximizing cardholder surplus. In their benchmark setting in which they assume a monopoly issuer (which is assumed to determine the interchange fee since acquirers are perfectly competitive and make no profit), they show maximizing the profit of the issuer amounts to maximizing consumers' surplus from using cards, which results in an inefficiently high interchange fee that is biased towards cardholders.

As Calvano (2011) notes, the key to the result of Bedre-Defolie and Calvano (2013) is the asymmetric nature of price discrimination. Their benchmark setup implies perfect price discrimination on one side (to cardholders) and no price discrimination to the other (merchants). In reality, price discrimination possibilities are likely to be significant on both sides. Indeed, Calvano notes there is actually substantial price discrimination on the merchant side, with MasterCard and Visa having dozens of different interchange fee categories in the United States depending on the nature of the retail business (e.g. airlines, gas stations, supermarkets etc., some of which are broken down further into the size of the business). Moreover, issuers are unlikely to be able to perfectly extract all consumers' surplus from usage through two-part tariffs (or other instruments). In this regard, it is instructive to see what happened to the range of issuers' plans in Australia following the regulation of interchange fees there. According to the two-part tariffs in Bedre-Defolie and Calvano's theory, following a substantial decrease in interchange fees, rewards should have fallen (or usage fees increased) one-for-one, and fixed fees reduced somewhat to reflect the lower consumer surplus from using cards. Instead, what happened is that benefits decreased by much less than one-for-one, and instead, the main effect was that the issuers' fixed fees actually *increased* (Chang *et al.*, 2005). This outcome suggests issuers' pricing behaved very differently from that predicted by Bedre-Defolie and Calvano's theory. Whether the extent of price discrimination is greater on one side or another is an open question. Bedre-Defolie and Calvano's theory helps point out how an asymmetry in the extent of price discrimination can matter, and it remains for empirical research to help determine whether the predictions of the theory play out.

Bedre-Defolie and Calvano (2013) provide a new theory of asymmetric internalization different from the merchant internalization story in the previous literature. One would imagine that allowing for both types of asymmetric internalization would reinforce their results, although that too remains an open question.

Section 6. Regulation

Regulators and competition authorities have intervened in the setting of interchange fees in various jurisdictions, including in the United States, the European Union, and Australia. Policy-makers have argued that interchange fees are too high based on any reasonable cost measure. They have also pointed to the rewards that cardholders often get for using payment cards as a sign that interchange fees must be too high. Major merchants complain that they would like to shift consumers to use other payment instruments due to the high cost of card acceptance but cannot (i.e., the must-take argument). Finally, policy-makers

have raised a concern that high interchange fees (and so merchant fees) act as a tax on users of other payment instruments such as cash, by raising merchants' prices. The European Commission published a Green Paper (European Commission, 2012), that contemplates a number of directions for future regulatory intervention, such as requiring merchants and card networks to disclose fees related to interchange, considering whether surcharging might create incentives for consumers to use more efficient payment means, separating the payment processing market from the card market, and increasing access to card systems by non-bank entities.³⁰

So what is the economic case for regulatory intervention? The usual approach would be to argue there is a serious market failure, significant enough that the costs of any intervention would be warranted, and that a regulatory solution exists that is likely to increase the desired objective (say total welfare). With this in mind, in this section we address the extent to which the existing economic literature provides reasonable grounds for regulatory intervention, and if interchange fees are to be regulated, how should this best be done?

Is intervention warranted?

Based on the existing economic theory of interchange fees, there is little reason to think that the private market always obtains the social optimum; and there is a basis for arguing that it might be systematically distorted towards excessive interchange fees. Furthermore, further entry or increased competition may not reduce any such distortion (indeed, intersystem competition could well result in greater upward pressure on interchange). In addition, there is no simple behavioral remedy, such as making interchange fees more transparent to cardholders and merchants, or banning a contractual restriction such the “no-surcharge” requirement that clearly addresses any such distortion. However, in order to support regulatory intervention, we require more than these conditions. The cost of regulatory intervention must not outweigh the benefits, and these costs include not only administrative costs but also the potential for mistakes in fee setting, as well as unintended implications for issues such as investment and innovation. Section 5 showed that any bias in privately set interchange fees is somewhat subtle. While some market failures have been proposed in the literature, they rest on fairly strong assumptions. Furthermore, these theories have not yet been empirically tested. Thus, in our view, regulatory intervention can easily go awry.

In looking for an underlying reason for privately set interchange fees to be excessive, the most promising candidate appears to come from the tendency of merchants to take into account the surplus their customers obtain from using cards when deciding whether to accept cards (i.e., merchant internalization). This is likely to be important for many different types of merchants and, as argued in the previous sections, can result in excessive interchange fees in quite a wide range of settings.³¹ Merchant internalization implies that merchants will accept cards even if the cost savings from doing so are less than the merchant fees they pay. Thus, it explains the must-take perspective put forward by merchants. It also implies there can be excessive use of cards at those merchants

³⁰ The European Commission has the objective of common market integration in addition to its objective of economic efficiency. As a result, the Green Paper also discusses issues such as requiring interchange fees to be the same across countries and how to encourage cross-border acquiring, which are viewed to contribute to integration, and may or may not enhance efficiency. We do not explore such issues in this paper.

³¹ Likewise, for closed card platforms, merchant internalization tends to imply merchant fees will be set excessively high.

accepting cards. However, the mechanism by which merchant internalization leads to a distortion in interchange fees is generally quite subtle, and is especially so in the realistic case in which merchants are heterogeneous. Regulatory intervention based on such a subtle source of possible market failure is without obvious precedent and this makes it controversial. While there are a number of economic settings in Industrial Organization in which fairly subtle market failures arise, few, if any, cause regulators to intervene.³²

Both the potential market failure arising from merchant internalization and that put forward by Bedre-Defolie and Calvano (2013) (i.e., based on the asymmetric choice between consumers and merchants), rely on a failure of card platforms to properly internalize the effects of their fee structure on the average merchant that accepts cards and the average consumer that uses cards, resulting in the interchange fee being set too high. The problem is that a card platform will focus on marginal rather than inframarginal users. However, this is a standard problem in industrial economics that, by itself, does not normally warrant regulation. In many economic settings, firms do not account for average or inframarginal users.³³ Furthermore, in a two-sided market setting, the point is even more subtle, since it is only the asymmetry in the extent to which the platforms fail to take into account the surplus of average users across the two sides, that drives the distortion in interchange fees.

Even if a substantial market failure can be established, regulating interchange fees for open card platforms and leaving closed card platforms unregulated risks tilting the industry's business model towards the closed model. As Tirole (2011, p.149) notes "*Whatever regulation (or lack thereof) one advocates, neutrality with respect to business organization should be the rule, so as to let the most efficient organizational forms emerge.*" It is notable that in the wake of the mandated decrease in the Australian interchange fee, three of the top four Australian banks signed up agreements to issue American Express or Diners Club cards (Tirole, 2011). Chan *et al.* (2012) also report that "*there has been a noticeable increase in the combined market share of the American Express and Diners Club schemes*" since banks began issuing American Express companion cards in late 2009. Thus, if regulatory intervention were warranted, it might be preferable to regulate merchant fees rather than the interchange fee, since this could be equally applied to all card systems. However, merchant fees are technically much harder to regulate, since they are set by hundreds of banks or, in many cases, processors that serve as intermediaries between banks and merchants; and such competitive end-service pricing can include a number of additional, bundled or specialized services depending on the particular merchant, processor or acquirer.

A proposed alternative to regulating interchange fees is to eliminate the no-surcharge rules that platforms (open or closed) have adopted. We have already discussed the no-surcharge rule in the context of antitrust. If the no-surcharge rule is inefficient but is not an antitrust

³² For instance, it is well known that due to a business-stealing effect, there will be excessive entry by firms in industries that produce fairly homogenous products (Mankiw and Whinston, 1986). This does not result in regulators attempting to limit the number of competitors in such industries.

³³ We know that firms will generally not choose the socially optimal level of quality of their products due to their inability to fully internalize the effects of quality choice on the average consumer they sell to. They instead focus on the effects on the marginal consumer. Yet, there is no widespread regulation of minimum quality standards except in certain industries (e.g. restaurants, hospitals etc) where quality cannot be easily observed by consumers, so there may be other types of market failures, and there are associated health or safety concerns.

violation, it could be subject to regulation. Relative to interchange fee regulation, this approach preserves competitive neutrality across open and closed platforms. Zenger (2011) establishes the equivalence between perfect surcharging and the Baxter (or tourist-test) interchange fee when merchants are homogenous. As above, the problem with this theory, however, is that in reality, merchants will not surcharge perfectly, with some merchants not surcharging at all and others surcharging excessively in order to extract fees from consumers with high demand, or perhaps with little information. Allowing surcharging is therefore, at best, only going to partially solve the problem. In other words, allowing surcharging may be a relatively weak way to address the concern of inflated interchange fees and may have other unintended consequences.

One wonders how much of the decision to intervene in payment cards has to do with the fact that interchange fees were historically set collectively by bank-owned joint ventures. Had the same fee structures emerged from pure closed card platforms, would regulators and competition authorities have gotten involved as readily as they have? We have the sense that some policy-makers in this area start with a strong prior belief that interchange fees are too high and, if economics cannot provide clear guidance about how to set rates, they might be satisfied with any tool that lowers interchange fees, regardless of its support in the economic literature. We are certainly sympathetic to the frustrations of these policy-makers with regard to the difficulty in extracting clear policy prescriptions from the literature on two-sided markets. It is a challenging literature, and there are many complications. We view the difficulty in determining appropriate fees as an important reason for regulation to proceed cautiously. With these thoughts in mind, determining the optimal fee is the focus of the next subsection.

What regulatory cap to set?

Supposing interchange fees are to be regulated, how should the regulatory cap be set? Existing regulators (such as the RBA in Australia and the Federal Reserve Board in the U.S.) have used narrowly defined issuers' variable costs to determine the cap on allowed interchange fees. The problem with this approach is it is not supported by any economic theory. None of the existing models that work out optimal interchange fees (either those maximizing welfare or consumer surplus) imply interchange fees based on issuers' costs will be optimal, or will indeed increase welfare relative to unregulated fees. This is the consensus reached in surveys of the literature. For instance, Prager et al. (2009, p. 48) write "*More importantly, the economic theory underlying the efficient interchange fee provides no rationale for either a strictly cost-based interchange fee or an interchange fee of zero.*" The main rationale given for these standards then is just that they are measurable objective standards.

In contrast, as discussed in Section 4, the Baxter interchange fee does have some grounding in economic theory. In some standard theoretical settings, it corresponds to the socially optimal interchange fee absent issuer margins. And, with positive issuer margins, it can be justified either as providing a lower bound for the socially optimal interchange fee or being the interchange fee which maximizes total user surplus (or consumer surplus). Rochet and Tirole (2011) have developed a related alternative to the Baxter interchange fee in terms of a merchant indifference test, which they called the tourist-test (as discussed in Section 4). It implies the same interchange fee as Baxter's except that it is lower than

Baxter's to the extent acquirers make positive margins. The European Commission is proposing to use this tourist-test interchange fee to determine the regulatory cap.³⁴

Whether this approach is considered a reasonable benchmark depends on whether one thinks total welfare or consumer surplus is the relevant objective. To the extent that issuers obtain positive margins, a welfare standard would require an interchange fee above the Baxter level in order to offset these margins. Thus, a Baxter interchange fee would be a lower bound to the desired interchange fee. Some policy-makers and economists have criticized a welfare standard arguing it would be equivalent to subsidizing a monopolist in order to reduce its price to cost (see Vickers, 2005 in particular), certainly an unusual government intervention. If one rejects the welfare standard in favor of consumer surplus, then the Baxter interchange fee remains the theoretically relevant standard even in the presence of issuer margins. Note that if regulators adopt a consumer-surplus standard, then in case there are any acquiring margins, these should also not be used to decrease interchange fees (by a symmetric argument), which is one reason why the Baxter interchange fee might be preferred over interchange fees based on merchant indifference.

The relevance of the Baxter interchange fee for regulation stems from the fact it is potentially measurable. It requires working out the avoided cost to merchants of accepting cards (in terms of the cost savings from not having to accept the relevant alternative, say cash, net of any costs the merchant incurs from accepting cards), and subtracting from this the cost the acquirer faces in processing the card transaction. While far from straightforward, this is something that can potentially be calculated from detailed studies of merchants' costs. Indeed, the European Commission claims to be embarking on such studies, with the purpose of constructing a merchant indifference compliant interchange fee. If done properly, such studies might provide a benchmark interchange fee that can be used to guide regulation, or a way of testing whether unregulated interchange fees are excessive or not based on the corresponding theory. However, the assumptions, methodology, data inputs, and potential biases of any particular study would need to be carefully considered.

All of this suggests that more theoretically-oriented research is needed in order to understand how the tourist-test interchange fee or the Baxter interchange fee should be adjusted under various generalizations of assumptions (following the work of Rochet and Tirole, 2011). The aim would be to obtain a better understanding of what one would ideally want to measure in any such standard. Some progress has been made in this direction already. Wright (2003b) shows how the Baxter interchange fee generalizes to merchant heterogeneity, showing that one needs to calculate the average Baxter fee across merchants, but averaging only over those merchants accepting cards. Verdier (2010) looks at how to modify Baxter's interchange fee in the case in which interchange fees affect the incentives of issuers and acquirers to invest in quality-improvement. In addition, more empirical research on the costs of accepting cash and other payment instruments would certainly be of use.

Some issues remain largely unexplored. A major issue in determining merchants' avoided cost from accepting cards is determining what is the relevant alternative for a particular type of card. Rochet and Wright (2010) have analyzed this issue in the context of debit and credit

³⁴ Neelie Kroes (European Commissioner for Competition Policy), speech/09/165, April 2009.

cards, showing how to calculate the Baxter interchange fee (or tourist-test interchange fee) for credit cards separately from debit cards. They extend existing models to allow a separate role for the credit functionality of credit cards. They emphasize the merchants' net avoided cost from not having to provide credit themselves, which can help regulators properly interpret the tourist-test threshold and apply it in the case of credit card interchange fees. The regulatory cap they propose implies credit card interchange fees would be set as a weighted average of the merchants' net avoided cost from not having to accept cash and the merchants' net avoided cost from not having to provide credit, with the weights being the proportion of each type of transaction (ordinary purchases versus purchases where credit is required). Future work might consider a broader range of possible payment instruments, as well as how to calculate the equivalent of Baxter interchange fees in such settings. For instance, existing theory does not cover the case in which the relevant alternative to the open platform's cards is a closed platform's cards. Nor does it cover the case of Internet transactions in case these make use of only payment cards and no other payment instruments. Finally, extending the analysis to handle ad-valorem instead of (or perhaps as well as) fixed per-transaction interchange fees, is an important consideration for the theory of Baxter-like interchange fees.³⁵

Having a better understanding of consumers' substitution patterns may be needed to construct a more complete Baxter-like interchange fee. Some empirical work exists on substitution patterns. For instance, Klee (2008) looks at data from a grocery chain to measure how payment choice (cash, card and check) differ based on transaction size. Koulayev, Rysman, Schuh and Stavins (2015) study both adoption and usage of payment instruments using annual survey data, focusing particularly on substitution patterns away from debit in light of recent regulation.

Another important practical concern in implementing a Baxter interchange fee or tourist-test interchange fee is to determine how to handle fixed and common costs. For instance, suppose the relevant alternative to accepting a payment card for a transaction is considered to be cash. Some costs of cash, such as putting in the first cash register, are likely to be fixed (they don't vary with the number of card versus cash transactions, unless all cash transactions are eliminated altogether). The concept of a merchant indifference test is based on a theory that assumes no fixed costs to merchants of accepting cards or cash, and that the extra costs of one more transaction is constant. It therefore focuses on marginal costs (i.e., the additional costs to the merchant if a cardholder uses cash rather than card for a single transaction). However, in reality there are some lumpy costs (e.g. the cost saving of eliminating one more armored vehicle transfer if cash use decreases enough). These types of costs should presumably still be averaged per transaction and included since they ensure the regulated interchange fee provides the right long-run incentives. On the other hand, including truly fixed costs would distort marginal incentives. Some costs may be difficult to classify. When employees spend time at the end of the day reconciling receipts and filling and emptying cash registers, there may be a tendency to treat this as a fixed cost (since from the manager's point of view, if employees don't spend time doing this they may otherwise be idle). However, such costs are better thought of as variable. Over the longer run, the manager will likely be able to reoptimize their allocation of time if he is given

³⁵ Shy and Wang (2011) provide both a positive and normative theory of why a platform should use an ad-valorem fee rather than a fixed per-transaction fee, although their analysis assumes the neutrality of such a fee.

an incentive to do so, productively using that time for other purposes, or reducing their hours and cutting the firm's wage bill.

These examples illustrate the need for a thoughtful study of merchant's costs, guided by the need to calculate the correct Baxter-like interchange fee based on economic theory. While there are a number of existing works in the literature that provide calculations of the relative costs to merchants of accepting different payment instruments (see Layne-Farrar, 2011 for a recent study and references therein), to the best of our knowledge, these are all based on surveys of merchants, such as those commonly cited by the Food Marketing Institute (1998).³⁶ These surveys are likely to suffer from two common flaws, both of which will tend to lead them to understate the true cost to merchants of accepting cash.

First, these surveys are often conducted only with large retail firms, firms that will tend to have lower costs of accepting cash due to the economies of scale involved with cash handling (as opposed to cards, which are fairly scale-invariant). Second, asking retail managers to report costs is unlikely to give unbiased estimates of the costs of different instruments. This is true for two reasons. One is that merchants, especially large merchants, may have an interest in stating relatively low costs of cash or high costs of cards if they think this will help bring down the fees charged to them for accepting cards. The other reason why the cost of cash may be understated, and the most important reason in our view, is that many of the costs of cash (such as having employees counting cash at the end of the day, and making sure accounts are balanced and double checked, as noted above) are not as salient to management as merchant fees. Indeed, many managers will likely not treat these as costs given that they perceive there to be zero opportunity cost to their employees' time in carrying out these activities. These issues suggest that better estimates of the true costs of handling cash (or other instruments) may be obtained only by finding out directly how much time employees spend on various activities, and then imputing costs from these.³⁷

Finally, a useful project would integrate the various forces that we outline in this paper into a unified model. Such a model could be used to compute a socially optimal interchange fee. In our understanding, such an academic study has not been attempted. Doing so would require many parameters to be realistically calibrated. It would be a major undertaking that perhaps would generate more empirical and theoretical questions long before it led to a useful guide for policy-makers on the optimal interchange fee. But such a contribution, and the related academic research, would be a welcome addition to a literature with many elements to it already.

A Government Product?

³⁶ Some studies try to look at the bigger picture and compare the social costs and benefits of different means of payments across all parties (consumers, merchants, issuers and acquirers, government etc). The most comprehensive in this regard is Garcia-Swartz et al. (2006) who find that cash and checks are more costly than earlier studies suggest, and that the increased use of payment cards appears to increase economic welfare.

³⁷ A similar issue arises with respect to the cost to merchants of depositing their cash and having it managed by their bank. While the acquirers' fees for payment cards are very salient to the merchant manager, the implicit costs of having their bank handle cash (e.g. in terms of interest foregone due to non-market rates being offered, or various other cross-subsidies being in place) are not.

Most discussions of policy intervention into card markets focus on the use of antitrust or regulation. However, the government has another option, which is to enter into the market with its own product and to set the interchange fee directly. While it might sound far-fetched, our view is that we are closer to this outcome than is commonly acknowledged. First, note that in China and India this is exactly what has happened, each with their own national payment card (respectively, China UnionPay and RuPay) sponsored by the government. The economic implications of government-sponsored card schemes and policies that are designed to support these schemes over international competitors have yet to be even discussed in the economics literature.³⁸

In the United States, the Federal Reserve Bank is already heavily involved in the cash market and even more so in the checking market, where it clears checks at par for banks. An interesting development in the payments industry is the introduction of decoupled debit cards. A decoupled debit card is a debit card that does not come from the consumer's bank. The card is marketed by a third party, and typically uses the Automated Clearinghouse House to transfer funds from the consumer's bank account to the merchant. The ACH system is owned by the Federal Reserve Bank and clears transactions at par.

Decoupled debit can be interpreted as the government's entry into this market with a card system that uses an interchange fee of zero. It seems feasible for a group of large merchants to take interest in such a card. However, up until now, decoupled debit has not gained a large market share, and it is unclear if it ever will. As we have pointed out earlier in this survey, the effect of increased system competition on interchange fees is theoretically ambiguous. The specific effects of the government setting the interchange fee for one competitor among multiple competitors remains unstudied.

Section 7. Conclusion

This paper surveyed the economics of payment card markets, particularly from the perspective of the economics of two-sided markets. We discussed the rationale for interchange fees in open systems. We contrasted privately set interchange fees in monopoly and competitive environments with socially optimal interchange fees. We found little scope for antitrust prosecution of card systems based on collusion charges. Our discussion of regulation is more complex. We find that numerous distortions may exist, for instance, due to asymmetries in the way issuers and acquirers compete, due to merchant internalization and due to the inability of merchants to perfectly steer consumers to their preferred means of payment. However, the importance of each concern and their collective importance are difficult to evaluate. Thus, we emphasized caution and further research as part of any regulatory initiative.

Our paper highlighted a number of areas in which further research would be both useful and feasible. While there currently may exist research on some of these areas, there is substantial scope for more, and with superior data. We briefly review them here:

³⁸ There is a further issue of determining the objective function of the government. Some rules that governments impose might be interpreted as protectionism.

1) Surcharging: Theoretically, why might card systems restrict surcharging, and what is the effect of doing so? Empirically, how many and which types of firms surcharge and, when they do so, how do their charges reflect costs or other factors? How do their cash prices change with surcharging? What is the impact of surcharging on consumer payment choice?

2) Pass-through: Can we accurately measure pass-through rates of changes in interchange fees? How do they differ among acquirers and issuers, and within specific retail markets?

3) Substitution patterns: If fees increase for consumers for one payment instrument, how do they substitute and to what products? Do they “un-adopt” the instrument, or decrease usage? Does a debit user facing higher fees switch to cash or a credit card? How does this differ across income levels, and how should we evaluate policy intervention in this light? Do merchants respond?

4) Regulatory standards: Extending the work of Rochet and Tirole (2011), theoretically justifiable regulatory standards (e.g. the Baxter interchange fees or merchant-indifference criterion) need to be tested for their optimality in the face of richer market scenarios, some of which we have discussed. In order to better calculate Baxter or merchant-indifference interchange fees that can serve as benchmarks for regulation, it would also be helpful to conduct more detailed and careful studies of the costs to merchants of accepting cards versus cash. We outlined a number of practical issues in this measurement, and biases that exist in the existing approaches, and how some of these can be avoided.

5) Testing theories: The assumptions of existing models of interchange fee determination need to be empirically validated and the predictions of these models tested. For example, merchant internalization is an important assumption to validate. How much of what a merchant pays to accept cards comes from the ability of the merchant to raise its price or quantity sold as a result of card acceptance, or is something else at work?

6) Unintended consequences: Can policy interventions meant to improve pricing in payments markets, such as surcharge or interchange fee regulation, lead to implications for other issues, such as innovation or the success of closed card systems?

6) Optimal interchange fees: So much has been written about how different economic phenomena bias privately-set interchange fees in one direction or another. What would a “grand synthesis” of these models, realistically calibrated, find for an optimal interchange fee? What research questions need to be answered to make such a study credible?

7) Institutional design: What are the implications of the recent shift by MasterCard and Visa to investor-owned companies that set interchange fees unilaterally rather than as joint ventures of banks? What are the objectives of the card systems in setting interchange fees? Absent centrally set interchange fees, what other mechanisms and combination of instruments can still achieve desirable outcomes, if any? What is the role of switch fees?

Overall, the literature on two-sided markets and payment cards is growing, but there is still a great deal to be learned.

References

- Armstrong, M. 2006. "Competition in two-sided markets," *RAND Journal of Economics*, 37: 668-691.
- Balto, D. 2000. "The problem of interchange fees: Costs without benefits?" *European Law Review*, 4: 215-224.
- Baxter, W.P. 1983. "Bank interchange of transactional paper: Legal perspectives," *Journal of Law and Economics*, 26: 541-588.
- Bedre-Defolie, O. and E. Calvano. 2013. "Pricing payment cards," *American Economic Journal-Microeconomics*, 5: 206-231.
- Cabral, L. 2005. "Market power and efficiency in card payment systems: A comment," *Review of Network Economics*, 5: 15-25.
- Calvano, E. 2011. "Note on the economic theory of interchange," mimeo, February, Bocconi University, Department of Economics and IGIER.
- Carlton, D. and A. Frankel. 1995. "The antitrust economics of credit card networks," *Antitrust Law Journal*, 63: 903-915.
- Chakravorti, S. and R. Roson. 2006. "Platform competition in two-sided markets: The case of payment networks," *Review of Network Economics*, 5: 26-44.
- Chakravorti, S. and T. To. 2007. "A theory of credit cards," *International Journal of Industrial Organization*, 25: 583-595.
- Chan, I, S. Chong and S. Mitchell. 2012 "The personal credit card market in Australia: Pricing over the past decade," *Reserve Bank of Australia Bulletin*, March, 55-64.
- Chang, H. and D. Evans. 2000. "The competitive effects of the collective setting of interchange fees by payment card systems," *The Antitrust Bulletin*, Fall, 641.
- Chang, H., D. Evans, and D. Garcia Swartz. 2005. "The effect of regulatory intervention in two-sided markets: An assessment of interchange-fee capping in Australia," *Review of Network Economics*, 4: 328-358.
- Coase, R.H. 1960. "The problem of social cost" *Journal of Law and Economics*, 3: 1-44.
- Ellison, G. 2005. "A model of add-on pricing," *Quarterly Journal of Economics*, CXX, 585-638.
- European Commission. 2006. "Interim Report I. Payment Cards," April 12th.
- European Commission. 2012. "Green paper: Towards an integrated European market for card, internet and mobile payments," COM (2011) 941 final.

- Evans, D. and R. Schmalensee. 1995. "Economic aspects of payment card systems and antitrust policy towards joint ventures," *Antitrust Law Journal*, 63: 861-901.
- Evans, D. and R. Schmalensee. 2005. "The economics of interchange fees and their regulation: An overview" in the conference volume "Interchange fees in credit and debit card industries: What role for public authorities" May 4-6, 2005, An International Payments Policy Conference. Sponsored by the Federal Reserve Bank of Kansas City.
- Farrell, J. 2006. "Efficiency and competition between payment instruments," *Review of Network Economics*, 5: 26-44.
- Food Marketing Institute. 1998. *A Retailer's Guide to Electronic Payment Systems Costs*. Washington, D.C.
- Frankel, A. 1998. "Monopoly and competition in the supply and exchange of money," *Antitrust Law Journal*, 66: 331-361.
- Frankel, A. and A. Shampine. 2006. "The economic effects of interchange fees," *Antitrust Law Journal*, 73: 627-673.
- Gans, J. and S. King. 2003. "The neutrality of interchange fees in payment systems," *Topics in Economic Analysis & Policy*, Volume 3. Article 1.
- Garcia-Swartz, D., R.W. Hahn, and A. Layne-Farrar. 2006. "The move toward a cashless society: Calculating the costs and benefits," *Review of Network Economics*, 5(2). Article 1.
- Genakos, C. and T. Valletti. 2009. "Testing the 'waterbed' effect in mobile telephony," *Journal of the European Economic Association*, 9: 1114–1142.
- Guthrie, G. and J. Wright. 2007. "Competing payment schemes," *Journal of Industrial Economics*, 55: 37-67.
- Hayes, R. 2010. "Is price regulation of payment card associations effective? Evidence from a dramatic policy experiment," mimeo, Melbourne Business School - University of Melbourne.
- Katz, M. 2005. "What do we know about interchange fee and what does it mean for public policy?" *Proceedings of Payments System Research Conference*, Federal Reserve Bank of Kansas City, May.
- Klee, E. 2008. "How people pay: Evidence from grocery store data," *Journal of Monetary Economics*, 55: 526-541.
- Klein, B., A. Lerner, K. Murphy and L. Plache. 2006. "Competition in two-sided markets: The antitrust economics of payment card interchange fees," *Antitrust Law Journal*, 73: 571-626.
- Koulayev, S., M. Rysman, S. Schuh, and J. Stavins. 2015. "Explaining adoption and use of payment instruments by U.S. consumers," Unpublished Manuscript, Boston University.

- Layne-Farrar, A. 2011. "Are debit cards really more costly for merchants? Assessing retailers' costs and benefits of payment instrument acceptance," mimeo.
- Mankiw, G. and M. Whinston. 1986. "Free entry and social efficiency," *RAND Journal of Economics*, 17: 48-58.
- Prager, R., M. Manuszak, E. Kiser and R. Borzekowski. 2009. "Interchange fees and payment card networks: Economics, industry developments, and policy issues," *Finance and Economics Discussion Series*, Board of Governors of the Federal Reserve System, 2009-03.
- Rochet, J.-C. and J. Tirole. 2002. "Cooperation among competitors: Some economics of payment card associations," *RAND Journal of Economics*, 33: 549-570.
- Rochet, J.-C and J. Tirole. 2003. "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1: 990-1029.
- Rochet, J.-C. and J. Tirole. 2011. "Must-take cards: Merchant discounts and avoided costs," *Journal of the European Economic Association*, 9: 462-495.
- Rochet, J.-C. and J. Wright. 2010. "Credit card interchange fees," *Journal of Banking and Finance*, 34: 1788-1797.
- Rysman, M. 2007. "An empirical analysis of payment card usage," *Journal of Industrial Economics*, 55: 1-36.
- Schmalensee, R. 2002. "Payment systems and interchange fees," *Journal of Industrial Economics*, 50: 103-122.
- Schuh, S., O. Shy, and J. Stavins. 2010. "Who gains and who loses from credit card payments? Theory and calibrations," *Public Policy Discussion Paper No. 10-3*, Federal Reserve Bank of Boston.
- Shy, O. and Z. Wang. 2011. "Why do payment card networks charge proportional fees," *American Economic Review*, 101: 1575-90.
- Small, J. and J. Wright (2002) "The bilateral negotiation of interchange fees in payment schemes," working paper, University of Auckland.
- Tirole, J. 2011. "Payment card regulation and the use of economic analysis in antitrust," *Competition Policy International*, 7: 137-158.
- Verdier, M. 2010. "Interchange fees and incentives to invest in payment card systems," *International Journal of Industrial Organization*, 28: 539-554.
- Verdier, M. 2011. "Interchange fees in payment card systems: A survey of the literature," *Journal of Economic Surveys*, 25: 273–297.

- Vickers, J. 2005. "Public policy and the invisible price: Competition law, regulation and the interchange fee," *Competition Law Journal*, 4: 5-16.
- Wang, Z. 2010. "Market structure and payment card pricing: What drives the interchange?" *International Journal of Industrial Organization*, 28: 86-98.
- Weiner, S. and J. Wright. 2005. "Interchange fees in various countries: Developments and determinants," *Review of Network Economics*, 4(4), Article 3.
- Weyl, G. 2010. "A price theory of multi-sided platforms," *American Economic Review*, 100: 1642-1672.
- Wright, J. 2003a. "Optimal card payment systems," *European Economic Review*, 47: 587-612.
- Wright, J. 2003b. "Pricing in debit and credit card schemes," *Economics Letters*, 80: 305-309.
- Wright, J. 2004. "Determinants of optimal interchange fees," *Journal of Industrial Economics*, 52: 1-26.
- Wright, J. 2010. "Why do merchants accept payment cards?" *Review of Network Economics*, 9, Article 1.
- Wright, J. 2012. "Why payment card fees are biased against retailers," *RAND Journal of Economics*, 43: 761-780.
- Zenger, H. 2011. "Perfect surcharging and the tourist test interchange fee," *Journal of Banking and Finance*, 35: 2544-2546.