The Bilateral Negotiation of Interchange Fees in Payment Schemes

John Small and Julian Wright*

January, 2002

Abstract

This paper considers the consequences of changing one feature of card payment schemes: the centralised setting of the default interchange fee, a fee banks pay each other for debit and credit card transactions. The level and setting of the interchange fee is currently under investigation in a number of jurisdictions. Policy makers and competition authorities have expressed concern over the apparent "agreement between competitors" that arises when the banks that compete to supply services to cardholders and merchants also determine the interchange fee paid between themselves. Our analysis focuses on a market-based alternative to setting interchange fees centrally that avoids the "agreement between competitors" allegation – the bilateral setting of interchange fees. We show that a move to bilateral negotiation of interchange fees would expose acquiring firms to hold-up by small issuers and lead to an escalation of interchange fees.

Keywords: Payment Systems, Credit Cards, Interchange Fees, Network Effects

* The authors are with the Network Economics Consulting Group (NECG) and the University of Auckland. The authors are grateful to NECG and Visa International for financial support for an earlier version of this paper. However, the views expressed here are strictly those of the authors, as is the responsibility for any errors. Address for Correspondence: Julian Wright, Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand (jk.wright@auckland.ac.nz).

1. Introduction

The setting of interchange fees in open debit and credit card payment systems (such as those offered by MasterCard and Visa) has recently attracted substantial interest from policy makers, banks, and merchants alike.[1] The interchange fee, the fee paid from the merchant's bank (the acquirer) to the cardholder's bank (the issuer) for each card transaction, has been the subject of discussion and debate by regulatory authorities in Australia, Canada, Israel, Poland, South Africa, the U.K., and the U.S., among other jurisdictions.[2] In Australia the central bank has moved to regulate lower interchange fees, claiming high interchange fees have caused excessive use of credit cards (see Reserve Bank of Australia, 2001).

Currently, open payment schemes set interchange fees through a committee comprised at least partially of representatives of the card issuers and acquirers. Since some of these banks compete with each other in respect of the merchant acquiring function, it has been argued that the joint determination of the interchange fee has the effect of controlling charges set for merchant services by competing banks. By setting the interchange fee cooperatively, it is argued, banks are effectively engaged in price-fixing.

The issue of price-fixing arose in the Nabanco case in which National Bancard Corporation (NaBanco) unsuccessfully sued Visa in the early 1980s on the grounds Visa's interchange fee was a *per-se* illegal agreement and violated the rule of reason. More recently, in Australia, the Australian Competition and Consumer Commission (ACCC) commenced legal proceedings alleging that the centralised setting of interchange fees was a price-fix. The ACCC has since dropped the action in light of the proposed regulation of interchange fees by the Reserve Bank of Australia. The European Commission has also investigated the use of interchange fees, releasing a statement of objections to Visa International in relation to its interchange fee on the grounds the fee is a restrictive collective price agreement (European Commission, 2000). The European Commission subsequently adopted a favourable position on Visa's proposed adjustment to its setting of interchange fees in Europe, which will allow Visa to continue setting the default interchange fee in Europe, while introducing a cap on the level of the fees it sets (European Commission, 2001). In Germany, debate has arisen concerning the banking system's intention to adopt a common interchange fee for all debit card payments.

A number of papers, including those by Baxter (1983), Chang and Evans (2000), Gans and King (2001a), Gans and King (2001b), Rochet and Tirole (2000), Schmalensee (2001), Wright (2001a) and Wright (2001b) analyse the optimal level of the interchange fee in card associations. These papers provide

[1] An open payment scheme is an association of financial institutions, each of which offers card services to merchants and/or cardholders. The association members compete against each other for cardholders and merchants, and the association itself competes against other payment systems. The existence of open payment schemes therefore allows both intra-network competition and inter-network competition. In contrast, closed payment schemes, such as American Express and Diner's Club, do not allow for intra-network competition. For a comprehensive account of the development of open and closed payment schemes see Evans and Schmalensee (1999).

[2] The scale and growth rate of payment card transactions is at least partly responsible for this level of interest, with U.S. consumers alone charging almost one trillion U.S. dollars to payment cards in 1998 (Chang and Evans, 2000). In Australia, the growth in the number of credit card payments per capita has exceeded 25 per cent in both 1998 and 1999 (ACCC, 2000). The success of open card associations relative to closed systems (such as those offered by American

some guidance on the economic role of the interchange fee, and its appropriate level, suggesting that primarily the interchange fee is a balancing instrument that attempts to strike the right trade-off between promoting cardholder demand and merchant acceptance of cards. In contrast, Frankel (1998) argues that if issuers and acquirers are allowed to set interchange fees collectively, they will set fees 'too high'. This will have the effect of taxing cash-paying consumers through high merchant service fees which are passed through into retail prices. According to this argument, the proceeds of the tax are used to induce consumers to use cards, through various rebate and reward programs. Balto (2000, p.222) puts forward a similar argument.

Rather than focus on what is the right level of the interchange fee, in this paper we consider the effect of adopting a market-based alternative to centrally set interchange fees. Balto (2000) suggests the interchange fee should be negotiated bilaterally rather than set centrally. The idea is that competition between individual issuers and between individual acquirers will cause interchange fees to be driven down to the underlying cost of issuing. However, given the network nature of a credit card payment system, it is not at all clear how this process of competition would work in practice, let alone whether cost-based interchange fees are a desirable objective.

We consider this decentralised alternative, analysing the bilateral negotiation of interchange fees between individual issuers and acquirers that compete for cardholders and merchants in the retail market. We explore the dynamics of merchant fees and card usage resulting from these bilateral negotiations. In doing so we preserve two rules of open credit-card payment systems.

The first rule is known as the "no-surcharge" rule, and says merchants are not allowed to set higher prices to customers who pay using their debit or credit card. The no-surcharge rule has been studied by Gans and King (2001a), Rochet and Tirole (2000), Schwartz and Vincent (2000), Wright (2000) and Wright (2001b). All these models assume that the elimination of the no-surcharge rule will lead to widespread cost-reflective surcharging. However, evidence from the Netherlands and Sweden suggests that even absent this rule, most merchants will not surcharge cardholders (ITM, 2000 and IMA, 2000). Based on this finding the European Commission formally cleared the use of the no-discrimination rule in the EU (European Commission, 2000).[3] Thus, whether or not the no-surcharge rule is imposed, the assumption that merchants do not surcharge may be a reasonable one to maintain.[4]

The second payment system rule we maintain is the "honour all cards" rule, which prevents merchants (and their banks) that accept one brand of card from refusing to accept cards of the same brand issued by another bank. Thus, if a merchant accepts a Visa credit card issued by bank A, it must also accept all other Visa cards, regardless of whom issues them, and what type they are (gold card, debit, credit etc). This rule contributes to the integrity and ubiquity of a card network. It has the pro-competitive effect

---

Express and Diner's Club) must also be a factor (Visa and MasterCard collectively accounted for 72 percent of the U.S. payment card market in 1998), since the policies of the closed systems are not in dispute (Chang and Evans, 2000).

[3] The no-discrimination rule prevents merchants from providing a discount for cash purchases or a surcharge for card purchases. The no-surcharge rule only prevents merchants from surcharging for card purchases.

[4] Frankel (1998) presents supporting historical evidence of "price convergence" whereby the costs of alternative payment methods, being small relative to transaction sizes, are not reflected in merchant pricing.

of ensuring that no issuer can be foreclosed from the network by the unilateral or concerted actions of acquirers. Even very small issuers can attract cardholders on terms that are similar to their larger rivals in respect of the acceptability of their cards. Without the honour all cards rule, the value of membership to cardholders would be dramatically reduced, since they could not be sure that their cards would be accepted. The confidence cardholders have in this aspect of the card system is particularly important when they are travelling outside their home region.[5] In our view this rule plays an essential role in achieving efficient outcomes in open payment schemes. For this reason, we maintain the honour all cards rule.

With these rules taken as given, we identify a number of costs that could arise if centralised setting of interchange fees were abandoned in favour of bilateral negotiations. These costs, which include a destabilisation of the network, suggest that the centralised setting of the interchange fee is an important element of an open card network. The costs we identify arise from the following problems:

(1) *Transaction cost problem*: The transaction costs of bilateral setting of interchange fees could be substantial given the vast number of separate issuers and acquirers in any payment scheme.

(2) *Hold-up problem*: Under the honour all cards rule, bilateral bargaining allows banks specialising in card issuing (issuers) to "hold-up" those specialising in merchant services (acquirers) by extracting very high interchange fees during bargaining. Even though this jeopardises the viability of the whole network, banks acting in their own private interests have a strong incentive to engage in such hold-up.

(3) *Escalation problem*: Given the hold-up power of issuers, there will be an escalation of interchange fees such that the payment network will shrink in size, and cardholders and merchants will experience a loss in value. Although every issuer individually manages to secure a very large interchange fee with acquirers, the issuers will not end up being better off in aggregate, and in fact are likely to be worse off as the network becomes imbalanced and implodes. Nevertheless, this does not stop each individual issuer further negotiating higher interchange fees with acquirers, as it takes advantage of its hold-up power to try to secure higher interchange fees than its rivals. Acting individually to maximise its own profit without regard for the other members of the network, each bank ends up contributing to the implosion of the very network that it relies on.

In open payment schemes the centrally set interchange fee plays a balancing role, allocating resources to whichever side of the network requires most promotion, and thereby maximising the number of feasible transactions between cardholders and merchants. Requiring that the fee be set bilaterally would cause other pressures to drive the interchange fee, thus preventing it being used for this purpose. In

[5] ACCC (2000), while highly critical of some aspects of credit card associations' activities, acknowledged the important role of the honour all cards rule in these networks. The honour all cards rule has also been formally accepted by the European Commission. (European Commission, 2000).

particular, we show that in moving from a centrally set interchange fee to the bilateral setting of interchange fees, interchange fees will increase, so that if policy makers believe socially optimal interchange fees are less than privately set interchange fees, the even higher interchange fees that result from bilateral negotiations are likely to further lower welfare. This is not to say that all payment systems require a centralised interchange fee.[6] Rather, given the need for the system to promote the network in a particular fashion, and given the need for the honour all cards rule, bilateral bargaining is likely to lead to inferior outcomes.

Payment systems compete with one another by setting the policies (rules) governing the supply of services to consumers. The rules governing each network therefore form a coherent and interdependent foundation that supports the payments system itself. The unilateral isolation and removal of one of these rules is akin to removing one of the cornerstones of the system. Without further adjustments, there is a risk of serious consequences. Our analysis suggests that when the rule to be removed is the centralised setting of interchange in an open payment card network, this risk is indeed high.

To the best of our knowledge, ours is the first formal analysis of what would happen in an open payment scheme with bilaterally negotiated interchange fees. Baxter (1983, p.577) provides one hypothesis, predicting the network would greatly diminish or collapse without a centrally set interchange fee. His reasoning is that when a single issuer increases its interchange fee, the decrease in demand by merchants only partially impacts on this issuer, since merchants cannot (or will not) price discriminate based on the type of card, and so any decrease in demand by merchants is spread across all card issuers. Facing such inelastic demand, each issuer will set very high fees. He argues the centralised setting of the interchange fee is necessary to avoid such exploitation of the system. Our modelling of the escalation of interchange fees reveals a similar logic in a model where interchange fees are set through bilateral bargaining, competition between banks is imperfect, and customers' and merchants' participation in the system is determined within the model.

The rest of the paper proceeds as follows. Section 2 starts by briefly reviewing the key features of interchange fees in open card payment systems. Section 3 describes exactly how the hold-up problem arises under the honour all cards rule using a formal model of negotiation over interchange fees. Section 4 explains the dynamics of the escalation problem that plagues the bilateral setting of interchange fees using a model of inter-bank competition for issuing and acquiring. Section 5 provides some brief conclusions.

2 A review of interchange in open card networks

There is no universally accepted interpretation of the interchange fee. Visa appears to view it as a tool for balancing and growing the network through the internalisation of network externalities, while MasterCard apparently sees interchange as a fee for services provided by issuers to acquirers (Cruickshank, 2000). While anti-trust concern, and our interest in this paper, is centred exclusively on interchange in

[6] In many countries interchange fees for on-line debit networks, which do not have an honour all cards rule, are negotiated bilaterally between banks.

open networks, it is worth observing that the proprietary (closed) payment card networks also have an implicit interchange fee.

American Express and Diner's Club are examples of proprietary payment card systems based on a closed network structure. This means that a person wishing to hold an American Express card, for example, has no choice about which organisation to arrange this through. A single organisation issues cards, acquires the value resulting from card usage at merchant outlets, and collects the necessary payment from cardholders.[7] Where a closed scheme sets a high merchant service fee and uses this to fund interest free benefits and reward schemes for cardholders, the effect is the same as a positive interchange fee set by open schemes.

In contrast, open payment networks are established through a set of contracts embodying the rules by which the network will operate. Amongst other things, these rules govern the terms and conditions on which cards will be issued, value will be acquired, and settlements will be effected. The members of the open payment schemes set a default interchange fee, although individual members are free to negotiate other bilateral interchange fees if they choose. By the setting of a default rate, members ensure an open network whereby any institution that complies with the standard rules can obtain the centrally set rate for interchange. This approach has the effect of promoting competition between issuers and between acquirers, and thereby of expanding the market for the underlying payment system.

The level of a centrally set interchange fee places a cap on the benefits that issuers can supply to cardholders, assuming that the system covers its own costs. The higher the interchange fee, the more heavily issuers can invest in attracting and retaining cardholders, either through lower cardholder fees or greater cardholder benefits. High interchange fees reduce the advantages of membership to merchants, however, because these fees are recovered from the merchant discounts. The network participants therefore need to set the interchange fee at a level that balances the benefits enjoyed by cardholders against the need to attract and retain merchants.

Based on this rationale for the interchange fee there are compelling reasons for setting the interchange fee at the network level. The card service is jointly produced by its member banks, none of whom could provide the service without the co-operation of others. Because the service is jointly provided and jointly consumed, it is natural to centralise the allocation of the costs of the service. Most of the service costs arise on the issuing side, but many of the benefits flow to merchants. Centralised determination of interchange allows the acquiring side of the network to be allocated some of the costs thereby internalising network externalities. These costs are only loosely related to the cost structure of any given merchant or acquirer since these do not include the cost of promoting the issuing side of the service. Thus, setting interchange at the network level is consistent with promoting balanced network development.

An important advantage of a centrally set interchange fee is that it creates a standardised and open market with minimal transaction fees. If each bank that is a member of the an open scheme has to negotiate separately with every other bank to determine interchange fees, then the cost of the necessary bargaining,

network, since when each individual bank negotiates it does not take into account the network externalities that arise from its decision to negotiate an agreement. For smaller banks, these bargaining costs represent fixed costs, which are likely to act as a barrier to entry. With N issuers and M acquirers in a particular payment system, the number of interchange fees that will need to be negotiated for a single payment system will be N x M.[8] At the international level, bilateral bargaining would require many millions of separate negotiations given Visa alone has more than 21,000 members internationally.

A simplistic solution to the problem of how to set interchange fees, is to rule them out altogether. Along these lines, Frankel (1998) argues that interchange fees should be regulated to zero. With zero interchange fees issuers could not cover their cost of funds except through setting higher cardholder fees or delivering lower cardholder benefits. More importantly, with higher cardholder fees or lower cardholder benefits, customers will not want to hold cards as much, and fewer cardholders will reduce the benefits to merchants of accepting the card (assuming there is a net positive externality running from cardholders to merchants). The net result would be a smaller and less valued network.

When the costs of many separate negotiations are sufficiently high, it is natural for a private organisation to develop a way to coordinate these negotiations. This is especially true when the terms being negotiated fundamentally relate to ensuring maximal value is created for the whole network, rather than any specific member of the network. As will be shown below, the fact that the private interests of members do not coincide with the broader objectives of network growth is the main reason that centralised interchange is required, and that bilateral bargaining is counter productive.

3. Bilateral bargaining and the hold-up problem

In this section we consider what would transpire if bilateral bargaining were mandated, while maintaining the other rules of open card networks discussed above ("honour all cards" and "no surcharge"). The outcome of bilateral bargaining is as much about what happens if two parties cannot agree as it is about what happens if they agree. Under the honour all cards rule, if two banks cannot agree on terms for interchange, the acquirer will not be able to offer merchant services. To see why, suppose that bank X issues cards and bank Y offers an acquiring service to merchants. If bank X and bank Y cannot reach an agreement, then there can be no interchange service between bank X and bank Y, and bank X's cardholders cannot use their cards at the merchants serviced by bank Y. However, if bank X's cardholders cannot use their cards at these merchants, then under honour all cards, bank Y's merchants will not be able to accept *any* cards, in which case bank Y will no longer be able to acquire.

Since any issuer, no matter how small, has the ability to hold-up an acquirer – destroying its business if it does not agree with the issuer's terms, this leaves issuers in a very strong bargaining position.

---

[7] Closed schemes have recently shifted towards a more non-unitary structure, whereby third parties issue cards on their behalf in return for a fee.

[8] To simplify the analysis we will assume financial institutions specialise in either issuing or acquiring. The logic of the price-fixing allegation which motivates our interest in bilateral negotiations suggests that financial institutions would need to specialise in either issuing or acquiring to avoid acting illegally (agreeing with competitors), so we

Although each issuer still wants to reach an agreement, so as to be able to offer an issuing service, the amount it stands to lose from failing to reach an agreement is far less than the acquirer with which it is negotiating. Suppose, there are a total of N issuers, N acquirers, all issuers and acquirers are the same size, and that issuer X is negotiating with acquirer Y over the terms of interchange. If acquirer Y does not agree with the issuer's terms, it stands to lose all its acquiring business, which is proportional to N, the number of issuers it deals with. However, if issuer X cannot reach an agreement it stands to lose only the business with acquirer Y. In other words, in negotiation, the acquirer stands to lose N times as much business as the issuer. This leaves the acquirer in a very poor bargaining position. It will be willing to accept a very high interchange fee with issuer X, in order to maintain its business.

To model this hold-up problem we provide a formal model of bilateral bargaining over the interchange fee. The emphasis in this section is to model the bargaining process in detail, leaving the details of how the resulting interchange fee feeds back into profits and affects competition to be analysed in the next section.

We suppose initially every issuer i has n_i cardholders and there are N such issuers, while each acquirer j has m_j merchants and there are M such acquirers. Thus, we suppose the total number of card transactions is

$$\sum_i \sum_j z n_i m_j$$

where z is a scaling parameter. We suppose each transaction generates profit to the acquirer j of $r_{i,j,} - a_{i,j}$, where $r_{i,j}$ is revenue per transaction earned by the acquirer j from transactions between cardholders belonging to issuer i who transact with merchants belonging to acquirer j, and $a_{i,j}$ is the existing interchange fee between issuer i and acquirer j. Likewise, we suppose each transaction generates profit to the issuer i of $a_{i,j} - c_i$, where c_i captures the fact that for given fees, if it was not for interchange revenue, issuers incur losses rather than profits on each transaction.

We consider the case of two banks, one being an issuer (denoted k) and the other an acquirer (denoted l), which are negotiating an interchange fee. If they both agree on an interchange fee of a^*, then the issuer will have profits of

$$\Pi_k = \sum_{j \neq l} z n_k m_j (a_{k,j} - c_{k,j}) + z n_k m_l (a^* - c_{k,l})$$

and the acquirer will have profit of

$$\Pi_l = \sum_{i \neq k} z n_i m_l (r_{i,l} - a_{i,l}) + z n_k m_l (r_{k,l} - a^*)$$

assume this is the case. This is also consistent with the approach adopted in the existing literature on interchange fees in payment schemes, which treats all banks as either specialist issuers or acquirers.

Under honour all cards, if they cannot reach an agreement, then the acquirer can no longer continue offering its service, and their respective profits are

$$\Pi_k^0 = \sum_{j \neq l} zn_k m_j (a_{k,j} - c_{k,j})$$

$$\Pi_l^0 = 0$$

We use the Nash Bargaining Solution (NBS) to determine the outcome of negotiating given these payoffs. The Nash bargaining solution (Nash, 1953), says, in general terms, that bilateral bargaining will lead to profits being split in relation to the difference between what each firm can achieve through an agreement and what they will get if they disagree. The NBS is the value of a that maximises:

$$NBS = [\Pi_k - \Pi_k^0][\Pi_l - \Pi_l^0]$$

Denoting the value of a which maximises NBS, a^*, it is straightforward to show that

$$a^* = \frac{\sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l})}{2zn_k m_l} + \frac{r_{k,l} + c_{k,l}}{2}$$

so that interchange revenue paid by acquirer l to issuer k is

$$\frac{\sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l})}{2} + zn_k m_l \left[\frac{r_{k,l} + c_{k,l}}{2} \right]$$

Bank k secures interchange revenue that not only pays for half of its costs of providing business with bank l and half of bank l's revenue from these transactions, but also half of bank l's profit arising from all its other acquiring business. The issuer's profit, resulting from bargaining, is then

$$\Pi_k^* = \sum_{j \neq l} zn_k m_j (a_{k,j} - c_{k,j}) + \frac{1}{2} \sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l}) + zn_k m_l \left[\frac{r_{k,l} - c_{k,l}}{2} \right]$$

while the acquirer's profit is

$$\Pi_l^* = \frac{1}{2} \sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l}) + z(n_k) m_l \left[\frac{r_{k,l} - c_{k,l}}{2} \right]$$

From the above profit results, it is clear that the issuer captures half of the profit that the acquirer derives from its trading with other issuers, plus half of the revenue (less cost of issuing) that the acquirer earns from business with the issuer's cardholders. If an acquirer does not accept the terms put forward by the issuer, the acquirer stands to loose all its existing acquiring business. When negotiating with a small

issuing firm, this leaves it in a particularly poor bargaining position. Note that the issuer k necessarily earns more profit than the acquirer l, regardless of their market shares or profitability prior to the bilateral bargain. Moreover, the interchange fee agreed through bilateral bargaining is increasing in the number of merchants the acquirer has, and is decreasing in the number of cardholders the issuer has. Thus a small issuing firm extracts half of the acquiring firm's profit by setting a very high interchange fee.

Implicit in the analysis above is that the other variables, such as the number of customers each bank secures remains constant, as does their revenue and cost per transaction. In practice, any increase in the interchange fee could have a variety of complicated effects. For the purposes of demonstrating the nature of the hold-up problem that arises from bilateral bargaining, these effects are very much second order. However, to show that taking them into account, might if anything accentuate the results already obtained, assume that the gross revenue acquirers earn per transaction increases in the interchange fee, but that as a result of a decrease in demand by merchants, the number of merchants it has decreases. As a result of a higher interchange fee we also assume that the issuer can offer more attractive terms to cardholders and thus increases its number of cardholders. Assuming, to a first approximation, that the total number of transactions between the cardholders and merchants remains the same[9], we can then derive the interchange fee which maximises *NBS* above:

$$a^* = \frac{\sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l})}{(2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}) zn_k m_l} + \frac{r_{k,l} + c_{k,l} - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}$$

Since both derivatives with respect to the interchange fee a^* (in the above expression) are positive, the interchange fee is generally higher once these two effects are taken into account. This has the effect of further increasing the asymmetry in bargaining power between issuers and acquirers. Bank k's profit from issuing is now:

$$\Pi_k^* = \sum_{j \neq l} zn_k m_j (a_{k,j} - c_{k,j}) + \frac{1}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} \sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l})$$
$$+ zn_k m_l \left[\frac{r_{k,l} + c_{k,l} - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} - c_{k,l} \right]$$

Bank l's profit from acquiring is now:

[9] To be precise, we assume that $zn_k n_l$ remains constant over the range of a^* considered.

$$\Pi_l^* = \frac{1 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} \sum_{i \neq k} zn_i m_l (r_{i,l} - a_{i,l})$$
$$+ zn_k m_l \left[r_{k,l} - \frac{r_{k,l} + c_{k,l} - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}}{2 - \frac{\partial c}{\partial a^*} - \frac{\partial r}{\partial a^*}} \right]$$

For all reasonable parameter values, the issuer's profit is even higher and the acquirer's profit is even lower, compared to before.

This analysis shows that, using the most commonly used economic model of how agreements are determined under bilateral bargaining (the Nash bargaining solution) acquiring firms are indeed left in a very poor bargaining position when negotiating interchange fees bilaterally with issuers.[10] In the case at hand, this implies issuer X will not only be able to capture half the revenue (less its cost) generated from business between itself and bank Y, but also half the profit that bank Y earns through all its other acquiring business. This outcome is further skewed once we consider the fact not all banks are equally sized. Very small issuers will still be able to capture half the profit that a large acquirer earns, which they do through negotiating even high interchange fees.

We have demonstrated a serious hold-up problem arises from bilateral bargaining under the honour all cards rule. The asymmetric bargaining power that arises from issuers having hold-up power is used in the next section to show how this leads to an escalation of interchange fees.

4. The escalation of interchange fees

In the previous section we demonstrated that under bilateral bargaining between an issuer and an acquirer and given the maintenance of the honour all cards rule, the issuer will have a much stronger bargaining position which it will exploit by securing a high interchange fee. In this section we explore the consequences of this asymmetric bargaining position, when all banks are setting their interchange fees bilaterally, and when the dynamic effects of bank competition and demand feedbacks are allowed to play out.

To model this escalation of interchange fees we use a setting with imperfectly competing issuers and acquirers. The model, unlike other models of interchange setting in the literature, allows for

- imperfect competition between issuers
- imperfect competition between acquirers
- heterogeneous consumers, not all of whom will choose to hold cards

[10] Using a simpler model of bilateral bargaining, in which the issuer makes a "take-it-or-leave-it" offer, leads to even more skewed outcomes, and even higher interchange fees.

- heterogeneous merchants, not all of whom will choose to accept cards
- benefits to cardholders that increase with the number of merchants who accept cards
- benefits to merchants that increase with the number of cardholders
- endogenous determination of the merchant discount, issuers' fees (net of any rebates), market shares, and customer and merchant participation

To make the model tractable we assume there are only two issuers and two acquirers. In the model there are a fixed number of potential cardholders (consumers) and potential suppliers who accept cards (merchants), who place different values on using cards from different banks, as well as on not using cards at all. To model this we adopt the standard additive random utility discrete choice model.[11] In particular, we assume there are a continuum of potential cardholders and potential card-accepting merchants, each of total measure normalized to 1. Each consumer attains utility $u_i=v_i+e_i$ from joining the card network with issuer i, where v_i is the expected benefits of joining the card network through bank i, and e_i is a random variable that reflects idiosyncratic taste differences across the two issuers (i=1,2). Consumers can also choose not to hold a card from either bank (they will use alternatives such as cash), in which case they get utility $u_0=v_0+e_0$ where v_0 is normalized to 0. Merchants face a similar choice, of either joining bank 1, bank 2, or not accepting cards. To distinguish the decision facing merchants from that facing consumers, we denote merchants' profit $\pi_j=w_j+e_j$ where j=1 and 2 represents the two acquirers. If they do not accept cards from either bank they get $\pi_0= w_0+e_0$ where w_0 is again normalized to 0.

Consumers and merchants each choose the alternative which gives them the highest utility or profit. Following Anderson *et al.* (1992) we assume the variable e_i in each case is independently and identically distributed according to the double exponential distribution. This implies the measure of cardholders belonging to issuer 1 and 2 are respectively

$$n_1 = \frac{e^{v_1/\rho}}{1+e^{v_1/\rho}+e^{v_2/\rho}} \quad \text{and} \quad n_2 = \frac{e^{v_2/\rho}}{1+e^{v_1/\rho}+e^{v_2/\rho}} \tag{1}$$

Similarly, the measure of merchants belonging to acquirer 1 and 2 are respectively

$$m_1 = \frac{e^{w_1/\rho}}{1+e^{w_1/\rho}+e^{w_2/\rho}} \quad \text{and} \quad m_2 = \frac{e^{w_2/\rho}}{1+e^{w_1/\rho}+e^{w_2/\rho}} \tag{2}$$

The parameter ρ represents the extent of competition between the alternative banks and is similar to the transportation cost parameter in the standard Hotelling model of horizontal differentiation. As $\rho\to 0$, the alternatives become homogenous, and banks will price as Bertrand competitors. Alternatively, as ρ

[11] See section 2.6 of Anderson *et al.* (1992).

gets large, banks will retain higher profits. We could easily allow ρ to differ across issuers and acquirers, although according to the examples we have considered, doing so does not affect the qualitative nature of our results.

Cardholders then get expected benefits of

$$v_i = (b_B - p_{I,i})(m_1^e + m_2^e), \quad i=1,2 \tag{3}$$

if they make transactions using cards from issuer i, where bank i is assumed to charge cardholders a fee per-transaction of $p_{I,i}$, and where consumers form expectations of the number of merchants there will be when making their joining decision. Similarly, merchants get expected benefits of

$$w_j = (b_S - p_{A,j})(n_1^e + n_2^e), \quad j=1,2 \tag{4}$$

if they accept card transactions through acquirer j, where bank j is assumed to charge merchants a price per-transaction of $p_{A,j}$ and where merchants form expectations of the number of cardholders there will be when making their joining decision. These benefit functions capture the fact whatever benefits cardholders get from using cards, their total expected benefits will be proportional to the number of merchants that they expect will accept cards. Similarly, whatever benefits merchants get from accepting cards, their total expected benefits will be proportional to the number of consumers that they expect will use cards.

The profits of issuer *i* are

$$\Pi_{I,i} = n_i m_1 (p_{I,i} + a_{i,1} - c_{i,1}) + n_i m_2 (p_{I,i} + a_{i,2} - c_{i,2})$$

while an acquirer *j* is assumed to earn profits of

$$\Pi_{A,j} = n_1 m_j (p_{A,j} - a_{1,j} - c_{1,j}) + n_2 m_j (p_{A,j} - a_{2,j} - c_{2,j}),$$

where $a_{i,j}$ denotes an interchange fee per-transaction charged by issuer i to acquirer j.[12] Profits for issuers arise from the (net) collection of fees from cardholders and the interchange fee received on transactions less the costs of servicing each cardholder. The cardholder fees charged, the interchange fee received and the costs of transactions are all measured per transaction. Each of these terms is then multiplied by the corresponding number of transactions. Profits for acquirers arise from the merchant discount charged less the interchange fee paid and the costs of servicing merchant transactions. Again, each of these terms is then

[12] These profit functions are consistent with the model of hold-up in Section 3 if the fee $p_{I,i}$ set by issuers is set at zero (or where positive, subtracted from the costs incurred by issuers in order to work out net costs), while the fee set by acquirers $p_{A,j}$ not only covers their costs, but some additional component which we denoted $r_{i,j}$ previously.

multiplied by the corresponding number of transactions. Implicit in the profit function is the assumption that the number of transactions between n_i cardholders and m_j merchants is simply $n_i m_j$, or that each cardholder makes one transaction with each merchant. In practice consumers only make purchases with a small fraction of the potential merchants. Normalizing the number of transactions to any particular number will simply scale the results we present.

We suppose at each round of bargaining, banks agree on interchange fees. We discuss how this negotiation process takes place shortly. First, we note that for any given interchange fee, banks set their price to maximise their profits, taking the other banks' prices as given. In doing this we look for a fulfilled-expectations subgame perfect Nash equilibrium, as introduced by Katz and Shapiro (1985). This implies the equilibrium prices and number of cardholders and merchants is determined by the simultaneous solution to the following four first-order conditions, together with the expressions (1)-(4), and the assumption that in equilibrium consumers' and merchants' expectations of participation rates must equal the actual levels of participation.

$$n_1 - \frac{e^{v_1/\rho}(1+e^{v_1/\rho})}{\rho(1+e^{v_1/\rho}+e^{v_2/\rho})^2}\left[m_1(p_{I,1}+a_{1,1}-c_{1,1})+m_2(p_{I,1}+a_{1,2}-c_{1,2})\right]=0 \tag{5}$$

$$n_2 - \frac{e^{v_2/\rho}(1+e^{v_2/\rho})}{\rho(1+e^{v_1/\rho}+e^{v_2/\rho})^2}\left[m_1(p_{I,2}+a_{2,1}-c_{2,1})+m_2(p_{I,2}+a_{2,2}-c_{2,2})\right]=0 \tag{6}$$

$$m_1 - \frac{e^{w_1/\rho}(1+e^{w_1/\rho})}{\rho(1+e^{w_1/\rho}+e^{w_2/\rho})^2}\left[n_1(p_{A,1}-a_{1,1}-c_{1,1})+n_2(p_{A,1}-a_{2,1}-c_{2,1})\right]=0 \tag{7}$$

$$m_2 - \frac{e^{w_2/\rho}(1+e^{w_2/\rho})}{\rho(1+e^{w_1/\rho}+e^{w_2/\rho})^2}\left[n_1(p_{A,2}-a_{1,2}-c_{1,2})+n_2(p_{A,2}-a_{2,2}-c_{2,2})\right]=0 \tag{8}$$

Clearly solving (1)-(8) simultaneously is a difficult non-linear problem which does not yield any closed form solution. This difficulty is compounded by the fact the interchange fees themselves have to be determined. According to the Nash bargaining solution, which we defined in Section 3, the bilaterally determined interchange fees which come out of the Nash bargaining depend not only on the above considerations, but also on the disagreement points, which have to be found using the model above also. To simplify, we follow the approach of Schmalensee (2001) and represent the solution to bilateral bargaining between issuer *i* and acquirer *j* as the interchange fee $a_{i,j}$ which maximises $\alpha\Pi_i^I + (1-\alpha)\Pi_j^A$, for some α greater than one-half.[13] We choose α greater than one-half to represent the fact that issuers have more bargaining power than acquirers, because of their hold-up power (as explained in Section 3). We have tried different values of α to capture various degrees of hold-up. Higher values of α lead to

[13] Schmalensee (2001) also assumed that the profit of the issuers gets more weight when determining the interchange fee which maximises the weighted sum of issuers and acquirers profit. In his model, there is just one interchange fee,

greater escalation of interchange fees, but qualitatively the results are unchanged. For this reason we restrict attention here to just the case α=0.75.

We start by picking the centralised level of interchange fee which maximises the banks' joint profit. We then consider repeated rounds of bargaining, where in each round of bargaining, each issuer reaches a separate bilateral agreement with each acquirer taking as given the interchange fees set in other agreements as those existing in the last round. This captures the idea that all issuers reach an agreement with all acquirers through simultaneous but separate agreements.[14] At the end of each round of bargaining we record the average interchange fee, the average price set by issuers and acquirers to their customers, the total number of card transactions for the network, and total bank profits. We repeat this process until the interchange fee converges to an equilibrium.

To illustrate the results of this model we consider two sets of parameter values. The first case we examine is one in which there is complete symmetry between consumers and merchants. Cardholders and merchants get the same benefits from using cards and issuers and acquirers face the same costs. We use the following parameter values: b_B=b_S=1, $c_{i,j}$=0.5, ρ=0.5. Table 1 reports the results after each round of bargaining until an equilibrium is reached.[15]

Table 1: Symmetric Case

	Round 0	Round 1	Round 2	Round 3	Round 4	Round 5
Average interchange fee	0	1.50	1.75	1.85	1.90	1.95
Average price to users	1.90	2.82	3.16	3.32	3.40	3.49
Volume of transactions	0.22	0.10	0.08	0.07	0.07	0.07
Total bank profits	0.60	0.46	0.43	0.42	0.42	0.42

Due to the symmetric set up, the joint profit maximising interchange fee is equal to zero. This maximises the number of card transactions and the level of banks' profits. However, if banks set their interchange fees bilaterally, the results in table 1 show they will set interchange fees that are substantially above zero. In fact, after only one round of bilateral bargaining, interchange fees will increase to 1.50. This not only increases the average price paid for the card service by cardholders and merchants (from 1.90 per transaction to 2.82 per transaction), but it reduces the volume of card transactions by more than half in the first round of negotiations. As a result of lower volumes, even though overall prices increase, bank profits are lower. Allowing bilaterally set interchange fees is unambiguously harmful because it leads to an

which either represents the rate set between a monopoly issuer and a monopoly acquirer, or between a group of issuers and a group of acquirers.

[14] We have also modelled the case that agreements between banks are reached sequentially taking the agreements of the other banks as given. This leads to the same qualitative results of escalation and an implosion of the network size, but means the order of negotiation between banks affects the exact path of interchange fees to the equilibrium.

escalation of interchange fees above the level that maximises the network's value, a lowering of the surplus of users as a result of higher prices, and a lowering of bank profits.[16] The network quite rapidly contracts, with the result being lower bank profits despite higher prices.[17]

Table 2 shows similar results apply when we start from an asymmetric situation. We assume $b_B=0$, $b_S=2$, $c_{I,i}=0.75$, $c_{A,I}=0.25$, $\rho=0.5$. In this case, the technological costs of card issuance are more than the technological costs of providing the acquiring service. However, more of the benefits of a card service are captured by merchants. In this case the interchange fee which maximises total bank profits is well above zero. This is because, facing an interchange fee equal to zero, cardholders would face higher prices than merchants, but yet provide an externality to merchants. A transfer from merchants to cardholders can ensure that cardholders face lower prices, thus internalising the externality. This will lead to greater card usage that ultimately benefits merchants. Thus, we start from a situation where the average interchange fee and price paid by users (if say measured as a percent of the average transaction), is not vastly different from that observed in practice.

Table 2: Asymmetric Case

	Round 0	Round 1	Round 2	Round 3	Round 4
Average interchange fee	1.26	2.71	3.01	3.11	3.16
Average price to users	1.90	2.77	3.18	3.34	3.42
Volume of transactions	0.22	0.10	0.08	0.07	0.07
Total bank profits	0.60	0.46	0.43	0.42	0.42

While an association which maximises the total profit of its members will seek an interchange fee which generally expands output, acting individually each issuer will opportunistically seek a higher level of interchange fees. Thus an escalation in interchange fees again occurs when banks move from setting the interchange fee centrally to setting it bilaterally. In this case, interchange fees more than double after just one round of bilateral bargaining. The result is a substantial increase in average price, with any reduction in cardholders' fees being more than offset by a rise in merchant fees. As a result of sub-optimal network pricing, the volume of transactions implodes, as do bank profits. The negative implications of moving to

[15] If issuers and acquirers set interchange fees taking into account the effects of rounds of negotiation, they will immediately set them at the levels implied by the equilibrium reached after convergence of the rounds of bilateral setting have taken place.

[16] While cardholders may initially benefit from higher interchange fees (and lower cardholder fees), the reduction in merchants that accept cards, will mean eventually both cardholders and merchants will be worse off, unambiguously lowering consumer surplus. Additionally, banks will compete less aggressively for cardholders as a result of reduced interchange revenue on each cardholder when the number of merchants that accepts cards declines.

[17] The solution converges to the equilibrium after five rounds, with subsequent rounds of bargaining not showing any change in interchange fees, or the implied prices, volumes, or profits.

bilateral bargaining over interchange fees remain when we start from a position where there are realistic asymmetries between cardholders and merchants.[18]

The results again reflect the implications of hold-up when competing issuers can engage in this behaviour. The hold-up problem will mean that each issuer can secure a higher interchange fee compared to the centrally set fee. When securing a higher interchange fee, an issuer faces the following two effects:

- increase in interchange revenue per cardholder

- reduced overall size of the market as a result of decreased demand from merchants

Starting from the point where joint profits are maximized, the first effect will tend to dominate the second effect in terms of increasing the issuer's profitability. To see why, note firstly that any reduction in demand from merchants caused by a single issuer raising its interchange fee reduces demand for all issuers in equal proportion to their size.[19] Thus, an issuer that increases its interchange fee gets all the benefits, but shares the costs, of such an action. This free-rider problem leads to banks, through bilateral negotiations, setting interchange fees about the level which is jointly optimal for them.

In addition, the reduced overall size of the market as a result of decreased demand from merchants has the effect of reducing the profitability to all issuers of attracting additional cardholders. This will reduce competition for cardholders, thereby increasing cardholder fees.

The much higher interchange fees that result from these effects suggests welfare will decrease. This is particularly likely where the joint profit maximizing interchange fee is already at, or higher than, the socially optimal level of the interchange fee.[20]

6. Conclusion

We have studied the impact of a prohibition on centralised setting of interchange fees in payment card networks, with particular reference to open credit card networks. Our analysis was motivated by the concerns of competition authorities in a number of jurisdictions. Our main result is that bilateral negotiations when applied in the context of the "honour all cards" rule and a lack of merchant surcharging lead to some very undesirable outcomes. Our models predict that moving to bilateral bargaining would induce substantial transaction costs, that issuers would hold-up acquirers, and that the resulting dynamic would result in interchange fees which are even higher than the level set by an association of card issuers and acquirers that maximises the joint profits from the card service. While it is true that these results would

[18] The solution converges to the equilibrium under bilateral bargaining after four rounds.
[19] This would not necessarily be the case if merchants price discriminated across different card types – this is ruled out under the no-surcharge rule or by merchants' unwillingness to surcharge in this way.
[20] Rochet and Tirole (2000) argue the centrally set interchange fee chosen by an issuer controlled card association is either at the same level or higher than the socially optimal interchange fee. Schamensee (2001) and Wright (2001a) provide some conditions under which the privately set interchange fee coincides with the socially optimal level.

not necessarily arise if the honour all cards rule is abandoned, it is widely accepted that this rule is a highly desirable one to maintain.

As a result of the escalation of interchange fees, and other problems caused by decentralised interchange fee setting, bilateral bargaining would diminish the size of the card network and its users would switch to other means of payment. While we did not model the ability of users to switch to rival networks (such as closed networks), presumably the ability to do so will speed up the decline of the existing open card systems, while reducing the welfare effects of the decline. Clearly, a tractable model of such inter-system competition could be of great benefit in considering the full effects on open payment systems of a move to decentralised interchange fees, as well as other suggested changes to payment system rules.

References

ACCC (2000), "Debit and Credit Card Schemes in Australia: A study of interchange fees and access," Australian Competition and Consumer Commission and Reserve Bank of Australia, October.

Anderson, S.P., A. de Palma, and J.F. Thisse. (1992) "*Discrete Choice Theory of Product Differentiation*," MIT Press: Cambridge.

Balto, David. (2000) "The Problem of Interchange Fees: Costs without Benefits," *European Competition Law Review*, 81(1), 50-81.

Baxter, W.F. (1983) "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics*, 26, 541-88.

Chang, H.H. and D.S. Evans (2000) "`The Competitive Effects of the Collective Setting of Interchange Fees by Payment Card Systems," *The Antitrust Bulletin*, Fall, 643-661.

Cruickshank, Don (2000) "Competition in UK Banking," HMSO: London.

European Commission (2000) "Commission Plans to Clear Certain Visa Provisions, Challenge Others," Press Release. DN:IP/00/1164. Brussels. October 16.

European Commission (2001) "Case COMP/29.373 – Visa International," *Official Journal of the European Communities*, 2001/ C 226 /10. August 11.

Evans, D. and R. Schmalensee (1999) "*Paying with Plastic: The Digital Revolution in Buying and Borrowing*," MIT Press.

Frankel, Alan (1998) "Monopoly and competition in the supply and exchange of money," *Antitrust Law Journal*, 66, 313-361.

Gans, J.S. and S.P. King (2001a) "The Neutrality of Interchange Fees in Payment Systems," mimeo. University of Melbourne.

Gans, J.S. and S.P. King (2001b) "Regulating Interchange Fees in Payment Systems," mimeo. University of Melbourne.

IMA (2000) "Study Regarding the Effects of the Abolition of the Non-discrimination Rule in Sweden for European Commission Competition Directorate General," Feb 29, http://europa.eu.int/comm/competition/antitrust/cases/29373/studies/

ITM (2000) "The Abolition of the Non-discrimination Rule," Amsterdam, March, http://europa.eu.int/comm/competition/antitrust/cases/29373/studies/

Katz, M. L. and C. Shapiro (1985). "Network externalities, competition, and compatibility," *The American Economic Review*, 75: 424 - 440.

Nash, J. (1953) "Two person co-operative games" *Econometrica*, 21, 128-140.

Reserve Bank of Australia (2001) "Reform of Credit Card Schemes in Australia," December.

Rochet, Jean-Charles and Jean Tirole (2000) "Cooperation among Competitors: The Economics of Credit Card Associations," mimeo, Toulouse.

Schmalensee, Richard (2001) "Payment Systems and Interchange Fees," Working Paper No. 8256, NBER.

Wright, Julian (2000) "An Economic Analysis of a Payment Card System" mimeo, University of Auckland.

Wright, Julian (2001a) "Determinants of Optimal Interchange Fees in Payment Systems," Department of Economics Working Paper No. 220, University of Auckland.

Wright, Julian (2001b) "Optimal Card Payment Systems" mimeo, University of Auckland.