National Australia Bank Limited
ABN 12 004 044 937

National
TAILORED FINANCIAL SOLUTIONS

General Manager, Payments Policy
500, Bourke Street
Melbourne
Victoria 3000
Telephone: 03 8641 3955
Facsimile: 03 8641 0109
peter_thomas@nag.national.com.au

2nd July 2001

Dr. John Veale
Head of Payments Policy Department
Reserve Bank of Australia
65 Martin Place
GPO Box 3947
Sydney NSW 2001

Dear John:

Designation of Visa, MasterCard and Bankcard

Thank you for your letter of 22nd May 2001.

You will by now have received the submission of the Australian Bankers' Association to which we were a party. The National Australia Bank believes that the ABA submission covers most of the issues of interest or concern to the RBA . We shall not, therefore, be making a separate submission, but instead will confine ourselves to matters on which we have particular views or which we feel may be of particular concern or interest to the RBA.

1. Guiding principles in relation to designation

In his letter to the Governor dated 5th April 2001, Frank Cicutto set out six principles which the National considered should guide the RBA's decision to designate and raised a number of other issues relevant to designation. The National believes that those principles and issues should also guide the RBA in deciding whether it should take any further action under the *Payment Systems (Regulation) Act*.

The six guiding principles are as follows:.

1. **The Australian retail payments system should be considered as a whole.**

2. **Regulation should facilitate competition between the various participants in the Australian retail payments system on a "level playing field".**
3. **Participation in the open credit card schemes should be open to all subject to appropriate prudential, technical and regulatory criteria.**
4. **Pricing should be transparent to consumers and based on user pays principles.**
5. **Cross-subsidisation should be minimised.**
6. **Consumers should be the beneficiaries of reform.**

A copy of the relevant sections of Frank Cicutto's letter is attached for completeness (the material relating to the now abandoned ACCC litigation has been excised).

In addition, we make the following specific points.

2. Interchange fee methodologies

In the course of this present inquiry into the setting of credit card interchange fees the National has had the opportunity to review numerous methodologies e.g. Baxter/Visa, MasterCard, Frontier Economics Residual Cost model, the ABA Avoidable Cost model, CoRE's Shared Avoidable Cost model (see below), a Par model, etc. Each of these models has been developed by respected academics or consultants and each has its own attractions. However, each has also been subjected to criticism by other respected academics or consultants.

In addition, it is apparent that all the methodologies are capable of producing a wide range of results, depending upon exactly how:

- **a particular cost or revenue caption is defined;**
- **each participating institution allocates costs and revenues; and**
- **it applies generally accepted cost accounting principles.**

The National agrees with the ABA that it would be preferable that the RBA not mandate any specific methodology, but rather, if it determines that regulation of interchange fees under the *Payment Systems (Regulation) Act* is warranted, it should approve a set of clearly defined principles through which it would be

possible to derive an upper limit to the interchange rate. The individual card associations would then set their rates within that overall cap.

It is the National's view that the principles contained in the ABA submission together with the ABA's Avoidable Cost methodology's interpretation of them would be as capable as any methodology of producing a cap or, to use the ABA's words an envelope, that would be in the public interest and satisfactory to card holders, merchants, credit card participants, future access seekers and the RBA. However, such an outcome would be highly dependant on the card association members and the RBA negotiating the cost captions to be included in the methodology and the exact definitions of those captions with good will and in the interests of all stakeholders in the credit card market. The National is, of course, most willing to enter into such negotiations.

3. "No surcharge" rule

In your letter to us of 22nd May you state in respect of the "no surcharge" rule that, " the National and Gans and King believe this rule should be abolished." You will note from the CoRE report (attached to this letter) that this does not appear to be the view of Gans and King. The National's position, is set out in Frank Cicutto's letter dated 5th April to the Governor, viz:

> **"Although there are some sound arguments for the "no surcharge" rule in some business sectors, the National as a practical matter does not oppose its abolition. However, if the rule is abolished it should follow that merchants are also allowed to impose a surcharge for other payment mechanisms, including cheques and cash, both of which have high handling costs. This may introduce an element of confusion to consumers at the check out counter."**

We would now add to this that the European Commission has recently decided that the rule is not anti-competitive and reports that in countries where the rule has been abolished, eg Holland and Sweden, such action appears to have had little or no effect on retail price setting by merchants. The European Commission's decision and the experience in countries such as Holland and Sweden are consistent with the National's view that the "no surcharge" rule is not anti-competitive.

4. End customers should be the beneficiaries of any reform

As we have stated before, should the result of the designation process be a lowering of the interchange fees, the National would wish to see any consequential savings flow through to end consumers. However, the National is concerned that the most likely outcome of any regulation of interchange fees is that any savings will be retained by the merchants and utilities. Moreover, a reduction in the level of interchange fees may cause card issuers to reduce reward programs and increase card fees. Consequently, the regulation of interchange fees may merely shift wealth from credit card holders to merchants and utilities. The RBA should ensure that end consumers, rather than merchants or utilities, are the beneficiaries of any regulation of interchange fees.

5. RBA question 13: credit card interest rates

We note that the Bankcard Association of Australia has already submitted its own responses to the RBA's 15 "credit card questions" addressed to the Bankcard Association of Australia and other market participants (but not the National).

We also note that, at Question 13 the RBA asks:

> **"Why are credit card interest rates around 3% higher than rates on other unsecured personal lending?"**

This question addresses a matter related to credit cards as a consumer borrowing instrument, rather than as a payment instrument, which is the object of the present enquiry. Nonetheless, we now provide a brief response. Credit cards and unsecured personal loans are very different instruments. The average unsecured loan is significantly larger than the average credit card outstanding , but the cost of maintaining a credit card account is substantially more than that of maintaining an unsecured loan. As a consequence the National believes that the return on both products is not dissimilar.

6. RBA question 15: multiple issuance

At Question 15, the RBA asks:

> **"Does the ability of issuing banks to offer a number of different brands of credit cards result in more competition between schemes than would be the case if they could only issue one brand?"**

The National considers that allowing issuing banks to issue a number of brands of credit cards would, prima facie, automatically lead to greater competition. In this regard we note that, in Australia, unlike the United States, issuers are able to issue American Express and other closed scheme cards as well as those of the open schemes. We are aware that Canada prohibits banks from issuing the cards of more than one of the open card schemes. We understand that four of the five major banks in Canada issue Visa, with Bank of Montreal being the only major bank to issue MasterCard. We further understand that Visa accounts for a high percentage of domestic transaction volume. This suggests to us that Canada's restrictions have lead to undue concentration in the market and has lessened the choice of Canadian consumers.

7. Gans and King reports

We are enclosing with this letter three reports prepared for the National by Professors Gans and King or their consulting firm CoRE Research, viz:

- **The Neutrality of Interchange Fees in Payments Systems - Draft dated 28th May 2001, Gans & King**
- **Regulating Interchange Fees in Payment Systems – Draft dated 14th June 2001, Gans & King**
- **Some Answers to the Reserve Bank of Australia Questions Associated with the Designation of Credit Card Systems in Australia - Dated 30th June 2001, CoRE Research**

The first two documents are technical papers, which Professors Gans and King may wish to publish in the relevant academic or professional journals in due course. These are being sent to you in the belief that you will find their contents interesting. However, the National has no comment to make on them.

The third document is the result of a request from the National that CoRE should address a number of the issues raised by in the RBA's 15 "Credit Card" questions. The National has not fully considered all of the matters raised by the document and we stress that it may well not always represent our views. Nonetheless, the National provides this document to you in the spirit of open communication so essential to the success of the designation process.

In particular you will note that CoRE has developed a methodology for the calculation of interchange fees. As you will be aware from their earlier published work, Professors Gans and King's do not believe that regulatory intervention in this area is in fact necessary. However, in order to develop their new methodology – Shared Avoidable Costs – they have assumed as premises that:

- **the retail market in Australia is not in fact competitive; and**
- **the "no surcharge" rule does have real economic consequences.**

Further the methodology seeks to meet the principles for the setting of interchange fees set out in the letter dated 21st February 2001 from the Australian Competition and Consumer Commission to the Review Banks. The National does not accept that either premise is correct. We are seeking advice on the CoRE methodology from other advisors, but suspect that there will be wide differences of opinion, as there has been on other issues in this area.

As discussed with you, we understand that the RBA intends to address the issues surrounding the eftpos and ATM networks once progress on the credit card issues is reasonably well advanced. We believe, if the case for credit card reform is established, that it would be highly desirable to effect reform of all three systems contemporaneously.

Accordingly we are recommending that the ABA now commission a full study of the eftpos and ATM issues in order that the industry be able to discuss reform of these systems with the RBA as soon as it is ready to consider these issues. The National reiterates that it believes these issues need to be addressed at an industry level and expects that it will be consulted throughout the RBA's process of considering reform.

Finally, no matter what form any regulation introduced by the RBA may take, it is the National's strong view that the RBA should assist the open credit card schemes and their members in procuring protection (either from the ACCC or the Commonwealth Government) from the risk that the *Trade Practices Act* might apply to conduct engaged in by the schemes and their members in compliance with the RBA's reforms.

We trust that these present submissions, the ABA's submissions and our own earlier letters will be of assistance to you. We look forward to working with you toward resolution of all these issues. If you have further queries we shall be pleased to receive them.

Yours sincerely,

(signed)

Peter J S Thomas

Extract of a Letter from Frank J Cicutto, Managing Director and Chief Executive Officer, National Australia Bank to the Governor of the Reserve Bank of Australia dated 5th April 2001.

Guiding principles in relation to designation

Should the RBA decide to designate, the National believes that its decision to do so and the scope of the designation, should be based on a number of broad principles, including the following:

1. **The Australian retail payments system should be considered as a whole.**

2. **Competition between the various participants in the Australian retail payments system should be on a "level playing field", consequently:**
   - **(a) the closed credit card schemes; and**
   - **(b) Australian dollar denominated credit cards issued by foreign institutions for use in Australia,**

   should fall within the scope of any RBA designation of the open credit card schemes.

3. **Participation in the open credit card schemes should be open to all subject to appropriate prudential, technical and regulatory criteria.**

4. **Pricing should be transparent to consumers and based on user pays principles.**

5. **Cross-subsidisation should be minimised.**

6. **Consumers should be the beneficiaries of reform.**

We set our comments on each of these principles below.

1. **The Australian retail payments system should be considered as a whole.**

   The efficiency, profitability and benefits to the Australian community of open credit card schemes should be reviewed in the context of the entire Australian retail payments system, including cash, cheques, the closed credit card schemes, direct debit and credit, and eftpos.

2. **Competition between the various participants in the Australian retail payments system should be on a "level playing field"**

   In the National's view, if the RBA designates the open credit card schemes, it should also designate similar systems within the Australian retail payments system. This will ensure that participants in the Australian retail payments system do not enjoy an advantage or suffer a detriment by reason of uneven regulation.

(a) The closed credit card schemes should fall within the scope of any RBA designation of the open credit card schemes

The National considers that any designation which applies to the open credit card schemes should also apply to closed card schemes. In this regard, we note that, for the purposes of designation, there seems to be no valid distinction between the open and closed credit card schemes. For example, VISA card "transactors" use their cards in the same way as American Express charge cardholders. Both schemes allow deferred payment without interest cost (provided the monthly repayment is made on the due date), both provide a monthly statement detailing usage and both may provide benefits such as loyalty programs.

The fact that American Express has an implicit interchange rate and VISA an explicit one, and that American Express is a closed scheme with no members (although we note the multiplicity of agency issuing agreements that American Express has entered into with local institutions) and VISA is an open scheme with membership rules, does not justify regulating one scheme but not the other.

Further, there is strong anecdotal evidence that the average merchant service fee charged by the closed schemes is higher than that charged by the open schemes' members.

Failure to include the closed schemes within the ambit of the designation will lead to the perverse result that the more expensive closed schemes will continue to thrive at the expense of the cheaper open schemes, through their ability to continue providing attractive benefits, such as loyalty schemes, to their cardholders.

[redacted]
.

(b) Australian dollar denominated credit cards issued by foreign institutions for use in Australia should fall within the scope of any RBA designation of the open credit card schemes

The National also considers that any designation which applies to the open credit card schemes should apply to Australian dollar denominated credit cards issued by foreign institutions for use in Australia.

The use of such cards in Australia would attract the international default interchange rate, which is higher than the current domestic interchange rate (and would presumably be higher than any interchange rate which resulted from the designation process). If this was the case, the foreign institutions that issue these cards would have a distinct advantage over the issuers of cards in designated schemes because they would be able to use their higher interchange rates to offer cardholders better loyalty programs.

3. **Participation in the open credit card schemes should be open to all subject to appropriate prudential, technical and regulatory criteria**

   Participation in the open credit card systems should be open to all, provided that all participants are required to meet the same degree of technical competence, financial soundness and regulatory oversight. This broad principle should apply equally to the open credit card schemes as to other parts of the Australian retail payments system. Changes to the membership rules of the open credit card schemes should be carefully structured so as not to introduce unwarranted risk into the Australian retail payments system.

4. **Pricing should be transparent to consumers and based on user pays principles.**

   Adoption of user pays and fully transparent pricing will ensure that the correct price signals are sent to consumers and this should in turn contribute to the economic efficiency of the system.

5. **Cross-subsidisation should be minimised**

   Cross subsidisation between different types of payment instruments and other financial services products should be minimised to the extent practicable. In particular, we note that for many years the open credit card schemes in Australia have subsidised the eftpos network.

6. **Consumers should be the beneficiaries of reform**

   Any cost savings which result from the proposed reforms to the Australian credit card system should flow through to the end consumers and should not be allowed to bolster merchants' profit margins. In this regard we shall be formally requesting the ACCC to monitor retail prices to ensure that cost savings are passed on to consumers in full.

Potential consequences of designation and other relevant issues

As the RBA considers the desirability of designating the credit card schemes there are a number of other issues which the RBA may wish to consider. These issues flow from the RBA's stated reform objectives, namely:

- **to broaden the membership of the international card schemes beyond regulated deposit taking institutions;**
- **to reduce interchange fees to a level that reflects only costs related to network considerations; and**
- **to eliminate the "no surcharge " rule.**

We have briefly set out a non-exhaustive summary of the potential consequences of designation and other relevant issues below:

1. **Foreign investment**

   If an overly prescriptive regime is introduced as a result of the designation process, this may act as a deterrent to foreign investment in Australia.

2. **Future investment in payment systems**

   Considerable time and a great deal of money has been invested in order to develop the Australian credit card business. An outcome that substantially reduces the profitability of the credit card business will act as a strong disincentive to make further investments in payment systems. In addition some investment is driven by prudential and regulatory requirements. For example, Triple DES in the credit card business and Continuous Linked Settlement in foreign exchange settlement.

3. **Smaller financial institutions**

   The National believes that a significantly lower interchange rate will have a far more serious impact on the smaller issuers of credit cards than the larger issuers. We have in mind particularly the credit unions and building societies which use credit cards as a means of providing low cost transaction accounts. The smaller organisations are also the principal issuers of VISA and MasterCard debit cards, a service to their customers that we believe provides them with significant income.

4. **Repricing of other services**

   A significant reduction in the interchange fees and an adherence to the principles of "user pays" and no cross subsidies, may lead to a repricing of credit card services and the elimination of some cardholder benefits. The same principles will also probably lead to a restructuring of the pricing for eftpos, particularly as regards merchants.

5. **Smaller merchants**

   The open credit card schemes provide the smaller merchants with the ability to provide delayed payment or extended terms to their customers. This service can be viewed as an outsourcing of the smaller merchants' credit ledgers. If as a result of the proposed reforms there is a significant shift away from credit card usage to eftpos this will have the possible consequences of:

   (a) reducing consumer spending with a consequent impact on the economy generally, and
   (b) providing the major retailers, which issue their own store cards, with a distinct advantage over smaller retailers.

6. No-surcharge rule

Although there are some sound arguments for the "no surcharge" rule in some business sectors, the National as a practical matter does not oppose its abolition. However, if the rule is abolished it should follow that merchants are also allowed to impose a surcharge for other payment mechanisms, including cheques and cash, both of which have high handling costs. This may introduce an element of confusion to consumers at the check out counter.

Conclusion – other matters

If a designation process is to occur, the National is keen to ensure that there is no risk that in giving effect to any regulation introduced by the RBA the National (and other participants) in any way contravene the *Trade Practices Act*. At the appropriate time, we would be very pleased to co-operate with the RBA to ensure that the designation process is not affected by this uncertainty.

In closing let me reiterate that the National is very willing to work with the RBA toward reform of the credit card system.

(Signed by FJ Cicutto, Managing Director and Chief Executive Officer)

DRAFT

The Neutrality of Interchange Fees in Payment Systems

by

Joshua S. Gans *and* Stephen P. King*
University of Melbourne
First Draft: 1st August, 2000
This Version: 28th May, 2001

There has been considerable public debate over the effect of interchange fees on credit card transactions. Regulators in Australia and Europe have argued that these fees can be set by banks to have an anticompetitive effect. In the US, it has been argued that these fees, together with a rule that prevents a surcharge for credit purchases, might create a cross subsidy between cash and credit customers. Academics have noted that, in particular circumstances, interchange fees have no real effects in the absence of such a no-surcharge rule.

This paper considers two aspects of credit card interchange fees. First, it provides a general neutrality result. We show that in the absence of a no surcharge rule, interchange fees can never have any real effects. This result does not depend on the degree or nature of competition at either the bank or the merchant level. Second, we consider the potential for a bank with market power to manipulate interchange fees in the presence of a no-surcharge rule in order to raise bank profits. We show that such cross subsidisation is not profitable if there is adequate competition from 'cash only' merchants. We then consider the interaction between imperfectly competitive merchants that accept credit cards. For the special case of a single merchant, we provide both necessary and sufficient conditions on demand and the merchant's profit function for fee manipulation to be feasible. We show how these conditions alter with multiple imperfectly competitive merchants. In particular, we show that the profitability of any cross subsidisation depends critically on the nature of merchants' competitive interactions. *Journal of Economic Literature* Classification Numbers: G21, L31, L42.

Keywords. credit card associations, payment systems, interchange fee, neutrality, no surcharge rule

---

* Financial support for this project came in part from the National Australia Bank. Responsibility for all views expressed lies solely with the authors and should not be attributed to the above organisation. We thank participants at the University of Melbourne for useful comments. All correspondence to: Joshua Gans, Melbourne Business School, 200 Leicester Street, Carlton Victoria 3053, Australia; E-mail: J.Gans@unimelb.edu.au. The latest version of this paper is available to http://www.mbs.edu/home/jgans/research.htm.

1. Introduction

Payment systems are an intriguing but under-studied area of economics. First, a payment instrument is used only if sufficient numbers of agents on both sides of a transaction – that is, merchants and customers – intend or are able to use a particular instrument. Thus, the use of a particular payment instrument – whether it be paper or plastic – is determined in large part by network effects. Second, many payment instruments are sponsored and come into being by the mutual cooperation of otherwise competing institutions. For example, the two main credit cards used – MasterCard and Visa – both are the result of joint venture arrangements among thousands of institutions.

For this latter reason, credit card associations and other payment associations have been regarded with suspicion by anti-trust and monetary authorities alike. Anti-trust authorities are concerned that the terms of the joint venture agreements might serve to diminish the interests of users of a payment instrument precisely because those agreements are structured in the mutual interest of members.[1] There is also a concern that the network effects associated with payment systems can themselves create barriers to entry that impede competition among different instruments; thereby raising the economy-wide cost of transacting. For monetary authorities, the concern is that some types of payment instruments increase systemic risk more than others. Credit cards are a form of debt whereas debit cards do not involve financial

[1] Recently, there have been investigations into the potential competitive concerns surrounding credit card associations in the U.K. (Cruickshank, 2000) and Australia (RBA/ACCC, 2000). In addition, over the past several decades there has been a series of antitrust cases in the United States (see Evans and Schmalensee, 1999).

institutions bearing additional credit risk. The concern is that the operation of card associations excessively promotes the use of risky payment instruments above riskless ones.

Despite this policy interest there have been relatively few analyses that explore the potential policy concerns raised here. A seminal paper is that of Baxter (1983) who first described the network and other externalities that have driven current interchange arrangements between financial institutions supporting particular payment instruments such as checks or credit cards. In addition, several others have focused on the anti-trust concerns regarding payment systems; in particular, credit card associations (Carlton and Frankel, 1995; Frankel, 1998; and Evans and Schmalensee, 1999). Finally, there have been historical studies about the development and evolution of payments systems (see Rolnick, Smith and Weber, 1997, for a good example).

In terms of formal analyses, however, there are only two recent examples designed to evaluate the policy concerns. Schmalensee (2001) considers the agency problems faced by card associations in setting incentives for members to pursue activities in the collective interest. A theme of that paper is the difficulty an association faces in setting optimal incentives given that it has to rely on a single instrument – the interchange fee – that is a payment from merchant acquirers to card issuers. Rochet and Tirole (2000) also examine the role of the interchange fee in considerable detail. In a highly specific model, they demonstrate that the interchange fee may be set in a way that encourages over-use of a credit card precisely because merchants are not allowed – by the rules of card associations – to set different prices for cash and credit card transactions. Consequently, a higher interchange fee causes

prices to rise to all customers allowing more rents to accrue to association members and also encouraging too many customers to hold and use credit cards.

The purpose of the current paper is also to focus on the impact of the interchange fee on the use of payment instruments. We too will consider the operation of card associations, although our model is general enough to be applied to any payment instrument. Our approach however is somewhat different from the previous theoretical analyses in terms of its objectives. First, we want to identify the minimum set of conditions that, if present, will mean that the interchange fee is neutral in terms of determining the use of a credit card. Thus, the emphasis is on the generality and robustness of our model rather than the goal of demonstrating that the interchange fee may have an impact or not within a specialised framework (Rochet and Tirole, 2000). After determining the conditions for credit card neutrality, we systematically analyse the role of the interchange fee when these conditions do not hold. We formalise the argument of Frankel (1998) and others, that banks with market power can raise the interchange fee to create a cross-subsidy from cash to credit customers and to raise bank profits, when there is coherence between cash and credit prices. For the special case of a single credit card accepting merchant, we provide necessary and sufficient conditions on the merchant's profit function for this argument to be valid. We then consider competition between credit card accepting merchants and show that the nature of this competition critically affects the incentives for banks to manipulate credit card fees.

Our analysis yields a number of important insights. If merchants can set separate cash and credit prices or if customers have competitive cash options for any

good or service that they wish to purchase, the interchange fee will be neutral.[2] This result occurs regardless of the degree of competition or heterogeneity among members of an association, their degree of integration over different services associated with credit cards, or the level of competition among merchants. If the conditions for neutrality are not satisfied then it might be possible for a bank with market power to raise its profits by systematically manipulating interchange fees. However, this critically depends on the nature of merchants' profit functions and on the form and extent of strategic interaction between merchants. When merchants sell imperfect substitutes for either cash or credit but each merchant sets only a single price, the ability of a bank to profitably manipulate credit card fees depends on the interaction between merchants' cash and credit sales. In this sense, any ability of banks to abuse market power by manipulating credit card fees is an empirical matter that rests on the nature of merchant competition.

A general conclusion can be derived from our research. Competitive concerns regarding credit card interchange lie not so much in the market power of the associations themselves, nor of their members, but instead in the market power and commercial interactions at the retail level. Ultimately it is both the existence of merchant market power and the nature of merchant competition that allows interchange fees to have real economic effects giving associations the power to systematically use these fees to distort customers cash and credit choice.

We describe the nature of four party credit card systems, and the basic intuition underlying our neutrality result in the next section. Section 3 sets up the model's structure. Section 4 then proves the general neutrality result formally while

[2] This possibility has been noted earlier by Frankel (1998), Rochet and Tirole (2000) and Wright (2000). However, their analyses are either informal or derived within the context of a very specialized model. Our contribution in this paper is to provide a very general treatment of the issue.

Section 5 explores what happens when critical conditions regarding merchant level competition are relaxed. This section focuses on and formalises the issues of cash and credit price coherence. A final section concludes.

2. Basic Intuition

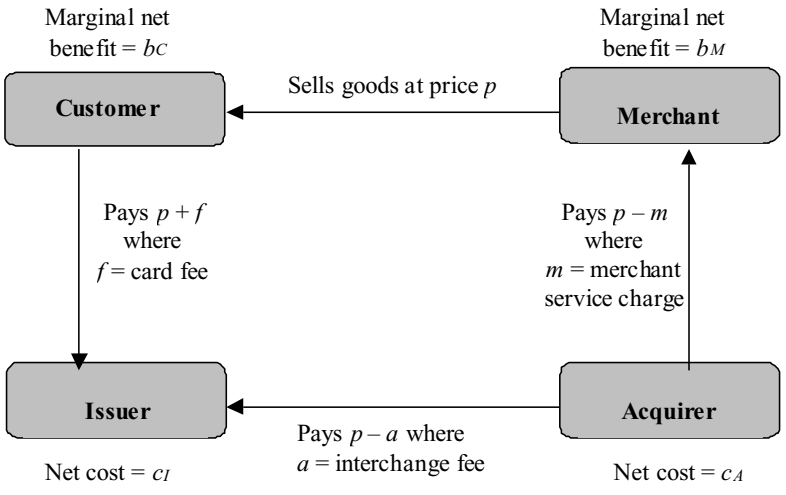
There are four parties to a credit card transaction. Card issuers are responsible for issuing cards and convincing customers to hold and use them. Customers then use cards to purchase goods and services from merchants who offer card facilities. If a merchant offers card facilities, such facilities are provided for by acquirers who are responsible for paying the merchant and themselves settling with the issuer for the amount of any transaction less the charges they keep for themselves and any interchange payments made to the issuer. The flow of funds is depicted in Figure One.

The basic intuition of previous research and our model can be illustrated by a simple example. Suppose that merchants derive a benefit from processing transactions via cards of b_M and this benefit is constant across merchants. Customers also receive a benefit from using cards. However, we suppose here that with probability β a customer's benefit is b_C otherwise it is 0. All customers wish to purchase one unit of a good only from merchants. We will assume, for the purposes of this example only, that issuers' and acquirers' costs are constant across transactions – being c_I and c_A respectively.[3] These benefits and costs are all assumed to net benefits and costs (relative to other payment instruments such as cash or check). As Baxter (1983) has

[3] These assumptions are made for illustrative purposes only and will not be assumed in the general model set-up in Section 3 below.

shown, the social desirability of card transaction will be positive if $b_M + b_C \geq c_I + c_A$, for customers who have an intrinsic value for using cards, and $b_M \geq c_I + c_A$ otherwise.

Figure One: The Four Party System



Implementing this optimum is not easy. First, all merchants should pay a merchant service charge, m, to acquirers equal to b_M. (Here we assume that merchants have no other costs.) Second, each customer should pay a tailored cardholder fee of $f = b_C$ or $f = 0$ depending upon their type. This requires knowledge of a customer's type but also, even if this is known, there is no reason why a particular bank might break even. For example, $b_M + b_C \geq c_I + c_A$ but it may be that $b_M < c_A$ or $b_C < c_I$. To ensure that the transaction goes ahead a payment – the interchange fee, a, – must be made from the profitable to non-profitable bank.

It has been demonstrated, however, that under perfect competition in the merchant, issuing and acquiring segments, this socially optimal outcome will be implemented (Baxter, 1983). In this situation, for a given interchange fee, $f = c_I - a$, $m = c_A + a$ and $p = m - b_M$ if a merchant offers card facilities (i.e., $m \leq b_M$ and 0

otherwise). Thus, all customers will use cards as $p = c_A + a - b_M \leq -f = -c_I + a$ or $b_M \geq c_I + c_A$. Otherwise, only a fraction, β, of customers will use cards with f and m as before but $p = c_A + a - b_M > -f = -c_I + a$ and $b_C - p = b_C - c_A - a + b_M \geq f = c_I - a$ or $b_M + b_C \geq c_I + c_A$. Cash customers end up paying a price equal to 0. Thus, under perfect competition, the social optimum is implemented.

However, Carlton and Frankel (1995) go further. They argue that perfect competition in all segments renders the interchange neutral in as far as real effects go; making the choice of *a* irrelevant for the association. That is, consider a situation where it is socially optimal for all customers to use credit cards. Then suppose that *a* is increased by Δ. Then, the merchant service charge will become $m = c_A + a + \Delta$ and cardholder fees will become $f = c_I - a - \Delta$ and the merchants' price becomes $p = c_A + a + \Delta - b_M$. Notice that the sum $p + f = c_A - b_M + c_I$; so any customer who used a credit card before the change will continue to use it afterwards as the total price of using a card, $p + f$, is unchanged. Moreover, for the same reason all merchants who offered card facilities will continue to offer them following the change. Hence, regardless of the level of the interchange fee, the outcome is the same in terms of the usage of the payment instrument.

Carlton and Frankel (1995) then go on to argue for the special nature of the perfect competition assumption and their beliefs regarding the circumstances under which the use of an interchange fee can have real effects. These circumstances include difficulties in charging customers rebates (or negative prices), imperfections in competition among banks that means that changes in wholesale prices influence margins as well as costs, and difficulties merchants have in charging customers

different prices depending upon whether they use credit cards or not. In this latter situation,

> [i]nterchange fees can be viewed as a way to raise costs to merchants who then pass those costs on to cash and credit customers alike by charging the same higher price to both. Cash customers are essentially being taxed to finance credit customers because the interchange fees eventually flow back to the card-issuing banks that will be forced by competition to give back at least part of the interchange fees in the form of rebates or lower fees to their credit card customers. Therefore, interchange fees allow credit card customers to impose a tax on cash customers. In such a setting, banks issue more credit cards and consumers use credit cards for more transactions than they would with no interchange fee. (Carlton and Frankel, 1995, pp.660-661)

Although they do go on to say that improvements in technology might cause this type of restriction to diminish over time. Nonetheless, card associations impose 'no surcharge' rules that restrict merchants from charging higher prices to card customers.

While it easy to imagine situations where banking segments are imperfectly competitive or a 'no surcharge' rule effectively applies, it is more difficult to imagine situations where customer rebates are not possible. Carlton and Frankel (1995) argue that a card association may have a rule that makes rebates impossible or issuers might have otherwise formed a cartel. However, even a monopoly issuer may wish to charge a customer rebate if interchange payments were sufficiently high. Thus, for our purposes we will continue to assume that customer rebates and indeed merchant rebates are possible.

Our goal is to explore the limits of the neutrality result as it arises in a perfectly competitive world and when a 'no surcharge' rule may or may not apply in order to understand whether such assumptions are critical or not. Rochet and Tirole (2000) investigate the role of the 'no surcharge' rule and demonstrate that, if merchants can charge different prices to cash and card customers, the interchange fee is again neutral even though the merchant sector is duopolistic while issuers have

market power.[4] This suggests that the application of the 'no surcharge' rule is of importance for neutrality results.

For this reason, we suppose here that the 'no surcharge' rule applies and instead focus upon the competitiveness of the issuing and acquiring segments. As mentioned earlier, we will demonstrate that the neutrality result continues to apply in the absence of perfect competition between issuers and acquirers. It is useful to give a flavour for our result by amending the above model and assuming there is a single issuer and a single acquirer. We will continue to assume they are not integrated. This monopoly case is, of course, at the opposite extreme from the perfect competition assumption employed by Carlton and Frankel (1995).

Consider first, the acquirer's problem. It will set *m* to maximise its profits given the level of cardholder use. The highest level of *m* is that which just allows merchants break even; that is, $m = b_M + p$.

The issuer sets *f* to maximise its profits. Notice that a customer will use a credit card so long as $b_C \geq p + f$ (for the high-types) or $0 \geq p + f$ for the low types. This is because each customer has an alternative cash purchase option. Given the monopoly profits, the issuer will choose either $f > -p$ and earn profits of $\beta(b_C - p - c_I + a)$ or choose $f \leq -p$ and earn profits of $-p - c_I + a$. We are assuming here that the issuer cannot price discriminate. Given this, there will be a high (low) level of card use if $a - p \geq (<) \frac{\beta}{1-\beta} b_C + c_I$.

Thus, it is the issuer's pricing policy that will determine the ultimate equilibrium. However, this introduces an additional constraint that the acquirer earn

---

[4] They assume perfect competition in acquisition. Hence, they need not consider the impact of integration.

non-negative profits. If the issuer chooses a high f resulting in a low level of card usage, this means that it cannot set f above $b_C + b_M - c_A - a$. In this case, the issuer earns $\beta\left(b_C + b_M - c_A - c_I\right)$. On the other hand, if the issuer chooses a low *f* resulting in a low level of card usage, it cannot set f above $b_M - c_A - a$. In this case, it will earn profits of $b_M - c_A - c_I$. Thus, it will choose a low *f* if and only if $b_M - c_A - c_I \geq \frac{\beta}{1-\beta} b_C$. Notice that neither this condition nor firm profits depends on the interchange fee. Notice also that the equilibrium is inefficient. That is, there is always under-provision of card services relative to the social optimum.

This example illustrates that there is nothing special regarding the degree of competition in issuing and acquiring that drives the neutrality result for interchange fees. Such neutrality can arise even where competition is absent and perfect price discrimination is not possible. What was critical in this example is the degree of merchant competition. In effect, there was an assumption of free entry that enabled customers to purchase their product without a credit card. This constrained the price, *p*, that merchants who offered card facilities could offer.

In essence the issue is that either in the absence of a no surcharge rule *or* with perfect merchant competition, the interchange fee represents a redundant price in a credit card network. In supplying payment services to customers, those customers care about the total prices they pay to merchants, issuers and acquirers. However, there are four prices in the system (payments to merchants, payments to issuers, payments to acquirers and the interchange fee). If only one price is changed in equilibrium, the other prices will all adjust in such a way as to leave the total prices faced by decision-makers exactly the same. In this respect, a change in the interchange fee alone will not have a real effect. This suggests that moves to regulate

interchange fees will, in the long-run, have no effect on competition or the actual operation of credit card associations.

3. A General Model of a Four Party Payments System

In this section, we present a general model of four party payments system. This model allows for any degree of both competition and integration in issuing and acquiring. It also allows for any form of merchant interaction. The model will be used in its general form to evaluate the neutrality of interchange fees. In a later section, the model will then be specialised to consider whether banks might have incentives to systematically distort credit card fees under a no surcharge rule.

Suppose that there are L customers, denoted by $l = 1,\ldots,L$; M 'cash-only' merchants who do not accept credit cards, denoted by $m = 1,\ldots,M$; and N 'credit-card' merchants who accept credit cards, denoted by $n = 1,\ldots,N$. The customers can choose to obtain a credit card from an issuing bank. There are $J + K$ such banks. Those denoted by $j = 1,\ldots,J$ only issue cards and do not supply services to merchants. Those denoted by $k = 1,\ldots,K$ are integrated and both issue credit cards and also supply credit-card facilities to merchants. The N credit-card merchants can choose to purchase their facilities from either one of the K integrated banks or from one of I acquirer-only banks, each of which is denoted by $i = 1,\ldots,I$.

Credit card purchases that require transactions between separate merchant and issuer banks involve an interchange fee, a. This fee might be set co-operatively between the banks. It is a charge paid by the merchant bank to the issuer bank as a percentage of the total transaction value; although we do not restrict it to be positive. We assume that there is only a single interchange fee. Given this fee, all banks that

issue credit cards set their customer charges. These might include an annual payment by the customer for the card and a per transaction fee. These charges are denoted by A_j and f_j respectively for a non-integrated bank and A_k and f_k for an integrated bank. The transaction fee is based on the size of a transaction and can be positive or negative. For example, reward schemes based on the volume of customer purchases represent a negative value of f. The merchant banks set the per transaction charges for the credit-card merchants, denoted by m_i and m_k for non-integrated and integrated banks respectively. This notation is summarised in Table 1.

Table 1: Summary of Notation

Agent	Indicator	Fixed Charge	Per Transaction Fee
Customers	$l = 1,\ldots,L$	p_m: price paid to cash-only merchants p_n: price paid to credit card merchants	
'Cash-Only' Merchants	$m = 1,\ldots,M$		$p_m q_m^l$
'Credit Card' Merchants	$n = 1,\ldots,N$		$p_n q_n^l$
Integrated Banks	$k = 1,\ldots,K$	A_k	m_k to merchants f_k to customers
Acquirers Only	$i = 1,\ldots,I$		m_i
Issuing Only	$j = 1,\ldots,J$	A_j	f_j

Given the bank fees, credit-card merchants choose the bank from which they purchase their facilities. We use an indicator function $\lambda_{i,k}^n$ to represent the bank chosen by merchant n. $\lambda_{i,k}^n = 1$ if merchant n purchases its credit card facilities from

bank *i* or *k*. Otherwise, it equals zero. If a merchant decides not to operate as a credit card merchant then it will set $\lambda_{i,k}^{n} = 0$ for all i, k.[5]

All merchants then set their prices. We make no specific assumptions about the nature of competition between the merchants but assume that each merchant n or m can only set a single price. In particular, credit-card merchants cannot set different prices for cash and credit-card purchases.[6]

Finally, given bank fees and merchant prices, each customer decides (a) whether to have a credit card; (b) if so, from which bank to gain a credit card and (c) which merchants to purchase from and the amount of purchases from each merchant. Again, we use an indicator function to represent each customer's credit-card choice. $\lambda_{j,k}^{l} = 1$ if customer *l* decides to gain credit card services from bank *j* or *k*. Otherwise, it equals zero. We denote purchases by *l* from a merchant *n* or *m* by q_n^l and q_m^l respectively where p_n and p_m are the prices paid to each merchant. A customer can use either cash and/or credit card to purchase from a merchant that offers credit card facilities, where Z_n^l represents the share of credit card purchases by customer *l* from merchant *n*. For convenience, we assume that each customer who chooses to use a credit card only obtains and uses one card. Similarly, each merchant that uses credit-card facilities only purchases such facilities from one bank.[7]

Each customer *l* makes their purchase decisions subject to a budget constraint,

---

[5] If such a merchant decided to continue in business as a cash-only merchant then it would simply be one of the *M* such merchants.

[6] This allows us to examine the situation both with and without a no surcharge rule. With such a rule a merchant is *either* a credit-card accepting merchant or a cash-only merchant. In the absence of such a rule a single merchant can be *both* a credit card accepting merchant and a cash merchant. Such a merchant would set two prices – a cash price p_m and a credit price p_n.

[7] The analysis can easily be extended to allow for consumers and/or merchants using more than one bank's facilities. However, it adds considerably to notation as each purchase must be distinguished by both the consumer, the merchant, the issuing bank and the merchant bank.

$$\sum_{j,k}\lambda_{j,k}^{l}\left[A_{j,k}+\sum_{n=1}^{N}Z_{n}^{l}p_{n}q_{n}^{l}\left(1+f_{j,k}\right)\right]+\sum_{n=1}^{N}\left(1-Z_{n}^{l}\right)p_{n}q_{n}^{l}+\sum_{m=1}^{M}p_{m}q_{m}\leq Y_{l} \qquad (1)$$

where Y_l is the customer's wealth and we use the subscript j,k to represent either non-integrated or integrated issuing banks.

Given the choices by all customers, the profits of credit card and cash-only merchants are given by:

$$\sum_{i,k}\lambda_{i,k}^{n}\sum_{l=1}^{L}\left(1-Z_{n}^{l}m_{i,k}\right)p_{n}q_{n}^{l}-C_{n}\left(\sum_{l=1}^{L}q_{n}^{l}\right) \qquad (2)$$

and

$$\sum_{l=1}^{L}p_{m}q_{m}^{l}-C_{m}\left(\sum_{l=1}^{L}q_{m}^{l}\right) \qquad (3)$$

respectively, where $C(.)$ represents the relevant merchant's cost function.

Given the choices by all customers, merchants and banks, the profit to a non-integrated issuing bank j from its credit card activities is given by:[8]

$$\sum_{l=1}^{L}\lambda_{j}^{l}\left[A_{j}+\left(f_{j}+a\right)\sum_{n=1}^{N}Z_{n}^{l}p_{n}q_{n}^{l}\right] \qquad (4)$$

Similarly, the profit to a non-integrated merchant bank i is given by

$$\sum_{n=1}^{N}\lambda_{i}^{n}\left(m_{i}-a\right)\left[\sum_{l=1}^{L}Z_{n}^{l}p_{n}q_{n}^{l}\right] \qquad (5)$$

Finally, the profit to an integrated bank k is given by

$$\sum_{l=1}^{L}\lambda_{k}^{l}\left[A_{k}+\left(f_{k}+a\right)\sum_{n=1}^{N}Z_{n}^{l}p_{n}q_{n}^{l}\right]+\sum_{n=1}^{N}\lambda_{k}^{n}\left(m_{k}-a\right)\left[\sum_{l=1}^{L}Z_{n}^{l}p_{n}q_{n}^{l}\right] \qquad (6)$$

[8] To avoid excessive notation we have set all bank costs equal to zero. It is easy to confirm that this assumption does not affect any results below.

Equations (1) to (6) define the four-party credit card system, including all agents, choice variables, prices and payoffs to each agent. We now consider the equilibrium that arises from the payments system. In particular we consider how equilibrium values of choice variables, prices and payoffs relate to the level of the interchange fee, *a*.

4. Neutrality of Interchange Fees

An equilibrium is a set of credit card charges, bank selections and purchases, $A_j^*, f_j^*, A_k^*, f_k^*, m_i^*, m_k^*, \lambda_i^{n^*}, \lambda_k^{n^*}, \lambda_j^{l^*}, \lambda_k^{l^*}, Z_n^{l^*}, q_n^{l^*}, q_m^{l^*}$, such that, given all choices made by other market participants, the relevant decision maker prefers its choice to any other feasible choice. For example, customers might maximise utility and banks and merchants might maximise profits. However, our results do not depend on either profit or utility maximisation. Rather, we impose a series of consistency axioms on all decisions:[9]

A1. When purchasing from a credit-card merchant, a customer *l* only cares about the total price of a unit purchase, $p_n(1+f_{j,k})$ and a customer's decision is invariant to changes in the components of this total price, so long as the total price remains unchanged.

A2. If a credit-card merchant chooses a bank to provide card services (and so decides to sell output) then it will choose its bank according to the relative charges $\{m_{i,k}\}_{\substack{i=1,\ldots,I \\ k=1,\ldots,K}}$. If all these fees change by the same proportion and/or by an identical constant then no credit-card merchant alters its choice of bank (although it may decide not to accept any bank's service and leave the industry).

[9] These axioms mirror those of Grant and King (1997) who derive a similar neutrality result in analysing shifts from an income to a consumption tax.

A3. Issuing banks only care about the total fees associated with providing credit card facilities for any dollar of transactions, $f_{j,k} + a$. They do not specifically care about the component parts of total per transaction fees.

A4. Acquiring banks only care about the total fees associated with providing credit card facilities for any dollar of transactions, $m_{i,k} - a$. They do not specifically care about the component parts of total per transaction fees.

These four consistency axioms are satisfied by standard utility and profit maximization assumptions, but are more general than these assumptions. They require that participants do not suffer from any 'money illusion', in that all participants only care about specific receipts or payments and not what the components of such receipts or payments are called. For example, if a customer pays $10 for a specific product at a specific store, then the customer does not care whether the payment involves $6 as a payment to the merchant and $4 as a payment for the merchant's bank, $8 to the merchant and $2 to the bank, or any other split that adds up to $10. The customer only cares about the total payment that they must make of $10.[10]

Given these axioms, we can prove the following proposition:

Proposition 1 (Neutrality). *Suppose that* $Z_n^l = 1$ *for all l, n. Then the prices charged by credit-card merchants and bank charges will change as the interchange fee changes. However, the value of the interchange fee does not affect customers' budget constraints or purchases, banks' profits or merchants' profits, or the total volume of transactions*.

The proofs of all propositions are in the appendix. The proposition provides a condition under which the interchange fee has no real effect on any market participant. In particular, all purchases from credit-card merchants must be made using credit cards.

[10] It is, of course, possible to think of situations where a participant might not satisfy these axioms. For example, if a customer simply disliked banks and would be willing to pay more for an identical product so long as the relevant bank received less. However, these situations would fall outside the bounds of standard economic analysis.

The logic behind Proposition 1 is straightforward. A rise in the interchange fee tends to increase the profits of issuing banks and decrease the profits of merchant banks. But each bank faces the same competitive options as before the change in interchange fees. The interchange fee does not alter the degree of bank competition, and so bank fees will move to offset the changes in the interchange fees. Further, as the interchange fee only indirectly affects merchants and customers, the change in bank fees only alter nominal variables – the credit-card merchant prices – and not any real variables. Thus, the changes in bank fees can completely offset the change in the interchange fee.

An alternative way to understand Proposition 1 is to note that, given the level of merchant competition, all changes in bank charges are 'passed through' to final customers. In this sense, the banks are engaged in a game against each other to maximise their objective (such as profits) given customer behaviour and a specific degree of merchant competition. Further, each bank has a price – their transaction specific credit card charges – that they can use as a strategic variable to affect customer behaviour. If an arbitrary transfer between the banks, such as the interchange fee, is altered, then this simply leads all banks to change their prices and offset the change in the interchange fee. Overall, charges and merchant prices alter, but no one is made better or worse off.

In this sense, the interchange fee is a redundant price. A credit card payment involves transfers from the customer to three parties – the merchant, the issuer and the acquirer. But there are four prices, p_n, $f_{j,k}$, $m_{j,k}$ and a. If one of these prices is altered then the equilibrium values of the other prices will change so there is no change in any real variable. For example, if the condition of the proposition holds,

any government attempt to fix the interchange fee or to eliminate the fee would have no welfare consequences and would be a waste of time.

Given its importance, it is useful to consider when the condition of Proposition 1 might reasonably arise. Here we identify two circumstances that naturally give rise it: (1) an ability to surcharge for credit card transactions and (2) perfect competition among merchants. Consider surcharging first. If there is no price constraint, such as a no surcharge rule or other transaction cost constraints, on credit card merchants, then this condition that $Z_n^l = 1$ for all (n, l) is trivial. A merchant who offers credit card facilities will also offer cash sales and will set separate cash and credit prices. A credit-card merchant who also sells for cash is simply an integrated version of one merchant n and one merchant m. This leads to the following corollary.

Corollary 2. *If there is no restriction on a merchant's ability to set separate cash and credit prices, then interchange fee will have no real effect on any economic variable.*

If, in contrast, credit-card merchants are unable to set different cash and credit prices, then the assumption that Z_n^l equals unity may arise as a result of market competition. To see this, suppose that all merchants sold a homogenous product and there was free entry. Suppose also that some merchants offer card facilities but other merchants do not. Then a cash customer will only buy from a credit merchant if the price set by that merchant is no greater than the price set by cash merchants. Suppose this was the case and Z_n^l is not equal to unity for some customer l and merchant n. By free entry, both cash and credit merchants must earn zero economic profit. Any rise in the interchange fee will lead to a rise in the credit price relative to the cash price and the cash customer will no longer purchase from the credit merchant in equilibrium. Any fall in the interchange fee will make the cash customer unprofitable for the credit merchant and the merchant will refuse to sell to that customer. Thus, any equilibrium

where Z_n^l is not equal to unity is non-generic and Z_n^l will equal unity for all customers and merchants for any other interchange fee. Put simply, if there is perfect merchant competition, a rule that limits merchants to a single cash and credit price will result in a market division between cash-only merchants and credit-only merchants. In such circumstances, the interchange fee is neutral.

5. The Incentive to Distort Interchange Fees

The previous section demonstrated that, in the absence of a no surcharge rule, or if merchants are perfectly competitive then the interchange fee has no real economic consequences. If, as suggested by Frankel (1998), banks have an incentive to systematically manipulate the interchange fee to raise profits, then this can only occur when there is *both* a no surcharge rule on credit merchants and imperfect merchant competition.

In this section, we consider situations where the condition required for Proposition 1 does not hold, so that Z_n^l is less than one for some consumers. This allows us to formally analyse the incentives that face banks to manipulate the interchange fee.

The 'cross subsidy' argument presented by Frankel (1998) states that banks would wish to raise the interchange fee under a no surcharge rule, or more generally what he terms as "price coherence."[11] The increase in the interchange fee can be

[11] Specifically, Frankel (1998, pp.316-317) writes: "A consequence of price equality across competing methods payment is the enhancement of any market power that might exist in the affected payment markets. More generally, if the price of a product moves in lock step with the price of a competing product despite changes in the relative cost of the products – a phenomenon I call "price coherence" – then a supplier with market power will be able to shift some of the incidence of its market power onto its competitors' customers. Price coherence constrains merchant choices. If the price of one brand

passed onto credit card customers by issuers through reductions in card fees. But the rise in the interchange fee will force merchants to raise the single price to both cash and credit customers. Rather than being neutral, the rise in the interchange fee will lower the relative credit card price and encourage increased use of credit cards – to the benefit of the banks.

There are two obvious problems with this simple argument. The first relates to the competition between cash and credit merchants. If we consider a single merchant *n*, the relative rise in the cash price following a rise in the interchange fee will tend to make more customers substitute to credit cards. This will tend to increase banks' profits. As such, if all merchants were credit-card merchants, it appears likely that banks would prefer higher interchange fees as this would encourage credit card adoption by customers and increase credit card transactions. However, the relative rise in the cash price from a credit card merchant will also make cash-only merchants appear to be relatively cheap for those customers that use cash as well as credit. A rise in interchange fees could lead to substitution away from credit-card merchants and, to the extent that this involves customers who previously made some but not all

---

increases, the merchant can drop the now more costly brand altogether, charge a different price for the brand than for competing products, or raise its prices for all products by the same amount. If the merchant chooses the latter course, the new price to consumers for any product will be based on a weighted average of the combined cost associated with all products. In that case, consumers will have no incremental incentive to choose the lower cost product. In other words, retail price coherence reduces the elasticity of demand facing the retailer's suppliers because any given wholesale price increase results in a smaller reduction in unit sales than would occur without price coherence.

In a market exhibiting price coherence, a supplier with market power will maximize profits at higher price levels than otherwise because the supplier can shift some of the economic burden (or "incidence") of its market power to customers who buy its competitors' products." He then goes on to posit, using a non-formal argument similar to that underlying the logic of this paper, that if retail prices could vary depending upon the payment instrument used, it would not be able to use the interchange fee as an instrument of market power (p.343).

purchases from such merchants using credit cards, it might reduce total credit card purchases.[12]

Second, if there is imperfect competition between credit card merchants, and these merchants make both cash and credit sales, then any change in the interchange fee will affect this competition. If merchants sell substitute products, a rise in the interchange fee will tend to weaken competition for both cash and credit customers under a no surcharge rule. Not only will merchants will raise their price directly due to a rise in the interchange fee but also they will raise their price as a response to the rise in the price charged by their competitors. A rise in the interchange fee might lead to new equilibrium prices that reduce both cash and credit sales and lower bank profits. This would potentially mitigate an association's incentive to set a high interchange fee.

The analysis in this section considers each of these weaknesses in turn. In order to consider the strongest argument for manipulation of credit card fees, we first consider the situation of a single merchant who accepts credit cards. This merchant can only set a single price and faces imperfect competition from cash merchants. We formalise the Frankel argument and provide both necessary and sufficient conditions for banks to find it profitable to systematically manipulate the interchange fee. These

[12] This argument is similar to the one presented in the model of Rochet and Tirole (2000). They demonstrate that, under a merchant duopoly competing in a Hotelling style model, merchants face strong incentives to adopt credit card processing when their rival does. This is because a merchant who does not process credit cards loses many customers to its rival while the adoption of credit card processing would be relatively attractive given the ability to raise prices to cash customers as well. Indeed, it is the externalities present in merchant's adoption decisions (in particular, the negative externality imposed by a rival's adoption) that generates potential rents for the association in raising interchange fees and encouraging over-use of credit cards.

conditions, in part, will depend on the credit merchant's interaction with cash merchants.[13]

The second part of this section extends the analysis to multiple credit card merchants who engage in imperfect competition. Again, we provide conditions for the banks to find manipulation of credit card fees to be profitable. We show how these conditions depend on the nature of inter-merchant competition. In particular, raising the interchange fee may harm bank profits when merchant prices are strategic complements.

The focus of our analysis is on bank manipulation of credit card fees under a no surcharge rule (or more generally 'price coherence') rather than simply bank monopoly pricing. In this sense, we address the specific issues raised by Frankel and ask under what conditions, if any, will banks with market power seek to raise interchange fees in order to raise merchant charges and merchant prices while *simultaneously* reducing direct fees to card holders?

The model used in this section is based on the general four-party payments system model presented in Section 3, albeit with some simplification in set-up to make the analysis tractable and the effects more transparent.

The Single Credit-Card Merchant Case

To capture the effect of bank market power, assume that there is a single monopoly issuer but that there is perfect competition in merchant acquiring. The issuer sets the customer charge f and the interchange fee a. Under perfect competition, the interchange fee is simply passed directly onto the merchant. As above, for

[13] This model, therefore, removes the strategic interactions that drive Rochet and Tirole's (2000) formalisation of the Frankel story and by so doing provides a potentially more general effect.

convenience we assume that the marginal cost of acquirers is zero so that the merchant services charge $m = a$ for all acquirers. As such, we can think of the acquirer as directly setting m through the interchange fee. To simplify calculations, we impose the credit card fees on a per-unit-of-sales basis rather than as a proportion of expenditure.[14]

Our starting point is market equilibrium in the absence of a 'no surcharge' rule. We begin by focusing on a single merchant n who offers credit card facilities to customers and faces imperfect competition for their product. There is perfect competition among all other merchants who set a cash price vector p. We hold p constant when considering the effect of the no surcharge rule on merchant n's behaviour. This allows us to focus on the effect of imperfect cash-only merchant competition on the issuers' incentives to manipulate the interchange fee.

For any given fees f and m, merchant n's profit is given by

$$\pi(p_n, p_n^c; f, m, p) = p_n Q + (p_n^c - m)Q^c - C(Q + Q^c)$$

where p_n and p_n^c are the cash and credit prices respectively, $Q = Q(p_n, p_n^c + f, p)$ and $Q^c = Q^c(p_n, p_n^c + f, p)$ the cash and credit sales respectively and $C(.)$ is the total cost function for the merchant.[15] We assume that Q, Q^c and C are all twice continuously differentiable with $Q_1 < 0$, $Q_2 > 0$, $Q_1^c > 0$, $Q_2^c < 0$, $Q_1 + Q_1^c < 0$ and $Q_2 + Q_2^c < 0$. The first four conditions simply mean that credit and cash transactions

[14] This type of pricing mirrors that of Baxter (1983) and Rochet and Tirole (2000). Proportional fees would mean that the issuer would like to impose a no surcharge rule if this raises total expenditure on credit card purchases. Per-unit-of-sales fees means that the issuer would like to impose the no-surcharge rule if this raises total credit card sales, as we show below. Analysing revenue changes rather than changes in sales does not alter the basic results presented below but does add to the algebraic complexity.

[15] Unlike the general model, we do not disaggregate sales down to the level of the individual consumer but deal with total cash or credit sales.

for merchant *n* are gross substitutes and have downward sloping demands. The last two conditions require that if *n* raises either its cash or credit price alone, then its total sales fall. We also make the standard assumption that *C* is increasing and convex.

The acquirers make no profit by assumption while the issuer makes profit $(f+m)Q^c - c_i(Q^c)$ where $c_i(Q^c)$ is the issuer's cost function, which depends on the volume of credit card transactions.

In the absence of a no-surcharge rule, merchant *n* will simply set its prices p_n and p_n^c to maximise profit given *f* and *m*. We assume that a unique solution to this profit maximisation exists for all *f* and *m* and denote the optimal level of credit card sales for *n* by $Q^{c*}(f,m,p)$. Given our assumptions, Q^{c*} is continuously differentiable and decreasing in both *f* and *m*.

The issuer will set the customer fee and the interchange fee (and hence the merchant fee) to maximise its profits, taking the merchant's pricing choice into account. Let the profit maximising fees be f^* and m^* with associated merchant prices p_n^* and p_n^{c*} and credit card sales $Q^{c*}(f^*,m^*,p)>0$. There is no reason why p_n^* and p_n^{c*} need to be equal. We assume that at these optimal bank fees and merchant prices $\pi > \underline{\pi}$ where $\underline{\pi}$ is the profit the merchant would gain if it refused to accept credit cards and only made cash transactions. Thus, *n* strictly prefers to accept credit cards than to refuse to accept those cards in equilibrium.

Imposing a no-surcharge rule on the merchant will clearly affect prices. For example, if $p_n^* > p_n^{c*}$ then a no-surcharge rule could lead to a rise in the credit card price and a fall in cash price. Our objective here, however, is not to consider if the no-surcharge rule will arbitrarily move merchant cash and credit prices but rather to see if a bank with market power can systematically alter fees under a no-surcharge rule to

raise its profit. In order to do this, we need a base-case where the no-surcharge rule, by itself, is neutral. From Proposition 1 and Corollary 2, we know that such a base case exists. In the absence of the no surcharge rule, the issuer can set credit card fees so that it maximises profits and $p_n^* = p_n^{c*}$.[16] We use this as our benchmark for analysing the potential for a cross-subsidy under the no-surcharge rule. In this situation, simply imposing the no-surcharge rule, given f^* and m^*, will have no effect on merchant prices – the merchant is already profit-maximising and trivially satisfying the no-surcharge rule.

A 'cross-subsidy' will arise at this benchmark under the no-surcharge rule if the issuer has an incentive to increase a (and thus m) while simultaneously reducing f by the same amount. In other words, a cross-subsidy will occur if there exists a $\Delta > 0$ such that the issuer strictly prefers fees $m^* + \Delta$ and $f^* - \Delta$ to f^* and m^* under the no-surcharge rule.

Under the no surcharge rule, demand for credit card transactions is given by $Q^c = Q^c(\rho, \rho + f^* - \Delta, p)$ while merchant n's profits under the no surcharge rule is denoted by $\hat{\pi}(\rho, \rho + f^* - \Delta; f^* - \Delta, m^* + \Delta, p)$ where $\rho = p_n = p_n^c$ is the common price set by the merchant under the no surcharge rule and where we restrict attention to bank fees that involve a potential 'cross subsidy.' Let $\hat{\pi}_1$ and $\hat{\pi}_2$ denote the partial derivates of $\hat{\pi}$ with respect to its first and second arguments respectively, with similar

---

[16] To see this note that Proposition 1 implies that, without a no surcharge rule, if a, and hence m increases by Δ while f simultaneously decreases by Δ, then the profit maximizing cash price for merchant n is unchanged while the profit maximising credit price rises by Δ. Further, this change has no real effects and changes no firm's profit. Thus, for any m, f and associated p_n, p_n^c where $p_n \neq p_n^c$, $p_n - p_n^c = \Delta$, we know that there exists prices and fees $m + \Delta$, $f - \Delta$ and associated p_n, p_n^c where $p_n = p_n^c$ and the profits for all firms is unchanged. Without loss of generality, we consider the optimal bank fees m^*, f^* such that $p_n^* = p_n^{c*}$.

notation for second-order partial derivatives. Similarly, let $\hat{\pi}_\rho$ and $\hat{\pi}_{\rho\rho}$ refer to the first order and second order derivatives of $\hat{\pi}$ with respect to ρ. We know that at $\Delta = 0$ there is a unique solution to the merchant's price setting problem so that $\hat{\pi}_\rho(\rho^*, \rho^* + f^*; f^*, m^*, p) = 0$ and $\hat{\pi}_{\rho\rho}(\rho^*, \rho^* + f^*; f^*, m^*, p) < 0$ where $\rho^* = p_n^* = p_n^{c*}$. At these prices $\hat{\pi}_\rho = \pi_1 + \pi_2$ and $\hat{\pi}_{\rho\rho} = \pi_{11} + 2\pi_{12} + \pi_{22}$. Further, note that $\hat{\pi} = \rho Q(\rho, \rho + f - \Delta) + (\rho - m - \delta)Q^c(\rho, \rho + f - \Delta) - c(Q + Q^c)$. As Δ only enters profit as $p_n^c - \Delta = \rho - \Delta$ then at $\Delta = 0$ and $\rho^* = p_n^* = p_n^{c*}$, $\hat{\pi}_\delta = -\pi_2$ and $\hat{\pi}_{\rho\delta} = -\pi_{\rho 2}$. We denote the optimal quantity of credit card sales for the merchant under the no-surcharge rule with f^* and m^* by $\hat{Q}^{c*}(f^* - \Delta, m^* + \Delta, p)$ and use the notation $\hat{Q}_1^{c*}$, $\hat{Q}_2^{c*}$, and $\hat{Q}_\Delta^{c*}$ to refer to the derivatives with regards to the first and second arguments and with regards to Δ.

We begin by showing that the monopoly issuer will have an incentive to raise Δ from zero if and only if this leads to an increase in the number of credit card transactions at merchant n.

Proposition 3. *The issuer's profits are increasing in* Δ *at* $(f, m, \Delta) = (f^*, m^*, 0)$ *if and only if* $\hat{Q}_\Delta^{c*} > 0$.

This proposition demonstrates that interchange fees will be set higher so long as this results in a greater volume of credit card transactions. This result reflects the issuer's market power. In the absence of the no surcharge rule, the issuer sets marginal card fees to customers and (through the interchange fee) to merchants above marginal cost. Raising the interchange fee while simultaneously lowering the customer fee does not alter issuer profits given the number of card transactions. But if re weighting the fees,

encourages more card transactions, the issuer's profits will rise.[17] If the issuer has constant marginal costs then it follows from the proof of Proposition 3 that profit maximising fee manipulation will be maximal. If the issuer finds it profitable to raise the interchange fee, it will keep raising this fee until the merchant is just indifferent between accepting credit cards or rejecting cards and relying solely on cash transactions.

How likely is it that a rise in the interchange fee will increase the volume of credit card transactions? The following proposition states necessary and sufficient conditions for this to be the case.

Proposition 4. $\hat{Q}_{\Delta}^{c*} > 0$ *if and only if* $\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right) < 0$. *In addition,* $\hat{Q}_{\Delta}^{c*} > 0$ *if any of the following conditions is satisfied:*

1. $\pi_{12} \leq 0$ *or* $\pi_{12} > 0$ *but* $\hat{\pi}_{\rho 1} = \pi_{11} + \pi_{21} < 0$ *and* $\hat{\pi}_{\rho 2} = \pi_{22} + \pi_{21} < 0$.
2. $\hat{\pi}_{\rho 1} < 0$ *and* $Q_1^c + Q_2^c < 0$.
3. $\hat{\pi}_{\rho 2} < 0$ *and* $Q_1^c + Q_2^c > 0$.

The first sufficient condition is relatively standard. It requires that the cross-partial derivatives of the profit function for the merchant with regards to cash and credit prices not be too large. The second condition is also intuitive. It requires that an equal rise in both the cash and credit prices for the merchant leads to a fall in the credit sales by the merchant. The third condition requires the opposite to hold and is less obvious.

The above proposition only considers a marginal increase in the interchange fee. However, if the relevant conditions on the merchant's profit hold globally then the monopoly issuer will have an incentive to continue to raise the interchange fee.

---

[17] This result is similar to Schmalensee (2001, p.11) who notes that in his framework, under certain conditions, the interchange fee that maximises banks' profit also maximises "total system output." Schmalensee's result is driven by double marginalisation between separate monopoly issuers and acquirers and the symmetric nature of demand for credit card transactions in his model. Our result reflects the integrated nature of card transactions. The issuer gains at the margin by encouraging these transactions because it receives benefits directly from consumers and indirectly from merchants via the interchange fee.

Proposition 5. *Suppose* $c_i(\hat{Q}^{c*}) = c_i\hat{Q}^{c*}$ *and one of the conditions from Proposition 5 is satisfied globally for merchant n. Then the monopoly acquirer will raise* Δ *until profits equal* $\underline{\pi}$.

The issuer will continue to have an incentive to raise the volume of credit card transactions so long as $f^* + m^* - \frac{\partial c_i}{\partial \hat{Q}^{c*}} > 0$. For example, in Proposition 5, we assume that there are constant marginal costs for the issuer so that this condition always holds as the volume of credit card transactions expand. As a consequence, if the volume of credit card transactions increase in Δ then the issuer will seek to raise Δ until the merchant is just indifferent between accepting or rejecting credit cared transactions.

The issuer has an incentive to raise Δ under the no-surcharge rule because this forces the merchant to 'average' over cash and credit prices. To see this, we can consider an alternative implementation of the no surcharge rule. Suppose that the issuer chooses fees f^* and m^* so that $p_n^{c*} > p_n^*$. Further, suppose that $\pi_{12} < 0$ for all p_n, p_n^c. Then the imposition of the no surcharge rule will lead to a uniform cash and credit price ρ^* where $p_n^{c*} > \rho^* > p_n^*$.[18] However, given that $Q_1^c > 0$ and $Q_2^c < 0$, the averaging of the cash and credit price under the no-surcharge rule will lead to greater credit card sales and profit for the monopoly issuer.

Proposition 4 shows that a monopoly issuer, *even when facing a single merchant that accepts credit cards*, might not find it profitable to manipulate credit card fees through the interchange fee. In fact, if the necessary and sufficient condition presented in Proposition 4 is violated it will pay the issuer to *lower* interchange fees. In other words, the issuer might find it profitable to lower the interchange fee while simultaneously raising customer fees in order to raise the volume of card transactions.

[18] This is formally proven in Proposition 7 below.

That said, the conditions for profitable manipulation are relatively weak in this single merchant setting. Manipulation of the credit card fees increases the number of credit card transactions. In fact, our results show that fee manipulation is only profitable for the issuer when it raises the volume of credit card transactions.

Manipulation of the credit card fees under a no-surcharge rule lowers the merchant's profits. Under the no-surcharge rule, the merchant is unable to respond to a rise in Δ by independently altering the credit card retail price. As a result, the merchant faces a constrained pricing choice and lower profits. In the extreme, the issuer will seek to distort credit card fees until the merchant is just indifferent between accepting credit cards or relying on cash-only transactions.

The effect of competition from the cash-only sector can be seen in our model. For example, if the merchant is a price taker with regards to the cash price then both Q_1^c and $\hat{\pi}_{\rho 1}$ are equal to zero. Any attempt to increase the cash price p_n will simply lead to a loss of all cash custom. There will be no increase in credit card sales or change in profit for merchant *n*. There is no benefit to the issuer from raising Δ. More generally, if Q_1^c and $\left|\hat{\pi}_{\rho 1}\right|$ are small, the ability for the issuer to increase credit card transactions by manipulating the interchange fee under the no surcharge rule will be limited. This suggests that, in general, competition from other payment instruments (including cash, check, debit cards and charge cards) will limit an association's incentive to manipulate its interchange fee.

In some credit card models (e.g., Rochet and Tirole, 2000) banks only gain from manipulating credit card fees by increasing the number of consumers who take-up a credit card. While this effect is included in our model, our analysis also shows that additional customer take-up of credit cards is not necessary to make fee manipulation profitable for the issuer. To see this, suppose that the cash and credit

card markets were independent in the sense that cash sales only depend on the cash price and credit card sales only depend on the credit price. Then so long as the merchant faces a downward sloping demand curve for cash and credit sales, the issuer will always find it profitable to manipulate credit card fees. Formally, the necessary and sufficient condition presented in Proposition 4 reduces to $-\pi_{11}Q_2^c < 0$ and this will always be satisfied.

Intuitively, if the cash and credit markets are independent, the no surcharge rule simply forces the merchant to raise its cash price whenever it wishes to raise its credit price. Interchange fee manipulation requires the merchant to raise the credit card price to maintain its profits from credit sales. But it can only do this by raising its cash price and lowering the profit from cash sales. The merchant will profit maximise by raising its credit price by less than it would in the absence of the no-surcharge rule in order to reduce the loss of profits from cash sales.[19]

Multiple Credit Card Merchants

The intuitive argument for the systematic manipulation of credit card fees often implicitly assumes the type of analysis considered in the single-merchant case. However, the case for distortion of credit card fees would be weak if it only held when applied to a single merchant. In this section we extend the above analysis to multiple merchants. We start by considering two merchants that are completely symmetric. We then consider asymmetric interactions between merchants.

---

[19] This effect is similar to a firm that is suddenly prevented from engaging in third degree price discrimination. Customers who paid a higher price under discrimination tend to pay a lower price after discrimination is prevented, while the opposite holds for customers who paid a lower price under discrimination. See Varian (1985).

Symmetric Merchants: Suppose two merchants, *n* and *s*, both accept credit cards. Both merchants face imperfect competition for their products while all other merchants set only cash prices and are perfectly competitive. As above, the cash prices set by all other merchants are given by the vector *p*. We assume that both merchants *n* and *s* are involved in symmetric but imperfect competition. In other words, the merchants each face identical cost and demand conditions and all equilibrium outcomes for one firm are the same as for the other firm. The merchants independently and simultaneously set both cash and credit prices, p_n, p_s, p_n^c and p_s^c. Merchant cash and credit sales are denoted by Q_n, Q_s, Q_n^c, and Q_s^c respectively.

Our analysis uses similar notation to above. Initially, there is no limitation on the ability of merchants to charge separate cash and credit prices. The issuer sets a single interchange fee and this is translated into a single merchant service fee. The issuer also sets a single fee for credit card users. As before, we denote the profit maximizing fees in the subgame perfect equilibrium by f^* and m^*. As above, without loss of generality we can consider the equilibrium fees so that $p_n^* = p_n^{c*} = \rho^*$ in the absence of a no surcharge rule. By symmetry, $p_s^* = p_s^{c*} = \rho^*$ also holds at these fees. Unlike above, however, (p_n^*, p_n^{c*}) and (p_s^*, p_s^{c*}) are mutual best responses.

As above, we consider the interaction between the no-surcharge rule and credit card fees by introducing Δ. The 'cross-subsidy' argument requires that the introduction of the no-surcharge rule will be associated with an increase in Δ by the monopoly issuer. For a single merchant this required that credit card sales were increasing in Δ. It is trivial to extend Proposition 3 to show that in the case of multiple symmetric merchants, the issuer will only find it profitable to raise Δ above zero if $\left(Q_n^{c*} + Q_s^{c*}\right)$ is increasing in Δ under the no-surcharge rule.

It is necessary to characterise the strategic interaction between the two merchants under the no-surcharge rule. Once the rule is in place, each merchant can only charge a single price, ρ_n and ρ_s. We assume that the merchants provide substitute products, in that $\frac{\partial Q_n}{\partial \rho_s} \geq 0$ and $\frac{\partial Q_s^c}{\partial \rho_n} \geq 0$. Thus merchant *n*'s cash and credit card sales are (weakly) increasing in single price set by merchant *s*. By symmetry, sales for merchant *s* are also increasing in *n*'s price. Further, as the merchants are selling substitute products we would expect prices to react as strategic complements. The profit maximising price for merchant *n*, ρ_n^*, will be increasing in ρ_s for any credit card charges, and vice versa. Thus $\frac{d\rho_n^*}{d\rho_s} > 0$.

It is useful to define the degree of strategic complementarity by θ where $\theta = \left(1 - \frac{d\rho_s}{d\rho_n}(1 + \frac{d\rho_s}{d\rho_n})\right)$. We assume that $\theta > 0$. This guarantees that the interaction between the two merchants is stable under variations in Δ. Note that if $\frac{d\rho_n^*}{d\rho_s} = 0$, then $\theta = 1$ so that under our assumption of strategic complementarity, $\theta \in (0,1)$.

For convenience, we focus on one of the conditions presented in Proposition 4 and assume that $\frac{\partial^2 \hat{\pi}_r}{\partial \rho_r \partial p_r} < 0$ and $\frac{\partial^2 \hat{\pi}_r}{\partial \rho_r \partial p_r^c} < 0$ for $r = n, s$. Under this assumption, the monopoly issuer would always wish to manipulate credit card fees if there was only a single merchant accepting credit cards. However, this need not occur when two merchants strategically interact. The following proposition provides a set of conditions relating to the degree of strategic complementarity and the demand interactions for the merchants.

Proposition 6. *If* $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right) \geq 0$ *at* $(f, m, \Delta) = (f^*, m^*, 0)$ *then the monopoly issuer will always find it profitable to set* $\Delta > 0$ *and manipulate credit card fees. If* $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right) < 0$ *at* $(f, m, \Delta) = (f^*, m^*, 0)$ *then there exists a* $\tilde{\theta} > 0$ *such that the*

issuer will only find it profitable to manipulate credit card fees if $\theta > \tilde{\theta}$. *If* $\theta < \tilde{\theta}$ *then the issuer will not find it profitable to set* $\Delta > 0$.

The proof of Proposition 6 is given in the appendix. It shows that if θ is relatively large, so that $\frac{d\rho_n^*}{d\rho_s}$ is small and strategic interaction between the two merchants is weak, then it always pays the issuer to raise Δ. But if θ is close to zero (i.e. $\frac{d\rho_n^*}{d\rho_s}$ is relatively large, albeit still well under unity to satisfy our assumption on θ) *and* if a certain cross derivative condition is satisfied then a rise in Δ will lead to a fall in total credit card sales. In other words, if merchant prices are strongly complimentary and $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right) < 0$, then any attempt to manipulate the credit card fees and exploit a 'cross subsidy' from cash customers will be unprofitable for the monopoly issuer.

Proposition 6 shows that, under certain conditions, the issuer will not find it profitable to raise Δ. It remains to show, however, that these conditions can be satisfied. In other words, we need to show that when θ is close to zero that $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right)$ can be less than zero.

To show that this is possible, note that it is reasonable that $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c}\right) < 0$. This condition simply requires that an equal increase in the cash and credit price for one merchant leads to a fall in that merchants credit sales. However, for substitutes, $\frac{\partial Q_n^c}{\partial p_s} > 0$. Hence, $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right)$ will be less than zero if $\frac{\partial Q_n^c}{\partial p_s}$ is a relatively small positive number. It remains to show that this is possible when θ is close to zero.

For θ to be close to zero, $\frac{d\rho_n^*}{d\rho_s}$ must be close to $\frac{1}{2}(\sqrt{5}-1)$ or approximately 6.18. Thus, it is only required that $\frac{d\rho_n^*}{d\rho_s}$ is close to a well-defined finite real number. From the total derivative of the first order conditions for merchant profit

maximisation, if $\frac{d\rho_n^*}{d\rho_s} > 0$ then $\frac{\partial^2 \hat{\pi}_n}{\partial \rho_n \partial \rho_s} > 0$. For simplicity, consider the special case where all inter-merchant second-order cross-price effects are zero and each merchant faces constant marginal costs.[20] Then $\frac{\partial^2 \hat{\pi}_n}{\partial \rho_n \partial \rho_s} = \frac{\partial Q_n}{\partial \rho_s} + \frac{\partial Q_n^c}{\partial \rho_s}$. This is positive as required. Further, if $\frac{\partial Q_n}{\partial \rho_s}$ is large then it is possible that $\frac{d\rho_n^*}{d\rho_s}$ will be relatively large even though $\frac{\partial Q_n^c}{\partial \rho_s}$ is relatively small. In other words, if $\frac{\partial Q_n}{\partial \rho_s}$ is relatively large then it is possible that θ is close to zero at the same time as $\left(\frac{\partial Q_n^c}{\partial p_n} + \frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_s}\right)$ is less than zero.

The problems that arise for the monopoly issuer when manipulating the credit card fees are illustrated by this special case. When the issuer raises Δ this has a direct effect of making credit card transactions relatively cheap compared to cash transactions. But it also leads to a rise in the single price set by each merchant. If merchant prices are strategic complements, these price changes lead to further rises in merchant prices. From merchant n's perspective, if $\frac{\partial Q_n}{\partial \rho_s}$ is relatively large and $\frac{\partial Q_n^c}{\partial \rho_s}$ is relatively small, the rise in the price of merchant s has a strong effect on cash sales but a relatively weak effect on credit card sales. In contrast, the rise in merchant n's own price can lead to a relatively strong decline in credit card sales. The gain in cash sales make it worthwhile for each merchant to raise their own price in response to a rise in the other merchant's price even though this may lead to a decline in credit card sales. While this trade off between cash and credit sales is worthwhile for the merchants, it is unprofitable for the credit card issuer. The issuer only gains profits on total credit card sales so if the mutual rise in merchant prices lead to a fall in these

[20] So that $C'' = 0$ and all terms like $\frac{\partial^2 Q_n^c}{\partial p_n \partial \rho_s}$ equal zero.

sales then issuer profits will decline. The issuer does not care about the even greater rise in cash sales – these do not help its profits.

Asymmetric Merchants: The analysis above provided conditions under which systematic manipulation of credit card fees can raise bank profits under a no-surcharge rule with either a single merchant or multiple symmetric merchants. The situation is significantly more complex with multiple asymmetric merchants that accept credit cards.

As above, suppose two merchants, *n* and *s*, both accept credit cards. We assume that $\frac{\partial Q_r^c}{\partial p_r} > 0$ and $\frac{\partial Q_r^c}{\partial p_r^c} < 0$ for $r = n, s$ and that the profit functions of both merchants are strictly concave with $\frac{\partial^2 \pi_r}{\partial p_r \partial p_r^c} \leq 0$ for $r = n, s$. Unlike above however, these merchants need not be symmetric.

Initially assume that there is no restriction on merchant pricing. Given the credit card fees, the merchants simultaneously set their prices in a Nash equilibrium. The acquirer sets the customer fee *f* and the interchange fee *a* (and hence the merchant service fee, *m*) to maximise its profit from the merchant pricing subgame. As before there is an extra degree of freedom available to the issuer in the absence of a no-surcharge rule. But with asymmetric competition there will not, in general, be profit maximising credit card fees that set cash and credit prices equal for *both* merchants. While credit card fees may align one merchant's prices, these same fees will not, in general, align the other merchant's prices. Without loss of generality, assume that at the optimal fees f^* and m^*, the equilibrium merchant prices have $p_n^* = p_n^{c*}$ and $p_s^* < p_s^{c*}$.

As the cash and credit prices for at least one merchant will differ in the absence of a no-surcharge rule, it is not possible to consider a marginal distortion of

credit card fees as above. In other words, we cannot simply consider whether it is profitable for the monopoly issuer to raise Δ. Rather, the following proposition considers (for relatively restrictive conditions) whether it is profitable for the issuer to introduce the no-surcharge rule.[21]

Proposition 7. *Suppose that each merchant's cash and credit demands are independent of the other merchant's prices. Then introducing the no-surcharge rule raises the issuer's profit.*

Proposition 7 does not allow for any strategic interaction between firms. As such, it provides a simple multiple firm analogue of the single merchant results and ignores the issues of strategic interaction raised in Proposition 6. More general results will depend on the exact nature of merchant interaction and are beyond the scope of this paper. As already shown above, even with symmetric firms, strategic interaction can undermine the profitability of manipulating credit card fees. With asymmetric competition, there is greater scope for any attempt by the monopoly issuer to force cash customer to cross subsidise credit-customers, to fail.

6. Conclusions

In this paper, we have formally analysed two key issues in the debate over credit card interchange fees and price restrictions. First, we considered the arguments surrounding the neutrality of interchange fees. Under what conditions on bank or merchant competition, customer behavior and pricing are interchange fees neutral. We showed that interchange fees are neutral regardless of the degree of bank or merchant

[21] Issuer profit in the absence of the no-surcharge rule is given by $(f^* + m^* - c_i)(Q_j^c + Q_k^c)$. Given the credit card fees, it will pay the acquirer to introduce the no-surcharge rule if this increases total credit card sales. Of course, the introduction of the no-surcharge rule will most likely change the optimal credit card fees for the issuer. But this can only further raise issuer profit. Thus, raising credit card sales given the credit card fees is a sufficient condition for the no-surcharge rule to raise issuer profit.

competition if no customer who uses cash pays a 'credit card' price. This immediately implies that, in the absence of any constraint (such as a no surcharge rule) that ties a merchant's cash and credit prices, interchange fees are always neutral.

Second, we considered the arguments surrounding the desire of associations to manipulate credit card fees when there is a no surcharge rule. We showed that, even if there is only a single merchant that accepts credit cards, an association might not want to raise the interchange fee to create a 'cross subsidy' between cash and credit customers. Further, if there are multiple credit card merchants who engage in imperfect competition, then the conditions for fee manipulation to raise bank profits become significantly stronger.

Our analysis of fee manipulation involved a monopoly issuer and perfectly competitive acquirers. However, our results do not depend on these specific assumptions. For example, if there was a monopoly acquirer and card issuers were perfectly competitive, then the acquirer would seek to engage in fee manipulation under a no surcharge rule in exactly the same way as shown above. The acquirer would raise m directly. To lower f the acquirer would also *raise* the interchange fee that it paid to the issuers. In other words, the acquirer would appear to raise its own costs and then to pass this through to the merchants. Of course, the rise in interchange fees is simply an indirect way for the acquirer to force the issuers to lower their fees to customers. The acquirer in fact gains greater profit after such a change in the credit card fees.

The Cruickshank (2000, p.81) noted an apparent willingness of acquirers to accept rises in interchange fees because they can be 'passed on' to merchants. The argument, however, that such pass through must reflect a weak bargaining position is shown to be fallacious by our model. Even a monopoly acquirer would appear to raise

interchange fees under our model. It is the interaction between the credit card fees and the no surcharge rule that drives the profit maximising fees in our model.

The results presented in this paper have significant policy implications. First, if authorities are concerned about banks manipulating credit card fees in the way suggested by Frankel (1998), then these concerns can be easily allayed. The authorities simply need to prevent card systems from requiring merchants to tie cash and credit prices together. In the absence of such a rule, the interchange fee is neutral regardless of the degree of bank and merchant competition. Interestingly, in a recent press release, the EU decided that a 'no discrimination' rule tying cash and credit prices created no competition concerns but that the cooperative setting of interchange fees represented an anti-competitive arrangement (European Commission, 2000). This conclusion is the exact opposite of that suggested by our analysis. In the absence of a no surcharge rule, cooperative setting of interchange fees cannot have any anticompetitive effect.

Even in the presence of a no surcharge rule, the setting of interchange fees only creates competitive concerns if there is inadequate retail level competition. In the presence of strong competition, any attempt to systematically distort interchange fees will simply split the market into competing cash and credit markets and will not raise banks' profits.

If there is both imperfect retail competition and a no surcharge rule linking cash and credit prices, then manipulating interchange fees can raise banks' profits. However, even for a single credit merchant, such manipulation will not always be profitable. With multiple merchants accepting credit cards, any change in banks' profits will depend on the nature of merchant interaction. It is quite possible that raising interchange fees will lower banks' profits.

In this sense, our paper shows that even with a no surcharge rule, the scope for any anticompetitive abuse of interchange fees is an empirical matter. It cannot be stated *a priori* that banks will or will not prefer higher interchange fees under a no surcharge rule. Specifically, when looking to the market power of associations the critical assessment lies not so much in the issuing and acquiring segments as in the level of market power among merchants.

Appendix

Proof of Proposition 1:

Consider the equilibrium $A_j^*, f_j^*, A_k^*, f_k^*, m_i^*, m_k^*, \lambda_i^{n*}, \lambda_k^{n*}, \lambda_j^{l*}, \lambda_k^{l*}, Z_n^{l*}, p_n^*, p_m^* q_n^{l*}, q_m^{l*}$ for an interchange fee a where $Z_n^l = 1$ for all n, l. Alternatively, let $A_j^{**}, f_j^{**}, A_k^{**}, f_k^{**}, m_i^{**}, m_k^{**}, \lambda_i^{n**}, \lambda_k^{n**}, \lambda_j^{l**}, \lambda_k^{l**}, Z_n^{l**}, p_n^{**}, p_m^{**} q_n^{l**}, q_m^{l**}$ represent the equilibrium for an interchange fee $a+\Delta$. We first show that $f_{j,k}^{**} = f_{j,k}^* - \frac{\Delta(1+f_{j,k}^*)}{1-a}$ and $A_{j,k}^{**} = A_{j,k}^*$ for each issuing bank j,k, $m_{i,k}^{**} = m_{i,k}^* + \frac{\Delta(1-m_{i,k}^*)}{1-a}$ for each merchant bank i,k, and $q_n^{l**} = q_n^{l*}$, $q_m^{l**} = q_m^{l*}$, and $Z_n^{l**} = 1$ for all l,m,n.

To see this, suppose that the banks did set these charges. Further, suppose that each cash merchant set an unchanged price under the new interchange fee and each customer still made all purchases from a credit-card merchant using credit card. Then, from each customer's perspective, any purchase from a credit card merchant at a specific price involves an effective price that is equal to $\frac{1+f_{j,k}^{**}}{1+f_{j,k}^*}$ times the effective price before the change in the interchange fee. Or, from the credit-card merchant's perspective, the change in customer credit-card fees is equivalent to them facing a new consumer demand function $\tilde{q}_n^l(\tilde{p}_n) = q_n^l\left(\frac{1+f_{j,k}^{**}}{1+f_{j,k}^*} p_n\right)$. Thus, from (2), each credit-card merchant's profit under these assumptions about banks' and cash-merchants' charging, can be written as $\sum_{i,k} \lambda_{i,k}^n \sum_{l=1}^{L} \left(1-m_{i,k}^{**}\right) p_n \tilde{q}_n^l - C_n\left(\sum_{l=1}^{L} \tilde{q}_n^l\right)$. Let $\rho_n = \frac{1+f_{j,k}^{**}}{1+f_{j,k}^*} p_n$. Then, the profit of a credit card merchant can be written as $\sum_{i,k} \lambda_{i,k}^n \sum_{l=1}^{L} \left(1-m_{i,k}^{**}\right) \rho_n \frac{1+f_{j,k}^*}{1+f_{j,k}^{**}} q_n^l(\rho_n) - C_n\left(\sum_{l=1}^{L} q_n^l(\rho_n)\right)$. But note that $\frac{1+f_{j,k}^{**}}{1-m_{i,k}^{**}} = \frac{1+f_{j,k}^*}{1-m_{i,k}^*}$ for all i,j,k at the postulated fees. So the credit-card merchant's profit can be written as $\sum_{i,k} \lambda_{i,k}^n \sum_{l=1}^{L} \left(1-m_{i,k}^*\right) \rho_n q_n^l(\rho_n) - C_n\left(\sum_{l=1}^{L} q_n^l(\rho_n)\right)$. This holds for all credit-card merchants. Hence, given our assumption about cash merchants, all credit-card merchants face an identical profit function to that faced before the interchange fee altered, with the exception that ρ_n has replaced p_n. But as p_n^* represented an initial equilibrium for all n, $\rho_n = p_n^*$ must represent an equilibrium under the new interchange fees. Or in other words, given our assumptions, $p_n^{**} = \frac{1+f_{j,k}^*}{1+f_{j,k}^{**}} p_n^*$. Substituting in for the fees, this means that $p_n^{**} = \frac{1-a}{1-a-\Delta} p_n^*$. By substitution, we can also show that $p_n^{**} = \frac{1-m_{i,k}^*}{1-m_{i,k}^{**}} p_n^*$.

Now, suppose that the credit card merchants did, in fact, set these prices p_n^{**} and all cash merchants set unchanged prices. From (1) the total price to any customer *l* of a purchase of one unit from a credit-card store n is given by $\left(1+f_{j,k}^{**}\right)p_n^{**}$. But, $\left(1+f_{j,k}^{**}\right)p_n^{**}=\left(1+f_{j,k}^{*}\right)p_n^{*}$ so that the budget set for each customer *l* is identical at the new prices and fees to the budget set before the change in interchange fee except for a change in the components of the price of a credit-card merchant. But under axiom 1 the customer only cares about the total price, not its components, so that no customer's decision problem is altered and $q_n^{l**}=q_n^{l*}$, $q_m^{l**}=q_m^{l*}$, $\lambda_{j,k}^{l**}=\lambda_{j,k}^{l*}$, $Z_n^{l**}=Z_n^{l*}=1$ for all customers *l*.

Returning to the cash-only merchants, given the postulated bank fees and new prices p_n^{**} for all credit-card merchants, customer behavior is unchanged with regards to the cash merchants. Thus, their profits, (3), are unchanged and setting their new prices at the original prices must remain an equilibrium. So $p_m^{**}=p_m^{*}$ is an equilibrium given the postulated bank fees.

Finally, returning to the credit-card merchant's choice of bank, under the postulated bank charges, $m_{i,k}^{**}=\frac{1}{1-a}\left(1-a-\Delta\right)m_{i,k}^{*}+\frac{\Delta}{1-a}$. Thus, under axiom 2, all credit-card merchants who choose a bank will make the same choice of bank after the interchange fee is changed as under the original interchange fee. Further, as profit for the credit card merchant is unchanged if they choose their original bank there is no reason for any credit-card merchant to alter their decision and exit the industry. Thus, $\lambda_{i,k}^{n**}=\lambda_{i,k}^{n*}$ for all credit-card merchants *n*.

So far we have shown that if $a+\Delta$, $f_{j,k}^{**}=f_{j,k}^{*}-\frac{\Delta\left(1+f_{j,k}^{*}\right)}{1-a}$, $A_{j,k}^{**}=A_{j,k}^{*}$, and $m_{i,k}^{**}=m_{i,k}^{*}+\frac{\Delta\left(1-m_{i,k}^{*}\right)}{1-a}$ then an equilibrium for customers and merchants involves $q_n^{l**}=q_n^{l*}$, $q_m^{l**}=q_m^{l*}$, $\lambda_{i,k}^{n**}=\lambda_{i,k}^{n*}$, $\lambda_{j,k}^{l**}=\lambda_{j,k}^{l*}$, $Z_n^{l**}=Z_n^{l*}=1$, $p_m^{**}=p_m^{*}$ and $p_n^{**}=\frac{1-a}{1-a-\Delta}p_n^{*}$. It remains to show that the postulated charges represent an equilibrium for the banks given the subsequent merchant and customer behaviour.

First, consider the merchant banks *i*. Suppose all other banks have set the postulated fees. Using the same change of variable as was used for credit-card merchants above, $\rho_n=\frac{1+f_{j,k}^{**}}{1+f_{j,k}^{*}}p_n$, and noting that $\frac{1+f_{j,k}^{**}}{1+f_{j,k}^{*}}=\frac{1-a}{1-a-\Delta}$ for all j,k, each merchant bank's profit can be written as $\sum_{n=1}^{N}\lambda_i^n\left(m_i-a-\Delta\right)\frac{1-a}{1-a-\Delta}\left[\sum_{l=1}^{L}\rho_n q_n^l\right]$. Let $\left(\tilde{m}_i-a\right)=\left(m_i-a-\Delta\right)\frac{1-a}{1-a-\Delta}$. Then each merchant bank's profit is $\sum_{n=1}^{N}\lambda_i^n\left(\tilde{m}_i-a\right)\left[\sum_{l=1}^{L}\rho_n q_n^l\right]$. But these profit functions are the same as the original profit functions except for a renaming of variables. So an equilibrium exists for all merchant banks where $\tilde{m}_i=m_i^{*}$ for all *i*. Thus, $m_i^{*}-a=\left(m_i^{**}-a-\Delta\right)\frac{1-a}{1-a-\Delta}$, or $m_i^{**}=\frac{1}{1-a}\left(m_i^{*}-am_i^{*}-\Delta m_i^{*}+\Delta\right)$. But this is just the postulated equilibrium value of the merchant bank fee for bank *i*.

Now, consider issuer banks *j* given that all other banks have set the postulated fees. When an issuer bank sets its fees it affects customer demand as seen by the merchants and thus prices. As customer demand for a given credit card merchant price p_n depends on the credit card adjusted price, for a general issuer bank fee we can write customer demand as $q_n^l\left(\left(1+f_{j,k}\right)p_n\right)$. Thus, issuer bank profit (5) can be written as $\sum_{l=1}^{L}\lambda_j^l\left[A_j+\left(f_j+a\right)\sum_{n=1}^{N}Z_n^l p_n q_n^l\left(\left(1+f_j\right)p_n\right)\right]$. Noting that p_n depends on the merchant banks' charges as well as the issuing bank fees, when $m_{i,k}=m_{i,k}^*$ the equilibrium charges for the issuing banks are A_j^* and f_j^*.

Given new merchant bank fees, the relationship between p_n and f_j will change. In particular, let $\tilde{p}_n=\frac{\left(1-m_{i,k}^{**}\right)}{\left(1-m_{i,k}^{*}\right)}p_n$. This equation is meaningful as $\frac{1-m_{i,k}^{**}}{1-m_{i,k}^{*}}=\frac{1+f_{j,k}^{**}}{1+f_{j,k}^{*}}=\frac{1-a-\Delta}{1-a}$ for all i,j,k. Given the new merchant bank fees and assuming that $Z_n^l=1$ for all n,l, credit card merchants' individual profits can be written as:

$$\sum_{i,k}\lambda_{i,k}^{n}\sum_{l=1}^{L}\left(1-m_{i,k}^{**}\right)p_n q_n^l-C_n\left(\sum_{l=1}^{L}q_n^l\right)$$
$$=\sum_{i,k}\lambda_{i,k}^{n}\sum_{l=1}^{L}\left(1-m_{i,k}^{*}\right)\tilde{p}_n q_n^l\left(\left(1+f_{j,k}\right)\frac{1-m_{i,k}^{*}}{1-m_{i,k}^{**}}\tilde{p}_n\right)-C_n\left(\sum_{l=1}^{L}q_n^l(.)\right)$$

Further, let $1+\tilde{f}_{j,k}=\left(1+f_{j,k}\right)\frac{1-m_{i,k}^{*}}{1-m_{i,k}^{**}}$ then the right hand side of merchant profits with $m_{i,k}=m_{i,k}^{**}$ is identical to profit with $m_{i,k}=m_{i,k}^{*}$ except for the merchant price and issuer bank fees being replaced by $\tilde{p}_n$ and $\tilde{f}_{j,k}$ respectively. Thus, the relationship between $\tilde{p}_n$ and $\tilde{f}_{j,k}$ under merchant bank fees $m_{i,k}^{**}$ is the same as the relationship between p_n and $f_{j,k}$ under merchant bank fees $m_{i,k}^{*}$.

The profit of an issuer bank *j* is given by $\sum_{l=1}^{L}\lambda_j^l\left[A_j+\left(f_j+a+\Delta\right)\frac{1-m_{i,k}^{*}}{1-m_{i,k}^{**}}\sum_{n=1}^{N}\tilde{p}_n q_n^l\left((1+\tilde{f}_j)\tilde{p}_n\right)\right]$. Noting that $\frac{1-m_{i,k}^{**}}{1-m_{i,k}^{*}}=\frac{1-a-\Delta}{1-a}$ and substituting in for f_j, issuer bank profit is $\sum_{l=1}^{L}\lambda_j^l\left[A_j+\left(\tilde{f}_j+a\right)\sum_{n=1}^{N}\tilde{p}_n q_n^l\left((1+\tilde{f}_j)\tilde{p}_n\right)\right]$. But given the construction of $\tilde{p}_n$, this is identical to the issuer banks' choice before the change to the interchange fee, so there is an equilibrium where all issuer banks choose $\tilde{f}_j=f_j^*$ and $A_j=A_j^*$. Substitution into the definition of $\tilde{f}_j$ means that $1+f_j^*=\left(1+f_j\right)\frac{1-m_{i,k}^{*}}{1-m_{i,k}^{**}}$ so that in the new equilibrium $f_j=f_j^{**}$ and $A_j=A_j^*$. Finally, we need to check the assumption that $Z_n^l=1$. But we have shown above that this will hold under the postulated fees.

Finally, considering the integrated banks, that $A_k = A_k^{**}$, $f_k = f_k^{**}$ and $m_k = m_k^{**}$ follows from noting that (6) is the sum of (4) and (5).

We have shown that $f_{j,k}^{**} = f_{j,k}^{*} - \frac{\Delta(1+f_{j,k}^{*})}{1-a}$ and $A_{j,k}^{**} = A_{j,k}^{*}$ for each issuing bank j,k and that $m_{i,k}^{**} = m_{i,k}^{*} + \frac{\Delta(1-m_{i,k}^{*})}{1-a}$ for each merchant bank i,k. Further, in showing this we have also shown that $\lambda_{i,k}^{n**} = \lambda_{i,k}^{n*}$ for all credit-card merchants *n* and that $\lambda_{j,k}^{l**} = \lambda_{j,k}^{l*}$, $q_n^{l**} = q_n^{l*}$ and $q_m^{l**} = q_m^{l*}$ for all customers *l*. Substitution shows that neither bank nor merchant profits nor customers' budgets alter between the original and the new equilibrium. As such, the change in interchange fees has no effect on bank, merchant or customer welfare.

Proof of Proposition 3:

The issuer's profit under the no-surcharge rule at $(f,m,\Delta) = (f^*,m^*,0)$ is given by $(f^* - \Delta + m^* + \Delta)\hat{Q}^{c*} - c_i(\hat{Q}^{c*})$. The derivative of issuer's profit with regards to Δ is given by $\left(f^* + m^* - \frac{\partial c_i}{\partial \hat{Q}^{c*}}\right)\hat{Q}_{\Delta}^{c*}$. But at $(f,m,\Delta) = (f^*,m^*,0)$, $\left(f^* + m^* - \frac{\partial c_i}{\partial \hat{Q}^{c*}}\right) > 0$. This follows as:

(i) both f^* and m^* are optimal fee choices for the issuer in the absence of the no-surcharge rule so that by the first order conditions for issuer profit maximisation without the no-surcharge rule, $Q^{c*} + \left(f^* + m^* - \frac{\partial c_i}{\partial Q^{c*}}\right)\frac{\partial Q^{c*}}{\partial f} = 0$ and $Q^{c*} + \left(f^* + m^* - \frac{\partial c_i}{\partial Q^{c*}}\right)\frac{\partial Q^{c*}}{\partial m} = 0$, and

(ii) Q^{c*} is strictly positive by assumption and $\frac{\partial Q^{c*}}{\partial f}$, and $\frac{\partial Q^{c*}}{\partial m}$ are strictly negative.

Thus, the issuer's profit is increasing in Δ at $(f,m,\Delta) = (f^*,m^*,0)$ if and only if $\hat{Q}_{\Delta}^{c*} > 0$.

Proof of Proposition 4:

Taking the derivative of the quantity of credit card sales,

$$\frac{dQ^c}{d\Delta} = -Q_2^c + Q_2^c \frac{d\rho}{d\Delta} + Q_1^c \frac{d\rho}{d\Delta}.$$

At $(f,m,\Delta) = (f^*,m^*,0)$ merchant *n* sets $\rho^* = p_n^* = p_n^{c*}$ to maximise profit so that $\hat{\pi}_{\rho}(\rho^*, \rho^* + f^* - \Delta; f^* - \Delta, m^* + \Delta, p) = 0$ and $\hat{\pi}_{\rho\rho} < 0$. Totally differentiating this first order condition gives $\hat{\pi}_{\rho\rho} d\rho - \hat{\pi}_{\rho 2} d\Delta = 0$. Thus $\frac{d\rho}{d\Delta} = \frac{\hat{\pi}_{\rho 2}}{\hat{\pi}_{\rho\rho}}$. Further, note that

$\hat{\pi}_{\rho\rho} = \hat{\pi}_{\rho 1} + \hat{\pi}_{\rho 2}$ so that $1 - \frac{d\rho}{d\Delta} = \frac{\hat{\pi}_{\rho 1}}{\hat{\pi}_{\rho\rho}}$. Hence, $\frac{dQ_n^c}{d\Delta} = \frac{1}{\hat{\pi}_{\rho\rho}}\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right)$ so that $\frac{dQ_n^c}{d\Delta} > 0$ if and only if $\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right) < 0$.

Turning to the sufficient conditions, note that $\hat{\pi}_{\rho\rho} = \hat{\pi}_{\rho 1} + \hat{\pi}_{\rho 2} < 0$ by the second order conditions for profit maximisation under the no surcharge rule and that $Q_2^c < 0$ while $Q_1^c > 0$. Thus, $\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right) < 0$ will be satisfied if both $\hat{\pi}_{\rho 1} < 0$ and $\hat{\pi}_{\rho 2} < 0$. Further, noting that $\hat{\pi}_{\rho 1} = \pi_{11} + \pi_{21}$ and $\hat{\pi}_{\rho 2} = \pi_{12} + \pi_{22}$ and that both $\pi_{11} < 0$ and $\pi_{22} < 0$ by the second order conditions for profit maximisation by the merchant in the absence of the no surcharge rule, the condition can only be violated if $\pi_{12} > 0$. Even then, the second order conditions for profit maximisation by the merchant in the absence of the no surcharge rule require that $\pi_{11}\pi_{22} - (\pi_{12})^2 > 0$, so that π_{12} cannot be 'too large.' Essentially, $\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right) < 0$ will only fail to hold if one of π_{11} and π_{22} is relatively large while the other is rather small. Even then, the condition may continue to hold depending on the relative size of $\left|Q_1^c\right|$ and $\left|Q_2^c\right|$. The first condition follows from this. For the second and third conditions, suppose either $\hat{\pi}_{\rho 1}$ or $\hat{\pi}_{\rho 2}$ is strictly greater than zero. Note that only one of these two can be positive. Suppose that $\hat{\pi}_{\rho 1} < 0$ and $\hat{\pi}_{\rho 2} > 0$. Then, $\frac{dQ_n^c}{d\Delta} = \frac{1}{\hat{\pi}_{\rho\rho}}\left(-\hat{\pi}_{\rho 1} Q_2^c + \hat{\pi}_{\rho 2} Q_1^c\right) = \frac{1}{\hat{\pi}_{\rho\rho}}\left(\hat{\pi}_{\rho\rho} Q_1^c - \hat{\pi}_{\rho 1}\left(Q_1^c + Q_2^c\right)\right)$ so that $\frac{dQ_n^c}{d\Delta} > 0$ if $\hat{\pi}_{\rho 1}\left(Q_1^c + Q_2^c\right) > 0$ which holds if $Q_1^c + Q_2^c < 0$. Alternatively suppose that $\hat{\pi}_{\rho 1} > 0$ and $\hat{\pi}_{\rho 2} < 0$ so that $\frac{dQ_n^c}{d\Delta} > 0$ if $\hat{\pi}_{\rho 2}\left(Q_1^c + Q_2^c\right) < 0$ which holds if $Q_1^c + Q_2^c > 0$.

Proof of Proposition 6:

Under the no-surcharge rule the profit of merchant *n* is given by $\hat{\pi}_n\left(\rho_n, \rho_n + f - \Delta; \rho_s, f - \Delta, m + \Delta, p\right)$. At $(f, m, \Delta) = (f^*, m^*, 0)$ merchants *n* and *s* simultaneously maximise profits by setting prices $\rho^* = p_n^* = p_n^{c*}$ and $\rho^* = p_s^* = p_s^{c*}$. In particular, $\frac{\partial \hat{\pi}_n}{\partial \rho_n} = 0$ and $\frac{\partial^2 \hat{\pi}_n}{\partial \rho_n^2} < 0$. Also, $Q_n^c = Q_n^c\left(\rho_n, \rho_n + f - \Delta; \rho_s, f - \Delta, m + \Delta, p\right)$ so that $\frac{dQ_n^c}{d\Delta} = -\frac{\partial Q_n^c}{\partial p_n^c} + \frac{\partial Q_n^c}{\partial p_n}\frac{d\rho_n^*}{d\Delta} + \frac{\partial Q_n^c}{\partial p_n^c}\frac{d\rho_n^*}{d\Delta} + \frac{\partial Q_n^c}{\partial p_s}\frac{d\rho_s^*}{d\Delta}$.

For all Δ, merchant *n*'s best response is characterized by $\frac{\partial \hat{\pi}_n}{\partial \rho_n} = 0$. Totally differentiating this first order condition gives

$$\frac{\partial^2 \hat{\pi}_n}{\partial \rho_n^2} d\rho_n + \frac{\partial^2 \hat{\pi}_n}{\partial \rho_n \partial \rho_s}\frac{\partial \rho_s}{\partial \rho_n} d\rho_n - \frac{\partial^2 \hat{\pi}_n}{\partial \rho_n \partial p_n^c} d\Delta + \frac{\partial^2 \hat{\pi}_j}{\partial \rho_n \partial \rho_s}\frac{\partial \rho_s}{\partial \Delta} d\Delta$$

By symmetry, $\frac{\partial \rho_n}{\partial \Delta} = \frac{\partial \rho_s}{\partial \Delta}$, so that

$$\frac{d\rho_n}{d\Delta}\left(\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}+\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial\rho_s}\frac{d\rho_s}{d\rho_n}+\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial\rho_s}\right)=\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n^c}$$

Noting that by symmetry, $\frac{\partial\rho_n^*}{\partial\rho_s}=\frac{\partial\rho_s^*}{\partial\rho_n}$ and substituting in the definition of $\frac{\partial\rho_n}{\partial\rho_s}$, this means that $\frac{d\rho_n}{d\Delta}=\frac{1}{\theta}\left(\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n^c}\Big/\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}\right)$. Note that given our assumptions on θ and on $\hat{\pi}$, $\frac{d\rho_n}{d\Delta}$ is positive and finite.

By substitution:

$$\frac{dQ_n^c}{d\Delta}=-\frac{\partial Q_n^c}{\partial p_n^c}+\frac{1}{\theta\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}}\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}\right)\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n^c} \tag{7}$$

Simplifying:

$$\frac{dQ_n^c}{d\Delta}=\frac{1}{\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}}\left(-\frac{\partial Q_n^c}{\partial p_n^c}\left(\frac{1}{\theta}\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n}+\left(1-\frac{1}{\theta}\right)\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}\right)+\frac{1}{\theta}\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_s}\right)\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n^c}\right) \tag{8}$$

First, suppose $\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}\right)\geq 0$ at $(f,m,\delta)=(f^*,m^*,0)$. Then from (7) and our assumptions, $\frac{dQ_n^c}{d\Delta}>0$ for all θ. But by symmetry, if $\frac{dQ_n^c}{d\Delta}>0$ then $\frac{d(Q_n^c+Q_s^c)}{d\Delta}>0$, so it is always in the monopoly issuer's interest to set $\Delta>0$.

Alternatively, suppose $\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}\right)<0$. Note that from equation (7),

$$\frac{\partial\left(\frac{dQ_n^c}{d\Delta}\right)}{\partial\theta}=-\frac{1}{\theta^2\left(\frac{\partial^2\hat{\pi}_n}{\partial\rho_n^2}\right)}\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}\right)\frac{\partial^2\hat{\pi}_n}{\partial\rho_n\partial p_n^c}>0$$

for all $\theta\in(0,1)$. Further, $\frac{dQ_n^c}{d\Delta}$ is continuous in θ. When $\theta=1$, equation (8) is identical to the single merchant case and $\frac{dQ_n^c}{d\Delta}>0$. From equation (7), for θ close enough to zero, the sign of $\frac{dQ_n^c}{d\Delta}$ is the same as the sign of $\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}$. Hence, for θ close enough to zero, $\frac{dQ_n^c}{d\Delta}<0$. As $\frac{dQ_n^c}{d\Delta}$ is monotonically increasing in θ, there exists a critical of $\theta>0$, denoted by $\tilde{\theta}$ such that for $\theta<\tilde{\theta}$ then $\frac{dQ_n^c}{d\Delta}<0$, but if $\theta>\tilde{\theta}$ then $\frac{dQ_n^c}{d\delta}>0$. Thus, if $\left(\frac{\partial Q_n^c}{\partial p_n}+\frac{\partial Q_n^c}{\partial p_n^c}+\frac{\partial Q_n^c}{\partial p_s}\right)<0$ and $\theta<\tilde{\theta}$, $\frac{dQ_n^c}{d\Delta}<0$ and by symmetry $\frac{d\left(Q_n^c+Q_s^c\right)}{d\Delta}<0$, so a monopoly issuer will not find it profitable to increase Δ above zero.

Proof of Proposition 7:

The profit maximising prices set by merchant *n* are independent of the prices set by merchant *s*. Thus, the no-surcharge rule is profitable for the issuer if it raises the credit card sales of merchant *s*.

The imposition of the no surcharge rule will lead to a uniform cash and credit price ρ_s^* for merchant *s*. Further, $p_s^{c*} > \rho_s^* > p_s^*$. To see this note that at $\rho_s = p_s^*$,

$$\frac{\partial \pi_s}{\partial \rho} = \frac{\partial \pi_s(p_s^*, p_s^{c*})}{\partial p_s^*} - \int_{p_s^*}^{p_s^{c*}} \frac{\partial^2 \pi_s}{\partial p_s \partial p_s^c} dp_s^c + \frac{\partial \pi_s(p_s^*, p_s^{c*})}{\partial p_s^{c*}} - \int_{p_s^*}^{p_s^{c*}} \frac{\partial^2 \pi_s}{\partial p_s^{c2}} dp_s^c .$$

But $\frac{\partial \pi_s(p_s^*, p_s^{c*})}{\partial p_s^*} = 0$ and $\frac{\partial \pi_s(p_s^*, p_s^{c*})}{\partial p_s^{c*}} = 0$ so if $\frac{\partial^2 \pi_s}{\partial p_s \partial p_s^c} \leq 0$ then $\pi_\rho > 0$ at $\rho = p_s^*$ and $\rho^* > p_s^*$ by concavity. A similar substitution shows that $\rho^* < p_s^{c*}$.

Given that $\frac{\partial Q_s^c}{\partial p_s} > 0$ and $\frac{\partial Q_s^c}{\partial p_s^c} < 0$, as $p_s^{c*} > \rho_s^* > p_s^*$, Q_s^c will rise after the no-surcharge rule is imposed and the issuer's profit will rise.

References

Ausubel, L.M. (1991), "The Failure of Competition in the Credit Card Market," *American Economic Review*, 81 (1), pp.50-81.

Baxter, W.F. (1983), "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics*, 26, pp.541-588.

Carlton, D. and A.S. Frankel (1995), "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 68, pp.643-668.

Cruickshank, D. (2000), *Competition in U.K. Banking: A Report to the Chancellor of the Exchequer*, The Stationary Office: London.

Economides, N. and S.C. Salop (1992), "Competition and Integration Among Complements, and Network Market Structure," *Journal of Industrial Economics*, 40 (1), pp.105-123.

European Commission (2000), "Commission plans to clear certain Visa provisions, challenge others," DN:IP/00/1164, 16 October, Brussels.

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

Frankel, A.S. (1998), "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66, pp.313-361.

Grant, S. and S.P. King (1997), "The Fiscal Dividend Myth of an Income/GST Tax Switch," *Australian Economic Papers*, 36, pp.167-178.

RBA/ACCC (2000), *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access*, RBA: Sydney.

Rochet, J-C. and J. Tirole (2000), "Cooperation Among Competitors: The Economics of Payment Card Associations," *mimeo*., Toulouse, April.

Rolnick, A.J., B.D. Smith and W.E. Weber (1997), "Lessons from a Laissez-Faire Payments System: The Suffolk Banking System (1825 – 1858)," *Working Paper* No.584, St. Louis Fed.

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

Varian, H. (1985), "Price Discrimination and Social Welfare," *American Economic Review*, 75, pp.870-875.

Wright, J. (2000), "An Economic Analysis of a Card Payment Network," *mimeo*., NECG and Auckland.

Regulating Interchange Fees in Payment Systems[*]

by

Joshua S. Gans *and* Stephen P. King
University of Melbourne

First Draft: 12th May, 2001
This Version: 12th June, 2001

DRAFT ONLY

[*] This research has been funded (in part) by the National Australia Bank. All views stated are those of the authors and are not necessarily held by the National Australia Bank.

1. Introduction

In recent times, competition authorities in several jurisdictions have begun investigating the rules and practices of credit card associations. This includes the United Kingdom (Cruickshank, 2000), Australia (ACCC/RBA, 2000) and the European Commission. In addition, there has been a historic and on-going set of antitrust cases in the United States concerning credit card associations. In each case, investigations were triggered by natural suspicions that arise when otherwise competing banks cooperate through credit card associations.

Two particular aspects of card associations have raised competition concerns. The first is the collective setting and the levels of interchange fees that govern the terms of settlement between issuers and acquirers in credit card associations. The second are rules preventing surcharges being imposed by merchants on credit card transactions. The combination of the two rules has led competition authorities to suggest that interchange fees ought to be regulated and that 'no surcharge' rules be removed. Controversy exists regarding the need for such regulation and the precise role of 'no surcharge' rules (see Gans and King, 2001a). However, to date, little in the way of formal economic analysis has been conducted to investigate such issues.[1]

The goal of this paper is to provide a simple model that illustrates the role played by 'no surcharge' rules in credit card associations and to provide a framework to analyse socially optimal interchange fees. We demonstrate that a 'no surcharge' rule is a means of preventing price discrimination by merchants with market power. In the absence of such a rule, merchants will raise credit card prices relative to cash

[1] Exceptions include Rochet and Tirole (2000), Schmalensee (2001), Wright (2001) and Gans and King (2001b).

prices. The use of a credit card becomes a self-selection mechanism for customers and enables merchants to identify relatively high value customers and to raise profits.[2] At the same time, such price discrimination lowers credit card transactions and is socially undesirable.

When there is a 'no surcharge' rule in force, the customer faces a single cash or credit price and chooses the relative use of each payment instrument. This choice, however, only depends on the customer's own relative benefits and costs of each payment instrument. The customer ignores any costs or benefits to the merchant. We demonstrate that the social efficiency of the payments system is maximised by setting interchange fees to internalise this externality. We present a general rule for optimal interchange fees based on the relative marginal benefits from credit to both customers and merchants. If the marginal benefits to both customers and merchants from using credit relative to cash are approximately the same, then the rule takes a particularly simple form. Interchange fees should be set to so that issuers and acquirers face the same marginal cost associated with each credit card transaction. This suggests a simple rule could be implemented to regulate such interchange fees in situations where it is determined that competition between payment instruments is otherwise insufficient.

This paper proceeds as follows. Section 2 presents the general model while section 3 shows how a merchant with market power will engage in price discrimination. Section 4 shows the effect of a 'no surcharge' rule, and notes that the addition of such a rule, while eliminating price discrimination, may raise or lower transactions efficiency. Section 5 then considers the socially optimal interchange fee

[2] Such self-selection schemes represent second-degree price discrimination. See Tirole (1988).

while section 6 considers a number of extensions to the analysis. A final section concludes.

2. The Model

We begin by modeling the interaction between a representative consumer and a representative merchant. The representative consumer has demand curve $Q(p)$, takes merchant prices as given and seeks to minimise the total cost of their purchases. The consumer can use cash or credit card or any combination to pay for total purchases. We consider the costs and benefits of credit card use relative to cash. Using the credit card involves an additional fee of f per unit purchased. The banks that issue credit cards set this fee. The consumer can save transaction expenses by using a credit card rather than cash. We denote these savings by $\int_0^{Q^c} b_c(q)dq$ where Q^c is the consumer's total credit card purchases. Marginal credit card benefits for the consumer are given by $b_c(.)$ and we assume that $b_c(0) > f$ and $b_c' < 0$. In other words, if cash and credit card retail prices are identical, it always pays the consumer to make some credit purchases. However, the relative benefits of such purchases over cash decline as the total amount of purchases rises. If the consumer makes both cash and credit card purchases, this implies that the customer will purchase on credit card until $p = p^c + f - b_c(Q^c)$, where p and p^c refer to the cash and credit card retail prices respectively. To avoid trivial outcomes we only consider situations where the consumer makes both cash and credit purchases.

The representative consumer can be interpreted in two ways. First, consider the purchases of an individual. Often credit card purchases are 'higher value' items. For example, a customer might use credit card for the weekly grocery shopping at a

supermarket, but might use cash to purchase just milk or bread at the same supermarket. Our representative consumer model captures this effect to the idea that credit card purchases are inframarginal while cash purchases are marginal. Alternatively, credit cards tend to be used more by higher income consumers. Such consumers will tend to have higher levels of willingness to pay for an item and this is captured in our framework.

The merchant may also receive benefits from credit card sales relative to cash sales. The marginal merchant benefit is denoted by $b_m(Q^c)$ where $b_m' \leq 0$. The merchant pays a merchant-services fee of *m* for each unit sold to a customer using a credit card. Total merchant profit is given by

$$\pi = Q^c p^c + \int_0^{Q^c} b_m(q)dq - mQ^c + (Q - Q^c)p - c(Q) \qquad (1)$$

where *Q* refers to total sales by both cash and credit and *c*(.) is the merchants cost function. We assume the standard restrictions on both *Q*(*p*) and *c*(*Q*) for a solution to the merchant's profit maximisation problem to be both well defined and unique.

The merchant simultaneously sets both the cash and credit prices. The customer then chooses both their total purchases and how to divide their purchases between each payment instrument. Given the credit card fees *f* and *m* and total purchases *Q*, total transactions costs are minimised at $Q^c = Q^{c*}$ such that $f - b_c(Q^{c*}) + m - b_m(Q^{c*}) = 0$. In other words, if the consumer makes both cash and credit card purchases then the optimal split of total purchases between payment instruments occurs when the total marginal benefit of credit card purchases to both the customer and the merchant equals zero.

3. Market Outcomes and Price Discrimination

We first analyse the retail market outcome given the credit card fees. The merchant will set both the cash and credit card price and will seek to divide cash and credit sales to maximise profit. However, the merchants' desired split of total sales between cash and credit must be consistent with the consumer's choice of payment instrument. Thus, the merchant will set p^c, Q^c, and p to maximise π subject to $p = p^c + f - b_c(Q^c)$. From (1), the first order conditions for the merchant's profit maximisation problem with respect to p^c, Q^c, and p respectively are given by:

$$Q^c + \lambda = 0 \tag{2}$$

$$p^c + b_m(Q^c) - m - p - \lambda b_c'(Q^c) \tag{3}$$

$$Q - Q^c + Q'(p)p - c'(Q)Q'(p) - \lambda = 0 \tag{4}$$

where λ is the Lagrange multiplier on the constraint imposed by the customer's choice of cash and credit purchases.

Substituting $\lambda = -Q^c$ from (2) into (4), the optimal value for p is simply the standard profit-maximising price for a monopoly seller. This reflects that when a consumer makes both cash and credit purchases, the cash price determines total purchases while the difference between the cash and credit prices determine the infra-marginal split between cash and credit sales. We denote this profit maximising cash price by p^m.

Equation (3) determines the relationship between the cash and credit prices. By substitution, $(p^c - p^m) + (b_m(Q^c) - m) + Q^c b_c'(Q^c) = 0$. But, by the customer's optimal choice of payment instruments, we know that $b_c(Q^c) - f = p^c - p$. Thus, the merchant will set the credit card price so that

$$(b_c(Q^c) - f) + (b_m(Q^c) - m) + Q^c b_c'(Q^c) = 0 \tag{5}$$

Comparing (5) with the socially optimal rule, and noting that $b_c' < 0$, the merchant will make too few credit card sales from a social perspective. While transactions costs are minimised when $(b_c(Q^c) - f) + (b_m(Q^c) - m) = 0$, the merchant will set prices so that $(b_c(Q^c) - f) + (b_m(Q^c) - m) > 0$. Proposition 1 immediately follows.

Proposition 1. *A profit-maximising merchant will not minimise total transactions cost and will have credit card sales Q^c strictly less than the socially optimal level Q^{c*}.*

In the absence of any pricing restriction, the merchant will use credit cards as a form of second-degree price discrimination. Credit cards are more likely to be used by either consumers with a relatively high willingness-to-pay or for relatively high value purchases. By setting a relatively high credit card price, the merchant is able to discriminate between these high value sales and other sales. To maximise profits, the merchant will trade off the transactions cost benefits of increased credit card sales, as measured by $(b_c(Q^c) - f) + (b_m(Q^c) - m)$, and the benefit from raising profits by raising credit card prices. The ability to raise credit card prices is limited by the ability of the customer to switch to cash purchases at the margin if credit card prices are too high. This is captured by the term $Q^c b_c'(Q^c)$.

The merchant's discrimination against credit card use is reflected in the relative cash and credit card prices. It does not mean that the credit card price is either higher or lower than the cash price in absolute terms. The exact relationship between the two prices will depend on whether customer's or merchant's net marginal benefits from credit card usage are higher at Q^{c*}. For example, suppose that $(b_c(Q^{c*}) - f) > 0$. Then to minimise total transactions cost it is socially desirable to have $p^c > p$. The higher credit card price allows the customer to 'compensate' the merchant who faces marginal credit card costs above the marginal benefit at the socially optimal level of

credit card transactions. In this case, the merchant's incentive to price discriminate leads to a credit price that is even higher than the socially optimal level.

Conversely, if $(b_m(Q^{c*}) - m) > 0$ then it is socially optimal to have the credit card price below the cash price. Price discrimination will tend to raise the credit card price and may even involve the merchant setting $p^c > p$.

Unlike other models (e.g., Rochet and Tirole, 2000 and Wright, 2001), the (socially undesirable) tendency of merchants to limit credit card sales here does not depend on any network or other externality. Rather it is simply a device for price discrimination. The merchant will tend to lower credit card sales and raise the credit card price because this allows them to identify high value consumers and high value transactions.

4. The Effect of a No-Surcharge Rule

A 'no surcharge' rule or a 'no discrimination' rule is often imposed on merchants by credit card associations. This rule means that the merchants are constrained in their ability to set different cash and credit card prices. A simple version of that rule would require that the merchant set the same price for cash and credit sales. In this section, we consider the effect of such a rule on merchant behaviour.

If the merchant can only set a single price, then the division of sales between cash and credit card will be determined completely by the consumer. The merchant will set the price *p* to maximise $\pi = \int_0^{Q^c} b_m(q)dq - mQ^c + pQ - c(Q)$ where $Q^c = \tilde{Q}^c$ is chosen by the customer so that $f - b_c(\tilde{Q}^c) = 0$. Assuming that the merchant continues to make both cash and credit card sales, the merchant will simply set the single profit-

maximising price at the same level as the cash price in the absence of a no surcharge rule, $p = p^m$.

The introduction of the no surcharge rule prevents price discrimination against credit card customers. However, it may raise or lower the quantity of credit card transactions relative to unconstrained merchant pricing. Denoting the unconstrained profit maximising credit card price by $\tilde{p}^c$, if $\tilde{p}^c > p^m$ then the no surcharge rule will tend to lower the credit card price and raise credit card sales. But if $\tilde{p}^c < p^m$, the no-surcharge rule will raise the credit card price and lower credit card sales.

Similarly, if we compare the outcome under a no surcharge rule with the socially optimal level of credit card transactions, the no surcharge rule can either move pricing closer to the social optimum or further away from the social optimum. The exact effect will depend on the relationship between the socially optimal price, the price set by the merchant under price discrimination $\tilde{p}^c$, and the cash price p^m. The no surcharge rule will only perfectly minimise transaction costs if $b_c(Q^{c*}) - f = b_m(Q^{c*}) - m = 0$ so that $\tilde{Q}^c = Q^{c*}$.

5. Optimal Interchange Fees

The analysis above showed that an unconstrained merchant will tend to set credit card prices too high relative to socially optimal prices. A no surcharge rule eliminates this price discrimination but, so far, we cannot say whether such a rule will tend to raise or lower the volume of credit card transactions. From a social perspective, the no surcharge rule could 'make things worse.'

To take our analysis further requires a formal model of issuer and acquirer banks. Issuing, acquiring or both functions could be characterised by imperfect

competition. We assume, however, that both functions are characterised by competition in two-part pricing. In other words, issuing and acquiring banks compete by setting both fixed charges to their customers or merchants and by setting per transaction fees such as *f* and *m*. Merchants and customers will choose their banks according to the total benefit that they receive. Profit maximising behaviour in such circumstances will lead banks to set transaction fees that reflect the true marginal cost of credit card transactions. In other words, banks have no incentive to distort marginal prices but rather seek to maximise profits by seizing the surplus from merchants and customers.[3]

For simplicity, suppose that the per-transaction cost to an issuing bank is constant and given by c_I while the per-transaction cost to an acquiring bank is constant and given by c_A. There might also be an interchange fee between issuers and acquirers. We denote the per-transaction interchange fee by *a* and adopt the convention that this fee is paid by acquirers to issuers (although that fee may be positive or negative). Thus, competition between different issuers and acquirers will lead to credit card fees $f = c_I - a$ and $m = c_A + a$.

In this section, we consider the socially optimal interchange fee *a* given the no surcharge rule.[4] Changing the interchange fee will alter the balance between cash and credit sales. However, the interchange fee will not affect total sales or the price set by the merchant, p^m. The optimal level of credit card sales will minimise total transaction costs, $\left(c_I Q^c - \int_0^{Q^c} b_c(q)dq\right) + \left(c_A Q^c - \int_0^{Q^c} b_m(q)dq\right)$. The first order condition for this

[3] Perfect competition is simply an extreme case of this competition where there are no fixed costs in either issuing or acquiring.

[4] It is easy to see that the interchange fee will be irrelevant in the absence of a no surcharge rule by noting from (5) that the merchant's credit card price only depends on the sum of *f* + *m*. Changing the

minimisation problem is given by $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}) = 0$. This is the same condition as identified by Baxter (1983).

Under the no surcharge rule the consumer determines the level of credit card purchases. The consumer makes credit card purchases so that $f - b_c(\tilde{Q}^c) = 0$. The socially optimal interchange fee will set $Q^{c*} = \tilde{Q}^c$. Noting that $f = c_I - a$, this means that $a^* = b_m(Q^{c*}) - c_A$.

The socially optimal interchange fee is intuitive. Under the no-surcharge rule, the customer chooses the level of credit card transactions according to their own marginal costs and benefits. They ignore the marginal costs and benefits of credit card purchases to the merchant. Thus, the customer's choice of an extra credit card purchase imposes an externality on the merchant. This externality is positive if $c_A < b_m(Q^c)$ and negative if $c_A > b_m(Q^c)$. The interchange fee acts to internalise this externality. The fee is positive if there is a marginal benefit to the merchant from an additional credit card transaction at the socially optimal level of transactions. The interchange fee is negative otherwise.

The optimal interchange fee will depend on the relative marginal benefits from additional credit card transactions to merchants and customers. It is convenient to define a variable α to capture these relative benefits. Thus, at the socially optimal level of credit transactions, $\alpha b_c(Q^{c*}) = (1-\alpha) b_m(Q^{c*})$. Proposition 2 calculates the socially optimal interchange fee.

Proposition 2. *The socially optimal interchange fee is* $a = \alpha c_I - (1-\alpha) c_A$.

---

interchange fee does not alter this sum and will not affect merchant pricing in the absence of a no surcharge rule.

PROOF: $a^* = b_m(Q^{c*}) - c_A$. By substitution, $a^* = \frac{\alpha}{1-\alpha} b_c(Q^{c*}) - c_A$. But from the order condition for transaction cost minimisation, $b_c(Q^{c*}) = (1-\alpha)(c_A + c_I)$. By substitution, $a^* = \alpha c_I - (1-\alpha)c_A$.

QED

In some circumstances, it might not be possible to estimate the relative marginal benefits from additional credit card transactions to merchants and customers with accuracy. In such circumstances, a reasonable starting assumption is that these marginal benefits will be relatively similar. Under this assumption, the socially optimal interchange fee takes a particularly simple form, $a^* = \frac{1}{2}(c_I - c_A)$.[5] In other words, if merchant and customer marginal benefits are relatively symmetric, the socially optimal interchange fee results in equal per transaction credit card fees for both merchants and customers with $m = f = \frac{1}{2}(c_I + c_A)$. Again, this accords with intuition. The interchange fee leads to merchant service charges and customer charges that reflect the marginal benefits of an additional credit card transaction to each of these parties.

6. Extensions

We now consider a number of extensions to check the robustness of our basic results; in particular, we consider the impact of competition at the merchant level and the effect of linearity in issuer and acquirer prices.

[5] Schmalensee (2001) derives a privately optimal interchange fee that has a similar basis. He appears, however, to make an algebraic error in deriving his fee for the 'symmetric demands case.' Correcting that error would yield essentially the same fee as that derived here.

Competing Merchants

The analysis above involved a single representative merchant. If there are multiple competing merchants then this alters the analysis in two ways. First, if there is a fixed number of imperfectly competing merchants, each of who gains benefits from credit card transactions, then such transactions can become relatively more valuable. This tends to raise the socially optimal interchange fee. Second, to the extent that there is free entry to the merchant sector, lowering the interchange fee can encourage entry and lead to lower retail prices.

To see these effects, suppose that a representative customer can purchase from *n* identical merchants. Each of these merchants is identical to the representative merchant above and the customer divides their purchases symmetrically over merchants. In equilibrium, the merchants charge a single cash or credit price *p* that depends on their marginal costs and the nature of merchant competition. In particular, the price *p* does not depend on credit card fees as all credit sales are infra-marginal.

Given the single price, the customer chooses both their total purchases and their mix of credit and cash purchases. As above, the customer chooses payment instruments to lower their total cost of transacting so that $f - b_c(Q^c) = 0$. As before, assume that issuers and acquirers compete in two-part tariffs so that $f = c_I - a$ and $m = c_A + a$. The total cost of transacting is given by $\left(c_I Q^c - \int_0^{Q^c} b_c(q)\,dq \right) + \left(c_A Q^c - n\int_0^{Q^c/n} b_m(q)\,dq \right)$. The first order condition for minimising the social costs of transacting is given by $c_I - b_c(Q^{c*}) + c_A - b_m\left(Q^{c*}/n\right) = 0$. Because the marginal benefits of credit card use diminish for merchants as credit transactions increase and credit transactions are divided over *n* merchants, the relevant marginal benefit attached to merchants is the

benefit each individual merchant receives on the marginal credit sale. This is the only change to the socially optimal level of credit card transactions relative to the single merchant case. Corollary 3 thus follows immediately from Proposition 2.

Corollary 3. *Suppose* $\alpha b_c(Q^{c*}) = (1-\alpha)b_m(Q^{c*}/n)$. *Then the socially optimal interchange fee is* $a^* = \alpha c_I - (1-\alpha)c_A$.

While introducing a fixed number of competing merchants does not change the nature of the optimal interchange fee, it does change its value. This is shown by the following proposition.

Proposition 4. *The optimal interchange fee,* a^* *is an increasing function of the (exogenously fixed) number of merchants, n.*

PROOF: To show this, we show that as *n* increases, the value of α such that $\alpha b_c(Q^{c*}) = (1-\alpha)b_m(Q^{c*}/n)$ is increasing. As $\frac{\partial a^*}{\partial \alpha} > 0$, this implies that the optimal interchange fee is increasing as *n* increases.

Note that for any *n*, transactions costs are minimised if $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}/n) = 0$. Totally differentiating this first order condition shows that $\frac{dQ^{c*}}{dn} = \frac{b'_m \frac{Q^{c*}}{n^2}}{b'_c + b'_m \frac{1}{n}} > 0$. Thus, as *n* increases the socially desirable volume of credit card sales increases and $b_c(Q^{c*})$ decreases. Further, totally differentiating the first order condition for socially optimal transactions shows that $\frac{d\left(Q^{c*}/n\right)}{dn} = \frac{-b'_c \frac{Q^{c*}}{n}}{b'_c + b'_m} < 0$. In other words, as *n* increases the optimal sales by each merchant falls and $b_m(Q^{c*}/n)$ increases. Thus, as *n* increases, $b_c(Q^{c*})$ decreases and $b_m(Q^{c*}/n)$ increases so that the value of α such that $\alpha b_c(Q^{c*}) = (1-\alpha)b_m(Q^{c*}/n)$ increases. *QED*

The result that the optimal interchange fee should increase as the degree of merchant competition increases reflects the fact that merchant benefits from credit card use grow in this model as the number of merchants grow. An alternative formulation would involve each merchant's marginal credit card benefit falling proportionately as *n* increases. In such a situation, the optimal interchange fee would remain unchanged as the exogenous value of *n* changes.

The value of the optimal interchange fee is further complicated if *n* is endogenous. Suppose merchants are imperfect competitors but that entry into the retail sector occurs until each merchant makes zero profit. In equilibrium, for any interchange fee *a*, the value of *n* will be given by $\pi = \int_0^{\frac{Q^c}{n}} b_m(q)\,dq - (c_A + a)\frac{Q^c}{n} + \frac{Q}{n}p - c\left(\frac{Q}{n}\right) - F = 0$ where *F* is an individual merchant's fixed costs (including any fixed costs associated with accepting credit cards). Note that in general *p* will be decreasing in *n*.

Totally differentiating the firm profit condition shows that for any interchange fee

$$\frac{dn}{da} = \frac{\frac{Q^c}{n} + \frac{1}{n}\left(c_A + a - b_m\left(\frac{Q^c}{n}\right)\right)\frac{\partial Q^c}{\partial a}}{\left(c_A + a - b_m\left(\frac{Q^c}{n}\right)\right)\frac{Q^c}{n^2} + \left(\frac{\partial p}{\partial n} - \frac{p}{n}\right)\frac{Q}{n}} \tag{6}$$

In general this can be positive or negative depending on the interchange fee. However, suppose that we are at an equilibrium for retail entry where the interchange fee is set so that (for the equilibrium value of *n*) the total cost of transacting is minimised. Then $a = b_m(Q^c/n) - c_A$ and $\frac{dn}{da} = \frac{Q^c}{Q\left(\frac{\partial p}{\partial n} - \frac{p}{n}\right)} < 0$. In other words, lowering the interchange fee below the value that minimises total transactions costs will lead to increased entry into the retail sector. Thus, the interchange fee can be used as a means of promoting increased retail competition, albeit at the expense of inefficiently reducing total credit transactions.[6]

[6] The use of lower regulated input prices to encourage retail competition is well known in the literature on access pricing. See, for example, Armstrong, Cowan and Vickers (1994). While increased entry will lower the retail price it will also involve extra expenditure on fixed costs and may be socially undesirable. See Mankiw and Whinston (1986) for a general discussion on socially optimal entry under imperfect competition.

Linear Credit Card Fees

The analysis above assumes that issuers and acquirers compete through setting two-part tariffs for credit card users. This assumption accords with reality as many credit card schemes do allow members to charge such tariffs. However, it is interesting to consider how the optimal interchange fee might alter if issuers and acquirers can only set linear credit card fees.

To analyse this situation, consider the representative firm and consumer model presented above. As before, the customer will choose the volume of credit purchases so that $f - b_c(Q^c) = 0$ and total transaction costs are minimised when $c_I - b_c(Q^{c*}) + c_A - b_m(Q^{c*}) = 0$. However, in general with linear pricing, $f > c_I - a$ and $m > c_A + a$, with f increasing in a and m decreasing in a. Let $f = f(c_I, a, \gamma_I)$ where γ_I is a parameter capturing the degree of issuer competition and the size of issuer fixed costs. A higher value of γ_I represents a lower level of issuer competition or a higher level of issuer fixed costs so that the mark-up of f above marginal cost is higher. Thus, the customer fee f is increasing in γ_I but decreasing in a. Similarly, let $m = m(c_A, a, \gamma_A)$ where γ_A is a parameter capturing the degree of acquirer competition and the size of acquirer fixed costs. The merchant fee m is increasing in both γ_A and a.

Substitution into the equations for the consumer's purchasing decision and the minimum transactions costs shows that the optimal interchange fee solves $c_I - f(c_I, a^*, \gamma_I) + c_A - b_m(Q^{c*}) = 0$. This immediately leads to the following proposition.

Proposition 5. *The socially optimal interchange fee when issuers and acquirers can only set linear fees is independent of the degree of acquirer competition. However, the*

socially optimal interchange fee increases if there is either a decrease in issuer competition or an increase in issuer fixed costs.

> PROOF: This follows immediately from $c_I - f\left(c_I, a^*, \gamma_I\right) + c_A - b_m\left(Q^{c*}\right) = 0$. As this equation does not depend on γ_A, the optimal interchange fee cannot depend on the level of acquirer competition. As f is increasing in γ_I but decreasing in a it immediately follows that a rise in γ_I must be associated with a fall in the interchange fee to maintain a socially optimal mix of transactions. *QED*

Although Proposition 5 might initially appear surprising, the intuition behind it is simple. Acquirer competition has no effect on the optimal interchange fee because it is customers, not merchants, who determine the mix of cash and credit purchases. A reduction in acquirer competition will make the representative merchant worse off, but it will not change either the socially optimal mix of transactions or the customer's transaction choice. In contrast, issuer competition directly affects the actual mix of cash and credit transactions. A reduction in issuer competition raises the mark-up of card fees over true marginal transactions costs and discourages credit card transactions. To offset this tendency towards insufficient use of credit when the issuer segment is not competitive, it is desirable to raise the interchange fee. Raising this fee lowers issuers' costs and, for any level of competition, tends to reduce customer charges.

A clear limitation of this analysis is the limitation on merchant competition. If there were endogenous entry of merchants (as discussed earlier in this section) then a reduction in acquirer competition and a rise in *m* would tend to reduce the number of merchants. A similar effect could occur if the interchange fee was increased to offset decreased issuer competition. In both cases, there would be an effect on the retail price as well as the split of transactions.

7. Conclusion

This paper has developed a simple model of payment systems designed to explore two issues. First, we consider the determinants of a socially optimal interchange fee designed to minimise the cost of transacting. Second, we consider the role of 'no surcharge' rules often imposed by credit card associations.

We find that the socially optimal interchange fee is set so as to ensure that issuers and acquirers take into account the full system impact of their actions. In particular, if merchants and customers tend to shared equally the benefits of card transactions, the socially optimal interchange fee is based on the difference in issuer and acquirer marginal (or avoidable) costs. This form of the interchange fee is not dependent on the level of merchant competition although, in general, as merchants become more competitive, the socially optimal interchange fee will rise.

We also find that the no-surcharge rule can play an important and potentially socially desirable role in reducing price discrimination by merchants to their customers. This is because, when surcharging is permitted, merchants can use customer's card decisions as a sorting device for potentially high willingness to pay customers. A no surcharge rule prevents such discrimination.

Our model here is a simple one and captures many elements of card associations. In reality, however, the impact of non-linear cost structures, an endogenous model of customer and merchant benefits as well as dynamic considerations are areas for fruitful future research.

References

Armstrong, M., S. Cowan and J. Vickers (1994),

Baxter, W.F. (1983), "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics*, 26, pp.541-588.

Carlton, D. and A.S. Frankel (1995), "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 68, pp.643-668.

Cruickshank, D. (2000), *Competition in U.K. Banking: A Report to the Chancellor of the Exchequer*, The Stationary Office: London.

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

Frankel, A.S. (1998), "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66, pp.313-361.

Gans, J.S. and S.P. King (2001a), "The Role of Interchange Fees in Credit Card Associations: Competitive Analysis and Regulatory Options," *Australian Business Law Review*, 29 (1), pp.94-122.

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

Mankiw and Whinston (1986)

RBA/ACCC (2000), *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access*, RBA: Sydney.

Rochet, J-C. and J. Tirole (2000), "Cooperation Among Competitors: The Economics of Payment Card Associations," *mimeo*., Toulouse, April.

Schmalensee, R. (2001), "Payment Systems and Interchange Fees," *Working Paper*, NBER.

Tirole, J. (1988), *The Theory of Industrial Organization*, MIT Press: Cambridge (MA)>

Wright, J. (2001)

Some Answers to the Reserve Bank of Australia Questions Associated with the Designation of Credit Card Payment Systems in Australia

by

Joshua Gans and Stephen King

This paper was commissioned and funded by the National Australia Bank. All of the views expressed and recommendations made are those of CoRE Research Pty Ltd (ACN 96 869 760). Consequently, the contents of this report do not necessarily reflect the views of the National Australia Bank nor should they be construed as so doing.

www.core-research.com.au

30th June 2001

Contents

Page

1 Background 1

2 Interchange Fees 4

2.1 The Necessity of a Collectively Set Interchange Fee 4

2.2 Competitive Determinants of the Interchange Fee 14

3 Principles for the Setting of Efficient Interchange Fees in Payment Systems 25

3.1 The Network View of Payment Systems 26

3.2 Basic Principles for the Setting of Interchange Fees 29

3.3 The Shared Avoidable Cost Approach 35

3.4 Governance Issues in Interchange Fee Revisions 40

4 No Surcharge Rule 45

4.1 Is a no surcharge rule material? 45

4.2 Private benefits of the no surcharge rule 48

4.3 The social costs of the no surcharge rule 49

4.4 Conclusion 50

5 Competing Payment Systems 51

5.1 Proprietary Card Systems 51

5.2 Competitive Interactions 52

5.3 Conclusion 54

6 Networks and Network Effects 55

7 Loyalty Points and Nonlinear Pricing 57

Contents

Page

8 References..... 59

1 Background

The purpose of this paper is to collate notes that answer the questions posed by the Reserve Bank of Australia (RBA) to the Bankcard association on the 30th April, 2001.

Those questions were as follows.

1. Is an interchange fee necessary to the functioning of an open credit card scheme? If so, why? If not, what are the alternatives?

2. In the open card systems operating in Australia, are there competitive forces that generate an equilibrium interchange fee? If so, what are they?

3. How do you think interchange fees for the card schemes operating in Australia should be determined in practice? Please spell out the advantages and disadvantages of your proposal.

4. How frequently should interchange fees be revised? Please detail the arguments for and against your proposal.

5. What specific risks do acquirers bring to a scheme, independent of their status as issuers?

6. What specific risks do self-acquirers bring to a scheme? Please note where those risks are different from the risks for third party acquirers.

7. In the presence of an interchange fee and membership fees, what is the justification for net issuer penalties? How large are such penalties?

8. Do you agree that the no surcharge rule is integral to the success of the open credit card systems? If so, why and if not, why not?

9. If the no surcharge rule were removed from the scheme regulations, do you think it would be removed from merchant agreements?

10. Which payment instruments do open card schemes compete with? How do they compete in each case?

11. Open credit card schemes appear to have much larger card bases and wider acceptance than three party schemes. Why is this so?

12. How are schemes promoted: (a) At the scheme level? (b) At the individual bank level?

13. Why are credit card interest rates around 3 percentage points higher than rates on other unsecured personal lending?

14. Which of your cards offer loyalty points? What is the role of loyalty points? What evidence is there that they achieve the issuers' objectives?

15. Does the ability of issuing banks to offer a number of different brands of cards result in more competition between schemes than would be the case if they could only issue one brand? Please detail your analysis.

We consider most of these questions in this document although the paper is organised according to issues rather than questions. This paper is not designed to be comprehensive but rather brings together in a single document the current economic thought on credit cards and identifies areas for further analysis.

The paper is organised as follows:

- Section 2 considers the issue of interchange fees. We first discuss the rationale and social benefits that arise from the collective setting of interchange fees. Then we examine the competitive pressures on interchange fees (drawing upon earlier work from Gans and King, 2001a).

- Section 3 then considers issues associated with the regulation of interchange fees. We outline a proposed methodology for the setting of interchange fees that is based on a 'network view' of payment systems. This methodology is new and

largely distinct from others that have been implemented in the past. It is based on the technical research contained in Gans and King (2001c). In addition, we provide a brief discussion of governance issues in this regard.

- Section 4 then evaluates the rationale and consequences of the no surcharge rule. We find that there are several beneficial elements of such a rule in curtailing market power in the banking sector as well as preventing undue price discrimination in the retail sector.

- Section 5 then considers the competitive interaction among credit card associations and between credit card associations and alternative payment systems.

- Section 6 then clarifies some confusion arising from the definition and application of term 'network effects.'

- Section 7 then evaluates the consequences of the introduction of loyalty points and why an association may wish to encourage these.

Some issues are dealt with more completely than others. In reality, this difference in emphasis reflects the current emphasis of economic research as opposed to the importance or otherwise of individual issues in terms of the designation of credit card payment systems in Australia.

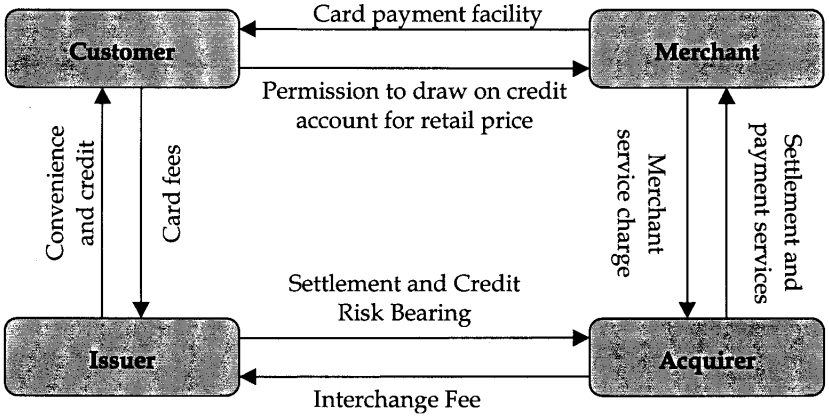
2 Interchange Fees

2.1 The Necessity of a Collectively Set Interchange Fee

2.1.1 Interchange Under Current Arrangements

Credit card associations are sometimes referred to as 'four party' systems. This is because up to four parties can be involved in processing a transaction. Aside from a *customer* and *merchant* who are parties to all payments transactions, the customer pays their account via their card *issuer* while an *acquirer* pays the merchant. The issuing and acquiring parties deal with one another to settle the transaction; thus, completing the loop. This loop is depicted in Figure One.

Figure One: Parties to a Credit Card Transaction



What drives the economics of the operation of credit card systems is that the value derived from the system is in facilitating transactions between customers and merchants. In effect, there is no

single party making a decision as to whether to use a credit card. Rather, that decision is jointly made by customers and merchants in the sense that a cardholder cannot use their card if merchants do not accept the card while a merchant cannot accept a card unless a customer is a cardholder. This means that the credit card pricing terms to both customers and merchants (in terms of cardholder fees and merchant service charges (MSC)) drive whether a particular credit card is used as a means of payment. This, in turn, determines whether cardholders receive the benefits of liquidity and transaction ease and whether merchants receive the benefits from increased security and electronic account processing associated with a card transaction.

For all credit cards that currently exist, a way of coordinating the issue of the jointness of the consumption of credit card services has been resolved. 'Closed loop' credit card systems (such as American Express and Diners Club) operate both the issuing and acquiring functions of credit card transactions and, importantly, set cardholder fees and merchant services charges jointly. This allows them to react in a centralised way to the links between cardholder and merchant decisions. As a result, closed loop card providers internalise the fact that a higher cardholder fee will diminish overall card usage unless it was accompanied by a corresponding reduction in merchant services charges. This enables them to use both cardholder fees and merchant services charges in tandem to maximise profits by controlling overall usage of their card.

Most credit card transactions however are undertaken through 'open loop' systems. Such associations are distinguished by the fact that there is no single issuer/acquirer and indeed, many banks and other participants can offer to issue cards, acquire from merchants or both. In so doing, each issuer and acquirer can develop their own branding, pricing and other marketing instruments and also be responsible for their own costs incurred in providing card services. So not only are cardholder fees and merchant services charges set in a decentralised manner, there are competitive forces operating on both the issuing and acquiring sides of the credit card system.

Participants in open loop systems must deal with the joint nature of credit card transactions. A credit card transaction depends on the activities and performance of both an issuer and an acquirer who are likely to be distinct firms. There needs to be an agreement from each issuer to each acquirer that issuers will guarantee payment to the acquirer and hence, to the merchant if an issuer's card is presented at

that merchant. If all of these agreements are forthcoming then a card association benefits from being able to support an 'honour all cards' rule; whereby each merchant who agrees to process that associations credit card transactions also agrees to accept any card (not matter who it is issued by) so long as that card carries the 'mark' of the association. An 'honour all cards' rule is automatic for a closed-loop system but it must be established by open loop systems in order for the system to operate effectively.

The existence of the 'honour all cards' rule makes a credit card more valuable for both customers and merchants alike. Customers know that, regardless of the specific issuer, any merchant who accepts the relevant association's cards will accept their card. For merchants, the 'honour all cards' rule allows them to look only at the card system for the guarantee of payment and not to consider the specific financial position of each issuer. Moreover, the rule allows acquirers to offer to process all cards issued under the system, when marketing to merchants. Thus, the 'honour all cards' rule is critical in ensuring the adoption of the credit card as a payment instrument by both customers and merchants alike.

The interchange fee needs to be considered in the context of the 'honour all cards' rule. When issuers and acquirers become members of an association they agree to uphold the 'honour all cards' rule. In particular, acquirers need to be able to acquire for all issuers and issuers need to rely upon this rule when marketing to potential cardholders. Thus, agreeing to the 'honour all cards' rule is an agreement to engage in the process of interchange. As depicted in Figure One, interchange begins when a customer who has a credit card requests that means of payment from a merchant who offers credit card facilities. That payment request gives the merchant the right to draw on the credit from the customer's issuing bank for the retail price agreed on by the customer and merchant. The merchant then passes the right on to its acquiring bank that settles the debt in return for a MSC (usually, some discount on the retail price). The acquirer then settles the debt with the customer's card issuer. Finally, the issuer collects the debt from the customer having earlier received other fees from the customer for the right to use a credit card. This completes the interchange process.

In order for this process to take place, the issuer and acquirer banks must agree to any financial payments that might pass between them. Under current arrangements, the interchange fee is paid from

acquirers to issuers as the greater proportion of costs associated with credit card transactions are currently borne by issuers relative to what they receive in revenues from cardholders. These costs include both processing costs and the costs associated with the issuer's guarantee of payment to the merchant. In principle, however, even if no financial payment passed from acquirer to issuer (i.e., the interchange fee was zero), the issuer and acquirer would have to *agree* not to impose a charge on the other. Without any agreement, as will be discussed in more detail below, issuers and acquirers alike remain vulnerable to expropriation by other parties who may command a greater relative bargaining position in the payment system.

Seen in this light, a collectively set interchange fee that applies equally to all card transactions (or a given type of transaction), regardless of the relevant issuer or acquirer, is a means of facilitating the issuer-acquirer agreement that is needed for the 'honour all cards' rule. It is also a means to aid expansion of the card system. An open loop card association will be more likely to be successful if it can easily accommodate new issuers and acquirers. Such accommodation can be achieved if new issuers and acquirers know they will face the same interchange terms that exist for other issuers and acquirers. A collectively agreed interchange removes the need for new issuers and acquirers to negotiate new agreements with existing issuers and acquirers. It also removes the potential risk for new issuers and acquirers of not being able to compete on the same terms as others on their side of credit card transactions.

This completes the description of the current operation of interchange arrangements and identifies the areas where a collectively set interchange fee facilitates the operation of card associations. To support the 'honour all cards' rule, issuers and acquirers in open loop systems need to agree upon fees (if any) that would be exchanged between them. A collectively determined interchange fee is one form of agreement. Further, to support the continuation of the system in the face of new entrant membership and other changed circumstances, uniformity of the interchange fee is important. Collective setting of the interchange fee allows for this uniformity. The key issue here, however, is whether alternative means would result in greater net public benefits. Those alternatives are evaluated in the next section.

2.1.2 Alternatives to Collective Setting of Interchange Fees

A collectively set interchange fee has two characteristics. First, it is a uniform fee that applies equally regardless of the particular issuer and acquirer that are parties to a given transaction; although it may vary according to transaction type (e.g., card not present and electronic transactions). Second, revisions to the fee require collective agreement or the use of some collectively agreed upon mechanism.

There are two alternatives to the collective setting of interchange fees. The first is that issuers and acquirers simply set charges for each other if a transaction goes between them. This is the *posted price option* whereby an issuer and an acquirer post a price per transaction or transaction value and if a transaction takes place each pays each other those prices. Second, issuers and acquirers can engage in *bilateral negotiations* of the interchange fee. In this situation, prices result from an agreement rather than from the price setting behaviour issuers and acquirers. Each alternative is evaluated in turn.

Posted Prices

Posted prices are the most familiar way by which transactions in the economy are settled. When purchasing an apple from the supermarket, the buyer accepts that they will have to pay the listed price. The buyer then decides whether or not to buy the apple.

A similar mechanism could be used to settle transactions in a credit card association. An issuer whose card was used by a merchant could simply post a price that the acquirer would have to pay. Similarly, an acquirer whose merchant processed a transaction by a customer could simply charge the customer's issuer an amount for that service. The net difference between these two charges would be the closest analogue to the interchange fee, as we know it today. In each case, the issuer and acquirer are aware of the posted price of the other party when both permitting a transaction and when setting prices to their respective clients (the customer and merchant respectively).

Posted prices lead to conflicts with issuer and acquirer obligations as members of a card association. For instance, suppose an acquirer must uphold the 'honour all cards' rule. This means that it cannot easily dissuade merchants from accepting card transactions from customers whose issuers are charging high prices. Consequently, individual issuers will be able to raise their prices and be

unconstrained by normal forces of a reduction in transactions processed. Similarly, an issuer would find it difficult to dissuade customers from using their cards at a particular merchant whose acquirer charged a high fee to the issuer. This lack of a demand response (i.e., making individual issuer and acquirer demands highly inelastic) would lead each to set relatively high charges.

The likely consequence of this is that there would be a significant incentive for issuers and acquirers to encourage 'on us' transactions that were not subject to interchange payments. 'On us' transactions occur where an issuer and acquirer are the same bank and effectively amount to partial vertical integration in the system – mimicking the response of a closed loop system. Ultimately, this would make entry by specialist issuers and acquirers almost impossible and would reduce the overall competitiveness in those respective functions. Cardholder fees and MSC would be higher than otherwise, to the detriment of all consumers. The end result would be a higher cost of transacting in the economy.

It should be noted that posted prices could involve a demand-side response if member's obligations under the system could be amended, for example by limiting the 'honour all cards' rule. But the pressures towards vertical integration would remain in this context, as 'on us' transactions would be priced significantly more efficiently than other transactions.[1]

Bilateral Negotiations

An alternative to posted prices is for each issuer and acquirer to engage in bilateral negotiations. This would not be unprecedented in banking circles as such negotiations form the basis of pricing for other transactions between banks (for example, electronic clearing). In those circumstances, however, there is no 'honour all cards' type of rule being supported and, as will be argued here, it is difficult to see how such a rule could be adequately and efficiently supported by bilateral negotiations.

[1] At its heart this is another version of the 'double marginalisation' problem in economics (see Economides and Salop, 1992). This arises because the pricing decisions of complementary input suppliers impose externalities on one another. When such prices are set independently, they are likely to give rise to inefficiencies to the detriment of both firms and consumers alike.

To begin this discussion, we must specify precisely what banks would be negotiating over if they were to negotiate bilaterally. If an issuer and acquirer were not integrated into each other's functions, this negotiation would simply be over the interchange fee paid between the banks.

However, let us explore the case of non-integrated issuers and acquirers further. Compared with the uniform fee that is currently set for interchange, bilateral negotiations are unlikely to generate such uniformity. What will be the impact of this on issuer and acquirer behaviour? Let us focus on the acquirer case. In this situation, an individual acquirer will face a different fee (and hence a different cost) depending on the issuer associated with a specific transaction. An acquirer then faces several choices. First, the acquirer might bear these differences by setting a uniform MSC despite facing differential interchange fees. The MSC would tend to reflect an average of the interchange fees. In such circumstances, in negotiations, an issuer would benefit more from having a higher fee than the acquirer would lose. To the extent that an increase in any specific issuer's interchange fee is simply absorbed by the acquirer or diluted as part of a uniform MSC, each issuer gains all the benefits from raising its interchange fee, but shares the costs of such a rise in the fee. These costs include, for example, fewer merchants agreeing to accept the credit card as the MSC rises. Thus, in this circumstance, pressure from issuers will tend to lead to interchange fees that are undesirably high from the perspective of the card association.

Alternatively, when faced by different interchange fees under bilateral negotiations, an acquirer might choose to set different MSCs according to the issuer associated with any transaction. But this simply shifts the problem of differential charges from the acquirer to the merchant. If the merchant cannot distinguish between different types of credit cards when charging customers, then the merchant will base its price on an average MSC. Again, each issuer will have an incentive to raise its interchange fee as it will get all the gain but share part of the cost with other association members. Alternatively, the merchant could try to pass the different MSCs onto the consumers. The merchants would have to try and set different prices depending on the issuer of the credit card used by a customer. This is likely to be impractical and will undermine the universal acceptance of the credit card; even in the absence of a no surcharge rule. Overall, the aim of the association - to have a widely accepted credit card that operates seamlessly regardless

of the specific issuer and acquirer - would be fundamentally undermined.

Of course, a similar pressure would exist on the issuer side as a result of differential interchange fees. On one level, this would provide a mitigating force on the upward pressure on interchange fees described above. On another, however, there are two reasons issuers may be in a superior position to manage these impacts; hence, shifting bargaining power towards them and away from acquirers. First, issuers may be more easily able to provide incentives for their cardholders to favour merchants who are serviced by specific acquirers. Hence, the impact on their cardholders from a specific interchange negotiation could be tied to transactions specific to that relationship rather than diluted in the issuer's overall pricing strategy. Secondly, there is a sense in which the 'honour all cards' rule might weigh more heavily on acquirers than issuers. For example, Small and Wright (2000) argue that this rule would mean that an acquirer would be unable to perform its function at all if negotiations with any single issuer broke down. This would give issuers tremendous bargaining power in any particular negotiation leading to a rapid escalation of interchange fees. Indeed, the more competitive the issuing market, the more likely is such an escalation to occur to the detriment of the card association.

In reality, if members of a card association were forced to negotiate bilaterally, then the card association would have to modify its 'honour all cards' rule to prevent the type of destructive escalation identified by Small and Wright. How (and whether) associations would achieve this is an open question. Regardless, a scaling back of the 'honour all cards' rule would fundamentally alter the competitiveness of credit cards from 'open loop' systems as payment instruments. Indeed, one possible response would be for those associations to limit issuer and acquirer activity and act more like closed loop systems. As such 'closed loop' systems typically have higher cardholder fees and MSCs, such a response would reduce the overall level of competition in the market for payment instruments to the detriment of consumers and social welfare alike.[2]

[2] See Rochet and Tirole (2000) for a comparison of the differing pricing strategies pursued by closed and open loop credit card systems.

A lack of a uniformly set interchange fee compromises the openness of 'open-loop' card associations. If new acquirers or issuers want to be part of the system, they would have to accept or negotiate interchange arrangements with *all* other issuers and acquirers in order to be able to issue a card capable of being honoured by all merchants who accept the association's cards. This will likely delay and complicate entry into issuing. Indeed, the RBA / ACCC Joint Study was concerned about this very feature of interchange arrangements for debit card and ATM networks.[3]

In summary, bilateral negotiations will likely involve the following:

- *Complexity*: each issuer and acquirer will have to engage in negotiations with one another. This will involve a considerable commercial burden; especially on smaller issuers. Indeed, there will be 234 such negotiations between the 26 issuers and 9 acquirers in Australia in each association; totalling over 500 negotiations. Issuers and acquirers would have to formulate pricing strategies to accommodate the differing interchange fees arising from such negotiations.

- *Issuer Bargaining Power*: issuers are likely to command a superior bargaining position in such negotiations, leading to higher interchange fees than is currently the case.

- *Dilution of the 'Honour all Cards' Rule*: the 'honour all cards' rule would probably be unable to operate as effectively as it does now, reducing the competitiveness of 'open loop' credit cards as payment instruments and increasing the costs of transacting in the economy.

On these grounds, bilateral negotiations would involve public detriments as compared with the current system of a collective uniform interchange fee and, in addition, would create entry barriers in both issuing and acquiring as well as reducing the effectiveness of 'open

[3] Interestingly, the ACCC has in the past appeared to encourage centralised price setting in both telecommunications – by accepting access pricing undertakings – and in electronic payments systems where in its draft determination of the CECS association encouraged it to set inter-bank fees rather than have them determined in bilateral negotiations. See Gans and Scheelings (1998) for a discussion.

loop' card systems as competitors in the overall payment instrument market.

However, there is a sense in which the above discussion is only the 'tip of the iceberg' insofar as costs associated with bilateral negotiations are concerned. First, the above analysis assumes that the bilateral negotiations were to apply only to transactions between domestic issuers and acquirers. There is, however, an international context to the setting of interchange fees. As mentioned above, non-uniformity in interchange fees may undermine the association's ability to have merchants 'honour all cards'. However, one of the chief benefits of holding a MasterCard or Visa credit card is that such cards – even if issued by an Australian bank – are accepted overseas. If Australian issuers had non-uniform interchange rates for all of the thousands of acquirers that exist internationally this may compromise the ability of Australian members of MasterCard and Visa to remain part of these international joint ventures. The same is true for Australian acquirers who process transactions by overseas cardholders who visit or otherwise purchase from Australian merchants.

Second, for large issuers and acquirers (and perhaps all such issuer/acquirers), agreeing even bilaterally to an interchange fee may still raise competition concerns.[4] These banks would be direct competitors in the issuing and acquiring segments. Nonetheless, there would be concerns regarding the reciprocity of interchange arrangements.[5] As in the posted price analysis above, this possibility would mean that 'on us' transactions are likely to involve more efficient pricing terms than other transactions. In this case, there would be pressures on issuers and acquirers to in fact become integrated. This would favour larger banks at the expense of smaller ones.

2.1.3 Conclusion

In summary, in the absence of a collectively set interchange fee arrangement, issuers would have an incentive to set excessively high interchange fees and 'free ride' on other members of the association. If

[4] Small and Wright (2000) argue that this would in fact be no different from the collective setting of interchange fees and a system of bilateral negotiations would entail each issuer/acquirer to divest one of its functions.

[5] This reciprocity has led to concerns in telecommunications where competing networks negotiate mutual interconnection arrangements. See Laffont and Tirole (1999) and Gans and King (2001d).

differential interchange fees led to differential MSCs, then the universal acceptance of the card regardless of the identity of the issuer would be undermined, reducing the value of the association and potentially causing it to collapse. Without a commonly set uniform interchange fee an 'open loop' system is unworkable.[6] Moreover, it is very difficult to see how the common costs of the joint venture as well as its international aspects could operate in the absence of uniformity. Thus, in our opinion, uniformly set interchange arrangements are an integral part of card associations designed to allow for many distinct issuers and acquirers and to consequently have all cards issued under the association's brand and standards accepted by all merchants.[7]

2.2 Competitive Determinants of the Interchange Fee[8]

There are two dimensions to competition in credit cards. At one level, within an association, issuers and acquirers compete for customers and merchants respectively. Second, credit card associations compete in what has been termed 'systems competition' with other credit cards and payment mechanisms. Both of these dimensions of competition could affect and be affected by the size of the interchange fee. It is this issue that is examined here.

Concerns are often expressed that the interchange fee can be used to soften competition and inflate prices or alternatively that the fee may be used as a device that leads to over-use of a particular credit card or credit cards in general. On one level, these two arguments are somewhat contradictory. If the interchange fee were a potential collusive instrument designed to raise MSCs, such increases in the 'price' of card transactions would lead to their under-use rather than over-use. The over-use argument relies on the notion that an interchange fee can be used to extract rents from non-credit as well as credit card users; thereby, making customers more likely to use credit

[6] "[E]ven if each member could afford to issue its own card, universality of acceptance - the key to a national payment system - could not be guaranteed absent prearranged interchange rules." (*National Bancard Corp v Visa* USA, 779 F.2d 592, 1986 at p.602).

[7] Indeed, while we do not analyse the issuer here, the lack of uniform pricing arrangements probably accounts for the lack of take up of debit transactions and perhaps may have reduced the efficiency of operations of ATM and EFTPOS systems.

[8] This section draws heavily upon Gans and King (2001a).

cards at the margin. Nonetheless, each argument is similar in that each relies on the possibility that interchange fees are set too high.

The purpose of this section is to evaluate the impact of the interchange fee on competition and the efficiency or inefficiency of credit card use. It is concluded that the arguments for inefficiency rest on a specific set of assumptions that are unlikely to be valid across a wide number of sectors in the economy. Hence, we argue that concerns about the use of the interchange as an instrument of market power are, at worst, implausible and, at best, highly overstated.

2.2.1 Interactions Between Prices

As a starting point it is useful to consider the interaction between various prices that make up a credit card transaction. These prices include the card fee and charges, the MSC, retail prices for card transactions and the interchange fee.

Consider the MSC set by acquirers. The costs of acquiring include the direct costs borne by acquirers and the interchange fee acquirers pay to issuers. An increase in the interchange fee – by raising acquirer costs – is likely to lead to an increase in the MSC. This, in turn, will raise the costs faced by merchants. If, as is usually the case, merchants charge the same price to customers regardless of the payment mechanism, this increase in MSC will also increase retail prices.

Issuers will set fees and other payment terms to card holding customers. Operating costs, including the direct costs associated with processing and risk bearing, will be an input into the prices set by issuers, but the interchange fee they receive from acquirers mitigates these costs. Therefore, an increase in the interchange fee lowers the net costs of issuers and is likely to lead to lower card fees and charges. This in turn increases card adoption rates among customers.

If the overall 'price' of credit card transactions is made up of both card fees and the MSC, the effect of a change in the interchange fee is ambiguous. Increasing the interchange fee merely shifts the balance in that price towards merchant charges and away from customer charges. Decreasing it would have the opposite effect. In this sense, an increase in the interchange fee cannot simply be likened to increasing the price of an input into the service. For credit card systems, the fee represents both a cost to one complementary supplier

and a cost-offset to another; implying no simple relationship between the interchange fee and the overall price of the service.[9]

Complications arise to the degree that many issuers are also acquirers. For these members, the interchange fee is only a cost for transactions that are not 'on-us;' that is, for transactions where the issuer and acquirer are distinct entities. In contrast, when a transaction is 'on-us' no interchange fee is paid and the issuer / acquirer apportions costs according to their own internal procedures. In principle, these procedures should reflect the actual costs associated with a transaction and may well differ from the interchange fee. In this respect, the interchange fee is not simply a floor on MSCs but part of the weighted costs involved in setting a MSC.[10]

Clearly the claim that a rise in the interchange fee will raise the 'price' of credit cards transactions has no basis in economics. The claim ignores the joint nature of the credit card process and neglects the fact that a cost to acquirers is a benefit to issuers and the cardholders who are their customers.

---

[9] Similar trade-offs arise in other industries. Take, for example, telecommunications and the pricing of fixed to mobile calls. A critical input into the price of a call from a fixed line to a mobile phone is the termination charge the fixed line network pays the mobile network for completing the call. As the caller pays the retail price of the call, an increase in that termination charge would increase the retail call price. However, from the perspective of a mobile phone network, this makes the attraction of additional subscribers more profitable as they also attract termination revenues. In effect, these termination revenues offset the costs of signing up an additional mobile phone customer and that, in turn, will be reflected in the price customers pay for their mobile phone subscription. Thus, a change in the termination charge (effectively a wholesale price) has an ambiguous effect on call prices overall. See Gans and King (2000) and Gans and King (1999) for a more detailed discussion of this trade-off and the role of competition in such pricing behaviour.

[10] Even where issuing and acquiring is relatively concentrated, the proportion of 'on-us' transactions can be small. Consider the following table that derives an estimate of these for the Australian industry.

Bank	Issuer Shares of Customer Spending	Share of Acquiring Revenue	Probability of 'On-Us' Transactions (Percent)	Net Revenue from Interchange
ANZ	28	19	5.32	Positive
CBA	19	31	5.89	Negative
NAB	21	28	5.88	Negative
WBC	19	16	3.04	Positive
Other	13	6	0.78	Positive

Source: Merrill Lynch (2000), Table 4. Last two columns calculated assuming 'symmetric' customer and merchant types. This table includes all card associations and, therefore, slightly distorts true numbers.

2.2.2 Incentives to Raise the Interchange Fee

As just described, a rise in the interchange fee is likely to raise MSCs but lower card fees. The Cruickshank and RBA/ACCC reports were explicitly concerned that card associations would have an incentive to increase the interchange fee with precisely this effect. Both argued that in card associations there would be limited resistance to higher interchange fees because some acquirers were also issuers who would benefit from higher fees and also because acquirers "are able to pass interchange fees on to their customers, the retailers, safe in the knowledge that all of their competitors face the same cost base" (Cruickshank p.264). Furthermore:

> Inflated interchange fees ... raise the cost to retailers of card payments. This reduces the acceptance of particular payment methods. If interchange rates were lower, credit and debit cards would be likely to be accepted in a wider range of retail outlets (such as smaller retailers) or in more non retail contexts (such as paying bills). Higher interchange fees raise retail prices generally, as retailers pass on their inflated costs to their customers. This in turn leads to a reduction in output and economic welfare. (Cruickshank p.81)])

Thus, not only are there supposed incentives for associations to inflate interchange fees, the consequences of this will be detrimental to overall economic efficiency.

What is interesting about this line of logic is that it can be completely turned on its head to demonstrate that associations would have incentives to choose a low rather than high interchange fee. The reverse argument would go like this: *a lower interchange fee would be unlikely to be resisted by issuers for two reasons. First, many issuers are also acquirers who would benefit from the reduction in interchange fees. Second, issuers "are able to pass interchange fees on to their customers … safe in the knowledge that all their competitors face the same [net] cost base."* Following this argument through, according to the second step in the Cruickshank logic, the lower fees would raise merchant adoption incentives and lower retail prices in general as merchants pass on their lower costs to customers, leading to an increase in output and economic welfare.

The problem here is not that these two alternative lines of argument are contradictory but that each is incomplete as each focuses solely on the role of the interchange fee on a single side of the credit

card transaction; Cruickshank on the merchant-acquirer side, our reformulated version on the customer-issuer side. This emphasises the need to apply a reasoning based on the potential choices made by all four parties to the transaction together.

A more sophisticated line of argument is that articulated by Frankel (1998) and essentially formalised by Rochet and Tirole (2000); it was also recognised implicitly in the Cruickshank report. The argument is that, because merchants do not vary retail prices to customers based on the form of payment (e.g., credit card versus cash),[11] cash and cheque customers are implicitly cross-subsidising card customers. The reason for this is that retail prices are based on the merchant's average MSC payments. As there is a mix of customers, this means that a rise in the MSC is passed on to both cash and card customers even though cash customers are not receiving the potential benefits from card transacting. This, in turn, impacts positively on customers' decisions to adopt credit cards. Recall from the previous section that the decision to adopt a card is based on the direct benefit a customer will receive from having a card as compared with the benefits from cash. If retail prices were higher for credit card users than cash users, on average customers would value credit cards less than if the price was the same regardless of the payment mechanism.

The problem is that, under certain conditions where the net social value of credit cards is relatively small (Rochet and Tirole, 2000, Proposition 3), it is possible that this potential for cross-subsidy may lead to an over-provision of credit card transactions from a social efficiency perspective. Essentially, the ability to cross-subsidise makes merchants more willing to bear higher MSCs as they derive an indirect benefit from this cross-subsidy. The association, that sets the interchange fee to maximise the overall profits of its members, then finds it desirable to raise the interchange fee choosing, on balance, to earn its revenues from merchants rather than customers. This is because the higher MSC, while increasing prices on average, does not change the relative prices paid by card versus cash users. Hence, up to a point, by increasing the interchange fee a credit card association with market power is able to extract rents from cash customers.

[11] This could be because of a no-surcharge rule imposed by the association or because of the convenience or transactions costs associated with providing such variation. We will discuss this in more detail below.

Economists are generally concerned about the distortionary impact of cross subsidies. While the beneficiaries, in this case cardholding customers, prefer this state of affairs, customers who do not have cards suffer a detriment. On balance, it would be preferable to remove the cross subsidy.

However, the proponents of the 'cross-subsidy' view overstate their case by assuming away mechanisms that would diminish or eliminate that subsidy. The force of retail competition as well as the fact that retail price variations are possible act to reduce the cross-subsidy and consequently, the incentive for an association to inefficiently raise its interchange fee to exacerbate such distortions.

2.2.3 Merchant Retail Price Competition

The force of the cross-subsidisation argument for inefficiently high interchange fees is contingent on *all* merchants in a given retail sector having card facilities and being unable or unwilling to offer card surcharges or cash discounts. This means that cash customers have no choice but to bear the higher retail prices.

In reality, in many retail sectors, merchant competition is stronger. This means that if retail prices for cash customers are high because all merchants offer card facilities, there is an incentive for one merchant to offer a cash price only or alternatively for a new entrant to offer cash-only prices. Only when there are high entry barriers (and products are not close substitutes in the eyes of customers) will no merchant wish to do this (Rochet and Tirole, 2000). In a highly competitive merchant segment, all cash customers will go to new entrants.

Gans and King (2001b) contains a technical analysis of this possibility. The main result is that under conditions of high retail price competition, merchants will segment themselves into card adopting and pure cash merchants respectively with card customers going to card merchants and cash ones to cash merchants. Because merchant competition is high, there will exist two prices – a cash and card price – whose differential is precisely the merchant's net costs associated with offering card facilities.

If the interchange fee rises, this will lead to an increase in the MSC. With strong retail competition, however, this will mean higher prices for cardholders only as non-card carrying merchants will not

bear any of the increased costs and, if there is perfect competition among cash merchants, they will not increase their retail price. The higher price for cardholders diminishes their willingness to pay for the credit card in the first place, putting downward pressure on card fees. Gans and King (2001c) demonstrate that all these price changes net out. That is, the interchange fee, while causing a rebalance in card fees and MSCs, does not change the overall usage of cards or level of effective retail prices in the merchant sector. Moreover, this result holds regardless of the degree of competition in the issuing and acquiring segments or the level of integration among issuers and acquirers.[12]

What this means is that competition in the merchant market is likely to neutralise any *allocative* effects from higher interchange fees. From a competition perspective, we need not be concerned about the setting of interchange fees where merchant segments are reasonably competitive; especially where there exist cash (or other payment type) retail prices that are themselves determined competitively. In our experience, most retail segments conform to these characteristics rather than ones where *all* merchants have card facilities. Ultimately, however, the validity of either assumption is an empirical matter.[13]

2.2.4 Lack of Retail Price Variation

Where retail market power is an issue or there is a high degree of product differentiation among merchants, our previous conclusion – that retail competition will neutralise the pricing effect of the interchange – will not necessarily arise. Nonetheless, the above analysis has assumed that it is impossible for a merchant offering credit card facilities to vary its price to customers based upon whether they use a card or not. Recall that in this situation, inefficiencies may arise because cash customers cross-subsidise card customers, giving the association an incentive to raise the interchange fee.

---

[12] Others had noticed that when all segments – issuing, acquiring and the retail segment – were perfectly competitive, the interchange fee did not matter. One reason for this is that each issuer and acquirer would make zero economic profits regardless and so would be indifferent as to the choice of fee (Rochet and Tirole, op cit note 25). However, the other reason was that competition would force retail price variation as the interchange fee changed; also neutralising its effect. See Carlton and Frankel (1995) at pp.656-9. Those authors argued (p.660) that if there were imperfect competition among issuers or acquirers, then this neutrality would not hold. Gans and King (2001b) demonstrates that this concern was unfounded and that neutrality rests critically on the existence of a competitive cash retail price in merchant segments.

[13] Interestingly, Aldi– the large scale supermarket entrant in Australia in 2000 – does not accept credit or debit card in order to focus on keeping its prices low.

One reason why merchants do not vary retail prices is because some credit card associations impose on merchants a condition that they do not charge a surcharge to credit card users. In Australia and the United States, however, the 'no-surcharge' rule does not prevent merchants from offering discounts to cash or cheque customers (or even EFTPOS customers). From an economist's perspective, this means that the 'no-surcharge' rule is relatively innocuous and merchants would be effectively free to vary prices according to payment mechanism; the only difference between a cash discount and a card surcharge being how you set the original retail price.

In this situation, the same outcome as in the case of retail price competition will arise. That is, merchants will not find it optimal to support the cross-subsidy (even if they were monopolists) as to do so would reduce their own profits, for any level of the MSC. This would mean that an increase in the interchange fee leading to higher merchant service fees would be precisely offset by lower credit card usage and lower customer fees for cards themselves.[14] On the plus-side for the association, there would be limited merchant resistance in adopting card facilities as they would not have to fear the potential profit-reducing detriment of supporting a cross subsidy from their cash customers.

In reality, however, many merchants choose not to vary their retail price according to card usage. Interestingly, it is probably only where the items concerned are 'big ticket' that there is merchant pressure on customers to chose payment methods other than credit card. This suggests that the cross subsidy imposed by the lack of retail price variation is small. Indeed, it is more plausible that the association will choose its interchange fee to balance the incentives of acquirers and issuers rather than exploit even a small cross-subsidy. In any case, the social inefficiency may not be large. Once again, this is ultimately an empirical matter to assess the magnitudes of any cross-subsidy as well as its impact on credit card usage and retail prices.

2.2.5 System Competition

A final mitigating effect on the potential anti-competitive effects of the use of the interchange fee by credit card associations is that there

[14] See Rochet and Tirole (2000) for a formal analysis of this situation.

is competition among associations and with other payment mechanisms. In Australia, competition between MasterCard, Visa, Bankcard, American Express, Diners Club and various store cards provides a competitive restraint on the fees and charges levied on both sides of a credit card transaction. In addition, the use of point of sale electronic transfer as well as traditional methods of payment, such as cash and cheques, provide competing payment options in the eyes of customers and merchants.

It is difficult to assess the strength of competition among these different payment methods. Evans and Schmalensee (1998) believe competition to be quite intense – even among associations with the same members – and point to expenditures on advertising and card promotion. Of course, the targets for such advertising could be closed card systems such as American Express rather than Visa or MasterCard. Evans and Schmalensee however do point to various advertising campaigns where Visa and MasterCard tend to target each other. This said, some commentators believe that credit card associations command market power.[15]

The key point is that credit card systems are not monopolies and competitive pressure will limit the extent to which those associations can manipulate interchange fees away from efficient levels to extract rents from either customers – card or cash – or merchants. If a credit card raises the interchange fee by too high a level, it will face difficulties in getting merchant acceptance or invite retail price variation. Indeed, the high MSCs of American Express have long been a problem in them gaining increased merchant acceptance.[16]

2.2.6 Conclusion

Arguments regarding the use of interchange fees as a device to simply raise prices such as MSCs are highly incomplete as they could equally be an instrument whereby such prices are lowered. The more sophisticated concern is that cash retail customers cross-subsidise

[15] See Ausubel (1991). Ausubel conducted a large empirical study of competition in credit cards in the United States and concluded that despite large numbers of participants, price patterns did not appear to reflect those that would arise from vigorous competition. Evans and Schmalensee, *op.cit*., rebut some of these suggestions; although the academic controversy remains.

[16] In Boston, a group of high quality restaurants, upset over American Express's high fees boycotted the card. The so-called 'Boston Fee Party' led eventually to a special merchant service fees for that group.

credit card customers and the interchange fee could be used as an instrument to capture more rents from cash customers thereby promoting over-use of credit cards.

However, this is only a concern if there is a lack of retail price competition, there are restrictions on retail price variation, *and* there is limited competition from alternative payment instruments. If any of these conditions did not hold, the usefulness and hence, impact of the interchange fee as a device for socially inefficient credit card usage is removed.

Moreover, it should be recognised that even where these conditions do hold, inefficient overprovision of credit card transactions only arises when the social net benefits from having credit card transactions relative to alternative payment instruments is small. When those net benefits are high, then Rochet and Tirole (2000) have demonstrated that it is socially efficient to have a high interchange fee and encourage customer as opposed to merchant adoption.

In general, economists find it desirable to unbundle particular dimensions of a transaction and would prefer prices to reflect true costs, even if that variation in costs arises from fees charged by a card association.[17] Thus, if there is a perceived problem relating to excessive interchange fees, a preferred solution might be to increase the transparency and variability of pricing.

> If credit card merchant discount fees were completely unbundled, so that merchants posted a schedule of discount rates and charged these fees as a separate line item in the same manner in which sales tax is added to the transaction amount, then there likely would be more intense and direct price competition among the various payment systems and subsidies from cash paying customers would cease. Unbundling such discount fees, however, might increase transaction costs significantly for some merchants. Therefore, while it might not be advisable to *require* unbundling, merchants should be given the freedom to pass payment system costs along to consumers through whatever surcharges, rebates, or multi-tier pricing systems they choose, as long as there is full disclosure to consumers of their pricing policies. Frankel (1998) p.348; emphasis in original.

[17] Elsewhere we have suggested that regulators play a role in encouraging customer understanding of price variations among telecommunications carriers. See Gans and King (2000).

Encouraging retail price variation and customer awareness of price differentials will help offset any potential competitive problems and can be socially desirable.

3 Principles for the Setting of Efficient Interchange Fees in Payment Systems

In 2001, the Reserve Bank of Australia used its new payment system powers to designate the credit card associations of Bankcard, MasterCard and Visa in Australia. This designation gives the RBA the ability to set standards for interchange in those associations; effectively regulating interchange fees. The designation decision was based on a concern that credit card associations were not operating in a way that would maximise economic efficiency – especially in terms of how interchange fees are currently set.

In the current debate regarding interchange fees, there are two reasons given why credit card transactions may not be efficient. First, imperfect competition in issuing or acquiring may mean that cardholder fees and/or MSCs are relatively high producing inefficient under-utilisation of credit cards. Second, the no surcharge rule in combination with a high interchange fee means that cash customers subsidise card customers of merchants leading to inefficient over-utilisation of credit cards.

Our earlier paper (Gans and King, 2001b) dismissed this second objection on the basis that competition among retailers would likely unravel the adverse consequences of no surcharge rules. However, that argument also implies that interchange fees are neutral and thus does not in itself suggest what might be a socially optimal interchange fee. Here, we take as an assumption that no surcharge rules, in combination with retailer market power, produce real economic effects from the setting of interchange fees and, consider what a socially optimal interchange fee might be in this environment.

The goal of this section is two fold. First, we consider the concept and guiding principles as to how an interchange fee may be set in an environment where it is not neutral. Second, we offer a specific methodology for the calculation of such fees – the shared avoidable cost approach – that satisfies these principles and, we believe, offers a theoretically rigorous and practicable means of determining interchange fees in credit card associations and other four party payment systems.

3.1 The Network View of Payment Systems

In the recent debate surrounding the nature of the operation of credit card associations, the setting of interchange fees has often been considered as analogous to the setting of access prices for the use of essential facilities. However, the economics of payment systems is different and distinct from the economics of essential facility access. The principles of access pricing for essential facilities are based on a vertical chain view of production. In contrast, payment systems operate according to network principles. As a result, the concepts and ideas fundamental to essential facility access simply cannot be applied to payment systems.

In what follows, and throughout this paper, we will use credit card associations as our example of a payment instrument. This reflects the focus of the current policy deliberations. However, it should be emphasised that all of the arguments and recommendations in this paper could equally apply to *any four party payment system* with collectively set or regulated interchange fees between issuers and acquirers (e.g., credit cards, debit cards and checks).

3.1.1 The Difficulty with a Vertical Chain Approach

As a starting point, we show why the vertical chain view of production does not directly apply to payments systems. Under a vertical chain approach to credit cards, the interchange fee is a payment from acquirers to issuers for services provided by issuers to the acquirers' customers, the retail merchants. This approach to credit card associations involves a one-way flow of service from issuers to merchants (through acquirers). The vertical chain view considers the interchange fee as the wholesale price paid by acquirers to issuers. As acquirers need an input provided by issuers under this vertical chain view of the payment system, the approach leads to a set of pricing principles with the primary goal of regulating the pricing behaviour of issuers. For example, the vertical chain view might lead to the recommendation that the interchange fee should simply reflect issuers' costs.

The problem with this one-way approach to payment systems becomes obvious when we note that issuers also earn revenues from customers and market to those customers. Moreover, those customers interact directly with the merchants who (under the vertical chain

approach) are seen as the end point in the consumption of credit card services. This interaction means that interchange fees cannot simply be viewed as a wholesale price.

The key assumption being implicitly made to justify a vertical chain approach is that merchants are the sole beneficiaries of the services of a credit card. In reality, both merchants and their customers benefit from the use of credit cards as opposed to other payment instruments. Both merchants and their customers receive some surplus from the use of credit cards and face real, inter-linked decisions regarding adoption of the payment instrument. That is, when customers determine whether or not to adopt a particular card, they will look to the level of merchant adoption of that card as well as the prices charged by those merchants relative to those who do not accept that card. Thus, customers and merchants have distinct benefits from a card's use, and these benefits are taken into account when determining the retail price of the goods and services a merchant provides to a customer. It is not possible to simply separate out merchant benefits from the prices paid by customers and hence, from customer benefits.

The network of interactions mean that interchange prices set under a vertical chain approach will not usually be socially desirable. For example, if the interchange fee is set so as to reflect issuer costs, then acquirers will consider the total costs (issuing and acquiring) associated with a card transaction. The price acquirers set to merchants – the MSC– will reflect those total costs. In contrast, issuers will not be concerned about any costs imposed on acquirers when signing up additional customers and encouraging them to use a credit card. If customer-merchant interactions are such that merchant net benefits associated with credit card use are not separated (in some manner) from other aspects of retail pricing, then it is likely that customers will use cards excessively from the merchant's perspective. This is because customers do not fully internalise the costs associated with credit cards that are borne by merchants.

3.1.2 Taking a Network View

A network view of payment systems views customers and merchants as the joint beneficiaries of any payment instrument. One cannot simply separate out the activities of issuers and acquirers because these activities interact with each other. The activities of issuers and acquirers together support the effective operation of the payment system.

Under the network view, the interchange fee is a payment made between acquirers and issuers designed to allow their respective consumers – merchants and customers – to more properly gauge the impact of their decisions to use a credit card on the total costs of both issuers and acquirers. This implies that customers should pay fees (or receive other benefits) based on the total costs of issuing and acquiring.

As a result, an interchange fee set at zero would be no more appropriate than one that reflects only issuer costs. Unless issuing and acquiring costs are identical, a socially desirable interchange fee will be based on both issuer and acquirer cost information (Although, as we will demonstrate below, this does not mean that the interchange fee would be increasing in both cost components).

In a network view, the interchange fee is not a simple wholesale price but a transfer payment designed to ensure that issuers and acquirers take a balanced view of the costs imposed on the system by their respective activities. This, in turn, means that issuers' and acquirers' customers will face the correct incentives to use the relevant payment instrument. In effect, just as wholesale prices allow wholesale costs to be reflected in the prices of final goods, an interchange fee should allow total system costs to be reflected in the respective payments made by both customers and merchants alike. That is, an interchange fee should respect that fact that the use of a credit card involves value at two consumer points rather than one.

3.1.3 Regulatory Concerns

The network view provides a different perspective on regulatory concerns regarding the level of interchange fee to that provided by the vertical chain view. The vertical chain view would be concerned about the interchange fee being used to restrict the activity of acquirers; thereby raising MSCs and restricting the overall use of credit cards. In contrast, the idea that interchange fees can be used to leverage monopoly power into the acquiring segment cannot be sustained under a network view that respects customer-merchant interactions and customer adoption decisions.

The regulatory concern that arises from a network view is that there may be distortions imposed by interchange fee levels on the use of one payment instrument over another. In particular, as argued by Frankel (1998), when there is a lack of price coherence between credit cards and cash (such that customers do not face a range of retail prices

contingent on the instrument used), then the usage pattern of payment instruments is unlikely to be socially optimal. In fact, if a card association commands some market power in the market for payment instruments, that card may be overused. Interchange fees might be set so that customers and merchants perceive the relevant credit card as being cheaper at the margin relative to other instruments. Rochet and Tirole (2000) qualify this statement and demonstrate that overuse will only arise where the issuing segment is sufficiently competitive. If there is a large degree of issuer market power, then any tendencies for overuse mitigate the higher costs associated with that market power.[18]

As a result, under the network view, the regulatory concern becomes one of overuse of credit cards rather than the lack-of-use conclusion from a vertical chain approach. Interchange fees – to the extent that they are regulated – should be constrained so as to limit their potential to be used as an instrument that encourages such overuse. For this reason, it is important – as we will argue below – that interchange fees only reflect issuer and acquirer costs, and not revenue elements.

3.2 Basic Principles for the Setting of Interchange Fees

We now turn to consider the basic principles that should be part of any methodology designed to set interchange fees in a socially desirable manner. It is important that the interchange fee be set with respect to basic regulatory concerns regarding distortions in the relative costs of payment instruments, competitive interactions in issuing and acquiring (i.e., the fact that the overall effectiveness of a payment instrument depends on issuers' and acquirers' activities, including pricing), and competition between payment instruments. Here we consider the principles related to each in turn.

3.2.1 Minimising the Cost of Transacting

The total cost associated with using a credit card is the cost to issuers and acquirers less the direct benefits to merchants and customers from its use. The average cost of transacting in the economy

[18] We have argued elsewhere that this argument depends critically on the level of market power among merchants as well the effective operation of a no surcharge rule (Gans and King, 2001b).

is the weighted sum of transaction costs associated with all payment instruments (including cash, cheque, credit, debit, barter etc.). If regulators are concerned that interchange arrangements for credit cards result in private costs that are lower than the social costs (say because cash customers 'cross subsidise' card customers at the merchant level or because one particular form of payment is subsidised by government), this suggests that their basic concern is that the cost of transacting is not being minimised. As such, the overall goal should be to set interchange fees that minimise the costs of transacting in the economy.[19]

It is, therefore, important that the interchange fee be set with regard to factors that are related to the social but not private cost associated with credit cards. For instance, Rochet and Tirole (2000) argue that merchants may adopt credit card services partly because the costs of adopting and using credit cards are passed through to all their customers and not just those using credit cards. This factor reduces merchant resistance to card adoption and thus may enable a card association to raise its profit by raising the interchange fees paid by acquirers. Thus, while in some circumstances it may be appropriate to take into account direct merchant benefits from processing card payments (e.g., increased security and customer attractiveness), it is important that overall merchant benefits (as say measured by acquirer and even issuer revenues) that include implicit subsidies from cash or other customers not be taken into account in the setting of the interchange fee.

Given this we can establish our first principle:

> **Principle 1:** *Interchange fees should be based only on issuer and acquirer costs and information relating to the direct net benefits customers and merchants receive from using a particular payment instrument.*

This principle excludes the possibility of using issuer and acquirer revenues in calculating the interchange fee as is done in some practiced and proposed methodologies. This implies that an association would be unable to set an interchange fee to exploit reduced merchant resistance afforded by their ability to average payment instrument costs across cash and card customers.

[19] Baxter (1983) and Rochet and Tirole (2000) use these criteria in evaluating the operation of payment systems.

The goal that interchange fees should be set so as to minimise the costs of transacting also assists us in refining further the nature of issuer and acquirer costs that should be included in interchange fee calculations. Credit cards, for example, offer two functions bundled together in a single product. First, they offer payment functionality – allowing customers to purchase goods and services. Second, they offer a line of credit – allowing customers to purchase goods and services on credit without prior approval. Notice that, for credit cards, it is not easy to decouple these functions. A reason why credit cards offer a means of payment for consumers is precisely because consumers do not have to worry immediately about the balance of funds in their bank account. They are extended credit for several weeks and indeed longer if, because of short-term cash flow problems, they cannot settle transactions immediately. Therefore, the value of payment functionality resulting from credit cards as opposed to other payment instruments is precisely because of the liquidity it affords.

That said, once credit is extended beyond say, a couple of months, payment functionality is less of an issue. This is because cardholders always have the option of exploring other means of financing outstanding card balances. It is only those cardholders who cannot secure alternative means of financing for which the extended time credit line is a critical function.

The relevant cost components for setting the interchange fee will relate to its functionality as a payments instrument. This is summarised by our second principle.

> **Principle 2:** *Issuer and acquirer costs should not be included in the determination of the interchange fee if they relate purely to an extended line of credit (e.g., beyond one quarter). On the other hand, costs related to payment functionality should be considered in the determination of interchange fees.*

This principle excludes cost components related to extended lines of credit but allows those that enhance payment functionality (and its liquidity value) even where these involve, in part, credit line functions.

3.2.2 Respect the Nature of Competitive Processes in Issuing and Acquiring

It is important to recognise that when setting interchange fees, the remaining prices associated with credit card transactions – cardholder fees and MSCs – will be determined by the nature of

competitive processes in issuing and acquiring. This impacts further on the type of issuing and acquiring costs that should be taken into account in determining interchange fees; especially given the goal of minimising transaction costs above.

Most critically, this means that we have to be careful to distinguish those items that may be part of the accounting costs of an issuer or acquirer but that are in reality a transfer to their respective consumers. An example of this is the cost of rebates paid by acquirers to certain merchants that are related to their adoption of services related to credit card transactions (e.g., other banking services or debit card processing). Another example is the cost of loyalty programs that represent in effect a discount or payment to customers for credit card usage. Such transfers are really price discounts rather than costs associated with payment functionality and hence, like cardholder fees and merchant services charges should not be used in determining interchange fees. This implies our third principle:

> **Principle 3:** *Issuer and acquirer costs that represent transfers to customers and merchants should not be included in the determination of interchange fees.*

While elements such as loyalty programs represent real costs – in that they are paid for by issuers – in light of the nature of competition between issuers they are a discretionary item (i.e., an issuer could offer reduced cardholder fees or a direct rebate instead) and part of the net revenues received by issuers arising out of the competitive process in issuing. To do otherwise would mean regulation of the competitive process itself.

Issuers and acquirers have revenue streams independent of their interchange arrangements. Consequently, some cost components will be recovered as part of the competitive process amongst issuers, in particular, those not related to the on-going costs of transacting. When setting interchange fees to minimise the costs associated with credit card transactions relative to other payment instruments, the interchange fee should, in principle, reflect those costs at the margin. However, establishing that margin is difficult in practice. This is because interchange fees typically relate to the transaction or value of transaction while customer and merchant decisions to use credit cards are only partly transaction based. If one uses a credit card, one does so because it is convenient at the time but also because one has made a prior decision to carry a card. Therefore, the full economic decision of

whether to hold a card involves not only marginal transactions but all transactions a customer is considering making.

Nonetheless, the idea that cost components should in some sense be marginal does exclude certain issuer and acquirer costs. Most importantly, when issuers and acquirers first set up their businesses there are many one-off investment costs that, once incurred, are never recovered. These sunk costs do not impact on the issuer and acquirer's on-going decisions regarding pricing and marketing their products and hence do not play a role in customer and merchant card adoption decisions. Consequently, accounting for or ensuring a return on capital for sunk investment expenditures associated with issuer and acquirer set-up should not be included in interchange fee calculations.

> **Principle 4:** *Interchange fees should only be based on those cost components of issuers and acquirers that would be avoided (over the long-term) if credit card services were no longer offered.*

This suggests that, as a practical matter, utilising an avoidable cost methodology is likely to be an appropriate means of ensuring that only costs related to credit card transactions at the margin (appropriately construed) are included.

An objection is often raised with respect to cost-based price regulation that it excludes an allowance for an appropriate return on sunk investment expenditures.[20] If regulated prices do not take into account investment expenditures, this sends a signal to future investors that they will not earn a return or recover those costs and hence, reduces investment in the industry.

This argument, however, while extremely important and legitimate in situations where the price of the service offered is completely regulated, is not necessarily a concern where regulated firms have alternative revenue sources. If interchange fees for issuers are reduced because their sunk investment costs are not taken into account, issuers will be able to recover those costs from other revenue sources. This is precisely because all issuers receive a common interchange fee and so changes in that fee would, for the most part, not impact on their profits in competition with one another.

[20] Indeed, we have been quite vocal in raising this objection in some regulatory environments. See Gans and Williams (1998), Gans (2000) and Gans (2001).

In effect, issuers and acquirers will earn their sunk entry costs in the marketplace rather than through the regulatory means. This is an appropriate way of ensuring that the goal of minimised costs of transacting is realised (by not distorting the interchange fee with very long-term expenditure items). As in all markets, firms will enter or exit issuing and acquiring only if they can earn a return on sunk expenditures and hence, there is no further need for regulatory underwriting of such returns.[21]

3.2.3 Allow for the Strategic Placement of the Payment Instrument

Interchange fees in credit card associations can be used as a means to encourage the development of the system in a particular way. For example, in order to encourage supermarket chains to accept credit cards, card associations in the United States offered special interchange fees for that purpose. Similar endeavours have been seen with regard to travel requirements as well as electronic transactions. In this respect, freedom to set the interchange fees to encourage transactions of a particular type or target specific customers and merchants may be an important instrument of competition between card associations.

In this respect, care must be taken in setting an interchange fee methodology to allow for changes in the strategic direction of the card association; subject of course to the earlier principles set out above.

> **Principle 5:** *Adjustments to interchange fees based on changes in customer or merchant mix or transaction type should be permissible and accommodated within the fee setting methodology.*

In this respect, a particular card association can respond to competitive and market forces and use its interchange arrangements as a means of signalling to issuers and acquirers the benefits of such change. For instance, if there is a situation where merchants are particularly resistant to adopting card services, then a reduction in interchange fee arrangements for that class of transactions, or perhaps overall, should be permitted. Alternatively, if there was customer resistance to using a form of payment – say for Internet transactions – then an appropriate

[21] We have argued that a similar logic should be applied in other network industries including telecommunications (Gans and King, 2000). It is important, however, to ensure that in other dimensions the price concerned is regulated appropriately so as to not critically diminish overall value created by the network; otherwise, concerns about regulatory takings could re-emerge.

increase in the interchange fee could be allowed to signal to issuers that they encourage customers to make such transactions.

Therefore, the interchange fee, while based ultimately on certain issuer and acquirer costs, should have scope to adjust to market circumstances. To do otherwise would be to remove an important dimension of competition between alternative payment instruments; especially where there is common membership between associations. Care, however, must be taken to ensure that such adjustments do not explicitly or implicitly violate Principle 1 above.

3.3 The Shared Avoidable Cost Approach

We now present a possible methodology that could provide the basis for calculating interchange fees. It is based on a formal model of payment systems (Gans and King, 2001c) that builds in assumptions of retail market power and an absence of cash and credit card price differentials so as to allow an optimal interchange fee to exist. We then solve for the interchange fee that minimises the costs of transacting and hence, satisfies the principles laid out above.

3.3.1 The Formula

We begin by stating our proposed interchange fee methodology. Having done this we will then discuss the rationale behind it. Our efficiency criterion is as outlined above: for a particular transaction to efficiently take place using a credit card rather than an alternative payment instrument (e.g., cash, cheque, charge card, debit card, promissory note or in-kind) the total net benefits of customers and merchants from the transaction must exceed the total costs associated with that transaction.[22] If this condition is not satisfied, another payment instrument would be preferred.

We propose to base the interchange fee on the basis of shared avoidable cost. Suppose we estimate that the average avoidable costs of issuing and acquiring are c_I and c_A respectively. Then we propose that the interchange fee, a, to be paid from acquirers to issuers be:

[22] This is the same criterion as specified by Baxter (1983) and Rochet and Tirole (2000).

$$a = \tfrac{1}{2}(c_I - c_A)$$

This means that the avoidable cost faced by an issuer (that is, $c_I - a$) is equal to the avoidable cost faced by an acquirer (that is, $c_A + a$). This means that when any issuer or acquirer is setting its relevant price, it is basing its choice on the same avoidable cost.

3.3.2 What do we mean by avoidable cost?

Strictly speaking, our proposed formula would be based on marginal cost. However, as mentioned earlier, in the context of credit card transactions, there are a series of margins that are potentially relevant including those at the level of transaction, cardholder or merchant adoptee. Even if, as a practical matter, we moved used average variable cost not marginal cost this issue would remain.

Practicalities require the use of the avoidable cost concept in relation to the calculation of the optimal interchange fee (something that has already been recognised in this context). The avoidable cost of issuing is the total amount that would be avoided, as of today, if no bank were to form an issuing function. Similarly, the avoidable cost of acquiring is the total amount that would be saved if no transactions were acquired. Notice that this notion of avoidable cost encompasses all on-going expenses with regard to each function but includes no attribution for previous sunk investments (that do not require renewal), thereby satisfying Principle 4. Strictly, it would also include some allocation of common costs, but as we will argue below, as a practical matter, under our particular methodology, not all of these would need to be considered.

As our formula is based on avoidable costs, loyalty points and other transfers are omitted (satisfying Principle 3). Also, avoidable costs relate to payment functionality. Consequently, costs solely attributable to extended credit lines – such as some types of credit losses - would be omitted (Principle 2). However, costs associated with the provision of short-term credit (say up to one quarter) would be included. These represent costs associated with providing liquidity in payment functionality – something that is integral to making credit cards valuable to customers and merchants.

Finally, the costs associated with a credit card payment system include shared costs between issuing and acquiring (when these are undertaken by the same bank) and shared costs between bank card

services operations and other banking functions. For each some of these costs will be related to the volume of issuer and acquirer activities and should be included in cost calculations. However, in some cases, a particular cost may be apportioned equally to the issuing and acquiring functions of banks. In this situation, that cost component need not be collected and included in the interchange fee as it would cancel out in the above formula.

3.3.3 Taking into Account Card Association Strategic Positioning

The above formula does, however, fail Principle 5 in that it offers little discretion for card associations to strategically position themselves in the market for payment instruments. Under our formula, interchange fees could only be varied if underlying costs associated with transactions or a class of transactions were to differ. It could not be responsive to demand conditions.

The reason for that is that it is based on an implicit assumption that, at the margin, customer and merchant benefits from using credit cards over other payment instruments are the same. In reality, however, the exact apportionment of total customer and merchant benefits would be the result of customer-merchant interactions. For example, if merchants were competitive, then all of the benefits would accrue to customers. On the other hand, if merchants have substantial marketing or bargaining power, then a large proportion of the benefits would accrue to merchants.

Suppose then, that the total benefit to the use of cards (summed over merchants and customers) is B. Suppose also that the portion of these benefits accruing to merchants is α and the portion to customers is $1-\alpha$. In this situation, we can readily derive the optimal interchange fee as:[23]

$$a = \alpha(c_I + c_A) - c_A$$

Note that if $\alpha = ½$, then we obtain the interchange fee formula as stated earlier as a special case of this one. If, on the one hand, customer

[23] Note that, at the margin (if acquirers use multi-part tariffs), αB will equal $c_A + a$. Moreover, if the costs of transacting are minimised, then the total marginal benefit, B, will equal the total cost per transaction $c_A + c_I$. Substituting this into the first condition, gives the stated formula.

bargaining power were very low (close to 1), then the interchange fee would reflect issuer costs only (as in the vertical chain view with its implicit assumption that merchants are the sole beneficiaries of the card services). However, if merchant competition were intense (close to 0), then the interchange fee would be negative and reflect only issuer cost. This would correspond to a situation where merchant resistance to offering card services was very high and the association's efforts would have to be directed towards encouraging them to process card transactions.

This generalised formula provides a means by which associations could differentiate themselves in a competitive sense. Associations may choose to target customer or merchant segments that are particularly resistant to the adoption of card services and hence, calculate their interchange fees accordingly. In addition, associations could gather evidence demonstrating whether customers or merchants were the greater beneficiaries from the use of cards. However, care would have to be taken in order to take into account the fact that the sharing of these benefits between customers and merchants largely takes place through variation in retail prices. To our knowledge, no study along these lines has ever been undertaken.

3.3.4 Theoretical Benefits of Our Methodology

Our methodology arises from a formal consideration of the economics of interchange in credit card associations.[24] Here, we summarise the features of our methodology that make it appealing from an economic point of view.

In economics, there is a strong presumption that efficient prices will reflect marginal costs. However, for open-loop payment systems, the joint nature of both consumption and production of card services makes identification of marginal costs a difficult matter. When a firm decides on its pricing (and respectively volume of sales) it needs to evaluate the additional revenue generated by an expansion in sales against the additional costs incurred. In credit card associations, interchange fees play an important role in the individual costs of issuers and acquirers, even though, for the association as a whole, they are really just a transfer payment. Nonetheless, when an issuer or

[24] That analysis is contained in Gans and King (2001c) and shares some similarity to Schmalansee (2001) although his model is more specialised.

acquirer attracts more card transactions this imposes both issuing and acquiring costs on the system. What our proposal does is equalise the cost impact of issuing and acquiring pricing decisions with the system-wide cost imposed.[25]

Although we have formally proved the desirability of our methodology, the idea of basing what is essentially a usage price (interchange fees) on avoidable costs has intuitive appeal to economists and reflects best practice and efficient pricing principles in other regulatory situations (e.g., telecommunications). As mentioned above, what it does not do is build in an explicit guarantee for issuers or acquirers of a rate of return on their investments. But then again it does not control the actual revenues that could be earned by issuers and acquirers. These are constrained by market and competitive conditions as in all deregulated markets and decisions regarding entry and exit will ensure that, in equilibrium, issuers and acquirers are earning at least a market rate of return on their investments. Hence, as the regulation of interchange fees does not regulate revenues of issuers and acquirers we do not require a provision to take into account fixed and sunk costs as may arise in more highly regulated industries.

3.3.5 Practical Benefits of Our Methodology

In practical terms our proposal has many advantages:

- It does not require measurement and apportioning of issuer and acquirer revenues (as in the Frontier Economics and some other currently practised methodologies). This removes a significant issue of gaming that concerned the ACCC.
- Loyalty points are no longer part of the basis of interchange calculations and are not restricted by it.
- Issuers and acquirers have incentives to achieve reductions in their own avoidable costs because interchange is based on industry averages only.
- Different interchange fees for different types of transactions (electronic, card not present or supermarket) arise naturally as each involves different marginal costs. Moreover, flexibility in

[25] This idea was first discussed by Schmalensee (2001).

interchange fee structuring will lead to different fee structures among associations even though on average they are likely to be the same.

- Interchange fees can potentially be a two-part tariff with one part related to costs based on transaction value and another on transaction numbers (as in the US). Each type of avoided transaction costs can be isolated and used to form the basis of such a tariff.

- Basing the interchange fee on and equalising avoidable costs, in our opinion, satisfies the ACCC concern that "the interchange component of the merchant service fee is related to the interchange based services which merchants consume."

Our proposal is admittedly, new and untested. However, given the relative simplicity of its make-up; we see this as a minor disadvantage.

Of all the proposals thus far, it is clearly focussed primarily on costs (being based only on avoidable costs associated with card transactions), allows issuers to recover some costs from merchants, allows for differential interchange fees (as there is no common cost attribution issue), and can be readily applied.

3.4 Governance Issues in Interchange Fee Revisions

Alongside the principle of collective action in setting interchange fees (outlined in Section 2 above) comes important issues as to the on-going governance of that process: that is, the means by which an association revises its interchange fees. At present, revisions to the interchange fees can, in principle, occur at any time but, as we understand it, would have to be triggered by some change in circumstances and a revision process put in place by members of an association.

In contrast, recent proposals have suggested changing this process to one very similar to that which would underlie changes in regulated prices in public utility industries. For example, Frontier Economics proposed that all three card associations in Australia (MasterCard, Visa and Bankcard) commit to a five yearly revision of interchange fees based on cost and other performance information

(both forward looking and historical) of association members at the time. In effect, the formula underlying interchange fees would be fixed as part of the authorisation of the collective setting of those fees.

The problem with this type of proposal for a formulaic approach to revisions of the interchange fee is that they suffer from all of the problems associated with regulation of prices in utilities industries. The 21st February letter from the ACCC to Review Banks highlighted these problems. That letter expressed concerns that:

- incentives on issuers would be weak because of the inclusion of revenue in interchange fee revisions;
- costs were not sufficiently forward looking;
- the basic principles for the inclusion of certain types of costs;
- cost allocation issues between network and non-network activities;
- the fact that such a revision formula has not and is not utilised in payments systems elsewhere.

While not attempting to address each of these criticisms here, we note that such criticisms are commonly levelled at regulators (including the ACCC) in public utility contexts and in many cases have a sound basis in economic and regulatory experience. There are ways of improving regulatory formulae to overcome such criticisms but, in general, no perfect formula exists that eliminates concerns regarding potential 'gaming' of regulated fees by market participants through either weakened incentives and the potential to distort reported information.[26]

To this broad set of criticisms, we should add another. Given the commonality of members among different associations, competition among them depends critically on the different choices they might make regarding their internal arrangements and how these react to market conditions. For instance, MasterCard has chosen to negotiate a set of arrangements for taxi companies that has allowed their cards to be adopted by those merchants whereas Visa has not. However, if a formulaic approach were committed to, it would be common among all

[26] See, Laffont and Tirole (1999) for a clear discussion of these issues.

associations and would remove a means by which different card associations distinguish their product and provide competitive reactions to changes in market circumstance. Hence, it would remove an aspect of competition in payments systems.

In conclusion, in our opinion, there are potential public detriments associated with the adoption of formulaic commitments to the revision of interchange fees as currently proposed. These are the potential for the emergence of 'regulatory costs' as part of the review process including weakened incentives and potential for gaming revisions of the interchange fee; and also, the removal of an instrument for competition between card associations. This is not a criticism laid specifically at the Frontier Economics proposal but more generally at *the entire concept of a formulaic approach to the revision of the interchange fee*. For these reasons, an authorisation proposal that includes a formulaic approach has a reduced likelihood of satisfying the net public benefits test; despite the benefits of collective action.[27]

3.4.1 The Alternative: Decoupling the Current Arrangements from the Revision Process

The alternative to a formulaic approach to revising the interchange fee would be a system of governance that allowed interchange arrangements to be revised only if there were a 'material change of circumstance.' The conditions by which there would be a material change of circumstance could be clearly spelled out in the association rules (including perhaps significant changes in issuer or acquirer costs, the emergence of new technologies, significant changes in the nature of transactions, changes in the strategies of other payment options, etc.). Then a revision process would occur by which some third party (perhaps an accounting firm) would be engaged to review the interchange fee according to certain pricing principles. For example, those principles might mirror those set out earlier in this section, including the types and nature of costs that would be taken into account, the relationship between those costs and merchant adoption decisions, forecasts of technological changes, and adherence

[27] To be sure, it is important to contrast a formulaic approach with the notion of an interchange fee formula – like the one we derived earlier. A formula can be used to base or benchmark a fee revision without having a formulaic approach that automates the revision process itself. We are arguing here that good regulatory practice as well as an acknowledgement of the competitive environment facing an association necessities a flexible revision structure.

to appropriate standards of accounting and transparency. Furthermore, those principles may include an equation relating these variables that allowed for some discretionary adjustments (such as our generalised interchange fee formula above).

In effect, such an approach is a more transparent, augmented version of the current system. It retains a chief benefit of having collective setting of the interchange fees – that such fees are set and are unlikely to change – allowing issuers and acquirers to compete effectively *without* an eye to future revisions of interchange fees. Moreover, in principle, such an approach is subject to regular reviews that would allow a response to changing technological and market circumstances and the potential for diverse strategies by card associations.

Of course, a criticism that could be levelled at this proposal is that it is very close to approving the current system and does not take into account concerns that such material changes in circumstances may have occurred (in say technology) that require a revision of the current level of the interchange fee. Moreover, it is perhaps unclear what the rationale is behind the current level of the interchange fee and whether it is consistent with the revision principles that would be outlined as part of the authorisation.

To answer this criticism, each association could undertake that revision process immediately – as if a material change in circumstances had occurred. An independent firm could be engaged by each association to review the interchange fee based on the pricing principles set out in the rules of the association and a new interchange fee would be set based on that process. This would answer the dual criticisms that current fees do not reflect current circumstances and that their determination is not transparent. This process would serve as evidence to the RBA that the amended revision process is both workable and results in outcomes that yield public benefits and do not give rise to anti-competitive effects.

3.4.2 Conclusion

Current proposals to 'set in stone' the revision methodology of credit card associations make it difficult to establish that the collective setting of interchange fees by those associations result in net public benefits. This is because formulaic approaches to interchange fee revisions carry all of the costs of regulatory systems and tie the hands

of associations in competition with one another and in response to changes in the nature of payment instruments.

In contrast, a suitably augmented process of revising the interchange fee based on 'material changes in circumstance' preserves the benefits of collective action (that is, certainty and clarity for members) while allowing revisions of the interchange fee to take place within associations that are transparent and less subject to both 'gaming' and the weakened incentives that would come with a formulaic approach.

Finally, to answer criticisms levelled at the current interchange fee, this review process could be put in place immediately and its outcome held up as evidence as to the workability and publicly beneficial effect of the augmented process.

4 No Surcharge Rule

The no surcharge rule prevents merchants from placing an explicit 'surcharge' on credit card transactions. It is a rule imposed by credit card associations (and indeed also by closed loop systems in many cases), so as a starting presumption, one expects that the rule serves a purpose in improving the operation or profitability of such associations. This raises an issue as to whether it is socially desirable in a context where card associations have some degree of market power.

Here we analyse the operation of the no surcharge rule and consider reasons why it may be of private and social benefit. In so doing, however, it is useful to consider whether association restrictions on surcharging are material, the role of association market power and the role of merchant market power.

4.1 Is a no surcharge rule material?

The no surcharge rule is a legal restriction imposed by acquirers (at the behest of associations) on merchants preventing surcharging on credit card transactions. However, this does not prevent discounts for cash.[28] Moreover, even when the rule applies, there is anecdotal evidence of deviations and negotiations over who bears MSCs – the customer or the merchant. In some cases, associations (e.g., MasterCard and American Express) have expressly allowed surcharging.[29]

The fundamental issue in determining if the no surcharge rule is material is whether the rule itself has led to a situation where retailers do not charge differential prices for credit card and cash transactions and for different card schemes. In countries where the no surcharge rule has been removed, retailers have continued to charge uniform

[28] For example, the University of Melbourne book room has an explicit 5 percent discount for cash transactions over both EFTPOS and credit card transactions.

[29] In taxis this has occurred with a 10 percent premium (plus GST) being placed on credit card transactions.

prices regardless of payment instrument used.[30] This experience suggests that the no surcharge rule may not be a material constraint on retailer behaviour. Retailers tend to charge one-price whether or not the no surcharge rule is in operation.

There are two reasons why the no surcharge rule may be immaterial. First, as Frankel (1998) argues, there are transactions cost reasons that lead to 'price coherence' among different payment instruments. 'Price coherence' does not necessarily mean that cash and credit card transactions occur at the same price; this is a special case. With price coherence, if there is a change in the costs of one payment instrument (say an increase in the MSC for credit card transactions), both prices will change leaving their relative price unchanged.

> Simple examples of price coherence might include checkout line candy or cigarette racks, or vending machines that could post efficiently only a single price point for any item sold. Sellers might be able to mitigate the effects of price coherence in some cases if they have tools other than price (e.g., shelf space allocations and marketing effort) to influence customers' purchase decisions. (Frankel, 1998, p.316)

Frankel then explains the importance of price coherence:

> Price coherence constrains merchant choices. If the price of one brand increases, the merchant can drop the now more costly brand altogether, charge a different price for the brand than for competing products, or raise its prices for all products by the same amount. If the merchant chooses the latter course, the new price to consumers for any product will be based on a weighted average of the combined cost associated with all products. In that case, consumers will have no incremental incentive to choose the lower cost product. In other words, retail price coherence reduces the elasticity of demand facing the retailer's suppliers because any given wholesale price increase results in a smaller reduction in unit sales than would occur without price coherence. (p.316)

[30] Below we consider reasons why prices may not actually change, however, it is also possible that when the no surcharge rule was removed, that associations and banks altered their pricing structure to accommodate it. This would certainly be possible if indeed, interchange fees were otherwise neutral as we contend in Gans and King (2001c). In this case, observations of no price change in response to the removal of no surcharge restrictions may not indicate the rule was immaterial at all but that, in fact, there was enough flexibility elsewhere to accommodate the change without altering retail pricing structures.

Frankel's argument suggests that, for many transactions, the removal of the no surcharge rule is unlikely to result in a material change.

The second reason why the removal of the no surcharge rule may be immaterial is related to Frankel's price coherence argument but relies less on transactions costs than on the nature of competition at the merchant level. If the majority of merchants have mostly cash customers or mostly credit card customers, a relaxation of the no surcharge rule is unlikely to have a sizeable effect. Moreover, when competition at the merchant level is strong, merchants will tend to be differentiated in terms of the payment instruments on offer (or used) and hence, the no surcharge rule is of little importance.[31]

When thinking about the economic impact of the no surcharge rule, it is important to consider the role that the rule plays in overcoming merchant resistance to the adoption of card services. On the one hand, merchants will be concerned about the competitive impact on sales from cash customers if they adopt card services under a no surcharge rule. If the no surcharge rule is eliminated, merchants can pass through credit card charges to card customers. On the other hand, under a no surcharge rule, if sales to cash customers are relatively low, while many customers carry and benefit from card transactions, then adoption becomes easier when some of the costs can be passed on to both cash and card customers. In this case, removing the no surcharge rule will increase merchant resistance. Moreover, it is exactly this situation in which there is likely to be either alternative tendencies towards 'price coherence' (why offer cash discounts if cash customers are relatively unimportant?) and retail market power. Hence, it is this situation, that the material impact of the no surcharge rule is possibly at its lowest.

In summary, there is reason to believe that the removal of the no surcharge rule, even if it leads to price differentials, will not necessarily produce a full decoupling of the costs of payment instruments to consumers. However, it is likely to have a larger impact in cases where merchant resistance to processing card transactions is already high (e.g., purchase of large items such as furniture or cars).

In analysing the no surcharge rule and its role, we next consider the relevant private benefits of the rule. Why might an association

[31] See Gans and King (2001c).

want to adopt a no surcharge rule? We then turn to consider whether these private benefits may omit important social costs. In so doing, we will make an assumption that the no surcharge rule is material and that merchants will de-couple cash and credit card prices if it is removed.

4.2 Private benefits of the no surcharge rule

As noted earlier, the no surcharge rule is likely to create some merchant resistance to the adoption of card services, even where the direct benefits the merchants receive from accepting the credit card exceed any charges they must pay. This is because merchants have to raise prices to their cash customers. In this respect, removing the no surcharge rule will immediately impact on merchant resistance and allow for the increased diffusion of credit card services across retail sectors.

However, the effect of this would also be to cause a price differential between cash and credit card customers (based on our assumption that there is no other constraint to differential pricing). In competitive retail segments, this means that credit card customers will bear the MSC. However, in those competitive segments, it is likely that such perfect pass through of card costs to customers would occur under a no surcharge rule through the separation between cash-only and credit card-only merchants. This means that a removal of the no surcharge rule has no real effects on competitive sectors.

Therefore, it is in segments where merchants have market power that there is likely to be some real effects from the removal of the no surcharge rule. For a given interchange fee, prices will rise for card transactors and fall for cash users.[32] Thus, there will be no incentive for the association to change the interchange fee.[33]

In retail segments characterised by retailer market power, the removal of the no surcharge rule is likely to result in a reduction in the diffusion of payment cards. When there is merchant market power, a removal of the no surcharge rule will likely lead to a price differential between cash and card transactions of a degree higher than the MSC.

[32] Rochet and Tirole (2000, Proposition 4).

[33] Gans and King (2001c).

This is because merchants can use a customer's desire to use a credit card as a means of price discriminating among different customer types. It is likely that cardholders are customers with greater wealth than non-cardholders and hence, have less elastic demand at the retail level (Gans and King, 2001d). Moreover, merchants with market power will set their card price based on the average benefit to cardholders rather than the marginal benefit (Rochet and Tirole, 2000, p.19). In addition, the decision to use a card at the point of sale is a decision taken by the customer, following an earlier decision by that customer to hold a card. In this situation, individual merchants will not internalise the benefit from having more cardholders and will set the card price at a level that extracts the cardholder's benefit at the time of sale (Wright, 2000). This will make customers more resistant to incurring the costs associated with holding cards.

For these reasons some cardholders will no longer find it attractive to hold cards so that there will be a reduction in the use of credit card services. Therefore, when there is imperfect retail-level merchant competition, the removal of the no surcharge rule has the unambiguous effect of reducing overall credit card usage as a payment instrument.[34] Essentially, the private benefit of the no surcharge rule is to curb the ability of merchants with market power to use credit cards as a means of price discriminating among their customers and therefore to increase the attractiveness of credit cards to potential cardholders.

4.3 The social costs of the no surcharge rule

The prevention of merchant level price discrimination has been identified as the reason why associations might find it desirable to have a no surcharge rule. Price discrimination is not necessarily socially undesirable, so it is difficult to use this factor as a means of identifying the social benefits of a no surcharge rule. Ultimately, the determination of the social costs and benefits is an empirical matter. As Wright (2000) notes, it depends on the extent of the 'hold-up' problem that emerges when cardholder adoption and usage decisions are decoupled. This said, where removal of the no surcharge rule leads to

[34] This result emerges in the model of Rochet and Tirole (2000), Wright (2000) and Gans and King (2001d).

customer hold-up, this is unlikely to reflect socially desirable price discrimination. In such circumstances the removal of the no surcharge rule is quite likely to result in additional social costs.

If there is a social cost from the no surcharge rule it comes from the 'cross subsidisation' by cash customers of card customers. Cash customers will reduce their purchases under a no surcharge rule and this will result in social losses. As noted above, this is due to both an absence of sufficient competition at the retail level as well as the no surcharge rule.

Rochet and Tirole (2000) also argue that the removal of the no surcharge rule is likely to raise the overall cost of transacting in the economy when there is a lack of competition in either (or both) the issuing and acquiring segments. In this situation, even with the no surcharge rule, there is an underprovision of card services. Removing the no surcharge rule will reduce the provision of credit card services further and hence, increase the cost of transacting.

It is only in situations where issuing and acquiring are very competitive (or there is close competition among different card systems) that the no surcharge rule may lead to an outcome where too many credit card transactions occur from a social perspective. In this situation, the removal of the no surcharge rule might lead to a desirable reduction in card transactions; although Rochet and Tirole (2000) cannot say this with certainty, as the magnitude of the reduction may be too large if there is merchant market power.

4.4 Conclusion

The discussion above suggests that the removal of the no surcharge rule, assuming that it results in material price changes, *is only likely to improve social welfare (in terms of reducing the costs of transacting in the economy) if issuing and acquiring segments are highly competitive. If it is believed that there is a lack of competition in either issuing or acquiring,*[35] *then the no surcharge rule serves a social function of curbing the adverse consequences of that lack of competition and increases credit card usage in a socially desirable manner.* In that case, a removal of the no surcharge rule may be detrimental to overall economic efficiency.

[35] Something that the Joint Study (2000) concluded there was.

5 Competing Payment Systems

The role of competing payment systems is important in analysing the potential impact of changes in interchange fee arrangements or the no surcharge rule. Arguments are made regarding the impact upon and the competitive response from proprietary card systems (such as American Express and Diners Club). In this section, we evaluate these arguments by considering first, the differences between proprietary and open systems in terms of their competitive behaviour and then examine the implications of one-off changes to card association interchange fees. As it is difficult to see how changes in the no surcharge rule would operate legally in this context, we do not examine this aspect here.

5.1 Proprietary Card Systems

Proprietary card systems differ from card associations in that they set MSCs and cardholder fees together rather than relying on the competitive interactions between issuers and acquirers. In effect, proprietary card systems operate as if they are a monopoly, integrated acquirer and issuer. For proprietary systems, any interchange fee is merely an internal transfer price and would not be expected to be of consequence for card service pricing behaviour.

A proprietary card system faces a choice as to how to price its mix of issuing and acquiring services. By setting a higher MSC, it will discourage merchant adoption of card services and hence, may choose to offset this by reducing cardholder fees and charges. If not, it would face a reduction in overall usage of its card facilities; although this may be profitable if the proprietary card system had some degree of market power. The extent of reduction in cardholder fees becomes a choice rather than something determined by the competitive interaction between issuers as with card associations. This suggests that a proprietary system will be more likely to favour a higher MSC when formulating its pricing policy; compared with a card association (Rochet and Tirole, 2000).

Given this, the fact that open card associations have more membership than proprietary systems likely reflects a high level of competition in the issuing segment of card associations. This competition means that despite interchange arrangements that may favour a lower MSC for those associations, interchange fees will be passed through to customers in the form of lower cardholder fees or other benefits. In this respect, the adoption of card services by merchants and customers is likely to be greater than that for proprietary systems.[36]

5.2 Competitive Interactions

Two conflicting arguments are sometimes advanced with regard to the interaction between a reduction in open loop card associations and proprietary card systems. The first is that a reduction in interchange fee arrangements will raise association cardholder fees and lower MSCs and that this will force proprietary card systems to respond in kind so as to maintain their merchant numbers. The second is that currently the cardholder fees of proprietary card systems are kept low by competitive pressure from associations and that this pressure will be relieved if there is a reduction in the interchange fee. In our opinion, these arguments conflict because they are incomplete. A better understanding of the nature of a card 'system' will allow the competitive implications to be more clearly defined.

To begin, it is important to emphasise once again that if there is no significant market power in retail sectors or the no surcharge rule (and more generally price coherence) does not apply, then interchange fees are neutral with regard to the real operation of card associations. In this situation, a change in the interchange fee of card associations will have no impact upon proprietary cards at all; even though it will change the mix of card association prices.

Suppose then that there is retailer market power and an effective no surcharge rule (or equivalent). In this case, changes in interchange fees change the real behaviour of association members and

[36] Another hypothesis is that the higher market share of credit card associations reflects the presence of network effects at a system level. However, this explanation does not accord with the fact that the card associations with the greatest share of transactions were late-comers to the Australian (and indeed international) market.

participants. The question is: how will this behaviour change? To answer this requires us to consider how the interchange fee came to be at its current level. We consider two alternative scenarios in which the fee is chosen to maximise the profits of association members or it is set at a level such that a reduction in the interchange fee will increase card association transactions.

A card association choosing interchange fees to maximise the profits of its members will seek to maximise the total volume of credit card transactions subject to limits on the desirability of merchants adopting card services and cost conditions.[37] Thus, if current interchange fees are set according to profit maximising principles, then a reduction in those fees will reduce the total volume of transactions being processed by card associations.

This reduction will reduce the competitive pressure on proprietary systems. This will increase the profits of proprietary card systems, although whether this will increase or reduce the volume of transactions processed by proprietary systems is unclear. However, the total volume of card transactions (both proprietary and association) will be reduced and the total cost borne by consumers (both cardholders and merchants) will be increased. This will only be socially desirable if policy-makers believed that the current level of credit and charge card transactions was too *high* relative to other payment instruments.

On the other hand, suppose that card associations had not set current interchange fees to maximise profits – say due to inertia. In this case, it is possible that a reduction in interchange fees would increase the volume of credit card transactions.[38] In this case, there would be a negative impact on the profits of proprietary systems and an increase in the overall volume of credit and charge card transactions. However, it should be noted that the total profits accruing to card association members might increase (at least in the short-run); although it may also encourage the entry of issuers and acquirers that compete some of these additional profits away. In this case, the change in the interchange fee will only be socially desirable if policy-makers believed

[37] See Rochet and Tirole (2000), Schmalensee (2001) and Gans and King (2001b). These results all rely upon the assumption that issuing and acquiring marginal costs are constant.

[38] It is also possible that such a reduction could lower the volume of credit card transactions even further (Gans and King, 2001b).

that the current level of credit and charge card transactions was too *low* relative to other payment instruments.

5.3 Conclusion

The impact of changes in interchange fees on proprietary card systems depends on the goal of the policy-maker. If the goal is to increase the volume of credit card transactions, then the impact on proprietary card system profits will be negative and their customers will benefit from the increase in competitive pressure. However, if the goal is to reduce the total volume of credit card transactions, then the impact on proprietary systems will be positive. In this respect, the interests of card associations and proprietary systems in terms of their view regarding regulation of interchange fees are conflicting.

6 Networks and Network Effects

Open (four party) payment systems are networks in the sense that they operate efficiently if each party to a transaction faces a price structure that links each party's respective decisions. Specifically, one could imagine a situation where issuers dealt with customers and acquirers dealt with merchants but there were no interchange arrangements or retail price variation to distinguish between credit card and cash transactions. In this situation, a payment system is unlikely to operate efficiently because issuer/customer and acquirer/merchant pricing decisions would be decoupled from one another. Externalities would be present and many efficient credit card transactions would not be taken.

As has been mentioned elsewhere, interchange arrangements and also retail price variation serve to allow networks to operate efficiently.[39] They provide a price link between all of the relevant decision-making parties to a transaction. Such pricing arrangements allow the network to function effectively. Essentially, they are a means by which the joint consumption and joint supply of payment system services are recognised and accounted for.

However, payment systems are also subject to what economists term *network effects* or *network externalities*. This arises because when one more merchant adopts card facilities or one more customer carries a credit card, this increases the value to all current (and future) customers and merchants of their own adoption decisions. Thus, the private value of an adoption decision is less than its social value. At its essence, this means that a payment system is likely to be of most social value when it is ubiquitous. Moreover, only payment systems that reach a critical mass of both customers and merchants will be sustainable in the long-term.

This means that credit card associations – at their early stages – will be concerned with growing to critical mass as quickly as possible. Doing so will mean that they will have to make an assessment as to

[39] Gans and King (2001a).

which party – customers or merchants – are most likely to resist adopting card services and structure their efforts accordingly. Moreover, they will be concerned about encouraging competition among issuers and acquirers to achieve this purpose. Hence, associations with more liberal access rules are more likely to survive.

The need to build on network effects helps us understand the existence of some of the historic rules in card associations in Australia. Take, for instance, the net issuer requirements that provide direct incentives for acquirers to become issuers. This was likely a response to concerns that while banks may have had strong incentives to become acquirers – as part of a range of services to merchants – their incentives to issue cards may have been more limited and risky; especially given an immature database on specific customer risks. This is reinforced by the fact that at the time card associations were established the potential entrants into issuing and acquiring remained the largest banks. Indeed, conventional wisdom continues to hold the issuing business as less competitive than acquiring. This is perhaps because of 'lock-in' effects on customers caused by customer habits and the like. Hence, there is an argument that some case for the net issuer rule may remain.

Network effects also help us understand when issuers in open loop associations are allowed to issue cards under multiple brands. The alternative would have been to prevent that. However, at the start-up phase of associations, this would have conferred a large advantage on the first-mover association – Bankcard. Hence, there was pressure, especially given the need for international linkages, to continue the US practice of duality in Australia. Given the existing concentration of the Australian banking sector, duality means that Australians (both merchants and consumers) can get easier access to the dominant international credit card systems; something that would be unlikely to be the case if multiple brand issuing were not possible.

In conclusion, it is important in evaluating rules and practices to distinguish between those that arise because payment systems are networks and necessitate pricing interaction between parties and those that arise because there are network effects associated with building and maintaining a critical mass of adoptions by customers and merchants. Different rules are based on the network versus the existence of network effects and it is important to clearly delineate between the two.

7 Loyalty Points and Nonlinear Pricing

Loyalty points represent a form of what economist's term as non-linear pricing. This means that customers' decisions to hold credit cards are based on the volume of transactions they expect to make as well as the intrinsic benefits of such transactions. A customer who holds a credit card is, therefore, more likely to actually use it with a given merchant because that customer is receiving an extra inducement from so doing.

Rochet and Tirole (2000) demonstrate that a move towards volume related discounts to customers (such as loyalty points) reduces merchant resistance to the adoption of card facilities. This is because, compared with a pricing structure without volume related discounts, cardholders will receive an even greater benefit from using cards than before. This means that merchants will find it more desirable to adopt credit cards and hence, MSCs can be correspondingly higher.

In this light, volume related pricing can improve the operation of card systems as cardholders face charges more closely tied to the costs they are imposing on the system with each transaction. Moreover, every issuer that offers a loyalty point scheme is increasing the ability of acquirers to offer terms that encourage merchant adoption of card services. Seen in this light, loyalty points are a positive development from a scheme's perspective.[40]

One might ask, however, whether a simple monetary inducement rather than, relatively illiquid points, would be a more effective means of non-linear pricing. This is, of course, an issue of controversy in economics. It is likely that loyalty points allow issuers to achieve more price discrimination among different user types. In effect, high usage customers are receiving a relatively greater benefit. In a network industry, this type of price discrimination assists the association in competition with other payment instruments. Moreover,

[40] Note, however, that loyalty points were first introduced by proprietary card schemes (American Express and Diners Club). In so doing, those schemes reduced merchant resistance to the adoption of their cards. Given this, introduction of loyalty points by issuers in open loop schemes was probably a competitive response to these developments. Moreover, this might also suggest a role for associations in setting interchange fees to encourage non-linear pricing by issuers.

because points are rarely tied to a particular association (i.e., they are used to purchase other goods), none of the usual concerns regarding switching costs apply.

8 References

Ausubel, L.M. (1991), "The Failure of Competition in the Credit Card Market," *American Economic Review*, 81 (1), pp.50-81.

Baxter, W.F. (1983), "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law and Economics*, 26, pp.541-588.

Carlton, D. and A.S. Frankel (1995), "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 68, pp.643-668.

Cruickshank, D. (2000), *Competition in U.K. Banking: A Report to the Chancellor of the Exchequer*, The Stationary Office: London.

Economides, N. and S.C. Salop (1992), "Competition and Integration Among Complements, and Network Market Structure," *Journal of Industrial Economics*, 40 (1), pp.105-123.

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

Frankel, A.S. (1998), "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66, pp.313-361.

Gans, J.S. (2000), "An Evaluation of the Draft Pricing Principles for Access to the Victorian Rail Network (Freight)," submission to the Office of the Regulator General (www.mbs.edu/home/jgans/papers/freight.pdf).

Gans, J.S. (2001), "Regulating Private Infrastructure Investment: Optimal Pricing for Access to Essential Facilities," *Journal of Regulatory Economics* (forthcoming).

Gans, J.S. and S.P. King (1999), "Termination Charges for Mobile Networks: Competitive Analysis and Regulatory Options," *Working Paper*, University of Melbourne.

Gans, J.S. and S.P. King (2000), "Mobile Network Competition, Customer Ignorance and Fixed-to-Mobile Call Prices" *Information Economics and Policy*, Vol.12, No.4, pp.301-328.

Gans, J.S. and S.P. King (2001a), "The Role of Interchange Fees in Credit Card Associations: Competitive Analysis and Regulatory Options," *Australian Business Law Review*, 29 (2), April, pp.94-122.

Gans, J.S. and S.P. King (2001b), "The Neutrality of Interchange Fees in Payment Systems," *unpublished paper*, Melbourne.

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

Gans, J.S. and S.P. King (2001d), "Using 'Bill and Keep' Interconnect Arrangements to Soften Network Competition," *Economics Letters*, 71, pp.413-420.

Gans, J.S. and R. Scheelings (1999), "Economic Issues Associated with Access to Electronic Payments Systems," *Australian Business Law Review*, 27 (5), pp.373-390.

Laffont, J-J. and J. Tirole (1999), *Competition in Telecommunications*, MIT Press: Cambridge (MA).

Merrill Lynch (2000), *Credit Cards: An Ace up the Sleeve*, Sydney.

RBA / ACCC (2000), *Debit and Credit Card Schemes in Australia: A Study of Interchange Fees and Access*, Sydney.

Rochet, J-C. and J. Tirole (2000), "Cooperation Among Competitors: The Economics of Payment Card Associations," *mimeo.*, Toulouse, April.

Schmalensee, R. (2001), "Payment Systems and Interchange Fees," *Working Paper*, NBER.

Small, J. and J. Wright (2000), "Decentralised Interchange Fees in Open Payment Networks: An Economic Analysis," *mimeo.*, Auckland and NECG.

Wright, J. (2000), "An Economic Analysis of a Card Payment Network," *mimeo.*, Auckland and NECG.