University of Auckland Working Paper No. 245

Competing Payment Schemes

Graeme Guthrie and Julian Wright*
Victoria University of Wellington and University of Auckland

February 14, 2003

Abstract

This paper presents a model of competing payment schemes. Unlike previous work on generic two-sided markets, the model allows for the fact that in a payment system one type of user (merchants) competes to attract users on the other side of the market (consumers who may use cards for purchases). It analyzes how competition between card associations affects the choice of interchange fees, and thus the structure of fees charged to cardholders and merchants. Implications of the analysis for the competitive neutrality, or otherwise, of proposals to regulate interchange fees are discussed.

1 Introduction

Recently several models of payment schemes have been developed in order to analyze the optimal structure of fees in payment schemes: How much is charged to cardholders versus merchants for card transactions? Policymakers in a number of jurisdictions have been concerned that merchants pay too much to accept credit card transactions, costs that in their view are ultimately covered by consumers who pay by other means (see Chakravorti and Shah, 2003 for a review of some of these concerns). In card associations, like MasterCard and Visa, the structure of cardholder and merchant fees is determined by the level of the interchange fee, a fee set collectively by the members of the associations. The interchange fee is paid by the merchant's bank to the cardholder's bank on each card transaction. In order to correct perceived distortions in the structure of fees, policymakers such as the Reserve Bank of Australia have moved to regulate lower interchange fees.

The policymakers' actions raise questions about the role of competition in determining the efficient structure of fees in payment card schemes. Do the alleged high interchange fees arise from a lack of competition between card schemes, or perhaps because of it? Will regulating lower interchange fees for

*We thank seminar participants at the University of Melbourne for helpful comments. Any errors are our own. Corresponding author: Julian Wright, Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand. jk.wright@auckland.ac.nz

bank card associations force unregulated proprietary schemes like American Express to also adjust their fee structure, or will it provide them with a competitive advantage? One needs a model of competing payment schemes to answer such questions.

This paper provides such a model, building on the recent literature on the economics of payment cards. Schmalensee (2002) presents a model of a card scheme and analyzes the optimal interchange fee as an instrument that allows card schemes to balance cardholder and merchant demand, taking into account that total card transactions depend on the product of the two types of partial demand. Rochet and Tirole (2002) develop a framework that takes into account that merchants themselves compete to attract customers, so that their demand for card acceptance depends on cardholders' benefits. As well as determining the welfare maximizing interchange fee, they also examine the welfare consequences of the no-surcharge rule, a rule used by card schemes to prevent merchants charging consumers more for purchases made with cards. In contrast to Schmalensee, their framework assumes merchants are all identical, and so does not consider the role of the interchange fee in balancing the demands of the two types of users. Wright (2001) provides a model of a card scheme which integrates the approaches of Schamalensee, and Rochet and Tirole. It allows merchants' transactional benefits of accepting cards to vary across industries, while allowing for merchant competition along the lines of Rochet and Tirole within each industry.[1]

All these models assume a single payment scheme that sets its own fee structure (through its interchange fee), with the alternative form of payment being an exogenously determined instrument such as cash. A natural question arises: How do fee structures change when there is competition between multiple card schemes, each of which can determine its own interchange fee? Does the competitive fee structure (or level of interchange fees) move closer or further away from the efficient level?

Rochet and Tirole (2001) have provided a generic model of competition between "two-sided" markets (or networks).[2] These markets have the property that there are two types of agents that wish to use a common platform, and the benefits of each side depend on how many users there are on the other side of the network. Like this paper, they are interested in the structure of fees between the two sides of the network (or, in the case of networks run by an association, the level of the interchange fee). Examples they give include Adobe Acrobat (Acrobat Reader and Writer), payment cards (cardholders and merchants), platforms (hardware/console and software providers), real estate (home buyers and sellers), shopping malls (shoppers and retailers), and Yellow Pages (readers and advertisers). In order to obtain a general yet tractable framework, they treat users on both sides as end users, abstracting from the fact that an important feature of some of these markets is that *one* side competes amongst itself to sell to the other side (e.g. merchants, software providers, retailers and advertisers).

[1] Several other papers analyze card schemes using formal models. Baxter (1983) provides an early treatment of interchange fees. Gans and King (2002) provide general conditions under which interchange fees will be neutral. Schwartz and Vincent (2002) analyze the no-surcharge rule, arguing it can be used to 'tax' cash customers. In contrast, Wright (2003) uses a similar model to Rochet and Tirole to consider the extremes in which merchants are monopolists, or merchants are perfectly competitive, showing these cases put important constraints on the ability of a card scheme to use interchange fees and the no-surcharge rule in harmful ways.

[2] See also Armstrong (2002), Caillaud and Jullien (2001), Parker and Van Alstyne (2000), and Schiff (2003) for other models of two-sided markets.

This paper provides a model of two competing payment schemes that takes into account the fact that merchants compete between themselves to attract customers, who may pay using cards. It also takes into account the observation that consumers typically face the same retail price regardless of the instrument they use for payment. These features are fundamental to any analysis of how competition affects the fee structure in payment schemes. To obtain sharp results, we focus on the benchmark case in which consumers view the two types of cards as providing identical transactional benefits for making purchases, and competing merchants view the two types of cards as providing identical transactional benefits for receiving payment. Essentially, ours is a model of perfect competition between payment systems.

A key determinant of the competitive fee structure (and interchange fee) is the extent to which consumers hold one or two cards when making purchases. When consumers hold only a single card there are two effects which determine the competitive fee structure. Taking the number of cardholders on each scheme as given, competing card schemes act exactly as though they were a single card scheme, attracting card usage and merchant acceptance for their base of cardholders. In addition, they will also seek to attract a greater base of cardholders in the first place, by charging less to card users and more to merchants (in the case of competing card associations, by setting higher interchange fees). Consistent with Rochet and Tirole (2002, p. 561), when merchants are homogenous there will be no scope to further raise merchant fees above the levels set by a single card scheme, since this would lead all merchants to reject cards. With heterogenous merchants, competition results in higher interchange fees than that set by a single card scheme, resulting in fewer card transactions and lower member profits. With consumers only able to hold a single card, card schemes compete by offering the maximal surplus to cardholders, which leads to a fee structure which is biased towards cardholders. Surprisingly, competition between payment schemes leads to a greater distortion in the fee structure compared to the case of a single (monopoly) card scheme (merchants are charged more and cardholders less). Section 3 analyzes this case, and provides some implications for the competitive neutrality of regulating interchange fees and the policy of duality.

Alternatively, when consumers hold both cards, merchants will reject a single card with unfavorable terms, causing customers to either switch to the card which is preferred by merchants or to switch merchants. Taking into account both effects, we show that competition between card associations results in an interchange fee which is lower than that set by a single card association and lower than the socially optimal interchange fee. Identical card schemes compete by offering the maximal surplus to merchants. This leads to a fee structure which, if anything, is biased towards merchants. This case is considered in Section 4, with some implications drawn for the competitive neutrality of regulating interchange fees and for the policy of honor-all-cards (the 'Walmart case').

Although we focus on competing card associations, we also show how our results apply directly to the case of competing proprietary schemes, like American Express and Discover card, that set their cardholder and merchant fees directly, and to the case of competition between a card association and a proprietary scheme. In each section we consider the cases with and without unobservable merchant heterogeneity. Section 5 provides a brief conclusion.

2 A single card scheme

We start by considering the simpler case of a single payment scheme where the only alternative to using cards is cash, before later introducing competing payment schemes. This allows our framework to be compared to that of Rochet and Tirole (2002) and Wright (2001). The essential modelling difference between our framework and theirs is that in our framework consumers are assumed to receive their particular draw of transactional benefits from using cards once they have chosen which merchant to purchase from. In Rochet and Tirole's model, consumers get their draw of transactional benefits before they choose which merchant to purchase from. Clearly our timing assumption is made for modelling convenience. We think it is a reasonable modelling approach for two reasons. First, with our setup, consumers still choose which merchant to purchase from taking into account the expected benefits from using cards versus the alternative payment instrument. By accepting cards, merchants will raise consumers' expected benefit from purchasing from them, since consumers will gain the option of using cards for purchases. In fact, this section shows the timing assumption does not alter the equilibrium conditions under which merchants accept cards or consumers use cards: they are identical to the conditions derived in Rochet and Tirole (2002) and Wright (2001). Thus, we do not think this particular assumption is driving the results we obtain. Second, the timing assumption can be motivated by the idea that consumers only learn of their particular need to use various types of payments once they are in the store.

A card association represents the joint interests of its members, who are issuers (banks and other financial institutions which specialize in servicing cardholders) and acquirers (banks and other financial institutions which specialize in servicing merchants). In such an open scheme, a card association sets an interchange fee a to maximize its members' profits.[3] The interchange fee is defined as an amount paid from acquirers to issuers per card transaction. Competition between symmetric issuers and competition between symmetric acquirers then determines the equilibrium fee per transaction for using cards, f, and the equilibrium fee per transaction for accepting cards, m.[4] In the case of a proprietary card scheme such as American Express, the scheme sets f and m directly to maximize its profits.

We assume a cost of c_I per transaction of issuing and c_A per transaction of acquiring. A proprietary scheme incurs the cost $c_I + c_A$ of a card transaction. Since we will be assuming perfect competition between payment schemes, we adopt the following simple model of competition between issuers and between acquirers in a card association. We assume issuers and acquirers can only set their fees in discrete units, so that Bertrand competition between identical issuers and between identical acquirers leaves some small margin to them each. It follows that the card fees in a card asssociation are set so that

$$f(a) = c_I - a + \pi_I \tag{1}$$

[3] As noted below, in our model this is equivalent to assuming instead that the card association seeks to maximize the total number of card transactions. Notably, MasterCard and Visa (the card associations) collect revenues based on the total number of card transactions on their systems.

[4] We assume away any joining fees or annual fees that consumers may face for obtaining a card in the first place. Such fees are no longer very important in the U.S. and are becoming less important in other jurisdictions. For simplicity, we also assume all fees are levied as per transaction rather than ad-valorem.

where π_I is some constant positive profit margin. Likewise, merchant fees are

$$m(a) = c_A + a + \pi_A, \tag{2}$$

where π_A is a some constant positive profit margin.[5] Card fees decrease and merchant fees increase as the interchange fee is increased. The implication is that the interchange fee only determines the structure of fees, not the overall level of fees. As a result, a card association will maximize its members' profits by choosing its interchange fee to maximize its volume of card transactions. The sum of cardholder and merchant fees per-transaction is denoted as $l = f(a) + m(a)$. Given our assumptions about bank competition, l is independent of the interchange fee a (it equals $c_I + c_A + \pi_I + \pi_A$).

As in Rochet and Tirole (2002), consumers get transactional benefits b_B from using cards as opposed to the alternative cash, and merchants get transactional benefits b_S from accepting cards relative to the alternative of accepting cash. We assume b_B is drawn with a positive density $h(b_B)$ over the interval $[\underline{b}_B, \overline{b}_B]$. Merchants are assumed to be unable to price discriminate depending on whether consumers use cards or not, so consumers will want to pay with cards if and only if $b_B \geq f$.[6] The quasi-demand for card usage is defined as $D(f) = 1 - H(f)$, which is the proportion of consumers who want to use cards at the fee f, where H denotes the cumulative distribution function corresponding to b_B.

The average convenience benefit to those consumers using cards for a transaction is then $\beta(f) = E[b_B \mid b_B \geq f]$, which is increasing in f . The net average convenience benefit to those who use cards for transactions is $\delta(f) = \beta(f) - f$. Note that $D(f)\delta(f)$, which captures the expected net benefits to a consumer from being able to use their card at a merchant, is positive and decreasing in f. Finally, we assume that

$$E(b_B) + b_S < l < \overline{b}_B + b_S \tag{3}$$

so as to rule out the possibility that there is no card use and to rule out the possibility that all consumers use cards.

We follow Rochet and Tirole (2002) and assume initially all merchants (sellers) receive the same transactional benefits b_S from accepting cards. This restriction is relaxed in Section 2.2.

2.1 No merchant heterogeneity

Initially, assume there are only two merchants, each with constant marginal costs d of production. Merchants are assumed to compete Hotelling style. In particular, consumers are uniformly distributed on the unit interval and the two merchants are located at either extreme. Consumers get utility v from the

[5] This assumption is the same as that used by Rochet and Tirole (2001) to analyze competition between platforms that are run as associations; see their section 3.4 for a more sophisticated justification of these functional forms. See also Hausman et al. (2003). Taking the limit as π_I and π_A tend to zero but assuming card associations still seek to maximize the total number of card transactions allows us to capture the case of perfect inter-system competition and (almost) perfect intra-system competition.

[6] This no price discrimination assumption can be motivated by the no-surcharge rules that card associations have adopted to prevent merchants from charging more to consumers for purchases made with cards. It can also be motivated by the observation of price coherence (Frankel, 1998) — that merchants are generally reluctant to set differential prices depending on the payment instrument used.

purchase of the good, which is assumed to be high enough that all consumers wish to purchase from one of the two merchants. A consumer located at x faces linear transportation costs of tx from purchasing from merchant 1 and $t(1-x)$ from purchasing from merchant 2. These transportation costs can be summarized by the function $T_i(x) = tx\,(2-i) + t\,(1-x)\,(i-1)$ where $i = 1$ corresponds to merchant 1 and $i = 2$ corresponds to merchant 2.

The timing of the game is summarized as follows:

(i) The payment card association sets the level of the interchange fee a.

(ii) Issuers and acquirers set fees f and m to cardholders and merchants respectively, according to (1) and (2).

(iii) Merchants decide whether or not to accept cards and then set their prices.

(iv) Consumers decide which merchant to buy from, based on their draw of x.

(v) Based on their individual realizations of b_B, consumers decide whether to use the card or cash for payment.

The game is first solved backwards to stage (iii), yielding merchants' equilibrium acceptance condition.

Proposition 1 *With a single card scheme, merchants accept cards if and only if* $b_S \geq l - \beta(f)$.

Proof. The last stage of the game is trivial to solve — given consumers face the same price for cash and card, consumers use cards whenever $b_B > f$. Now consider stage (iv) of the game. Let I_i denote an indicator variable which takes the value 1 if merchant i accepts cards and 0 otherwise. A consumer located at x will purchase from the merchant i for which $v - p_i + D(f)\delta(f)I_i - T_i(x)$ is the highest, where p_i is the common price set by merchant i regardless of the method of payment. Merchant i's market share is

$$s_i = \frac{1}{2} + \frac{1}{2t}\left(p_j - p_i + D(f)\delta(f)(I_i - I_j)\right) \tag{4}$$

and merchant i's profit is

$$\pi_i = s_i\left(p_i - d - D(f)(m - b_S)I_i\right) \tag{5}$$

since a fraction $D(f)$ of all sales will be via cards if the merchant accepts cards.

Solving for the Nash equilibrium in prices in stage (iii) we get

$$\begin{aligned} p_i &= d + t + \frac{1}{3}D(f)\delta(f)(I_i - I_j) + D(f)(m - b_S)\left(\frac{2}{3}I_i + \frac{1}{3}I_j\right) \\ p_j &= d + t + \frac{1}{3}D(f)\delta(f)(I_j - I_i) + D(f)(m - b_S)\left(\frac{2}{3}I_j + \frac{1}{3}I_i\right). \end{aligned}$$

With these prices, merchant i's market share is

$$s_i = \frac{1}{2} + \frac{D(f)}{6t}\left(b_S + \beta(f) - l\right)(I_i - I_j) \tag{6}$$

and its profit is $\pi_i = 2t(s_i)^2$. Therefore, merchant i accepts cards if and only if doing so increases its market share. Accepting cards is thus optimal if and only if

$$b_S \geq l - \beta(f), \tag{7}$$

completing the proof. ■

Notice card acceptance (that is, $I_i = 1$) does three things. First, it raises the demand faced by merchant i. It provides consumers with a valuable option from shopping at the merchant concerned, which is that they can use cards if doing so is convenient for them ($b_B > f$). The expected value of this option is measured by the term $D(f)\delta(f)$ in the firm's market share equation (4). Second, and by symmetry, it lowers the demand faced by the rival firm. Thus, in this model card acceptance has a business stealing effect. Third, as the firm's profit equation (5) reveals, card acceptance changes the merchant's costs, increasing them if $m > b_S$ for the merchant concerned.

Condition (7) defines the marginal merchant transactional benefits, below which both merchants will not accept cards and above which both merchants will accept cards. Interestingly, this is the same equilibrium condition that Rochet and Tirole (2002) and Wright (2001) obtained.[7]

We are now ready to analyze the effects of different interchange fees at stage (i) of the game. For a given value of merchants' transactional benefits b_S, consider the profit maximizing interchange fee $a^{\Pi}(b_S)$ and the welfare maximizing interchange fee $a^W(b_S)$. The card scheme's profit (that is, the total profit of the association's member banks) is

$$\Pi(f) = (\pi_I + \pi_A)\, D(f)$$

if both merchants accept cards, and is zero if both merchants reject cards. Then we have

Proposition 2 *A single card scheme sets its interchange fee $a^{\Pi}(b_S)$ so that $\beta(f) + b_S = l$.*

Proof. A single card association maximizes $\Pi(f)$ by maximizing $D(f))$ subject to the constraint in (7) which ensures merchants accept cards. Since $D(f)$ is increasing in a and $\beta(f)$ is decreasing in a, this implies

$$b_S + \beta(f) = l. \tag{8}$$

Note that $E(b_B) = \beta(\underline{b}_B) < \beta(f) < \beta(\overline{b}_B) = \overline{b}_B$ and so from (3) there is a sufficiently low interchange fee such that $\beta(f) > l - b_S$ and a sufficiently high interchange fee such that $\beta(f) < l - b_S$. It follows that there exists a unique interchange fee $a^{\Pi}(b_S)$ satisfying (8). ■

Assuming merchants accept cards, welfare in this model can be written as

$$W(f) = \int_f^{\overline{b}_B} (b_B + b_S - c_I - c_A)\, h\,(b_B)\, db_B$$

[7] To be precise, in Rochet and Tirole's model the term l on the right hand side is replaced by m. This difference only arises because in their model the consumers' fee f is effectively a fixed fee and so occurs regardless of the extent to which the card is used. Moreover, Rochet and Tirole also found another equilibrium was possible in which merchants with values of b_S slightly above this critical level both reject cards, although they rule out this second equilibrium. In their model, each firm's acceptance decision is affected by what the other firm does. Due to our slightly different timing assumption, each firm's acceptance decision is independent of what other firms are doing, and so we get a unique equilibrium.

since monetary transfers between cardholders, merchants, and banks do not matter given the unit demands assumption. Throughout the paper we ignore additional constant terms in the welfare function that arise from the average transportation costs incurred by consumers. As in Rochet and Tirole (2002), it follows that

$$a^W(b_S) = \min\left(b_S - c_A + \pi_I, a^\Pi(b_S)\right). \tag{9}$$

With no merchant heterogeneity, the welfare maximizing interchange fee is either the same as the privately chosen interchange fee set by a single scheme, or is lower.

2.2 Merchant heterogeneity

It is likely merchants in different industries get different benefits from accepting cards b_S. To model this we follow Wright (2001) and assume within each industry merchants are exactly as in the previous section, but across industries vary by their type b_S. The random variable b_S is drawn with a positive density $g(b_S)$ over the interval $[\underline{b}_S, \overline{b}_S]$, and a cumulative distribution denoted G. For the case of merchant heterogeneity we replace the condition (3) with its natural extension

$$E(b_B) + \underline{b}_S < l < \overline{b}_B + \overline{b}_S. \tag{10}$$

This rules out the possibility that there is no card use and the possibility that all consumers and all merchants use cards.

The particular draws of b_S are assumed to be unobserved by the card schemes. Specifically, draws of b_S are realized at stage (iii) of the game, prior to merchants deciding whether to accept cards or not. Consumers are exogenously matched to all the different industries, and so without loss of generality they buy one good from each industry. The merchants' (sellers') "quasi-demand" function which measures the proportion of merchants with transactional benefits above some level b_S is denoted $S(b_S) = 1 - G(b_S)$.

Using the above result on when merchants in a particular industry accept cards and when they reject cards, the number of card transactions will be $D(f)S(l - \beta(f))$. The card scheme will maximize its members' profits

$$\Pi = (\pi_I + \pi_A)\, D(f)\, S(l - \beta(f))$$

(and card transactions) at the interchange fee denoted a^Π. Welfare can now be written

$$W(f) = \int_f^{\overline{b}_B} \int_{l-\beta(f)}^{\overline{b}_S} (b_B + b_S - c_I - c_A)\, g(b_S)\, h(b_B)\, db_S db_B$$

which is maximized at the interchange fee denoted a^W.

A useful benchmark for later results is the special case in which quasi-demand functions are linear. Suppose b_B and b_S are both uniformly distributed. Then we have from Wright (2001) that

$$D(f) = \frac{\overline{b}_B - f}{\overline{b}_B - \underline{b}_B}, \tag{11}$$

$$S(b_S) = \frac{\overline{b}_S - b_S}{\overline{b}_S - \underline{b}_S}, \tag{12}$$

$$\beta(f) = \frac{\overline{b}_B + f}{2}, \tag{13}$$

$$a_U^{\Pi} = \overline{b}_S - c_A - \pi_A, \tag{14}$$

and

$$a_U^W = \pi_I + \frac{1}{3}\left(\overline{b}_B - c_I + 4(\overline{b}_S - c_A)\right) - \frac{2}{3}\sqrt{\left(3(\pi_I + \pi_A)^2 + \left(\overline{b}_B + \overline{b}_S - c_I - c_A\right)^2\right)}. \tag{15}$$

3 System competition when consumers hold only one card

Suppose there are two competing identical card systems. Identical systems not only have the same costs, they also provide the same transactional benefits to cardholders and to merchants. The only distinguishing feature of each card scheme is the fee structure it chooses. Specifically, each card association i sets an interchange fee denoted a^i.[8] Like the case with a single scheme, we assume that for competing card associations

$$f^i(a^i) = c_I - a^i + \pi_I \tag{16}$$

and

$$m^i(a^i) = c_A + a^i + \pi_A \tag{17}$$

so that the interchange fee determines the structure but not the overall level of fees. As before we define l as the sum of f and m, which from the fact that the issuers and acquirers in each scheme are assumed to have the same costs and margins, does not depend on i.

We suppose consumers have to join one of the two identical card schemes before learning of their preferences for particular merchants. Let λ be the fraction of consumers who hold card 1 only and $1 - \lambda$ be the fraction of consumers who hold card 2 only. To rule out multiple equilibria in subgames, we assume that if consumers are indifferent between holding each of the two cards, they will randomly choose one of the cards to hold. Alternative tie-breaking conventions are consistent with one card scheme attracting more cardholders than the other, but result in the same equilibrium fee structure. We first consider the case without merchant heterogeneity.

3.1 No merchant heterogeneity

The timing of the game is now as follows:

(i) Each payment card association sets the level of their interchange fee a^i.

(ii) Issuers and acquirers set fees f^i and m^i to cardholders and merchants according to (16) and (17).

(iii) Consumers decide which type of card to hold.

(iv) Merchants decide whether to accept cards (neither, one or both) and then set their prices.

[8] Superscripts are used to denote different card schemes.

(v) Consumers decide which merchant to buy from after getting their draw of location x.

(vi) Based on their individual realizations of b_B, consumers decide whether to use their card or cash for payment.

Note in working out the equilibrium below we assume (consistent with our timing assumption above) that merchants take as given consumers' choice of which card scheme to belong to (that is, they treat λ as a parameter) when deciding whether to accept cards or not. This seems reasonable given that there are many merchants in practice, each of which is likely to be too small to influence a consumer's decision to join a particular card scheme.

To solve the game we work backwards. The first proposition below characterizes when merchants will accept cards from each scheme. Merchants' acceptance condition at stage (iv) now involves working out whether they will reject both cards, accept one, or accept both. This turns on the value of the following simple function. Define ϕ^i as

$$\phi^i = D(f^i)(b_S + \beta(f^i) - l), \tag{18}$$

which is a measure of how much accepting cards contributes to a merchant's equilibrium profits, holding constant the acceptance decision of the rival merchant.

Proposition 3 *Merchants accept a card from scheme i if and only if $b_S \geq l - \beta(f^i)$.*

Proof. In stage (vi) consumers use cards if $b_B > f$ and merchants accept cards. Given this, in stage (v) a consumer at location x will choose the merchant for which

$$v - p_i + D(f^1)\delta(f^1)I_i^1 - T_i(x)$$

is the highest if they hold a card from scheme 1 and the merchant for which

$$v - p_i + D(f^2)\delta(f^2)I_i^2 - T_i(x)$$

is the highest if they hold a card from scheme 2, where I_i^j is the indicator variable that takes a value of 1 if merchant i accepts cards offered by card scheme j and 0 if merchant i does not accept cards offered by card scheme j. This implies merchant i's market share is

$$s_i = \frac{1}{2} + \frac{1}{2t}\left(p_j - p_i + \lambda D(f^1)\delta(f^1)(I_i^1 - I_j^1) + (1-\lambda)D(f^2)\delta(f^2)(I_i^2 - I_j^2)\right).$$

Merchant i's profit is then

$$\pi_i = s_i\left(p_i - d - \lambda D(f^1)(m^1 - b_S)I_i^1 - (1-\lambda)D(f^2)(m^2 - b_S)I_i^2\right).$$

Solving for the Nash equilibrium in prices in stage (iv) we get

$$\begin{aligned}
p_i &= t + d + \frac{1}{3}\lambda D(f^1)\delta(f^1)(I_i^1 - I_j^1) + \frac{1}{3}(1-\lambda)D(f^2)\delta(f^2)(I_i^2 - I_j^2) \\
&\quad + \lambda D(f^1)\left(m^1 - b_S\right)\left(\frac{2}{3}I_i^1 + \frac{1}{3}I_j^1\right) + (1-\lambda)D(f^2)\left(m^2 - b_S\right)\left(\frac{2}{3}I_i^2 + \frac{1}{3}I_j^2\right) \\
p_j &= t + d + \frac{1}{3}\lambda D(f^1)\delta(f^1)(I_j^1 - I_i^1) + \frac{1}{3}(1-\lambda)D(f^2)\delta(f^2)(I_j^2 - I_i^2) \\
&\quad + \lambda D(f^1)\left(m^1 - b_S\right)\left(\frac{2}{3}I_j^1 + \frac{1}{3}I_i^1\right) + (1-\lambda)D(f^2)\left(m^2 - b_S\right)\left(\frac{2}{3}I_j^2 + \frac{1}{3}I_i^2\right)
\end{aligned}$$

and substituting these into the equilibrium profits for merchant i, profits simplify to

$$\pi_i = 2t\left[\frac{1}{2} + \frac{1}{6t}\left(\lambda\phi^1(I_i^1 - I_j^1) + (1-\lambda)\phi^2(I_i^2 - I_j^2)\right)\right]^2.$$

In deciding which card to accept at stage (iv), merchant i compares $\lambda\phi^1$ if it accepts just card 1, $(1-\lambda)\phi^2$ if it accepts just card 2, and $\lambda\phi^1 + (1-\lambda)\phi^2$ if it accepts both cards. A merchant will therefore accept a card from scheme i if and only if $\phi^i \geq 0$ or equivalently, if and only if $b_S \geq l - \beta(f^i)$. ■

Even though there are two card schemes, the condition that determines whether merchants accept cards is identical to the case with a single card scheme. A merchant's decision about accepting cards of type 1 is independent of its decision about accepting cards of type 2. Each individual merchant does not expect to be able to influence the number of cardholders of each type, and so it acts as though these are two segmented groups of consumers. Where both schemes offer the same fee structure, merchants that accept one card will accept both. Consistent with this, MasterCard and Visa offer very similar fee structures, and merchants that accept one, accept both.

Given the symmetry of merchants in an industry we have that $p \equiv p_i = p_j$, $s_i = 1/2$, $I^1 \equiv I_i^1 = I_j^1$, and $I^2 \equiv I_i^2 = I_j^2$. Now consider stage (iii) where consumers choose which card scheme to join. Consumers get expected utility

$$v^1 = \int_0^{\frac{1}{2}} \left(v - p + D(f^1)\delta(f^1)I^1 - tx\right) dx + \int_{\frac{1}{2}}^1 \left(v - p + D(f^1)\delta(f^1)I^1 - t(1-x)\right) dx$$

if they join scheme 1 and expected utility

$$v^2 = \int_0^{\frac{1}{2}} \left(v - p + D(f^2)\delta(f^2)I^2 - tx\right) dx + \int_{\frac{1}{2}}^1 \left(v - p + D(f^2)\delta(f^2)I^2 - t(1-x)\right) dx$$

if they join scheme 2.

If both schemes set the same interchange fee (so $\phi \equiv \phi^1 = \phi^2$) then merchants will accept both cards if $\phi \geq 0$ and neither otherwise. When merchants accept both cards, consumers randomize over which card to hold, and the members of such card schemes get (in aggregate) profits of

$$\Pi^1 = \Pi^2 = (\pi_I + \pi_A)\frac{D(f)}{2}.$$

If card schemes act to maximize their joint profits they will set their interchange fees so that $\phi^1 = \phi^2 = 0$ which leads to the highest level of $D(f)$ such that merchants still accept cards. This is the interchange fee $a^{\Pi}(b_S)$ defined in (8). We now show this is also the equilibrium outcome from competition between the two identical schemes.

Proposition 4 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers choose to hold only one card and there is no merchant heterogeneity is equal to $a^{\Pi}(b_S)$, the interchange fee which maximizes the card schemes' joint profits.*

Proof. If both schemes set $a^{\Pi}(b_S)$, then $\phi^1 = \phi^2 = 0$. Neither scheme will set a higher interchange fee since this will cause all merchants to reject its cards. Suppose scheme 1 sets a lower interchange fee such that $\phi^1 > 0$ and $\phi^2 = 0$. Consumers know from proposition 3 that merchants will accept both cards. Consumers will therefore only hold cards of scheme 2 since $v^2 > v^1$ given $D(f^2)\delta(f^2) > D(f^1)\delta(f^1)$.

Thus, neither scheme will want to change its interchange fee from $a^{\Pi}(b_S)$. Moreover, at any other proposed equilibrium with $\phi^i > 0$, scheme i can always increase its transactions by increasing a^i until $\phi^i = 0$, causing consumers to hold only card i without leading merchants to reject card i. Thus, both schemes setting $a^{\Pi}(b_S)$ is the unique equilibrium. ■

Despite competition between identical schemes, they will each set their interchange fees as though they are a single scheme maximizing card transactions (and profits). When consumers hold only one card, the only effect of competition between card schemes is to make it more attractive for each card scheme to lower card fees since this has the added benefit of attracting more consumers to hold their card in the first place. However, since with no merchant heterogeneity a single scheme already sets the interchange fee to the point where merchants only just accept cards, there is no scope to further lower fees to cardholders by raising merchants' fees.

Competition between two identical proprietary schemes will drive the sum of their fees $f^i + m^i$ down to costs $c_I + c_A$. Subject to this constraint, the logic of proposition 4 implies competing proprietary schemes will choose f^i and m^i so that $b_S + \beta(f^i) = f^i + m^i$. The result will be the same structure of fees for cardholders and merchants as implied by the equilibrium interchange fee set by competing card associations, in the limit as $\pi_I = \pi_A \to 0$. In this case, the implicit interchange fee set by proprietary schemes can be defined as $m^i - c_A$. The result implies the implicit interchange fee set by competing proprietary schemes will be (approximately) equal to the explicit interchange fee set by competing card associations. The same structure of fees also arises from competition between a proprietary scheme and a card association, taking the limit as $\pi_I = \pi_A \to 0$.

An important implication of this result is to consider what happens if one scheme has its interchange fee regulated below the competitive level $a^{\Pi}(b_S)$. For instance, in 2002 the Reserve Bank of Australia proposed regulations of the card associations' interchange fees, while explicitly noting it will leave proprietary schemes fee structure unregulated. Since proprietary schemes, such as American Express, do not have to set interchange fees to achieve their desired price structure (they set cardholder and merchant fees directly), any regulation of interchange fees acts as a potential handicap to card associations. Our model implies such asymmetric regulation will not induce an unregulated proprietary scheme to alter its fee structure in this setting. Despite this, the proprietary scheme does benefit from the regulation. Competition between card schemes does not ensure competitive neutrality when one scheme has its interchange fee regulated below the competitive level. In fact, with perfect competition, the unregulated scheme attracts all card transactions away from the regulated scheme. More generally, the model predicts that attempts to make consumers pay a greater share of the cost of card payments (by regulating a lower interchange fee for card associations), will result in consumers switching to cards of unregulated proprietary schemes that will be able to offer lower card fees (or higher card rebates). Merchants will still accept these proprietary cards even though cards offered by regulated schemes involve lower merchant fees, since consumers will hold (and want to use) proprietary cards exclusively. Thus, even if $a^{\Pi}(b_S)$ is above the socially optimal level, it does not follow that regulating a lower interchange fee, but leaving proprietary schemes free to set their fee structures, will lead to any welfare improvements.

3.2 Merchant heterogeneity

The above analysis applies to a single industry. In this section we extend the above model to allow industries to vary in the transactional benefits they get from accepting cards. This is done exactly as in Section 2.2. Allowing for merchant heterogeneity (across different industries), we can apply proposition 3 to each industry. This solves stage (iv) of the game. Then solving stage (iii) of the game, consumers will now take into account the expected proportion of merchants that accept each type of card in deciding which card scheme to join. Since the option of using their card provides positive surplus, consumers will prefer to join a card scheme which, other things equal, they know will be accepted by more merchants.

Denoting the decision of merchants in an industry (of type b_S) to accept cards of scheme i or not as $I^i(b_S)$, consumers get expected utility

$$\begin{aligned} v^1 &= \int_{\underline{b}_S}^{\overline{b}_S} \left(\int_0^{\frac{1}{2}} \left(v - p(b_S) + D(f^1)\delta(f^1)I^1(b_S) - tx \right) dx \right) g(b_S) db_S \\ &+ \int_{\underline{b}_S}^{\overline{b}_S} \left(\int_{\frac{1}{2}}^1 \left(v - p(b_S) + D(f^1)\delta(f^1)I^1(b_S) - t(1-x) \right) dx \right) g(b_S) db_S \end{aligned}$$

if they join scheme 1 and

$$\begin{aligned} v^2 &= \int_{\underline{b}_S}^{\overline{b}_S} \left(\int_0^{\frac{1}{2}} \left(v - p(b_S) + D(f^2)\delta(f^2)I^2(b_S) - tx \right) dx \right) g(b_S) db_S \\ &+ \int_{\underline{b}_S}^{\overline{b}_S} \left(\int_{\frac{1}{2}}^1 \left(v - p(b_S) + D(f^2)\delta(f^2)I^2(b_S) - t(1-x) \right) dx \right) g(b_S) db_S \end{aligned}$$

if they join scheme 2.

Since merchants accept cards whenever $\phi^i \geq 0$ (proposition 3), scheme 1's profit is

$$\Pi^1 = (\pi_I + \pi_A)\lambda D(f^1) S\left(l - \beta(f^1)\right) \tag{19}$$

and scheme 2's profit is

$$\Pi^2 = (\pi_I + \pi_A)(1-\lambda) D(f^2) S\left(l - \beta(f^2)\right). \tag{20}$$

If λ is held constant, both (19) and (20) are maximized at the interchange fee a^{Π} defined in Section 2.2 which maximizes the number of card transactions for a single card scheme. Allowing λ to be determined endogenously, competition can lead to a higher or lower interchange fee, depending on whether consumers prefer to join card schemes which have low fees but few merchants accepting cards or which have high fees and many merchants accepting cards. At a^{Π}, this just depends on whether a higher interchange fee increases or decreases the average net benefit $\delta(f)$ of those consumers using cards.

Consumers will hold a card with the scheme that provides the highest value of v^i. Since

$$v^1 - v^2 = \delta\left(f^1\right) D\left(f^1\right) S\left(l - \beta(f^1)\right) - \delta\left(f^2\right) D\left(f^2\right) S\left(l - \beta(f^2)\right),$$

scheme i will not get any cardholders unless it sets its interchange fee to maximize consumers' expected benefits from holding a card $\delta(f^i)D(f^i)S(l - \beta(f^i))$. Assuming it is never optimal for a scheme to have all merchants accepting cards or consumers always using cards, it follows that

Proposition 5 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold only one card maximizes consumers' expected benefits from holding a card.*

and that

Proposition 6 *If $\delta(f)$ is decreasing in f then the equilibrium interchange fee resulting from competition between identical card schemes when consumers hold only one card cannot be below the joint profit maximizing interchange fee.*

Proof. Suppose both schemes set interchange fees at a^{Π}. This maximizes the total number of transactions $D(f)S(l-\beta(f))$ per cardholder. Let $T(f) = D(f)S(l-\beta(f))$ and $\theta(f) = \delta(f)T(f)$, so that $f^{\Pi} \equiv f(a^{\Pi}) = \arg\max_f T(f)$. It follows that

$$T(f) \leq T(f^{\Pi}) \quad \forall\, f \geq f^{\Pi}.$$

Since $\delta(f)$ is decreasing in f for all f, then

$$\delta(f) \leq \delta(f^{\Pi}) \quad \forall\, f \geq f^{\Pi}.$$

Since $T(f)$ and $\delta(f)$ are both nonnegative for all f, it follows that

$$\theta(f) = \delta(f)T(f) \leq \delta(f^{\Pi})T(f^{\Pi}) = \theta(f^{\Pi}) \quad \forall\, f \geq f^{\Pi}.$$

Therefore $\arg\max_f \theta(f) \leq f^{\Pi}$ and the equilbrium interchange fee cannot be less than the joint profit maximizing interchange fee. ■

The intuition behind the result is simple. Holding constant the fraction of cardholders on each scheme, card schemes will maximize their respective profits by acting as though they do not face any competition. The market is effectively segmented. At this point, a small change in a scheme's interchange fee will have only a second order impact on the number of card transactions. However, a slightly higher interchange fee may have a first order impact on the average net benefit consumers get from using cards. If the average net benefit to those using cards increases when the interchange fee is increased above a^{Π}, each card scheme will set interchange fees too high in an attempt to get consumers to switch to holding their cards exclusively, an effect which ends up reducing the total number of card transactions and their members' profits.[9] If the reverse assumption holds, then schemes will compete by setting an interchange fee that is too low for their own good.[10]

Two identical proprietary schemes will behave in the same way. Competition will drive the sum of their fees $f^i + m^i$ to costs $c_I + c_A$. Subject to this constraint, the schemes will set fees f^i and m^i to maximize $\delta(f^i)D(f^i)S(f^i+m^i-\beta(f^i))$ in an attempt to attract consumers to hold their cards exclusively.

[9] This condition is satisfied if the consumer quasi-demand for card use is linear (b_B follows the uniform distribution). Then $\delta(f) = \left(\overline{b}_B - f\right)/2$, so that $\delta(f)$ is decreasing in f.

[10] When only some consumers hold cards, lower card fees have the beneficial effect of inducing more people to hold cards in the first place, and so increase the total number of card transactions. However, even with partial cardholding, the effect identified here should hold. Relative to the interchange fee which maximizes total card transactions and the card schemes' joint profits, each card scheme will still want to set a higher interchange fee to get more consumers to sign up to its own cards, trading off an increase in share of cardholders with a reduction in the total number of card transactions.

Similarly, competition between a proprietary scheme and a card association results in the same structure of fees in the limiting case in which $\pi_I = \pi_A \to 0$.

Does regulating one scheme's interchange fee still raise issues of competitive neutrality in this environment? Any regulation of scheme 1's interchange fee that lowers its interchange fee below the competitive level will allow scheme 2 to attract all consumers to hold its cards exclusively by setting an interchange fee above the regulated fee imposed on scheme 1. Moreover, in the case in which the competitive level of interchange fees is above a^{Π}, regulating scheme 1's interchange fee below a^{Π} will lead scheme 2 to set its interchange fee (or fee structure) at the joint profit maximizing level (a^{Π}), thereby attracting all cardholders and the maximum number of card transactions. Regulating the interchange fee of card associations but leaving proprietary schemes free to set their fee structure is not competitively neutral.

The result above may also have some implications for duality, which is the practice of MasterCard and Visa having the same members setting their rules and interchange fees. To the extent duality means both schemes set interchange fees (or their fee structure) to maximize their joint profits, the above analysis suggests it could actually result in lower interchange fees rather than higher interchange fees. It is thus interesting that in Canada, where duality does not apply, interchange fees appear to be relatively high. For instance, according to a report by the Australian Banking Association, filed publicly with the Reserve Bank of Australia, Canadian interchange fees for electronic transactions were the highest of the eleven countries examined. It is also noteworthy that interchange fees are relatively high in the United States compared to other OECD countries, despite the fact competition between the four main credit card schemes there (Visa, MasterCard, American Express and DiscoverCard) appears to be relatively vigorous when compared to competition in other jurisdictions. Certainly, high interchange fees (or distortions in fee structures) are not necessarily caused by a lack of inter-system (or intra-system) competition.

4 System competition when consumers hold both cards

The previous model considered competition between two identical card schemes in which consumers could only hold one type of card. In this section we take the opposite extreme, and suppose consumers all hold both types of cards. After consumers have chosen which merchant to purchase from, they get their draw of b_B which is assumed to be same for both schemes (that is, cards from the two schemes provide identical transactional benefits). As in Section 3, we assume some tie-breaking conventions to rule out multiple equilibria in subgames. We assume that if merchants are indifferent between accepting two different cards they will accept both cards, and if consumers are indifferent between using two different cards, they will randomize over which card they use. Alternative assumptions are consistent with one card scheme attracting more transactions than the other, but result in the same equilibrium fee structure. The timing of the game is identical to that in Section 2.

4.1 No merchant heterogeneity

To solve the game we work backwards. The first proposition characterizes when merchants will accept cards from each scheme. The measure of how much accepting cards contributes to a merchant's equilibrium profits (the expression ϕ^i) is again critical in defining a merchant's optimal decision.

Proposition 7 *Merchants accept card i if $\phi^i \geq \phi^j$ and $b_S \geq l - \beta(f^i)$, and accept card i if $f^i > f^j$ and $b_S \geq l - \beta(f^i)$. Otherwise, they reject card i.*

Proof. Noting that consumers will want to use a card when $b_B \geq f$, consumers get expected utility from card usage of $D(f^j)\delta(f^j)$ if they can only use card j. If the merchant accepts both cards, given identical transactional benefits of the two cards, the consumer will always prefer the card with a lower fee f^j.

Define the variable L_i^j to capture the likelihood consumers will prefer to use scheme j at merchant i assuming they get a sufficiently high draw of b_B. If merchant i only accepts card j then let $L_i^j = 1$. Similarly, if merchant i does not accept card j then let $L_i^j = 0$. If merchant i accepts both cards then $L_i^1 = 1$ and $L_i^2 = 0$ if $f^1 < f^2$, $L_i^1 = 0$ and $L_i^2 = 1$ if $f^1 > f^2$, and $L_i^1 = L_i^2 = 1/2$ if $f^1 = f^2$.

In stage (iv) consumers will purchase from the merchant i for which

$$v - p_i + D(f^1)\delta(f^1)I_i^1 L_i^1 + D(f^2)\delta(f^2)I_i^2 L_i^2 - T_i(x)$$

is the highest, implying merchant i's market share is

$$s_i = \frac{1}{2} + \frac{1}{2t}\left(p_j - p_i + D(f^1)\delta(f^1)\left(I_i^1 L_i^1 - I_j^1 L_j^1\right) + D(f^2)\delta(f^2)\left(I_i^2 L_i^2 - I_j^2 L_j^2\right)\right)$$

and merchant i's profit is

$$\pi_i = s_i\left(p_i - d - D(f^1)(m^1 - b_S)I_i^1 L_i^1 - D(f^2)(m^2 - b_S)I_i^2 L_i^2\right). \tag{21}$$

To understand (21), note that if a merchant accepts none of the cards ($I_i^1 = 0$ and $I_i^2 = 0$), or just one type of card (either $I_i^1 = 1$ and $I_i^2 = 0$ or $I_i^1 = 0$ and $I_i^2 = 1$), then the profit margin is identical to the case with a single card scheme, as in equation (5). When the merchant accepts both cards ($I_i^1 = 1$ and $I_i^2 = 1$), consumers will only use the one with the lowest fee (say f^1), which implies merchants will face the corresponding merchant fee m^1 and card demand $D(f^1)$ since $L_i^1 = 1$ and $L_i^2 = 0$.

Then solving for the Nash equilibrium in prices in stage (iii) of the game, the equilibrium prices are

$$\begin{aligned}
p_i &= d + t - \frac{2}{3}\phi^1 I_i^1 L_i^1 - \frac{2}{3}\phi^2 I_i^2 L_i^2 - \frac{1}{3}\phi^1 I_j^1 L_j^1 - \frac{1}{3}\phi^2 I_j^2 L_j^2 + D(f^1)\delta(f^1)I_i^1 L_i^1 + D(f^2)\delta(f^2)I_i^2 L_i^2 \\
p_j &= d + t - \frac{2}{3}\phi^1 I_j^1 L_j^1 - \frac{2}{3}\phi^2 I_j^2 L_j^2 - \frac{1}{3}\phi^1 I_i^1 L_i^1 - \frac{1}{3}\phi^2 I_i^2 L_i^2 + D(f^1)\delta(f^1)I_j^1 L_j^1 + D(f^2)\delta(f^2)I_j^2 L_j^2
\end{aligned}$$

and so the equilibrium profits for merchant i are

$$\pi_i = 2t\left[\frac{1}{2} + \frac{1}{6t}\left(\phi^1\left(I_i^1 L_i^1 - I_j^1 L_j^1\right) + \phi^2\left(I_i^2 L_i^2 - I_j^2 L_j^2\right)\right)\right]^2. \tag{22}$$

Equation (22) implies that a merchant will reject any card i for which $\phi^i < 0$ (equivalently $b_S < l - \beta(f^i)$). If the merchant faces two cards both of which involve $\phi^i \geq 0$ (equivalently $b_S \geq l - \beta(f^i)$), the merchant will (weakly) prefer the card which has a higher ϕ^i. If it accepts both cards and the card with a lower ϕ^i has lower cardholder fees (say card 1), then its equilibrium profits will be determined through the use of card 1, and the merchant will be better off rejecting card 1. A merchant will only not mind accepting a card with lower ϕ^i if the card also has higher card fees, since then consumers will never use the card.

In this case, the merchant's acceptance of the card is an irrelevant decision. The other case in which merchants will accept both cards is when they both offer the same value of $\phi^i \geq 0$. ■

The proposition shows that merchants' decision to accept cards does not just depend on comparing the fees and transactional benefits across the cards. Interestingly, merchants will not necessarily prefer the card scheme with the lowest merchant fee even though consumers hold both cards. Merchants also care about the fees consumers will face from using cards (if cards are beneficial for merchants, they want consumers to use them more), and the benefits consumers obtain from using cards (reflecting the fact that this allows merchants to attract more customers by accepting cards). For example, merchants may prefer to accept American Express over Visa, despite the former generally having a higher merchant fee, if this provides their customers with greater benefits. The increase in demand for their product may exceed the additional merchant fees they face. Similarly, in choosing between two debit cards, merchants may prefer to accept a debit card with a higher interchange fee (higher merchant fees and lower cardholder fees) since then consumers will be incented to use it more, and this may benefit merchants if the alternative is that consumers use a more expensive instrument (say cash). Where both schemes offer the same fee structure, merchants that accept one card will accept both. However, if schemes charge too much, merchants will accept neither.

The implication of the proposition for card scheme competition is dramatic. Identical card schemes will compete by setting their interchange fee to maximize ϕ^i. The equilibrium interchange fee is characterized in the following proposition.

Proposition 8 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and there is no merchant heterogeneity is equal to the transactional benefits to merchants of accepting cards less the costs (and margins) from acquiring.*

Proof. Any scheme which sets an interchange fee that does not maximize ϕ^i can be 'undercut' by another scheme which sets a different fee structure. Schemes that do not maxmize ϕ^i will either be rejected by merchants (if another scheme offers a higher ϕ^i with the same or higher f^i), or will be accepted by merchants but not used by cardholders (if another scheme offers a higher ϕ^i with a lower f^i). The undercutting scheme will attract all the card transactions. As a result, identical card schemes will compete by setting their interchange fee to maximize ϕ^i.

Using (16) and (17), it follows that

$$\frac{d\phi^i}{da^i} = h(f^i)\left(b_S - l + f^i\right)$$

so that the competitive interchange fee for a given industry of type b_S is

$$a^C(b_S) = b_S - c_A - \pi_A.$$

Equation (3) implies f lies in the interior of the interval $[\underline{b}_B, \overline{b}_B]$. Since h is positive over its support, $d\phi^i/da$ is positive to the left of $a^C(b_S)$ and is negative to the right of $a^C(b_S)$, implying $a^C(b_S)$ corresponds to the unique global maximum of ϕ^i. The fact that merchants will be willing to accept cards follows from the fact that at $a^C(b_S)$, $\phi^i = D(f^i)(b_S + \beta(f^i) - l) > D(f^i)(b_S + f^i - l^i) = 0$. ■

In equilibrium, both schemes set their interchange fee at this level, and merchants will accept both schemes' cards. Each card scheme shares equally in the card transactions. These interchange fees can be compared to the interchange fee that maximizes the schemes' joint profits — the interchange fee $a^{\Pi}(b_S)$ defined in equation (8).

Proposition 9 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and there is no merchant heterogeniety leads to an interchange fee lower than that which maximizes the schemes' joint profit (or joint card transactions).*

Proof. Since at $a^C(b_S)$, $f^i\left(a^C(b_S)\right) = l - b_S$ it must be that $f^i(a^C(b_S)) = \beta\left(f^i(a^{\Pi}(b_S))\right)$. Since $f^i < \beta(f^i)$, it follows that $f^i(a^C(b_S)) > f^i(a^{\Pi}(b_S))$ which implies $a^C(b_S) < a^{\Pi}(b_S)$. ■

Thus, competition between card schemes results in an equilibrium in which card schemes are worse off, as there are fewer total card transactions compared to the case without competition. A more interesting comparison is whether the competitive interchange fee is higher or lower than the welfare maximizing interchange fee.

Proposition 10 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and there is no merchant heterogeneity leads to an interchange fee lower than that which maximizes overall welfare.*

Proof. Given (9), if $b_S - c_A + \pi_I \leq a^{\Pi}(b_S)$ then $a^W(b_S) = b_S - c_A + \pi_I$ and $a^C(b_S) - a^W(b_S) = -\pi_I - \pi_A < 0$. If $b_S - c_A + \pi_I > a^{\Pi}(b_S)$ then $a^W(b_S) = a^{\Pi}(b_S)$ and from proposition 5, $a^C(b_S) < a^W(b_S)$. ■

Competing card schemes end up setting an inefficient interchange fee (pricing structure). They charge merchants too little, and cardholders too much. The cause of the inefficiency in the pricing structure is the asymmetry in competition to get cardholders to use cards (with a high interchange fee) versus competition to get merchants to accept cards (with a low interchange fee). In a world in which all consumers hold both cards, merchants control which card is used by choosing which card to accept. Competition drives cards schemes to offer maximal surplus to merchants, in an attempt to have their card accepted exclusively, thus obtaining all card transactions. The resulting distortion in the pricing structure is limited by the fact in a model of Hotelling competition between merchants, merchants internalize their customers' average surplus in deciding whether to accept cards or not. Thus, offering maximal surplus to merchants amounts to maximizing social surplus, aside from the fact merchants ignore the surplus their card acceptance creates for issuers and acquirers (this is the term $\pi_I + \pi_A$).[11]

It is also straightforward then to determine what two identical proprietary schemes will do. Through competition over their fees f^i and m^i, they will compete $f^i + m^i$ down to $c_I + c_A$. Each scheme will then set f^i and m^i to maximize ϕ^i subject to this constraint since if they do not they can be 'undercut' by the other scheme that sets its fee structure to maximize ϕ^i and takes away all their business. This

[11] Note when both merchants accept cards they may get little aggregate benefit from doing so. This reflects their business stealing motive for accepting cards. It is their private incentive to accept cards that causes them to internalize their customers benefits. If instead, an interchange fee is set to only take into account merchants' aggregate interests, it will result in interchange fees being set too low, even assuming $\pi_I = \pi_A = 0$.

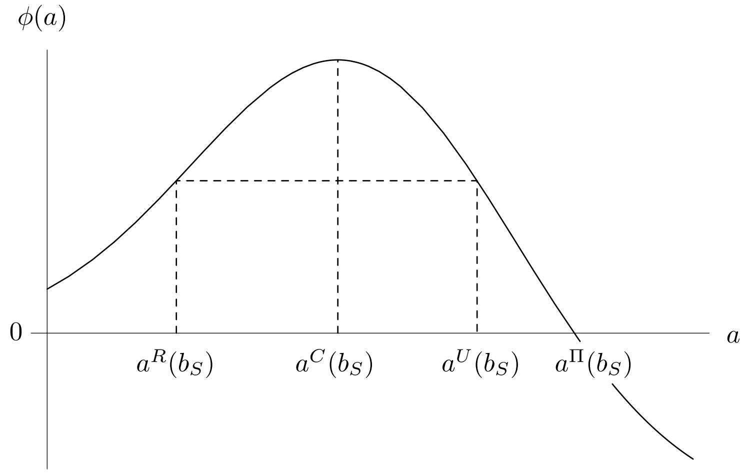

Figure 1: Effect of regulating one payment scheme

is equivalent to choosing f^i to maximize $D(f^i)(b_S + \beta(f^i) - c_I - c_A)$, which implies the same structure of fees for cardholders and merchants as implied by the equilibrium interchange fee set by competing card associations in the limit as $\pi_I = \pi_A \to 0$.[12] The same structure of fees also arises from competition between a proprietary scheme and a card association, taking the limit as $\pi_I = \pi_A \to 0$.

An interesting implication of this symmetric equilibrium is that if one card scheme is forced through regulation to set an interchange fee below the competitive level $a^C(b_S)$, this will lead the rival scheme to take the whole market by setting a higher interchange fee. Any scheme that is left unregulated will not want to match the cut in interchange fees imposed on the regulated schemes. Even if merchants continue to accept cards from the regulated schemes, since consumers now face higher fees for using cards from the regulated scheme, the regulated scheme will not attract any card transactions. Merchants will continue to accept cards from the unregulated scheme even though the merchant fees for the regulated scheme are lower, since the more expensive cards allow them to to attract customers from rivals who do not accept such cards. In fact, an unregulated scheme will actually have an incentive to increase its interchange fee in these circumstances. The reason is, if one scheme is forced to set an uncompetitive (low) fee structure, this gives a rival scheme some scope to also set a less competitive fee structure, but in the direction of increasing the number of its total transactions (and thus, profits). It does this by increasing its interchange fee above $a^C(b_S)$, although not beyond the interchange fee set by a single card scheme $a^\Pi(b_S)$. Thus, if card associations have their interchange fees regulated below the competitive level $a^C(b_S)$, the model predicts proprietary schemes will respond by setting *higher* merchant fees (and lower card fees), and will attract business away from card associations. Figure 1 illustrates this by noting the interchange fee chosen by the unregulated scheme, denoted $a^U(b_S)$, for a particular interchange fee that is imposed on the regulated scheme, denoted $a^R(b_S)$.

The above model provides some insight for competition between on-line debit cards offered directly by banks and off-line debit cards offered by the card associations. On-line debit cards offered by banks

[12]Note in the limit as $\pi_I = \pi_A \to 0$, the equilibrium interchange fee will be efficient provided $a^\Pi(b_S) \geq b_S - c_A$.

and off-line debit cards offered by the card associations are similar instruments both from the perspective of consumers and merchants.[13] Moreover, consumers that have access to debit cards offered by card associations typically also have access to an on-line debit card offered directly by their bank (this may also be their ATM card), suggesting the analysis of this section applies. The above results suggest competition between on-line debit and off-line debit should drive the interchange fees below the level that the schemes prefer, and to the extent of issuer and acquirer margins, below the efficient level. If card associations try to set their preferred pricing structure for off-line debit ignoring the existence of on-line debit (and thus set relatively high interchange fees), merchants will simply reject off-line debit knowing that such consumers will substitute by using on-line debit instead. One way for card associations to prevent this type of destructive competition would be to tie the acceptance of their off-line debit cards to acceptance of credit cards, assuming the interchange fee for credit cards was not subject to the same kind of competitive pressure. This provides one interpretation of MasterCard and Visa's tying behavior, behavior that has resulted in the 'Walmart case'.

Walmart, together with a large number of other merchants, are suing MasterCard and Visa for billions of dollars for tying off-line debit and credit cards together. If merchants accept Visa credit cards they must also accept Visa off-line debit cards. Off-line debit cards have an interchange fee (and merchant fee) close to the levels used for credit cards, and according to Chakravorti and Shah (2003) are about three to five times more expensive for merchants to accept than on-line debit cards. The analysis provides one channel by which it is possible for the card associations' tying behavior to improve welfare. It does this by allowing the scheme to impose a different pricing structure for off-line debit cards than that chosen by banks for on-line debit, a pricing structure which could be more efficient given that competition between schemes can result in interchange fees being set too low. This is especially likely to be the case if banks, in setting their pricing structure for on-line debit card transactions face additional incentives to set low interchange fees, for instance, if banks need to use low merchant fees to incentivise merchants to install costly pinpads.

4.2 Merchant heterogeneity

In this section we extend the above model to allow industries to vary in the transactional benefits they get from accepting cards. This is done exactly as in Section 2.2. An analysis of which merchants accept cards at stage (iii) depends on two cases.

In case 1, both schemes set the same interchange fees (so $\phi \equiv \phi^1 = \phi^2$). Then merchants will accept both cards if $\phi \geq 0$ and reject both cards if $\phi < 0$. Recall from (18) that $\phi = 0$ at $b_S = l - \beta(f)$. With equal interchange fees, consumers randomize over which card to use, and each card scheme receives half the card transactions. Scheme profits are

$$\Pi^1 = \Pi^2 = (\pi_I + \pi_A) \frac{D(f) S\left(l - \beta(f)\right)}{2}.$$

In case 2, scheme 1 sets a lower interchange fee.[14] Then $f^1 > f^2$, $D(f^2) > D(f^1)$, and $\beta(f^1) > \beta(f^2)$.

[13] An on-line debit card requires cardholders enter a pin number while an off-line debit requires a cardholder signature to verify the transaction. This is the meaning of the distinction between on-line and off-line.

[14] The case in which it sets a higher interchange fee follows by symmetry.

Define

$$b(f^1,f^2)=\frac{D(f^2)\beta(f^2)-D(f^1)\beta(f^1)}{D(f^2)-D(f^1)}. \tag{23}$$

For industries with $b_S > l - b(f^1,f^2)$, merchants prefer to accept cards from scheme 2 as $\phi(f^2) > \phi(f^1)$, but merchants accept cards from both schemes as they know consumers will only use cards from scheme 2 (as $f^2 < f^1$). Only merchants with high transactional benefits of accepting cards will be willing to accept the more expensive card knowing consumers will always use it. Merchants with b_S between $l - \beta(f^1)$ and $l - b(f^1,f^2)$ will only accept cards from scheme 1, the cheaper card to accept. Merchants with yet lower transactional benefits of accepting cards will not accept either card. Scheme 1's profit is thus

$$\Pi^1 = (\pi_I + \pi_A)D(f^1)\left[S\left(l-\beta(f^1)\right) - S\left(l-b(f^1,f^2)\right)\right]$$

while scheme 2's profit is

$$\Pi^2 = (\pi_I + \pi_A)D(f^2)S\left(l-b(f^1,f^2)\right).$$

The nature of the equilibrium, if any, depends on the specific distributions on b_B and b_S. A case in which there is an explicit solution is the case in which b_B and b_S are distributed uniformly, as introduced in Section 2.2. Then we get[15]

Proposition 11 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and quasi-demand functions are linear is for both schemes to set* $a_U^C = \frac{2}{3}\left(\overline{b}_S - c_A - \pi_A\right) - \frac{1}{3}\left(\overline{b}_B - c_I - \pi_I\right)$.

Proof. A symmetric equilibrium in which both schemes set the same interchange fee a^* requires that

$$\frac{D(f^*)S\left(l-\beta(f^*)\right)}{2} \geq D(f^1)\left[S\left(l-\beta(f^1)\right) - S\left(l-b(f^1,f^*)\right)\right] \tag{24}$$

for any lower interchange fee (so $f^1 > f^*$) and

$$\frac{D(f^*)S\left(l-\beta(f^*)\right)}{2} \geq D(f^1)S\left(l-b(f^1,f^*)\right) \tag{25}$$

for any higher interchange fee (so $f^1 < f^*$).

With the uniform distribution on b_B we have (11), (13) and so

$$D(f)\beta(f) = \frac{\overline{b}_B^2 - f^2}{2\left(\overline{b}_B - \underline{b}_B\right)}. \tag{26}$$

Substituting (11) and (26) into (23) implies

$$b(f^1,f^2) = \frac{f^1+f^2}{2}.$$

Similarly, given the uniform distribution on b_S we have that

$$S(b_S) = \frac{\overline{b}_S - b_S}{\overline{b}_S - \underline{b}_S}$$

[15] Two identical proprietary schemes or a card association and a proprietary scheme will compete the sum of f and m down to the sum of costs c_I and c_A, and imply the same implicit interchange fee as in proposition 11 where $\pi_I = \pi_A \to 0$.

and so

$$S\left(l-\beta(f^*)\right)=\frac{\overline{b}_S-l+\beta(f^*)}{\overline{b}_S-\underline{b}_S}$$

Now consider the limit of scheme 1's profits as its interchange fee approaches the proposed symmetric equilibrium at a^* from below. This implies $f^1>f^2=f^*$, so that $b(f^1,f^*)\rightarrow f^*$, and

$$D(f^1)\left[S\left(l-\beta(f^1)\right)-S\left(l-b(f^1,f^*)\right)\right]\rightarrow D(f^*)\frac{\left(\overline{b}_B-f^*\right)}{2\left(\overline{b}_S-\underline{b}_S\right)}.$$

Thus, (24) will be violated for a sufficiently small decrease in scheme 1's interchange fee if

$$f^*<\frac{1}{3}\left(\overline{b}_B-2\overline{b}_S+2l\right).$$

Alternatively, consider the limit of scheme 1's profit as its interchange fee approaches the proposed symmetric equilibrium at a^* from above. This implies $f^1<f^2=f^*$, so that $b(f^1,f^*)\rightarrow f^*$, and

$$D(f^1)S\left(l-b(f^1,f^*)\right)\rightarrow D(f^*)\frac{\left(\overline{b}_S-l+f^*\right)}{\overline{b}_S-\underline{b}_S}.$$

Thus, (25) will be violated for a sufficiently small increase in scheme 1's interchange fee if

$$f^*>\frac{1}{3}\left(\overline{b}_B-2\overline{b}_S+2l\right).$$

Thus, the only potential candidate for a symmetric equilibrium is when

$$f^*=\frac{1}{3}\left(\overline{b}_B-2\overline{b}_S+2l\right)$$

which corresponds to the interchange fee

$$a_U^C=\frac{2}{3}\left(\overline{b}_S-c_A-\pi_A\right)-\frac{1}{3}\left(\overline{b}_B-c_I-\pi_I\right). \tag{27}$$

To confirm this is an equilibrium, note that if scheme 1 sets a lower interchange fee, then its card transactions

$$D(f^1)\left[S\left(l-\beta(f^1)\right)-S\left(l-b(f^1,f^*)\right)\right]$$

will decrease since $D(f^1)$ will decrease and $S\left(l-\beta(f^1)\right)-S\left(l-b(f^1,f^*)\right)$ is independent of a^1. Alternatively, if scheme 1 sets a higher interchange fee (so $f^1<f^*$), then its card transactions

$$D(f^1)S\left(l-b(f^1,f^*)\right)$$

will decrease as

$$\begin{aligned}\frac{d\left(D(f^1)S\left(l-b(f^1,f^*)\right)\right)}{da^1} &= \frac{\left(\overline{b}_S-l+\frac{f^1+f^*}{2}\right)}{\left(\overline{b}_B-\underline{b}_B\right)\left(\overline{b}_S-\underline{b}_S\right)}-\frac{\left(\overline{b}_B-f^1\right)}{2\left(\overline{b}_B-\underline{b}_B\right)\left(\overline{b}_S-\underline{b}_S\right)}\\ &= \frac{2\overline{b}_S-2l+2f^1+f^*-\overline{b}_B}{2\left(\overline{b}_B-\underline{b}_B\right)\left(\overline{b}_S-\underline{b}_S\right)}\\ &< \frac{3f^*-\left(\overline{b}_B-2\overline{b}_S+2l\right)}{2\left(\overline{b}_B-\underline{b}_B\right)\left(\overline{b}_S-\underline{b}_S\right)}\\ &= 0.\end{aligned}$$

To see why there are no asymmetric equilibria, note that if scheme 1 sets a lower interchange fee than scheme 2, then as shown above, its card transactions

$$D(f^1)\left[S\left(l-\beta(f^1)\right)-S\left(l-b(f^1,f^2)\right)\right]$$

will decrease since $D(f^1)$ will decrease and $S\left(l-\beta(f^1)\right)-S\left(l-b(f^1,f^2)\right)$ is independent of a^1. This is true regardless of firm 2's interchange fee. Thus, scheme 1 will always want to set its interchange fee as close as possible to scheme 2's interchange fee (that is, an infinitesimal amount less than a^2). Clearly there can be no asymmetric equilibrium. ■

For this case we obtain the same conclusions reached in the simpler case in which merchants are identical. Specifically,

Proposition 12 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and quasi-demand functions are linear leads to an interchange fee lower than that which maximizes the schemes' joint profit (or joint card transactions).*

Proof. Comparing the competitive interchange fee (27) with the interchange fee (14) that maximizes the profit of a single scheme, it follows that $a_U^C < a_U^\Pi$ since $\overline{b}_B + \overline{b}_S > l$ from (10). ■

and

Proposition 13 *The equilibrium interchange fee resulting from competition between identical card schemes when consumers hold both cards and quasi-demand functions are linear leads to an interchange fee lower than that which maximizes overall welfare.*

Proof. Comparing the competitive interchange fee (27) with the interchange fee (15) that maximizes welfare, it follows that

$$a_U^W - a_U^C = \frac{2}{3}\left(\overline{b}_B + \overline{b}_S - c_I - c_A + \pi_I + \pi_A\right) - \frac{2}{3}\sqrt{\left(\overline{b}_B + \overline{b}_S - c_I - c_A\right)^2 + 3(\pi_I + \pi_A)^2}.$$

The expression is positive provided $\overline{b}_B + \overline{b}_S - c_I - c_A > \pi_I + \pi_A > 0$, which is true given (10) and the definitions of π_I and π_A. ■

Competition between two identical card schemes results in interchange fees being set too low from the point of view of the schemes. Interchange fees do not maximize the total number of card transactions, nor overall welfare. With merchant heterogeneity, different fee structures (interchange fees) are preferred by merchants in different industries. In industries with low transactional benefits of accepting cards, merchants will prefer relatively low interchange fees, and in industries with high transactional benefits of accepting cards, merchants will prefer relatively high interchange fees. When both schemes set the same interchange fee, they will share half the card transactions each. In this case, any merchants that accept one card will accept the other card, and consumers will randomize over card usage. Relative to this outcome, each scheme will want to offer a slightly different fee structure if by doing so it will be preferred by more than half the merchants. Thus, schemes compete by attracting merchants, since when consumers hold both cards it is merchants that decide which card will be used. For the case with linear quasi-demand functions, a unique equilibrium exists which involves both schemes setting the same fee structure. As with the case in which merchants are identical, the equilibrium fee structure favors merchants. The extent of bias in the fee structure is limited by the fact competing merchants also internalize their customers' benefits from using cards. However, since merchants do not internalize the surplus that issuers and acquirers obtain, the equilibrium interchange fee still contain some bias towards too low merchant fees relative to card fees.[16]

5 Conclusions

This paper extends existing models of payment systems to allow for competition between payment schemes (such as Visa and American Express). It examines how competition affects the choice of fee structure by card schemes, namely, how much to charge cardholders versus how much to charge merchants. There is a concern by policymakers that consumers face disorted incentives to use credit cards, as a result of low card fees (and other rebates) at the same time as merchant fees are set high.

This paper addresses one implicit assumption behind the policymaker's concern — that a lack of system competition explains why MasterCard and Visa can set high interchange fees, and thus why schemes (including proprietary schemes) set high merchant fees and low card fees. The paper shows that competition between payment schemes does not necessarily result in a more efficient *structure* of fees in payment systems, compared to the case of a single card scheme, even if greater competition does help ensure the overall level of prices is set at the efficient level. Competition can increase or decrease interchange fees.

We uncovered two biases in price structure that arise from competition between payment schemes. When consumers hold multiple cards, merchants tend to reject the more expensive one, causing schemes to compete by attracting merchants. They do this by lowering their interchange fees (or setting more balanced fee structures). Conversely, when consumers just hold a single card, merchants accept both cards, causing schemes to compete by attempting to attract additional cardholders from each other. In this case competition does not cause schemes to set a more balanced fee structure, and if merchant demand remains elastic, inter-system competition results in an even more unbalanced fee structure.

The main implication of our results for policy are two-fold. First, policymakers would be wrong to think the price structures observed in credit card schemes are necessarily the result of a lack of competition between payment card schemes — the analysis here suggests the choice of interchange fees (or fee structure) by a single card association could result in higher or lower interchange fees than those set by competing schemes. Which case prevails depends on the extent to which competition leads to a bias towards cardholders' interests (when consumers only hold one card) or a bias towards merchants' interests (when consumers tend to hold both cards). Second, regulating lower levels of interchange fees for payment card associations (such as MasterCard and Visa) while leaving proprietary card schemes

[16] In the case of two competing identical proprietary schemes there is no equilibrium. If the schemes set fees so that $f^i + m^i = c_I + c_A$, then they will earn zero profits, and one scheme can always do better by setting a different fee structure so it is strictly preferred by some merchants allowing it to set $f^i + m^i > c_I + c_A$. However, each scheme can always do (almost) twice as well as its rival by adopting the same card fee but slightly undercutting on the merchant fee, thereby stealing all card transactions from its rival.

(such as American Express) free to set their price structure will favor the proprietary schemes, resulting in increased card transactions and profit for such schemes at the expense of card associations.

References

Armstrong, M. (2002) "Competition in Two-Sided Markets," mimeo, Nuffield College, University of Oxford.

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

Caillaud, B. and B. Jullien (2001) "Competing Cybermediaries," *European Economic Review*, 45, 797–808.

Chakravorti, S. and A. Shah (2003) "Underlying Incentives in Credit Card Networks" *The Antitrust Bulletin*, forthcoming.

Frankel, A. S. (1998) "Monopoly and Competition in the Supply and Exchange of Money," *Antitrust Law Journal*, 66(2), 313–361.

Gans, J. and S. King, (2002) "The Neutrality of Interchange Fees in Payment Systems," mimeo, University of Melbourne, July 18th.

Hausman, J. and G. Leonard, and J. Tirole. (2003) "On Nonexclusive Membership in Competing Joint Ventures," *Rand Journal of Economics* 34(1).

Parker, G. and M. Van Alstyne (2000) "Information Complements, Substitutes, and Strategic Product Design" Working Paper No. 299, March. Tulane University and University of Michigan.

Rochet, J. and J. Tirole. (2001) "Platform Competition in Two-Sided Markets," mimeo, Toulouse University.

Rochet, J. and J. Tirole. (2002) "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics* 33(4) 549–570.

Schiff, A. (2003) "Open and Closed Systems of Two Sided Networks," *Information, Economics and Policy*, forthcoming.

Schmalensee, R., 2002, "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50, 103–122.

Schwartz, M. and D. Vincent, 2002, "*Same Price, Cash or Card: Vertical Control by Payment Networks*," Working Paper 02-01 Georgetown University.

Wright, J. (2001) "*Determinants of Optimal Interchange Fees in Payment Systems*," University of Auckland Working Paper No. 220.

Wright, J. (2003) "Optimal Card Payment Systems," *European Economic Review*, forthcoming.