Competing Payment Schemes[*]

Graeme Guthrie[†]

Victoria University of Wellington

Julian Wright[‡]

National University of Singapore

February 4, 2006

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 users on one side of the market (merchants) compete to attract users on the other side (consumers who may use cards for purchases). It analyzes how competition between card associations and between merchants affects the choice of interchange fees, and thus the structure of fees charged to cardholders and merchants. Implications for other two-sided markets are discussed.

[†]School of Economics and Finance, Victoria University of Wellington, PO Box 600, Wellington, New Zealand. Email: graeme.guthrie@vuw.ac.nz

[‡]Corresponding author. Department of Economics, Faculty of Arts and Social Sciences, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570. Email: jwright@nus.edu.sg

Competing Payment Schemes

1   Introduction

Policymakers in a number of jurisdictions are concerned that retailers pay too much to accept credit card transactions, costs that in their view are ultimately covered by consumers who pay by other means. To address such concerns, several models of payment schemes have been developed in order to analyze the optimal structure of fees in debit and credit card schemes.[1] They ask: How much will payment schemes charge to cardholders versus merchants for card transactions, and how do these charges compare to the socially optimal structure of fees? However, these papers all assume there is a single payment system. This paper relaxes this assumption.

In considering how competition between payment schemes determines the structure of fees between cardholders and merchants, this paper also falls within the recent literature on two-sided markets.[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. In comparison with the two-sided markets literature, this paper takes into account the specificities of payment schemes. This is important, since of the different examples of two-sided markets, it is the application to payment schemes that has generated the greatest policy interest. In particular, we allow consumers to make a decision about which card(s) to hold as well as which to use, while merchants just decide whether to accept each of the cards. In addition, we allow for the fact that merchants may compete amongst themselves so as to attract consumers, a feature common in a number of other two-sided market settings.

With consumers and merchants being able to choose which card(s) to hold or accept, not surprisingly, there is a plethora of equilibria depending on how consumers and merchants settle on a particular card scheme. We characterize the set of equilibria and the properties of the corresponding range of outcomes. In these equilibria, all transactions are conducted at fees corresponding to a single fee structure. Either both schemes set the same structure of fees, with at least one side involving all agents multihoming, or else one scheme attracts all card transactions exclusively.

We highlight several potential asymmetries between cardholders and merchants that influence the equilibrium structure of fees. A key asymmetry is that merchants accept cards to attract business, in addition to any convenience or transactional benefits they obtain. Even if cardholders and merchants play an otherwise symmetric role, this asymmetry causes platform competition to over-represent the interests of cardholders — once in competing for cardholders and once in competing for merchants who internalize their customers' card benefits. If merchants are homogenous, a single card scheme will already set its fee structure to the point where merchants are just willing to accept cards, so there is no scope for platform competition to result in higher charges to merchants. The model then predicts that platform competition cannot raise the charges to merchants. On the other hand, if merchants (like cardholders) are heterogenous, platform competition can result in merchants being charged more (and cardholders less) reflecting the over-representation of cardholders' interests when merchants accept cards to get more business.

Another potential source of asymmetry between cardholders and merchants is that one side may play

---

[1] Published papers include Gans and King [2003a, 2003b], Rochet and Tirole [2002], Schmalensee [2002], and Wright [2003a, 2003b, 2004], in addition to the earlier work of Baxter [1983].

[2] See, for instance, Armstrong [2006], Armstrong and Wright [2005], Caillaud and Jullien [2001, 2003], Hagiu [2006], Rochet and Tirole [2003], and Schiff [2003].

a greater role in determining which card scheme will be chosen in equilibrium. At one extreme, consumers choose their preferred card to hold knowing merchants will accept both cards in equilibrium and platform competition focuses on attracting cardholders uniquely. This can further bias the competitive fee structure towards cardholders. On the other hand, if cardholders always hold multiple cards, merchants can steer consumers to their preferred card network, meaning platform competition will focus on attracting merchants to accept their cards exclusively. In this case, despite the additional incentive merchants face to accept cards, platform competition can result in lower merchant fees and higher card fees compared to those set by a single scheme, although in our model not lower than those that are socially optimal.

Our work is closest to that of Rochet and Tirole [2002, 2003]. On the one hand, it can be thought of as an extension of Rochet and Tirole [2002] to the case of competing schemes. Rochet and Tirole [2002] analyze the privately and socially optimal interchange fee when there is a single card association, all merchants receive the same benefit from accepting cards and merchants compete to attract business from cardholders. To show our model is comparable to theirs, we start in Section 2 with the same three assumptions. The key difference is that we assume issuers are perfectly competitive, whereas in their model issuers enjoy some market power. Issuer market power can justify higher interchange fees, and reduces the divergence between the privately and socially optimal interchange fee that we find.

Moving to the case with competing card schemes, our model can be viewed as a variation on Rochet and Tirole [2003]. They study competition between differentiated platforms, whereas we restrict attention to the benchmark case of homogenous card schemes. Their model assumes the equivalent of buyers always holding both cards, which generates one extreme of the range of equilibria we study. They also assume sellers that are just like other end-users and so do not consider a seller's incentive to accept cards in order to attract additional business. When these assumptions are relaxed, so that consumers can choose which card(s) to hold and merchants accept cards to get more business, we find equilibrium interchange fees that are higher.

The contribution of our paper is to combine the key conditions considered in each of these papers, allowing for merchants to accept cards for strategic reasons (Rochet and Tirole, 2002) and for competition between platforms (Rochet and Tirole, 2003). The key finding from doing so is that competition between schemes does not generally eliminate the upward bias in privately-set interchange fees caused by merchants accepting cards to expand business. In fact, as we show, inter-system competition can accentuate the over-representation of the interests of cardholders with competing payment schemes sometimes setting interchange fees too high even for their own good.

In Section 3 the main analysis is presented in terms of a model of two competing card associations, such as MasterCard and Visa. Section 4.1 shows the same basic results also apply to the case of competing proprietary schemes, like American Express and Discover, that set their cardholder and merchant fees directly, or between a card association and a proprietary scheme. Initially, we follow Rochet and Tirole [2002] and assume merchants are homogenous with respect to the benefits they obtain from accepting cards. This assumption is relaxed in Section 4.2, where we follow Wright [2004] in allowing merchants in different industries to obtain different benefits from accepting cards. Implications for policy and for the analysis of other two-sided markets are discussed in Section 5, which also provides some concluding remarks.

2 A single card scheme

In this section, we start by reviewing the simpler case of a single payment scheme where the only alternative to using cards is cash. This will allow us to determine the impact of system competition on the equilibrium fee structure by providing a benchmark without system competition. It also allows our analysis to be compared to that of Rochet and Tirole [2002] in the case in which there is only a single payment scheme. One difference between our model and theirs is that in our model consumers are assumed to receive their particular draw of transactional benefits once they have chosen which merchant to purchase from. This assumption is made for modelling convenience. In the case of competing merchants, consumers still choose which merchant to purchase from taking into account the expected benefits of using cards versus cash. By accepting cards, merchants will still raise consumers' expected benefit, since consumers will gain the option of using cards for purchases. In fact, this timing assumption does not alter the equilibrium condition under which merchants accept cards; as we shall see, the condition we derive is equivalent to that derived in Rochet and Tirole [2002].

Initially, we focus on a card association rather than proprietary schemes. We assume the card association is made up of identical issuers (banks and other financial institutions that specialize in servicing cardholders) and identical acquirers (banks and other financial institutions that specialize in servicing merchants). We assume a cost of c_I per transaction of issuing and c_A per transaction of acquiring. Let $c = c_A + c_I$ be the total cost for each card transaction. In such an open scheme, a card association is assumed to set an interchange fee a. The interchange fee is defined as an amount paid from acquirers to issuers per card transaction. Competition between identical issuers results in card fees f equal to

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

while competition between identical acquirers results in merchant fees m equal to

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

Given perfect intra-system competition, members of the association make no economic profit and should be indifferent to the level of the interchange fee. In practice, card associations may still maximize their size (the volume of card transactions). For instance, Visa International, which is a private non-stock for-profit membership corporation, obtains revenues from a small levy on each card transaction which is collected from its more than 21,000 members.[3] This helps fund the activities Visa conducts. As such, Visa executives could well have an incentive to maximize card volume, which increases the revenue available for their activities.

The assumption of identical issuers and acquirers, and that the card scheme maximizes card volume, is made primarily for modeling convenience. Alternatively, we could have allowed for imperfect competition between issuers and assumed that the card scheme maximizes its members' collective profits. In the case of a single scheme, the results of this section can be modified in a straightforward way, as we will explain. In the case of competing identical schemes, provided the sum of issuer and acquirer margins is not affected by interchange fees, profit maximizing card schemes will still maximize card volume. To the extent issuers and acquirers do make positive margins, this can also make the welfare analysis more complicated, and we will discuss how to adjust for this where relevant.

Card fees can be negative to reflect rebates and interest-free benefits that banks offer cardholders based on their card usage. Card fees decrease and merchant fees increase as the interchange fee is increased.

---

[3]In our model, this can be incorporated as part of the costs c_A and c_I that members face per transaction.

The main implication of the above assumption about bank competition is that the level of the interchange fee affects only the structure of fees and not the overall level of fees: the sum of cardholder and merchant fees per-transaction is independent of the interchange fee a (it equals c).

Consumers (whom we will refer to as buyers) get transactional or convenience benefits b_B from using cards as opposed to the alternative cash and merchants (whom we will refer to as sellers) get transactional or convenience benefits b_S from accepting cards relative to the alternative of accepting cash. The benefits b_B are drawn with a positive density $h(b_B)$ over the interval $[\underline{b}_B, \overline{b}_B]$. The hazard function $h(f)/(1 - H(f))$ is assumed to be increasing in f, where H denotes the cumulative distribution function corresponding to b_B. All sellers are assumed to receive the same transactional benefits b_S from accepting cards (but this assumption will be relaxed in Section 4.2). We will refer to $b_B - f$ as the 'surplus' to buyers from using cards and $b_S - m$ as the 'surplus' to sellers from accepting cards. (Note, however, if sellers compete to attract buyers, they will also profit from accepting cards through a business stealing effect.) We assume that

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

so as to rule out the possibility that there is no card use and to rule out the possibility that cards are always used.

The model of competition between sellers follows that in Rochet and Tirole [2002]. It costs sellers d to produce each good and all goods are valued at v by all buyers.[4] There is a measure one of potential buyers (consumers) who wish to buy from two competing sellers. Seller competition is modeled in a standard Hotelling fashion. In particular, consumers are uniformly distributed on the unit interval and the two sellers are located at either extreme. A consumer located at x faces linear transportation costs of tx from purchasing from seller 1 and $t(1 - x)$ from purchasing from seller 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 seller 1 and $i = 2$ corresponds to seller 2. Consumers are assumed to each want to purchase one good (that is, v is assumed to be sufficiently high so that all consumers become buyers). Consumers observe whether each seller accepts cards or not, along with their price, when deciding which store to go to. As we will see, this makes sellers internalize their customers' transactional benefits from using cards when deciding whether to accept cards or not.[5]

The timing of the game is summarized as follows:

(i). The payment card association sets the level of its interchange fee a. Issuers and acquirers then set fees f and m to cardholders and merchants according to (1) and (2).

(ii). Buyers decide whether or not to hold the card. Simultaneously, sellers decide whether or not to accept the card.

(iii). Competing sellers set their retail prices p_1 and p_2. Observing these prices and whether they accept cards, consumers at each location x decide which seller to buy from.

---

[4]Given we allow for competition between merchants, assuming consumers have inelastic demand for goods greatly simplifies the analysis. The assumption is justified by the observation that the impact on retail prices of the choice of payment instruments will normally be very small, so that changes in final demand for goods can be neglected in a first-order approximation.

[5]An earlier version of the paper allowed for the possibility that some fraction of consumers choose which store to shop at ignoring whether the seller accepts card or not. In Guthrie and Wright [2003], we also considered the possibility sellers are monopolists. When consumers have unit demands, these changes reduce the business stealing incentive sellers have to accept cards and so lead to lower equilibrium interchange fees. However, once elastic consumer demand is considered, sellers can attract additional business by accepting cards even if there is no business stealing effect (Farrell, 2005).

(iv). Based on their individual realizations of b_B, buyers decide whether to use the card for payment (if they hold the card) or cash.

Throughout, sellers are assumed to be unable to price discriminate depending on whether buyers use cards or not, so buyers will want to pay with the card if and only if $b_B \geq f$.[6] Using this property, we can define a number of important functions. The quasi-demand for card usage is defined as $D(a) = 1 - H(f(a))$, which is the proportion of buyers who want to use cards at the fee $f(a)$. The expected convenience benefit to those buyers using cards for a transaction is $\beta_B(a) = E[b_B | b_B \geq f(a)]$, which is decreasing in a. Since $\beta_B(a) - f(a)$ is the expected surplus to a buyer from using a card conditional on the buyer wanting to use a card, the expected transactional surplus to a buyer from being able to use their card at a seller is

$$\phi_B(a) = D(a)\left(\beta_B(a) - f(a)\right),$$

which is positive and increasing in a. The expected transactional surplus to a seller from being able to accept cards from a buyer is

$$\phi_S(a) = D(a)(b_S - m(a)).$$

The expected joint transactional surplus for a buyer and seller where buyers can use a payment card is defined as

$$\phi(a) = D(a)(\beta_B(a) + b_S - c),$$

which, given that $f + m = c$, equals $\phi_B + \phi_S$.

The following lemma summarizes some useful properties of these functions and introduces three important levels of the interchange fee.

Lemma 1

1. There exists a unique interchange fee,

$$a_W = b_S - c_A, \tag{4}$$

that maximizes ϕ. It is also the unique interchange fee that solves $\phi_S = 0$.

2. There exist unique interchange fees, denoted $\underline{a}$ and $\overline{a}$, that maximize ϕ_S and solve $\phi = 0$ respectively.

3. The interchange fees satisfy $\underline{a} < a_W < \overline{a}$.

Proof. See the appendix. ∎

The results of this lemma are summarized in Figure 1. The range of interchange fees that can arise in equilibrium in this paper will turn out to lie between $\underline{a}$ and $\overline{a}$, with $\underline{a}$ maximizing sellers' transactional surplus and $\overline{a}$ being the highest interchange fee at which sellers are still willing to accept cards. These can be thought of as the lower and upper bounds for interchange fees of interest. Expected joint transactional surplus ϕ is maximized when buyers only use cards if and only if $b_B + b_S \geq f + m$.[7] Since buyers wish to use cards whenever $b_B \geq f$, this requires $b_S = m$, which defines the interchange fee a_W. This is also the interchange fee for which $\phi_S = 0$, since at $m = b_S$, sellers get no transactional surplus from accepting cards. Externalities arising from the buyer's choice of paying by card are fully internalized by buyers at the point at which sellers remain indifferent over the buyer's choice.

---

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

[7]Given our assumption that $f + m = c$, this is also the condition for maximizing welfare.

Figure 1: Important interchange fees
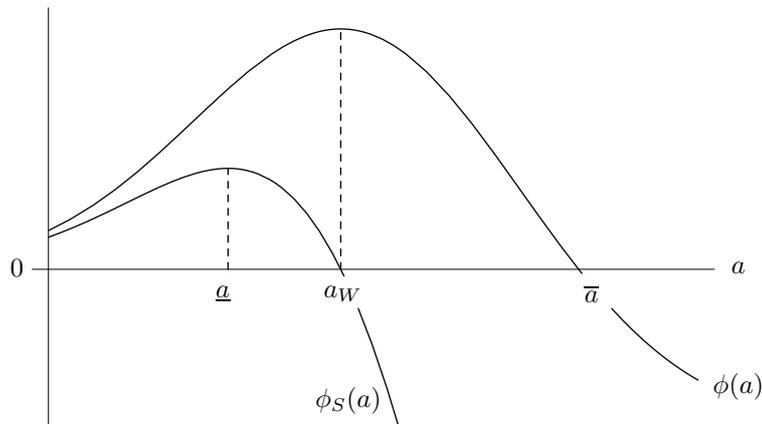
Notes. The interchange fee $\underline{a}$ maximizes $\phi_S(a)$, the expected transaction surplus to a seller from being able to accept cards, and a_W maximizes $\phi(a)$, the expected joint transaction surplus to buyers and sellers from card usage. $\phi_S(a) = 0$ at the interchange fee a_W and $\phi(a) = 0$ at $\overline{a}$.

We assume buyers and sellers do not face any other costs, benefits, or fees from holding or accepting cards. Since there is no cost of holding a card, but holding a card provides consumers with a valuable option if some sellers accept the card, consumers will all want to hold the card in the case of a single card scheme. The same is not necessarily true of sellers, since accepting a card requires that they accept the cards of all buyers, which may not be profitable for high merchant fees.

We start by characterizing the equilibrium in stage (ii) of the game, in which buyers and sellers decide whether to hold and accept cards. In the case of a single card scheme, this is relatively straightforward. Given sellers set the same price regardless of how buyers pay, if sellers did not accept cards for strategic reasons, their additional profit from card acceptance will be just their expected transactional surplus ϕ_S. Since consumers choose the seller to purchase from taking into account their policy on card acceptance, each seller will have an additional incentive to accept cards, which arises from its ability to raise price or gain market share. In deciding whether to accept a payment card, sellers will consider not only their transactional surplus ϕ_S, but also their customers' surplus ϕ_B from being able to use cards. This suggests that an individual seller's decision to accept cards will depend on $\phi_S + \phi_B$, which is the function ϕ. Proposition 1 shows this conjecture is correct. There is also a trivial equilibrium in stage (ii), in which even though $\phi \geq 0$, buyers do not hold cards since sellers do not accept them and vice-versa (the "chicken-and-egg problem"). To avoid such outcomes, in this paper we only consider Pareto undominated equilibria in the stage (ii) subgame, in which buyers and a single seller cannot both be made better off (with one side strictly better off) by moving to another equilibrium given the fees they face.[8]

Proposition 1 *Sellers will accept cards if and only if $\phi \geq 0$.*

---

[8]When two competing sellers jointly accept cards, their joint profit from doing so is determined by ϕ_S rather than ϕ since they internalize the business stealing effect between them. Our definition of Pareto undominated equilibria assumes sellers are not able to coordinate to eliminate this business stealing incentive for accepting cards.

Proof. See the appendix. ∎

The result implies a seller's incentive to accept cards depends on the joint transactional surplus that they and their customers attain. The card scheme maximizes the volume of card transactions, which equals D if $\phi \geq 0$ and is zero otherwise. Since D is positive and increasing in a, this implies the scheme will raise a until ϕ is driven to zero. We denote the resulting interchange fee a_M, indicating it is the interchange fee that a single (monopoly) card scheme will set. This implies that

Proposition 2 *A single card scheme sets its interchange fee so that sellers obtain no profit from accepting cards; that is, $a_M = \overline{a}$.*

The card scheme maximizes the volume of card transactions by increasing interchange fees to the point where sellers are just willing to accept cards. This is the interchange fee that maximizes the surplus of buyers from having cards, taking into account the participation constraint of sellers. If sellers did not internalize any of their customers' surplus from using cards, this would occur when the fee charged to sellers equals the transactional benefits they obtain; that is, when $b_S = m$. This ensures $b_B + b_S \geq f + m$ if and only if $b_B \geq f$, so that the buyers' choice of card usage would correspond to the joint surplus maximizing usage of cards. Since $f + m = c$, this is also the socially optimal level of the interchange fee a_W, so that buyers use cards if and only if $b_B + b_S > c$.[9] Given sellers instead internalize their customers' surplus from using cards, they will be willing to pay more to accept cards. This allows a single card scheme to set a higher interchange fee, so as to promote greater card usage. However, a higher interchange fee means buyers, facing lower fees, will sometimes use cards even though $b_B + b_S < f + m$. The result is over-usage of cards. As shown in Proposition 3, this implies the privately set interchange fee a_M exceeds the welfare-maximizing level a_W.

Proposition 3 *A single card scheme sets its interchange fee too high.*

Proof. From Proposition 2, $a_M = \overline{a}$. From Lemma 1, $\overline{a} > a_W$. Together these results imply $a_M > a_W$. ∎

These results are consistent with those obtained in Proposition 3 in Rochet and Tirole [2002]. The only difference is that in their model, with positive issuer margins, it is possible for the socially optimal interchange fee and privately set interchange fee to be equal. With positive issuer margins, cardholders will face higher fees and will not use cards enough unless the interchange fee is set above a_W. This calls for a higher interchange fee, but never higher than $\overline{a}$ given that sellers also have to be willing to accept cards. Having established our benchmark case is consistent with existing results, we next turn to extending the framework to incorporate competition between payment schemes.

3 Competition between identical card schemes

In this section, we modify the model of Section 2 by assuming there are two identical competing card systems. Identical systems not only have the same costs but they also provide the same benefits to buyers and sellers. The only distinguishing feature of each card scheme is the fee structure it chooses. Specifically, each card association $i = 1, 2$ sets an interchange fee denoted a^i.

---

[9]This is also the Baxter interchange fee. Like us, Baxter [1983] assumes away issuer and acquirer margins.

Like the case with a single scheme, we assume that for competing card associations

$$f^i(a^i) = c_I - a^i \qquad (5)$$

and

$$m^i(a^i) = c_A + a^i, \qquad (6)$$

so that the interchange fee determines the structure but not the overall level of fees. Thus, our model is one in which there is perfect intra-system, as well as inter-system, competition. It is assumed each scheme seeks to maximize the volume of its card transactions. Rochet and Tirole [2003] make a similar simplifying assumption when they consider competing associations in their framework.[10] A further motivation for this approach is that, as is shown in Section 4.1, it allows us to recover the equilibrium fees that result from competition between two identical (profit-maximizing) proprietary schemes, which set the fees f^i and m^i directly, or between one proprietary scheme and one card association. As before, the sum of f and m is the constant c.

The timing of the game is the same as before:

(i). Each payment card association sets the level of their interchange fee a^i. Issuers and acquirers then set fees f^i and m^i to cardholders and merchants according to (5) and (6).

(ii). Buyers decide which cards to hold (neither, one or both). Simultaneously, sellers decide whether to accept cards (neither, one or both).

(iii). Sellers set their retail prices. Observing these prices and whether they accept cards, consumers at each location x decide which seller to buy from.

(iv). Based on their individual realizations of b_B, buyers decide whether to use a card or cash for payment, and if holding multiple cards, which card to use.

Recall that in our model, buyers are assumed to not face any other costs, benefits or fees from holding cards. In an earlier version of the paper (Guthrie and Wright, 2003) we allowed buyers to face a fixed cost or intrinsic benefit of holding cards, so that some buyers disliked holding cards while others valued holding cards. The existence of buyers who will hold cards regardless of which cards sellers accept will mean that by only accepting the card with the lower interchange fee, sellers can steer card transactions to their preferred card. In equilibrium, the interchange fee must ensure sellers get at least this surplus otherwise another scheme can lower the interchange fee and attract all sellers exclusively. On the other hand, buyers who just want to hold a single card will drive card schemes to offer higher interchange fees in order to attract such buyers to hold their card exclusively.[11] Such scenarios are captured in our model by allowing the possibility that when buyers are indifferent between holding a card or not, they may either hold the card or not.

Taking into account that buyers who only hold card i (singlehome) will only be able to use card i, while buyers who hold both cards (multihome) will use the card that has the lower fee whenever sellers accept this card, implies the profit a seller gets from accepting card i now depends on whether it also accepts the other card, as well as which card has the lower fee, and which card(s) the buyers hold. This makes characterizing the equilibria in stage (ii) more complicated compared to the case with a single card

---

[10]We do not deal with issues of duality, in which banks may be members of both card associations. For a specific analysis of duality, see Hausman et al. [2003].

[11]Similar effects may arise when issuers charge annual fees.

scheme. There is the possibility of multiple consistent demand configurations (equilibria) in the stage (ii) subgame. In this case, we make the weak restriction, as before, that buyers and sellers settle only on Pareto undominated demand configurations (configurations that cannot be beaten by buyers and an individual seller). This only rules out outcomes involving very low or very high interchange fees, at which both buyers and individual sellers would prefer to switch to a card scheme with a more moderate fee structure. Using the remaining consistent demands, the following proposition then characterizes equilibria of the full game.

Proposition 4 *In any equilibrium, all card transactions occur at a single card fee corresponding to a single interchange fee that lies in the interval $[a_W, \overline{a}]$. Either both schemes set the same interchange fee, and they share card transactions with all agents on at least one side multihoming, or one scheme attracts all card transactions exclusively at this interchange fee.*

Proof. See the appendix. ■

Proposition 4 derives properties of equilibrium interchange fees. The proof is long and relegated to the appendix. Our approach is to first characterize equilibria in buyers' and sellers' decisions at stage (ii). Given it is costless to do so, buyers will hold card i whenever they expect some sellers will exclusively accept card i. They will also hold card i whenever sellers accept both cards and card i has a higher interchange fee (lower card fees). Otherwise, they will be indifferent about holding card i given they also hold card j. In contrast, accepting cards is a strategic decision for sellers. Since sellers accept cards to attract business, as in Proposition 1, each individual seller's decision to accept cards depends on ϕ rather than ϕ_S. Moreover, sellers compare the joint surplus created if they accept one card exclusively (which will involve some transactions being lost when not all buyers hold the accepted card) with the case they accept both cards (which may involve some buyers using the scheme with lower ϕ). With feasible and Pareto undominated demand configurations specified, we are able to rule out certain equilibria for the full game and characterize those that remain.[12]

An implication of the above proposition is that we only have to discuss a single competitive interchange fee, which we denote a_C. Even though there can be equilibria in which the schemes set different interchange fees, in these equilibria all card transactions will occur on a single scheme and so the interchange fee of this scheme is the one of interest. Interestingly, the features of the equilibrium implied by Proposition 4 appear consistent with pricing by the two main rival card associations in practice. The interchange fees of MasterCard and Visa are known to be very similar. Moreover, any retailer that accepts one of these cards also accepts the other, so that there is indeed full multihoming on one side (the seller side), an implication of Proposition 4 given the schemes share card transactions.

To pinpoint exactly where the competitive interchange fee a_C lies within the range of possible equilibrium interchange fees $[a_W, \overline{a}]$ is complicated by the fact that there are multiple consistent (Pareto undominated) equilibrium demand configurations in the stage (ii) subgame for any particular combination of interchange fees. For instance, if scheme 1 sets a relatively high interchange fee and scheme 2 sets a relatively low interchange fee, both within the range $[a_W, \overline{a}]$, then there is an equilibrium in the stage (ii) subgame where all buyers hold both cards and sellers exclusively accept the card from scheme 2, and

---

[12]For example, we can rule out an equilibrium where the schemes both attract transactions but at different interchange fees. For this to happen it must be that not all merchants accept both cards, otherwise cardholders would always hold and use the card with the higher interchange fee. Instead, it must be that there are some merchants that just accept the card with the lower interchange fee while others accept the card with the higher interchange fee. But then, all buyers will choose to hold both cards, which implies sellers are better off only accepting the card which generates a higher value of ϕ.

there is also an equilibrium where all buyers exclusively hold the card from scheme 1 and sellers either accept the same card or both cards. We need some rule to select the unique demand configuration from the set of possible equilibrium configurations.[13] The following two results show that each of the extremes of this range can arise as equilibria depending on how demands are determined in the stage (ii) subgame.

Corollary 1 *If buyers always hold both cards whenever it is an equilibrium for them to do so, competing card schemes set interchange fees at the socially optimal level ($a_C = a_W$).*

Proof. We start by noting that, given it is costless for buyers to hold both cards, it is always an equilibrium for them to choose to hold both cards in the stage (ii) subgame. Now suppose scheme 2 sets its interchange fee equal to a_W in stage (i). If scheme 1 sets its interchange fee at the same level, all buyers multihome and possibly some sellers multihome as well. In contrast, if scheme 1 sets its fee at a level different from a_W, sellers will not accept its cards: they can reject scheme 1's cards and, because buyers multihome, steer all card transactions onto scheme 2 (which gives them a higher surplus). Thus, scheme 1 cannot do better than setting its interchange fee equal to a_W, proving that $a_C = a_W$ is an equilibrium interchange fee.

Moreover, $a_C = a_W$ is the only equilibrium interchange fee that can arise. If $a_C > a_W$, a scheme can set a slightly lower interchange fee and attract all sellers exclusively. If $a_C < a_W$, it can set a slightly higher interchange fee and attract all sellers exclusively. In either case, a_C cannot be an equilibrium interchange fee. ∎

Quite plausibly, however, some buyers may not hold a second card when doing so gives them no additional benefit.[14] This opens up the possibility of the full range of other equilibria between a_W and $\overline{a}$. For instance, at the other extreme, we get[15]

Corollary 2 *If sellers always accept both cards whenever it is an equilibrium for them to do so, competing card schemes set interchange fees at the same level as a single card scheme ($a_C = a_M$).*

Proof. Recalling $a_M = \overline{a}$ from Proposition 2, we have to show $a_C = \overline{a}$. Suppose scheme 2 sets its interchange fee equal to $\overline{a}$. If scheme 1 sets its interchange fee at the same level, all sellers multihome (given it is an equilibrium to do so) and possibly some buyers multihome as well. If scheme 1 sets its fee at a level lower than $\overline{a}$, there is a unique equilibrium in the stage (ii) subgame in which sellers multihome. This involves buyers holding only scheme 2's cards (if any buyers hold card 1, sellers strictly prefer to only accept card 1). Given the equilibrium selection rule specified in the corollary, this equilibrium with multihoming sellers will be selected, implying scheme 1 attracts no demand. Finally, if scheme 1 sets its fee at a higher level than $\overline{a}$, clearly it attracts no demand since sellers will not accept its cards. Thus, scheme 1 cannot do better than setting its interchange fee equal to $\overline{a}$, proving that $a_C = \overline{a}$ is an equilibrium interchange fee.

---

[13]In the literature on network effects such rules are often interpreted as defining the expectations that buyers and sellers have.

[14]In reality, consumers may perceive some small cost to holding each card. Similarly, if issuers face per-customer costs, then competition between them may generate corresponding per-customer annual fees, which will make buyers prefer to hold just a single card.

[15]The result that follows also arises if we assume expectations "stubbornly favor one firm" (Farrell and Katz, 1998). This means each buyer and seller expects all others to join (exclusively) one of the card schemes whenever it is an equilibrium for them to do so in stage (ii). If scheme 1 is the favored scheme, scheme 1 will set its interchange fee as if it faces no competition.

It cannot be an equilibrium for the competitive interchange fee to be lower than $\overline{a}$, since then one scheme can raise its interchange fee to $\overline{a}$ and attract all demand (given it is then an equilibrium in the stage (ii) subgame for sellers to accept both cards and buyers to hold only the card with the higher interchange fee). Clearly, it is also not an equilibrium for the competitive interchange fee to be higher than $\overline{a}$, since sellers will then not accept cards. ∎

Between these two extremes, any interchange fee within $[a_W, \overline{a}]$ can also be supported as an equilibrium, reflecting the extent to which the interests of buyers versus sellers determine how card schemes get chosen given the multiple consistent demand configurations at stage (ii) of the game. For instance, if both buyers and sellers expect all agents to join (exclusively) the scheme that offers the higher value of $\omega \phi_B + (1 - \omega) \phi$ whenever it is an equilibrium for sellers to accept the scheme's cards, competition will lead schemes to set their interchange fees to maximize this function subject to $\phi \geq 0$. No card scheme can attract any demand if it increases or decreases its interchange fee relative to this level, while at any other interchange fee that does not maximize this function, one scheme can do better by offering a higher value of $\omega \phi_B + (1 - \omega) \phi$ and attracting all card transactions exclusively. Given ϕ_B captures buyers' interests when sellers always multihome and ϕ captures sellers' interests when buyers always multihome, this modeling approach provides a convenient way of characterizing the full range of possible equilibria, as we do in the following proposition.

Proposition 5 *If the equilibrium interchange fee a_C maximizes $\omega \phi_B + (1 - \omega) \phi$ subject to $\phi \geq 0$, we have:*

1. *For every $\omega \in [0, 1]$, the interchange fee a_C exists.*

2. *If $\omega = 0$, competing card schemes set interchange fees that are socially optimal ($a_C = a_W$); if $\omega > 0$, competing card schemes set interchange fees that are too high ($a_C > a_W$); and if $\omega = 1$, competing card schemes set the same interchange fee as a single card scheme ($a_C = a_M$).*

3. *If ω is less than $\omega^* = -\beta_B'(\overline{a})/(1 - \beta_B'(\overline{a})) > 0$, competing card schemes set lower interchange fees than a single card scheme ($a_C < a_M$).*

Proof. See the appendix. ∎

Whichever equilibrium is selected in practice, competition between payment schemes can never increase the equilibrium interchange fee (when sellers obtain identical benefits from accepting cards), reflecting that a single card scheme already sets interchange fees to the highest possible level consistent with sellers accepting cards.[16] As Proposition 5 shows, competition can lower interchange fees if sellers' interests are sufficiently important in determining the choice of payment scheme. However, in such cases, competition never lowers interchange fees below the welfare maximizing level. This reflects that sellers fully internalize the surplus of buyers, so that buyers' interests are over-represented, appearing both on the buyers' and sellers' side. In the extreme case where buyers always multihome, competition between payment schemes lowers interchange fees all the way to the socially optimal level a_W. At the other extreme, when sellers multihome (whenever doing so is an equilibrium), competition between payment schemes does not change interchange fees at all, which remain stuck at the monopoly level a_M. If reality is something in between these two extremes, interchange fees may or may not decrease with competition, but competitive interchange fees remain too high. As a result, competition between card schemes can

---

[16]When we relax the assumption of homogenous merchants in Section 4.2, this will no longer be true.

mitigate the bias against sellers resulting from buyers' interests being over-represented, but this need not be the case, and except in one extreme case, the resulting competitive interchange fee still contains some upward bias.[17]

4   Extensions

In Section 4.1, we show that the results of competition between two card associations also apply to competition between two identical proprietary schemes such as American Express and Discover, or one proprietary scheme and one card association. We then use this result to explore the implications of the asymmetric regulation of card schemes currently proposed by authorities. Section 4.2 extends the analysis of Section 3 to handle merchant heterogeneity.

4.1   Competition with proprietary schemes

In Section 3, we considered the case of competition between card associations, each of which sets an interchange fee to achieve its desired fee structure. It is also straightforward to determine what two competing proprietary schemes, which set the fees f and m directly, will do.

The profit of a proprietary card scheme i is

$$\Pi^i = (f^i + m^i - c)T^i, \tag{7}$$

where T^i is the number of card transactions on system i. Given identical schemes, homogenous Bertrand-type competition between them means that the results from the previous section directly apply to two competing proprietary schemes. The analysis from stage (ii) onwards is the same as before since at stage (ii) the fees f^i and m^i can be taken as given. At stage (i), any equilibrium involving competition between identical proprietary schemes will involve the sum of their fees $f + m$ being driven down to cost c. If a proprietary scheme tries to set its fees such that $f + m$ is above c, it will be possible for the rival scheme to set fees that are just slightly lower and attract all buyers and sellers exclusively, given that we assumed that buyers and sellers select only Pareto undominated demand configurations. Thus, of all possible fee structures and demand configurations, only those characterized in Proposition 4 can be equilibria. The equilibrium fees in the case of competing proprietary schemes then correspond exactly to those implied by the interchange fee resulting from competing card associations. For instance, if the competitive interchange fee is a_C, the competitive fee structure with competition between identical proprietary schemes is just $f^1 = f^2 = c_I - a_C$ and $m^1 = m^2 = c_A + a_C$.

The same logic also applies to competition between a proprietary scheme and an identical card association. If they both have identical costs and the proprietary scheme tries to set fees such that $f + m$ is above c, then given $f + m = c$ for the card association, it can always set an interchange fee so that it is strictly preferred by both buyers and sellers. This causes the proprietary scheme to set fees such that $f + m = c$ if it is to get any demand. Thus, the results of Section 3 continue to hold.

An important application of this model is to address what happens when, as is currently the case, regulations are imposed on card associations but not on the proprietary schemes, which set fees to cardholders and merchants directly. For instance, in 2003 the Reserve Bank of Australia regulated

---

[17]This result assumes issuers and acquirers are perfectly competitive. If issuers and acquirers retain positive margins, these will result in higher card fees, thereby reducing any overusage of cards. As we show in Guthrie and Wright [2003], this can then cause competing schemes to set interchange fees too low.

Figure 2: Effect of regulating one payment scheme

(a) $a_C = a_W$ (b) $a_C = a_M$
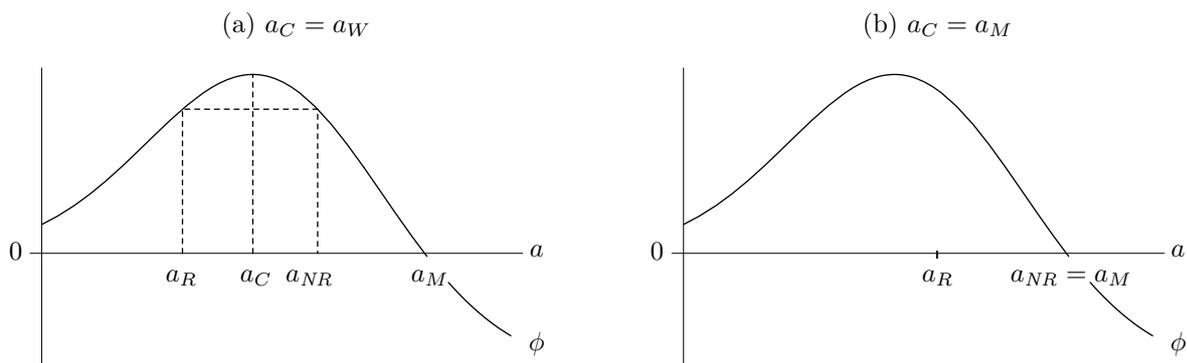
Notes. The equilibrium interchange fee in an unfettered market is a_C. The regulated interchange fee is a_R. The scheme that is not regulated will respond by setting its interchange fee at a_{NR}.

substantial reductions in the interchange fees of card associations that explicitly leave proprietary schemes free to set their fees. Since proprietary schemes do not have to set interchange fees to achieve their desired price structure, any regulation of interchange fees could act as a potential handicap to card associations.

The consequences of any asymmetric approach to regulation can be examined in the context of our model by starting at a competitive interchange fee a_C, which, from Section 3, lies within the range $[a_W, \overline{a}]$. To illustrate the range of possible outcomes, we consider each of the two extreme ways in which competitive interchange fees can be determined, corresponding to Corollaries 1 and 2. In the first case, where buyers always multihome, the competitive interchange fee equals a_W. Suppose regulators require one of the schemes to set its interchange fee at $a_R < a_W$. This could be because regulators erroneously think the competitive interchange fee is above the socially optimal level or because they seek to reduce the number of card transactions for other reasons. Given buyers continue to multihome, sellers will exclusively accept cards of the unregulated scheme if it does not change its fees. However, since the regulated scheme offers both buyers and sellers lower surplus, it relaxes the competitive constraint on the unregulated scheme, which will therefore want to change its fee structure.

If the unregulated scheme is a card association, it will respond by raising its interchange fee, which allows it to attract more card transactions. The result of the regulation will be that sellers pay more, buyers pay less and there will be greater use of cards. Figure 2(a) illustrates this situation, showing that card transactions now occur at a higher than optimal interchange fee and welfare (as indicated by ϕ) will be lower.

If the unregulated scheme is a proprietary scheme, it will respond by increasing the total amount charged to buyers and sellers together, enabling it to obtain higher profits. Specifically, the proprietary scheme will choose f_U and m_U to maximize $(f_U + m_U - c) D(f_U)$ subject to sellers being willing to accept its cards exclusively. Denoting the card fee corresponding to the regulated interchange fee a_R by f_R, the relevant constraint is that

$$D(f_U)(\beta_B(f_U) + b_S - f_U - m_U) > D(f_R)(\beta_B(f_R) + b_S - c). \tag{8}$$

The proprietary scheme will set f_U to the same level as prior to the regulation, which is $f_U = c_I - a_W$. This maximizes the sellers' incentive to accept its cards exclusively, which the proprietary scheme then exploits by setting a higher merchant fee — the highest merchant fee such that (8) still holds. The

13

interchange fee regulation will therefore have no effect on the fees buyers pay or cards usage, but instead will just raise the amount sellers pay. The unregulated scheme ends up taking the whole market.

When the competitive equilibrium is determined instead by sellers multihoming whenever it is an equilibrium for them to do so, Corollary 2 implies interchange fees will be too high at a_M. If the interchange fee is regulated to some lower level a_R where $a_W \leq a_R < a_M$, the response of the unregulated scheme will depend on whether it sets the fees f and m directly or not. In the case it is a card association, and so only sets an interchange fee, it cannot do better than to leave its interchange fee unchanged at a_M. Sellers will continue to multihome at this interchange fee, while buyers will switch to holding only the cards of the unregulated scheme given it sets a higher interchange fee. The unregulated card association will end up taking the whole market at the initial interchange fee a_M. Total welfare remains unchanged. Figure 2(b) illustrates this situation.

If the unregulated scheme is a proprietary scheme, it will choose f_U and m_U in order to maximize $(f_U + m_U - c) D (f_U)$ subject to sellers getting a non-negative surplus from accepting its cards and buyers preferring its cards to those of the regulated scheme (given sellers accept both). This implies it will follow the regulated scheme and increase its charge to buyers, charging buyers just slightly less than f_R so they strictly prefer to hold its cards. However, it will not lower fees to sellers by the corresponding amount. Given $f + m = c$ for a card association, the regulated scheme's merchant fees will decrease as a result of a lower interchange fee by the same amount its card fees will increase. In contrast, the unregulated proprietary scheme will set $m_U = \beta_B (f_U) - f_U + b_S$, the level of fees at which sellers are just willing to accept cards. Given $\beta_B (f_U)$ is increasing in f_U, the unregulated proprietary scheme will not lower fees to sellers by the full amount of its increase in f_U.[18] Thus, the proprietary scheme profits from the regulation, obtaining all card transactions and a positive profit. The number of card transactions and the surplus of buyers is lowered, while sellers are no better off given they always obtain the normal Hotelling profit of $t/2$ in equilibrium. Here the regulation increases total welfare by limiting the overusage of cards, although it is the unregulated proprietary scheme that captures the increase in surplus.

4.2   Merchant heterogeneity

In this section, we extend our model to handle merchant heterogeneity. A key simplifying assumption made up till now is that sellers obtain identical transactional surplus from accepting cards. This meant seller demand was inelastic up to some critical point, whereas buyer demand was assumed to be elastic over the full range of fees. Elastic seller demand for card acceptance arises naturally if sellers are heterogeneous with respect to their transactional surplus from accepting cards.

To model merchant heterogeneity, we follow Wright [2004] and assume there are many industries, each of which has a different value of b_S (in some industries, being able to accept cards is more useful than in others). 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. Apart from this form of heterogeneity, industries are all alike. Within each industry there are two sellers (both with the same draw of b_S) that compete as described in Section 3. The particular draws of b_S are assumed to be known to merchants but unobserved by the card schemes, and are assumed to be independent of all the other random variables. Consumers are exogenously matched to all the different industries and so, without loss of generality, they buy one good from each industry. The sellers' "quasi-demand" function, which measures the proportion of sellers with transaction benefits above some level b_S, is denoted $S(b_S) = 1 - G(b_S)$. The timing is the same as in the

---

[18]For instance, if b_B is uniformly distributed, m_U only decreases by half of the amount by which f_U is increased.

benchmark model.

We start by reconsidering the case with a single card scheme. For any given industry, the decision of sellers at stage (ii) is the same as before, so from the analysis in Proposition 1, sellers of type b_S will reject the card if $\phi < 0$ and will accept a card if $\phi \geq 0$. This implies sellers in industries with $b_S \geq b_S^m \equiv m - (\beta_B - f) = c - \beta_B$ will accept cards and others will not. There are therefore $S(b_S^m) = 1 - G(b_S^m)$ industries that accept cards. Analogous to β_B, we also define $\beta_S(a) = E[b_S | b_S \geq b_S^m(a)]$. A consumer's benefit from holding the card is then $\phi_B S$, since a consumer gets an expected surplus of ϕ_B from using their card at each seller that accepts cards, consumers buy one good from each industry and there are S industries that accept cards.

A single card scheme sets a_M to maximize the volume of card transactions, which is $T = DS$. Recalling that cards are assumed to be free to hold (but have a positive option value), we know all buyers will hold a card in any equilibrium. Buyers will use their card a proportion D of the time at a measure S of sellers. The corresponding first order condition is

$$S\frac{dD}{da} + D\frac{dS}{da} = 0.$$

At the margin, the output maximizing interchange fee balances the increase in consumer usage of cards resulting from lower card fees (this has to be multiplied by the proportion of sellers that accept cards to obtain the impact on total demand) with the decrease in seller demand for accepting cards resulting from higher merchant fees (this has to be multiplied by the proportion of times consumers use cards to obtain the impact on total demand). As Wright [2004] shows, this is also the profit maximizing interchange fee even if issuers and acquirers retain some profits, provided the pass through of costs to end users is the same on both sides of the market. (For instance, this condition holds if the issuing and acquiring sides are symmetric.)

The comparison with the socially optimal interchange fee is more subtle, and is studied in detail in Wright [2004]. Welfare can be written as

$$W = \int_f^{\overline{b}_B} \int_{b_S^m}^{\overline{b}_S} (b_B + b_S - c) g(b_S) h(b_B) \, db_S db_B = (\beta_B + \beta_S - c) DS.$$

To obtain specific results, we assume b_B and b_S are distributed according to the uniform distributions on $[\underline{b}_B, \overline{b}_B]$ and $[\underline{b}_S, \overline{b}_S]$ respectively. Then $b_S^m = c - (\overline{b}_B + f)/2$, so that

$$a_M = \overline{b}_S - c_A.$$

In comparison, the interchange fee that maximizes welfare W is

$$a_W = a_M - \frac{\Delta}{3}, \tag{9}$$

where $\Delta = \overline{b}_B + \overline{b}_S - c$, which is positive (otherwise cards can never provide a positive surplus). Clearly, $a_M > a_W$. Thus, at least with linear demands, a single card scheme sets the interchange fee too high, which reflects the fact sellers accept cards to attract additional business.[19] Thus, Proposition 3 continues to hold in this setting.

Proposition 6 *Assume b_B and b_S are uniformly distributed. A single card scheme sets its interchange fee too high.*

---

[19]When sellers do not accept cards to attract additional business, b_S^m becomes m, and so $a_M = a_W$.

Our interest is in how system competition alters this conclusion. With heterogeneous sellers, there is no longer any critical level at which all sellers are just willing to accept cards. However, facing given fees, buyers' and sellers' decisions at stage (ii) with respect to card adoption are the same as those in Section 3, except that the surplus to buyers now depends on how many sellers accept each of the cards. Thus, the objective function of buyers can be described by equations (A-9) and (A-10) in the appendix with the indicator variables I^i and M replaced by the measure of sellers that accept card i exclusively and the measure of sellers that accept both cards.

The range of possible equilibrium outcomes varies between the case in which buyers' interests alone determine the equilibrium interchange fee and, at the other extreme, the case in which sellers' interests alone determine the equilibrium interchange fee. The range of possible equilibrium outcomes can be characterized by supposing the competitive equilibrium interchange fee a_C maximizes

$$(\omega (\beta_B - f) + (1 - \omega) (\beta_S - b_S^m)) DS, \tag{10}$$

where $0 \leq \omega \leq 1$. When $\omega = 0$, choosing the interchange fee to maximize this objective function leads to the same outcome as if the competitive interchange fee is set assuming sellers get to select their preferred card to accept knowing buyers will always hold this card. For instance, this case arises when buyers always multihome. On the other hand, when $\omega = 1$, choosing the interchange fee to maximize this objective function leads to the same outcome as if the competitive interchange fee is set assuming buyers get to select their preferred card to hold knowing (participating) sellers will always accept this card. More generally, we get:

Proposition 7 *Assume b_B and b_S are uniformly distributed. When buyers' interests are not weighted too highly ($\omega < 1/3$), competing card schemes set lower interchange fees than a single card scheme. When buyers' interests are given more weight ($\omega > 1/3$), competing card schemes set interchange fees higher than that set by a single scheme. If $\omega > 0$, competing card schemes set interchange fees too high ($a_C > a_W$).*

Proof. Using the uniform distribution, maximizing (10) is equivalent to maximizing

$$\left(\omega \left(\overline{b}_B - f\right) + (1 - \omega) \left(\overline{b}_S - b_S^m\right)\right) \left(\overline{b}_B - f\right) \left(\overline{b}_S - b_S^m\right),$$

which gives the unique solution

$$a_C = a_M + \frac{\Delta}{3} \left(\frac{2\sqrt{1 - 4\omega + 7\omega^2} - (1 + \omega)}{3\omega - 1}\right). \tag{11}$$

It follows that $a_C < a_M$ if $0 \leq \omega < 1/3$, and $a_C > a_M$ if $1/3 < \omega \leq 1$. If $\omega = 1/3$, L'Hôpital's rule implies that $a_C = a_M$. Moreover, if $\omega = 0$, $a_C = a_M - \Delta/3 = a_W$ using (9), but otherwise, $a_C > a_W$. ∎

This result parallels the result in Proposition 5 with homogenous sellers. There, the critical weighting on buyers' interests that determined if competition between card schemes decreased interchange fees was $\omega^* = -\beta_B'(\overline{a})/(1 - \beta_B'(\overline{a}))$. Since with uniformly distributed cardholder benefits $\beta_B'(\overline{a}) = -1/2$, the finding here is just a special case of the earlier result. There is, however, one important difference. Previously, with homogenous sellers, the fact sellers all stopped accepting cards at the same point meant that a monopoly card scheme already set the highest feasible interchange fee. Competing schemes would never set a higher interchange fee. This is no longer the case with seller heterogeneity, which implies competition between card schemes can increase interchange fees above the level set by a single monopoly

scheme. Competition can lead card schemes to set interchange fees too high for their own good. In effect, each card scheme sets its interchange fee too high in an attempt to get buyers to switch to holding its card exclusively, an effect which ends up reducing the total number of card transactions as fewer sellers accept cards.

The welfare implications of competition between card schemes with heterogenous merchants are also similar to our earlier result in Proposition 5 with homogenous merchants. System competition accentuates the bias against sellers resulting from sellers internalizing the surplus of buyers. To see why this is the case, note that we can rewrite (10) as $(\beta_B - f + (1 - \omega)(\beta_S - m)) DS$ by using the fact that $b_S^m = m - (\beta_B - f)$. Thus, competing schemes behave like a single card scheme, except they take into account the weighted surplus of end users in addition to the volume of card transactions. However, except in the special case with $\omega = 0$, the surplus of buyers is always given greater weight in this calculation (reflecting that sellers already internalize buyers' transactional benefits in making their joining decisions). This biases the choice of interchange fee above the welfare maximizing level, which maximizes the same expression only when $\omega = 0$.[20] For instance, if the interests of buyers and sellers are given equal weight ($\omega = 1/2$), competition between schemes results in higher interchange fees being set than those chosen by a single scheme, thereby further distorting fees away from the socially optimal level in our model.

5   Policy implications and concluding remarks

Policymakers in some jurisdictions have claimed that competing credit card schemes set interchange fees too high. Policymakers have also charged that there is a lack of competition between card schemes. For instance, the Reserve Bank of Australia [2002, p. 8] states "In Australia, credit card interchange fees are not determined by a competitive market". As a result, regulators in Australia and in Europe have required that card schemes lower their interchange fees. The competition authority in the United Kingdom has reached a similar decision with respect to MasterCard, finding the MasterCard interchange fee gives rise to a "collective price restriction". This suggests authorities view a lack of competition between card schemes as a possible cause of high interchange fees.

Our results open up the possibility that, in fact, the reverse conclusion holds. For instance, we showed that competition between schemes can be the cause of high interchange fees.[21] Reducing system competition may not only reduce interchange fees, but it may move them closer towards the efficient level. For instance, to the extent competing (heterogeneous) merchants internalize their customers' benefits from using cards and to the extent cardholders are at least as important as merchants in determining which card will be adopted by both sides, system competition will generally drive interchange fees higher. The result highlights the dangers of using one-sided logic in making inferences in two-sided markets. Greater system competition may not lower merchant fees, and provided they are implemented across both card associations and proprietary schemes, regulations lowering interchange fees by a modest amount may actually benefit card schemes by raising the volume of card transactions.

Given the possibility that interchange fees are set too high under inter-system competition, does it make sense to limit such competition or regulate interchange fees to lower levels? Our analysis suggests

---

[20]This logic does not depend on the particular uniform distribution considered above. However, the upward bias in interchange fees need not hold if issuers and acquirers price above marginal cost. Then Guthrie and Wright [2003] show competing schemes can again set interchange fees too low.

[21]Interestingly, in the United States, where competition between credit cards is arguably quite strong (with American Express and Discover competing aggressively), interchange fees are higher than in many other countries where competition between credit card schemes seems weaker. Weiner and Wright [2005] provide further evidence consistent with this effect.

first one would need to understand the role consumer and merchant expectations play in determining which card will be adopted by both sides. This is by no means an easy exercise and certainly not one that has been conducted to date. Limiting inter-system competition could result in lost productive or dynamic efficiencies, which could easily dominate any allocative efficiency gains from a more efficient fee structure.

What about regulating interchange fees to lower levels instead? We think doing so solely on the basis of this paper's analysis would be a mistake. The tendency for card networks to set interchange fees too high in our model does not reflect any anticompetitive motive on their part. In fact, our model precludes such incentives by assuming perfectly competitive issuers and acquirers. Therefore, the standard basis for government regulation in industry does not apply here. Allowing for the possibility that issuers and acquirers retain positive margins opens up the possibility that competitive interchange fees are too low, as we have noted. Regulation also seems premature until we have a better understanding of why merchants do not steer their customers to pay by other means (for example, by offering discounts) if doing so makes their customers and themselves jointly better off. The possibly important role of impulse purchases in driving credit card usage and acceptance has also not been studied to date.

Even if one puts aside these concerns, our model provides no basis for the claim by policymakers that competition should drive interchange fees to cost or the suggestion that cost-based interchange fees are efficient or more desirable than privately set interchange fees. On the other hand, appropriate benchmarks for determining efficient interchange fees, such as those derived in this paper, would be more difficult to implement in practice. At a minimum they require estimating merchants' benefits from accepting cards. Knowing only that privately set interchange fees are too high, but not knowing how much to lower them by, does not seem like a sensible basis for regulation. Even if an appropriate benchmark could be established, a further practical problem arises since, as we have shown, the regulation of interchange fees gives proprietary schemes (which do not have to set interchange fees to achieve their desired structure of fees) a competitive advantage. This can ultimately undermine any regulation of interchange fees.

One of the main motivations of this paper was to extend the existing literature on two-sided markets to the case in which users on one side of the market compete amongst themselves to attract users on the other side. Many two-sided markets have this feature, including payment cards, shopping malls, Yellow Pages, video games, and so on. By developing a model with competition between sellers, we were able to discern how allowing one type of user to compete amongst itself affects the equilibrium structure of fees. Competition between sellers generally increases the privately and socially optimal interchange fees, meaning it is optimal to charge more to sellers and less to buyers. This is true with or without competition between platforms, although with seller heterogeneity, system competition can accentuate this effect. When one type of user (sellers) is in business to attract sales from the other type of user (buyers), the sellers tend to internalize the benefits buyers get from the platform. This makes it more desirable to set fees which favor buyers, since by offering more surplus to buyers the schemes will find it easier to attract sellers. At the same time, since buyers' interests are over-represented, platforms will tend to recover too much from sellers and not enough from buyers.

Applying the same logic to other two-sided markets such as consumer directory services suggests that Yellow Pages (or other directories) will recover too much from advertisers and not enough from readers. A difference between payment cards and other two-sided markets such as directory services is the possibility for card schemes to set negative prices for card usage without inducing unbounded consumption. Such pricing would be difficult to implement in many other two-sided markets. To the extent that prices are constrained to be non-negative on one side of the market, the equilibrium fee structure could involve

services being given free to this side of the market. In such situations, it is possible that the resulting (constrained) fee structure is efficient even though the unconstrained optimum involves charging sellers too much.

An obvious feature of the equilibrium in our model is that it is sensitive to how consumers and merchants settle on a particular card scheme. This raises the possibility that user beliefs or historical precedents may determine equilibrium fee structures. This could also underlie the fact that sometimes quite different fee structures can emerge in apparently similar two-sided markets. For example, rental agencies (which help match tenants and landlords) typically charge landlords exclusively for the service, but in some cities such as Boston and New York the tenant typically pays the entire fee (see Evans, 2003). This makes understanding the structure of prices in two-sided markets all the more challenging.

References

Armstrong, M., 2006, 'Competition in Two-Sided Markets,' *RAND Journal of Economics*, forthcoming.

Armstrong, M. and Wright, J., 2005, 'Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts,' mimeo, University College of London.

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

Caillaud, B. and Jullien, B., 2001, 'Competing Cybermediaries,' *European Economic Review*, 45, pp. 797–808.

Caillaud, B. and Jullien, B., 2003, 'Chicken & Egg: Competition Among Intermediation Service Providers,' *RAND Journal of Economics*, 34(2), pp. 309–328.

Evans, D. S., 2003, 'The Antitrust Economics of Two-Sided Markets,' *Yale Journal on Regulation*, 20, pp. 325–382.

Farrell, J., 2005, 'Efficiency and Competition Between Payment Instruments,' mimeo, University of California, Berkeley.

Farrell, J. and Katz, M., 1998, 'The Effects of Antitrust and Intellectual Property Law on Compatibility and Innovation,' *Antitrust Bulletin*, 43, pp. 609–650.

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

Gans, J. S. and King, S. P., 2003a, 'The Neutrality of Interchange Fees in Payment Systems,' *Topics in Economic Analysis & Policy*, 3(1). Article 1.

Gans, J. S. and King, S. P., 2003b, 'A Theoretical Analysis of Credit Card Reform in Australia,' *Economic Record*, 79, pp. 462–472.

Guthrie, G. and Wright, J., 2003, 'Competing Payment Schemes,' Working Paper No. 0311. Department of Economics, National University of Singapore.
http://nt2.fas.nus.edu.sg/ecs/pub/wp/index.asp

Hagiu, A., 2006, 'Pricing and Commitment by Two-Sided Platforms,' *RAND Journal of Economics*, forthcoming.

Hausman, J. A., Leonard, G. K. and Tirole, J., 2003, 'On Nonexclusive Membership in Competing Joint Ventures,' *RAND Journal of Economics*, 34(1), pp. 43–62.

Reserve Bank of Australia, 2002, 'Reform of Credit Card Schemes in Australia. IV. Final Reforms and Regulation Impact Statement,' Reserve Bank of Australia, Public Document, August.

Rochet, J-C. and Tirole, J., 2002, 'Cooperation among Competitors: Some Economics of Payment Card Associations,' *RAND Journal of Economics*, 33(4), pp. 549–570.

Rochet, J-C. and Tirole, J., 2003, 'Platform Competition in Two-Sided Markets,' *Journal of the European Economic Association*, 1(4), pp. 990–1029.

Schiff, A., 2003, 'Open and Closed Systems of Two Sided Networks,' *Information, Economics and Policy*, 15, pp. 425–442.

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

Weiner, S. and Wright, J., 2005, 'Interchange Fees in Various Countries: Developments and Determinants,' *Review of Network Economics*, 4, pp. 290–323.

Wright, J., 2003a, 'Optimal Card Payment Systems,' *European Economic Review*, 47, pp. 587–612.

Wright, J., 2003b, 'Pricing in debit and credit card schemes,' *Economics Letters*, 80, pp. 305–309.

Wright, J., 2004, 'Determinants of Optimal Interchange Fees in Payment Systems,' *Journal of Industrial Economics*, 52, pp. 1–26.

Proof of Lemma 1

We start by proving the existence of unique roots and maxima of ϕ_S. Since $\phi_S(a)/D(a) = b_S - c_A - a$ is decreasing in a, the unique solution to $\phi_S(a) = 0$ is $a_W = b_S - c_A$. Since

$$\phi'_S(a) = h(c_I - a)(b_S - c_A - a) - D(a),$$

$\phi'_S(a)$ takes the values $h(\overline{b}_B)(b_S + \overline{b}_B - c) > 0$ and $-D(b_S - c_A) < 0$ at $c_I - \overline{b}_B$ and $b_S - c_A$ respectively, and there exists an interchange fee $\underline{a} \in (c_I - \overline{b}_B, b_S - c_A)$ such that $\phi'_S(\underline{a}) = 0$. Given the hazard function $h(f)/(1 - H(f))$ is assumed to be increasing in f, it follows that $h(c_I - a)/D(a)$ is decreasing in a. Hence $\phi'_S(a)/D(a)$ is decreasing in a over the interval $[c_I - \overline{b}_B, b_S - c_A]$, and $\underline{a}$ is the only solution to $\phi'_S(\underline{a}) = 0$ in $(c_I - \overline{b}_B, b_S - c_A)$. Since $\phi'_S(a) < 0$ for all $a \in [b_S - c_A, c_I - \underline{b}_B]$, it follows that $\underline{a}$ is the unique maximum of ϕ_S.

We conclude by proving the existence of unique roots and maxima of ϕ. Since

$$\phi(a)/D(a) = \beta_B(a) + b_S - c$$

is decreasing in a and takes the values

$$\beta_B(b_S - c_A) + b_S - c > f(b_S - c_A) + b_S - c = 0$$

and

$$\beta_B(c_I - \underline{b}_B) + b_S - c = E[b_B] + b_S - c < 0$$

at $b_S - c_A$ and $c_I - \underline{b}_B$ respectively, there exists a unique solution $\overline{a} \in (b_S - c_A, c_I - \underline{b}_B)$ to $\phi(a) = 0$. The fact that the derivative of $D(a)\beta_B(a)$ with respect to a equals $f(a)D'(a)$ implies that $\phi'(a) = h(c_I - a)(b_S - c_A - a)$. The unique solution to $\phi'(a) = 0$ is therefore $a_W = b_S - c_A$.

Proof of Proposition 1

Assume to start with all buyers hold the card and look initially at the decision of sellers at stage (ii), whether to accept the card or not. Define I_i as the indicator variable which takes the value 1 if seller i accepts the card and 0 otherwise. Consumers take into account the transactional surplus ϕ_B they expect to get if shopping at a seller that can accept cards and the price such a seller will set. Given the Hotelling model setup, the market share of seller i is then

$$s_i = \frac{1}{2} + \frac{1}{2t} (p_j - p_i + \phi_B (I_i - I_j)). \tag{A-1}$$

Seller i's corresponding profit is

$$\pi_i = s_i (p_i - d + \phi_S I_i). \tag{A-2}$$

Solving for the equilibrium prices implies

$$p_i = d + t - \phi_S I_i + \frac{1}{3} (\phi_S + \phi_B) (I_i - I_j), \tag{A-3}$$

so that substituting (A-3) back into (A-1) and (A-2), and using that $\phi = \phi_S + \phi_B$, implies $\pi_i = 2ts_i^2$ where $s_i = 1/2 + \phi (I_i - I_j) / (6t)$. Regardless of what the other seller does, each seller will accept cards if doing so increases (or does not decrease) its equilibrium market share. Thus, each seller will accept cards if ϕ is non-negative but not otherwise.

It follows that provided $\phi \geq 0$, there is an equilibrium where both sellers will accept the card and all buyers will hold the card. The other equilibrium, in which buyers do not hold cards and so sellers do not accept them, is Pareto dominated since buyers are strictly better off holding cards and each individual seller acting alone is no worse off accepting them (whenever $\phi \geq 0$). If $\phi < 0$, the unique equilibrium involves sellers not accepting cards.

Proof of Proposition 4

We define $D^i = D(a^i)$, $\phi_B^i = \phi_B(a^i)$, $\phi_B^{12} = \phi_B(\max\{a^1, a^2\})$, $\phi_S^i = \phi_S(a^i)$, $\phi_S^{12} = \phi_S(\max\{a^1, a^2\})$, $\phi^i = \phi_S^i + \phi_B^i = \phi(a^i)$, and $\phi^{12} = \phi_S^{12} + \phi_B^{12}$. Let λ_B^i denote the measure of buyers who hold card i only (singlehome) and λ_B^{12} denote the measure of buyers who hold both cards (multihome). We solve the game by working backwards. Define I_i^1 as the indicator variable that takes the value 1 if seller i accepts card 1 exclusively and 0 otherwise, I_i^2 as the indicator variable that takes the value 1 if seller i accepts card 2 exclusively and 0 otherwise, and M_i as the indicator variable that takes the value 1 if seller i accepts both cards (multihomes). Consumers who hold card 1 take into account the transactional surplus ϕ_B^1 they expect to get if shopping at a seller that just accepts card 1 (and likewise for the case of card 2). Consumers who hold both cards also take into account the transactional surplus $\phi_B(\max\{a^1, a^2\})$ they expect to get if shopping at a seller that accepts both cards. This reflects the fact that when buyers hold both cards and sellers accept both cards, buyers will use the card with the lowest card fee (highest

interchange fee). Given the Hotelling model setup, the market share of seller i is then

$$s_i = \frac{1}{2} + \frac{1}{2t} \left(p_j - p_i + \left(\lambda_B^1 + \lambda_B^{12} \right) \phi_B^1 \left(I_i^1 - I_j^1 \right) + \left(\lambda_B^2 + \lambda_B^{12} \right) \phi_B^2 \left(I_i^2 - I_j^2 \right) \right.$$
$$\left. + \left(\lambda_B^1 \phi_B^1 + \lambda_B^2 \phi_B^2 + \lambda_B^{12} \phi_B^{12} \right) \left(M_i - M_j \right) \right). \tag{A-4}$$

A seller that accepts card 1 exclusively can accept cards from buyers who multihome, as well as those holding card 1 exclusively. A seller that accepts both cards will be able to accept cards from a wider range of buyers (including those holding card 2 exclusively), but as a result will lose control over how buyers holding both cards pay. Seller i's corresponding profit is

$$\pi_i = s_i \left(p_i - d + \left(\lambda_B^1 + \lambda_B^{12} \right) \phi_S^1 I_i^1 + \left(\lambda_B^2 + \lambda_B^{12} \right) \phi_S^2 I_i^2 + \left(\lambda_B^1 \phi_S^1 + \lambda_B^2 \phi_S^2 + \lambda_B^{12} \phi_S^{12} \right) M_i \right). \tag{A-5}$$

We solve for the equilibrium in stage (iii) by working out each seller's profit-maximizing choice of prices given (A-4) and (A-5). Solving the two sellers' best response functions simultaneously, rearranging, and using that $\phi^i = \phi_B^i + \phi_S^i$ implies the equilibrium price p_i can be written

$$p_i = d + t - \left(\lambda_B^1 + \lambda_B^{12} \right) \phi_S^1 I_i^1 - \left(\lambda_B^2 + \lambda_B^{12} \right) \phi_S^2 I_i^2 - \left(\lambda_B^1 \phi_S^1 + \lambda_B^2 \phi_S^2 + \lambda_B^{12} \phi_S^{12} \right) M_i$$
$$+ \frac{1}{3} \left(\left(\lambda_B^1 + \lambda_B^{12} \right) \phi^1 \left(I_i^1 - I_j^1 \right) + \left(\lambda_B^2 + \lambda_B^{12} \right) \phi^2 \left(I_i^2 - I_j^2 \right) \right.$$
$$\left. + \left(\lambda_B^1 \phi^1 + \lambda_B^2 \phi^2 + \lambda_B^{12} \phi^{12} \right) \left(M_i - M_j \right) \right). \tag{A-6}$$

Substituting (A-6) back into (A-4) and (A-5), and using again that $\phi^i = \phi_B^i + \phi_S^i$ implies

$$\pi_i = 2t s_i^2,$$

where

$$s_i = \frac{1}{2} + \frac{1}{6t} \left(\left(\lambda_B^1 + \lambda_B^{12} \right) \phi^1 \left(I_i^1 - I_j^1 \right) + \left(\lambda_B^2 + \lambda_B^{12} \right) \phi^2 \left(I_i^2 - I_j^2 \right) + \left(\lambda_B^1 \phi^1 + \lambda_B^2 \phi^2 + \lambda_B^{12} \phi^{12} \right) \left(M_i - M_j \right) \right).$$

Regardless of what the other seller does, each seller will accept cards at stage (ii) if doing so increases its equilibrium market share. Relative to a seller that does not accept either card, a seller that just accepts card i will obtain additional market share of

$$\frac{1}{6t} \left(\lambda_B^i + \lambda_B^{12} \right) \phi^i, \tag{A-7}$$

while a seller that accepts both cards will obtain additional market share of

$$\frac{1}{6t} \left(\lambda_B^1 \phi^1 + \lambda_B^2 \phi^2 + \lambda_B^{12} \phi^{12} \right). \tag{A-8}$$

Equations (A-7) and (A-8) describe the seller's objective function.

Noting that the two sellers' decisions will be identical in all cases, we can write $I^1 = I_i^1 = I_j^1$, $I^2 = I_i^2 = I_j^2$, and $M = M_i = M_j$. Any equilibrium in stage (ii) requires also characterizing what buyers will do given both sellers choose the same actions. Given sellers set the same price regardless of how consumers pay, a buyer who just holds card i can expect to obtain the transactional surplus from card usage of

$$\left(I^i + M \right) \phi_B^i. \tag{A-9}$$

A buyer who holds both cards can expect to obtain the transactional surplus from card usage of

$$I^1 \phi_B^1 + I^2 \phi_B^2 + M \phi_B^{12}, \tag{A-10}$$

22

while a consumer who holds neither card will obtain no transactional surplus from card usage.

Note that in terms of their joining decisions, the two groups (buyers and sellers) are symmetric except in two respects. First, since sellers accept cards for strategic reasons, a seller's card acceptance decision depends on $\phi_S + \phi_B$ rather than just its transaction surplus ϕ_S. In contrast, a buyer's card holding decision just depends on its transaction surplus ϕ_B. Second, where a multihoming buyer transacts with a multihoming seller, it is the buyer who determines which card to use, which explains why we always take the surplus evaluated at the maximum of the two interchange fees in this case.

Using the objective function of each of the parties, defined by (A-7)–(A-10), we can characterize some features of equilibria in the full game.

We start by noting that, in equilibrium, at least one of the schemes must set an interchange fee in the interval $[a_W, \overline{a}]$. For example, if both schemes set fees greater than $\overline{a}$, sellers will not accept either card since $\phi < 0$. Each scheme has an incentive to set an interchange fee at $\overline{a}$ instead and attract all card transactions. Similarly, if both schemes set fees less than a_W, since both buyers and individual sellers prefer an interchange fee of a_W (buyers prefer it because it lowers their card fees, while sellers prefer it because it raises ϕ), each scheme has an incentive to raise its interchange fee to a_W and thereby attract all buyers and sellers (and more transactions for each cardholder). Finally, if one scheme (say scheme 1) sets a fee below a_W and the other (say scheme 2) sets a fee above $\overline{a}$, scheme 2 will not attract any card transactions. It has an incentive to set an interchange fee at a_W, attracting all buyers and sellers as in the previous case. None of these three alternate scenarios are consistent with an equilibrium in stage (i), proving that, in equilibrium, at least one of the schemes must set an interchange fee in the interval $[a_W, \overline{a}]$.

Therefore, without loss of generality we can suppose that scheme 1 sets an interchange fee $a^1 \in [a_W, \overline{a}]$ in equilibrium. If scheme 2 sets the same interchange fee (so that $a^2 = a^1$), our assumption that buyers and sellers choose only Pareto undominated demand configurations requires either all buyers and sellers choose a single card scheme, or the schemes share card transactions with all agents on at least one side multihoming.

Suppose, instead, that scheme 2 sets $a^2 > a^1$, implying that $\phi^2 < \phi^1$. If all sellers accept only card 2 or both cards, buyers will only use card 2 since it has lower card fees. In this case, all card transactions will be using card 2. If some sellers accept only card 1, since there is no cost to holding cards, all buyers will hold card 1 (whether or not they also hold card 2). But this allows all sellers to reject card 2 and steer buyers to use card 1. In this case, all card transactions will be using card 1. It follows that if $a^2 > a^1$, either card 1 is always used or card 2 is always used. Notice, however, that if $a^2 > \overline{a}$ then $\phi^2 < 0$, all sellers reject card 2, and all card transactions will use card 1.

Finally, suppose that scheme 2 sets $a^2 < a^1$. If $\phi^2 \leq \phi^1$, both buyers and individual sellers prefer card 1, so the only Pareto undominated equilibrium involves buyers and sellers choosing card 1. On the other hand, if $\phi^1 < \phi^2$, either all sellers accept both cards and buyers will only use card 1, or all sellers reject card 1 and steer buyers to use card 2. (The explanation is analogous to that in the preceding paragraph.) It follows that if $a^2 < a^1$, either card 1 is always used or card 2 is always used. Notice, however, that if $a^2 < a_W$ then scheme 2 must attract zero demand — otherwise, scheme 1 would attract zero demand, giving it an incentive to deviate and choose $a^1 = a_W$.

Together, these results imply that in any equilibrium all card transactions occur at a card fee corresponding to a single interchange fee which lies in the interval $[a_W, \overline{a}]$. Either both schemes set the same interchange fee, and they share card transactions with all agents on at least one side multihoming, or one scheme attracts all card transactions exclusively at this interchange fee.

Proof of Proposition 5

We start by proving the existence of a_C. Since $\phi_B(a)$ is increasing in a for all a, and $\phi(a)$ is increasing in a for all $a \in [c_I - \bar{b}_B, b_S - c_A)$, it follows that $\omega\phi_B(a) + (1-\omega)\phi(a)$ is also increasing in a for all $a \in [c_I - \bar{b}_B, b_S - c_A)$. Therefore, we can restrict attention to interchange fees in the interval $[b_S - c_A, \bar{a}]$. The existence of an interchange fee that maximizes $\omega\phi_B(a) + (1-\omega)\phi(a)$ in this interval is guaranteed by the continuity of that function.

The result that $a_C = a_W$ if $\omega = 0$ follows immediately from Lemma 1. In contrast, if $\omega > 0$, the objective function $\omega\phi_B(a) + (1-\omega)\phi(a)$ is increasing at a_W, so that any equilibrium interchange fee must be strictly greater than a_W. When $\omega = 1$, the objective function becomes $\phi_B(a)$, which is increasing in a and is therefore maximized by choosing the highest interchange fee such that the constraint $\phi(a) \geq 0$ still holds. From Lemma 1, this is the interchange fee $\bar{a}$.

Finally, we prove that $a_C < a_M$ whenever $\omega < \omega^*$. Since ϕ_B is increasing in a and ϕ is decreasing in a at $\bar{a}$, $\omega < \omega^*$ implies that

$$
\begin{aligned}
\omega\phi_B'(\bar{a}) + (1-\omega)\phi'(\bar{a}) \;&<\; \omega^*\phi_B'(\bar{a}) + (1-\omega^*)\phi'(\bar{a}) \\
&=\; \omega^* D(\bar{a}) + (1-\omega^*)h(c_I - \bar{a})(b_S - c_A - \bar{a}) \\
&=\; \frac{-\beta_B'(\bar{a})}{1 - \beta_B'(\bar{a})}D(\bar{a}) + \frac{1}{1 - \beta_B'(\bar{a})}h(c_I - \bar{a})(b_S - c_A - \bar{a}).
\end{aligned}
$$

From the definition of $\bar{a}$, $\phi(\bar{a}) = 0$, we see that

$$
\beta_B(\bar{a}) + b_S - c = 0.
$$

Therefore

$$
b_S - c_A - \bar{a} = c_I - \bar{a} - \beta_B(\bar{a})
$$

and

$$
\omega\phi_B'(\bar{a}) + (1-\omega)\phi'(\bar{a}) < \frac{1}{1 - \beta_B'(\bar{a})}\left((c_I - \bar{a})h(c_I - \bar{a}) - \beta_B'(\bar{a})D(\bar{a}) - \beta_B(\bar{a})D'(\bar{a})\right).
$$

Since the derivative of $\beta_B(a)D(a)$ equals $(c_I - a)h(c_I - a)$, it follows that

$$
\omega\phi_B'(\bar{a}) + (1-\omega)\phi'(\bar{a}) < 0.
$$

Having shown that, for all $\omega < \omega^*$, $\omega\phi_B + (1-\omega)\phi$ is decreasing in a at $\bar{a}$, it is obvious that $\bar{a} \neq \arg\max_a\{\omega\phi_B(a) + (1-\omega)\phi(a) : \phi(a) \geq 0\}$. This completes the proof that $a_C < \bar{a}$ whenever $\omega < \omega^*$.