Cooperation among Competitors: The Economics of Payment Card Associations*

Jean-Charles Rochet †

Jean Tirole ‡

May 16, 2000

Abstract

The paper analyzes the cooperative determination of the interchange fee by member banks in a payment card association. The interchange fee is the "access charge" paid by the merchants' banks, the acquirers, to cardholders' banks, the issuers. The paper develops a framework in which banks and merchants may have market power and consumers and merchants decide rationally on whether to buy or accept a payment card. After drawing the welfare implications of a cooperative determination of the interchange fee, the paper provides a detailed description of the factors impacting merchant resistance, compares cooperative and for-profit business models, and makes a first cut in the analysis of system competition.

JEL Classification: G21, L31, L42
Keywords: Payment cards, association, discrimination, interchange fee.

*We benefitted from the comments of Bruno Biais, Howard Chang, David Evans, Patrick Rey and Richard Schmalensee. Research support from Citicorp and Visa USA is gratefully acknowledged. The views in the paper are solely those of the authors.

†Institut D'Economie Industrielle (IDEI) and GREMAQ (UMR 5603 CNRS), Toulouse.

‡IDEI and GREMAQ (UMR 5603 CNRS), Toulouse, CERAS (URA 2036 CNRS), Paris, and MIT.

1 Introduction

The rapid growth of payment cards[1] usage is a striking feature of modern economies. The payment activity, a fundamental dimension of the payment card industry, is characterized by the existence of strong network externalities: In a payment card transaction, the consumer's bank, called the issuer, and the merchant's bank, the acquirer, must cooperate in order to enable the transaction.[2] Two successful not-for-profit joint ventures, Visa and MasterCard,[3] have designed a set of rules to govern the "interconnection" between their members:

(1) *Interchange fee*: The acquirer pays a collectively determined interchange fee (the analog of an access charge in telecommunications) to the issuer. The issuer guarantees the payment.[4]

(2) *Honor-all-cards rule*: All affiliated merchants must accept the card of any issuing member.

(3) *No-surcharge rule*: Affiliated merchants are not allowed to impose surcharges on customers who pay with a card.[5]

Some of these institutional features have gained wide acceptance. The payment guarantee by the issuer can be motivated by considerations of delegated monitoring. And to see the benefits of a centrally-determined interchange fee cum the honor-all-cards rule, it suffices to envision the complexity of bilateral bargaining among thousands of banks as well as the cost for issuers (respectively, merchants) of informing consumers about the set of merchants (respectively, banks) with whom an agreement has been reached.[6]

In contrast, although they have not yet been successfully challenged in court,[7] two features

[1]Payment card is a generic name that includes credit cards, debit cards and charge cards. Since we focus here on the payment activity, these differences are immaterial for our purposes.

[2]Similar issues arise for ATM (Automatic Teller Machine) transactions, but they only involve (at most) three actors (the cardholder, the issuer and the bank who operates the ATM) instead of four.

[3]Visa and MasterCard are each owned jointly by thousands of banks and handle 75 percent of the total volume of general purpose payment card transactions. There also exist "proprietary systems" such as American Express, in which the issuer and the acquirer are the same firm. We refer to Evans-Schmalensee (1999) for an excellent overview and analysis of the industry.

[4]The interchange fee depends on the fraud-control devices installed at the merchant's premises.

[5]This is the US version of the rule. Interestingly, US merchants rarely offer discounts for cash payments, even though such discounts are not prohibited! In some European countries, payment card associations impose a stricter rule and prohibit any form of discrimination.

[6]For more on transaction costs, see Evans-Schmalensee (1995, p886, 887 and 890).

[7]See in particular National Bancard Corp. v. Visa USA, Inc. 596 F. Supp 1231 (SD Florida 1984).

of these interconnection rules have recurrently been viewed with suspicion by competition authorities and by some economists.[8] First, the no-surcharge rule is sometimes viewed as an attempt by payment card systems to leverage their market power by forcing more card transactions than is efficient. Second, the collective determination of the interchange fee is regarded by some as a potential instrument of collusion: Aren't the banks able to inflate payments to each other and, in fine, tax merchants and consumers? Shouldn't the access charge be regulated as in telecommunications?[9] Even if one accepts the existence of an interchange fee, one may still be legitimately concerned that it be set too high. At a general level, agreements among competitors can be anticompetitive and one must investigate whether this is indeed the case in our context. For example, a joint venture among competitors whose primary motive is to raise the price on the final good market by overcharging for a common input and redistributing the proceeds among its members is anticompetitive. Similarly, a high reciprocal access charge negotiated between rival telecommunications networks may in some circumstances be anticompetitive.[10] It is tempting to draw an analogy between such situations and that of a collectively-chosen interchange fee. As we will see, one should refrain from making such a quick transposition.

This paper analyzes the validity of these two concerns. In order to provide a policy analysis, it develops a normative framework of the determination of an efficient interchange fee and of the impact of the no-surcharge rule. The strength of our approach relative to the previous literature (in particular Baxter (1983)) reviewed in section 2 is that we endogenize consumer and merchant behavior and are therefore able to perform a proper welfare analysis. Baxter

[8] E.g., Frankel (1998) and Carlton-Frankel (1995).

[9] It is sometimes argued that the interchange fee is unnecessary and that its existence is evidence of a collective exercise of market power. In a famous US case, NaBanco, whose expertise lied principally in recruiting merchants and which was rewarded for this activity by keeping some or all of the merchant discount revenue, argued that it could not compete effectively with Visa members that issue cards to consumers and service merchants and therefore economize on the interchange fee for some of their merchant transactions. NaBanco's argument that the interchange fee was a per se violation of the Sherman Act was rejected by the court. A careful study explaining the mechanisms of the interchange fee is in order, both to study this argument and to build some intuition about the stumbling blocks that an economic regulation of the interchange fee would probably face. In section 6.4, we show that issuers do not gain from entering the acquiring business and generating "on-us" transactions if the acquiring segment is competitive.

[10] See Laffont-Rey-Tirole (1998 a,b) for a formalization of this argument and a number of qualifications to it.

focuses on the technological benefits to consumers and merchants brought about by the use of payment cards relative to alternative means of payment. Namely, he assumes that consumers and merchants adopt the card as long as the technological benefits exceed their payments to the banks. This ignores the fact that consumers and merchants are strategic players.

First, when (at least some) consumers know which stores take payment cards before they select which to patronize, or may leave the store when they discover the card is not accepted, card acceptance is used by merchants to attract customers. A merchant's total benefit, and thus its decision of whether to accept a card then depend not only on the merchant's technological benefit (fraud control, theft protection, speed of transactions, customer information collection,...), but also on the product of its increase in demand due to system membership and its retail markup.[11] Thus, Baxter in general overstates merchants' resistance to an increase in the merchant discount and therefore to an increase in the interchange fee. Second, when merchants are allowed to offer cash discounts, a consumer's decision to use a card depends not only on the technological benefit (convenience, theft and fraud control,...), but also on the extra charge for using a payment card. Third, when several payment card systems compete, the opportunity cost for a merchant of accepting a card is endogenous as long as some customers hold cards on multiple systems. For example, a merchant who turns down American Express may see the customer pay with Visa or Master Card rather than with cash or a check. Thus, Baxter's hypothesis understates merchants' resistance under system competition.

The paper is organized as follows. Section 2 briefly reviews the literature. Section 3 develops the model under the no-surcharge rule and no system competition. Section 4 compares the interchange fee selected by the payment card association with the socially optimal one. Section 5 analyses the impact of the no-surcharge rule. Section 6 studies the determinants of merchant resistance, compares different business models, analyses some elements of system competition, and discusses the robustness of the results. Section 7 summarizes the main insights and discusses some topics for future research.

[11] Furthermore, one cannot just set this retail markup to zero by assuming that merchants are undifferentiated Bertrand competitors since the very decision of whether to accept the payment card is a factor of differentiation among merchants.

2 Relationship to the literature

Economic research has only recently started studying the payment card industry. The theoretical and empirical analyses of the US credit card market were initiated by Baxter (1983) and Ausubel (1991), respectively. Similarly, ATM (Automatic Teller Machines) networks have been analyzed only recently by Gilbert (1990), Matutes and Padilla (1994), McAndrews and Rob (1996) and Kim (1998).

The formal literature on access pricing in the payment card industry is meager. The standard reference is Baxter (1983)'s analysis of a competitive payment card industry. To rephrase Baxter's argument, let us introduce some notation (summarized in figure 1). Baxter notes that a card payment is a service offered to two parties (the cardholder and the merchant) jointly by two other parties (the issuer and the acquirer). The total cost of this service is the sum of the issuer's cost c_I and the acquirer's cost c_A. Suppose that the benefit accruing to the cardholder (or buyer) for the marginal use of a payment card is equal to b_B. Similarly, the benefit to the merchant (or seller) of this marginal use of a payment card is b_S. The benefits b_i and costs c_i referred to above are *net* benefits and costs. The cardholder and the merchant must compare the utilities they get by using payment cards with those associated with alternative payment methods (cash, checks,...). At the social optimum, the total benefit of the marginal transaction, $b_B + b_S$, is equal to its total cost, $c_I + c_A \equiv c$.

To implement this social optimum, the cardholder must pay a customer fee f equal to b_B and the merchant must pay a merchant discount m equal to b_S. But if the cardholder and the merchant are serviced by two different banks, as is usually the case, there is no reason why both banks should break even on the transaction. [While $b_B + b_S = c_I + c_A$, in general $b_B \neq c_I$ and $b_S \neq c_A$]. One bank makes a profit and the other makes a loss equal to the former bank's profit. The money making bank must therefore compensate the money losing bank for facilitating the provision of the joint service. Baxter's theory says nothing about the sign of the socially optimal interchange fee, which can be defined as a (positive or negative) transfer $a = b_S - c_A = c_I - b_B$ from the acquirer to the issuer. That is, the transfer flow from the acquirer to the issuer depends on the magnitude of these costs and benefits. As Schmalensee (1999) notes, Baxter's analysis also predicts nothing concerning the choice of the interchange fee by payment card systems; for, if the perfect competition assumption is to be taken seriously, issuers and acquirers make no profit regardless of the level of the interchange fee.

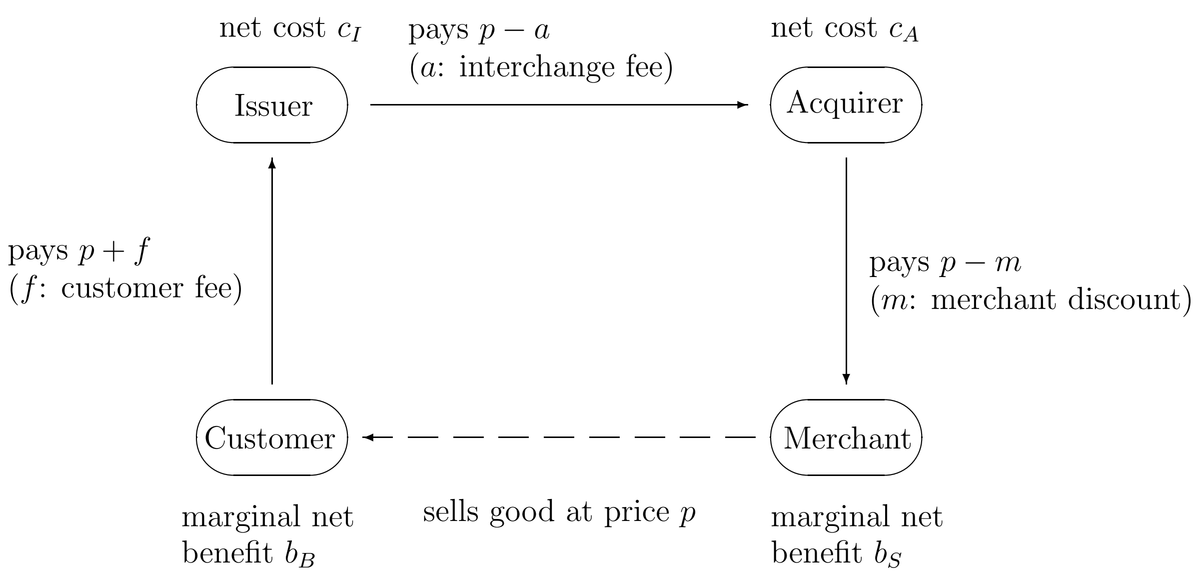

Figure 1

Schmalensee (1999), in an analysis complementary to ours, analyzes the provision of payment card services as a moral-hazard-in-teams problem. The number of payment card transactions is a function of the issuers' and the acquirers' efforts, with a complementarity between the two efforts.[12] Each side's effort is bidimensional: marketing effort as well as terms given to the banks' clients (merchant discount for acquirers, customer fee for issuers). The Nash equilibrium of the resulting "second stage" game depends on the interchange fee, which is determined in a first stage through bargaining between issuers and acquirers. Schmalensee solves for the outcome of this two- stage game for an arbitrary allocation of bargaining power[13]. Schmalensee argues that there is no support for a public policy of forcing interchange fees to zero.

[12]Schmalensee first analyzes the case of a monopoly issuer and a monopoly acquirer. He then generalizes the model to oligopolistic competition on both sides.

[13]As Schmalensee notes, in the US, banks' voting rights in Visa and MasterCard are more sensitive to issuing volume than to acquiring volume; this suggests that the bargaining power is on the issuing side.

Our paper, like Schmalensee's, analyzes market power issues. It follows Baxter in its emphasis on the determination of the efficient interchange fee; yet, by departing from Baxter's perfectly competitive paradigm and thus from the banks' indifference as to the level of the interchange fee, and by modifying his analysis to account for consumer and merchant rational behavior, our framework allows a comparison between the privately optimal interchange fee (the object of Schmalensee's analysis) and the socially optimal one. Furthermore, and because we derive the demand for payment card transactions from individual consumer preferences, we are able to analyze the impact of the no-surcharge rule, which has not yet been studied in the literature.

The question of the determination of interchange fees in ATM networks, studied for example in McAndrews and Rob (1996), has obvious similarities with our topic. However, there is a fundamental element of our analysis, namely merchant resistance, that is absent in the context of ATM networks. ATM transactions only involve three parties, as opposed to four in the case of card (or check) payments. Finally, McAndrews (1997) studies the impact of the direct presentment regulation, that prevent U.S. banks from charging each other an interchange fee for checks. His results argue in favor of lifting this regulation.

3 A model of the payment card industry under the no-surcharge rule

In our basic model, there is a single payment card association. In this section and the next, we assume that the payment card association prohibits merchants from imposing surcharges on customers paying with a card (as Visa and MasterCard currently do). Section 5 looks at the impact of preventing the card association from adopting this policy.

Our analysis makes two simplifying assumptions. Although both can easily be relaxed, these two assumptions fit well the payment card industry. First, we assume that acquirers are competitive while issuers have market power. The acquiring side involves little product differentiation as well as low search costs and is widely viewed as highly competitive.[14] In contrast, the issuing side is generally regarded as exhibiting market power. The cause and the extent of market power is highly country-specific. It may be due to innovation[15] or to other factors such as search costs, reputation, or the nature of the card.[16] In the model below, we assume that issuers have *some* market power: The issuing side is not perfectly competitive.

The second simplifying assumption is that customers have a fixed volume of transactions, normalized to one transaction. This implies that, from the point of view of the issuer-customer relationship, there is no difference between a fixed yearly fee and a per transaction customer fee (see section 6.1 for a discussion of this point). The analysis of optimal price discrimination and volume discounts offered by issuers is interesting in its own right, but it is somewhat orthogonal to the problem at hand, and we ignore it by assuming a fixed volume of transactions. There is endogeneity of the volume of *payment card* transactions, though, because consumers may choose not to have a card, or may be unable to use their card if the merchant refuses it.

As in our discussion of Baxter's model, let c_I and c_A denote the per transaction cost, incurred by an issuer and an acquirer, respectively. The interchange fee is denoted a and the merchant discount m. The number of transactions per customer is normalized at one; the customer pays a yearly customer fee equal to f. Last, let b_B and b_S denote the customer's and the merchant's (per transaction) benefit from using the card rather than an alternative payment method, say cash. In the basic version of our model, all merchants have the same benefit while consumers are heterogenous.

Consumers: Issuers' market power may impact social welfare if issuers cannot perfectly price discriminate. We therefore assume that consumers differ as to their benefit from using a payment card rather than an alternative payment method. For example, some customers have an easy access to cash or a low value of time of going to get cash before shopping, while others

[14] See, e.g., Evans- Schmalensee (1999, chapter 6).

[15] Attractive frequent user programs, payment facilities, co-branding and single bill offerings (telephone and payment card for example), corporate card, and so on.

[16] In France, payment cards are primarily debit cards; the payment is automatically debited at the end of the month from the customer's bank account (thus credit is limited to intra-month credit). Consumers therefore use debit cards issued by their banks. Yearly fees are high.

attach a high value to the convenience afforded by the use of cards. The benefit b_B is continuously distributed on an interval $[\underline{b}_B, \bar{b}_B]$. The fraction of consumers with benefit less than b_B is given by the cumulative distribution function $H(b_B)$, with density $h(b_B)$. The hazard rate $h/(1-H)$ is increasing, in order to guarantee concavity of the optimization programs. Let

$$E(b_B \mid b_B \geq b_B^*) \equiv \frac{\int_{b_B^*}^{\bar{b}_B} b_B h(b_B) db_B}{1 - H(b_B^*)}$$

denote the expected benefit enjoyed by an average *cardholder* (as opposed to consumers in general) when consumers with type $b_B \geq b_B^*$ purchase the card, and those with type $b_B < b_B^*$ do not.

Issuers: Each issuer has market power over its customers. We further assume in a first step that issuers are not in the acquiring business. In section 6.4 we will observe that, due to the competitiveness of the acquiring business, issuers are actually indifferent between entering the acquiring business and staying out at the equilibrium interchange fee (and so they may actually be in the acquiring business after all), and furthermore that they would not benefit from an interchange fee that creates a strict preference for them to enter the acquiring segment.

Assuming that the card is accepted by all merchants (an aspect which we will need to investigate in our equilibrium model), a customer with benefit b_B and facing customer fee f purchases the card if and only if

$$b_B \geq f.$$

For expositional simplicity, let us focus on a symmetric oligopolistic equilibrium, in which all issuers in equilibrium charge the same customer fee f. Let $D(f)$ denote the total demand for cards, and $\beta(f)$ the average cardholder benefit. That is

$$D(f) \equiv 1 - H(f),$$

and

$$\beta(f) \equiv E[b_B \mid b_B \geq f].$$

Note that the demand for cards, $D(f)$, decreases with the customer fee, and that the average cardholder benefit, $\beta(f)$, is increasing and bounded.

The net cost of a transaction for an issuer is equal to the difference between the "technological cost" c_I and the interchange fee a. Let $f = f^*(c_I - a)$ denote the equilibrium customer fee. We make the following mild assumption:

Assumption A: *The oligopolistic equilibrium fee, $f^*(c_I - a)$, is defined for all values of the interchange fee a (even $a > c_I$) and decreases with it. Each member bank's profit increases with the interchange fee a.*

As we will see, a negative issuer marginal cost (which arises when $a > c_I$) does not create the problem of unbounded consumption usually associated with negative marginal costs; for, even if the customer is offered cash-back bonuses or other inducements to use the card (see section 6.1), the customer must still pay for the merchant's good, that is cannot use the card "on a stand-alone basis". Let us provide a few examples satisfying Assumption A:

Example 1: Monopoly issuer

A monopolist chooses its fee so as to maximize

$$\Big[f + (a - c_I)\Big]\Big[1 - H(f)\Big]. \tag{1}$$

A simple revealed preference argument shows that this fee is a decreasing function of the interchange fee. That is, a monopoly issuer finds it more costly to restrict the number of payment cards and to exercise its market power if the interchange fee increases.[17] Moreover, from the envelope theorem, the issuer's profit decreases with its net cost, and therefore increases with the interchange fee.

Example 2: Symmetric Cournot oligopoly

Assumption A is satisfied in a symmetric Cournot oligopoly whenever the elasticity of demand exceeds one (Seade 1987).

Example 3: Symmetric differentiated Bertrand oligopoly

[17] The reader will here recognize the standard argument that a proportional subsidy to firms with market power reduces the distortion due to excessive margins.

With product differentiation, the relevant benefit for the consumer is the intrinsic benefit corrected by the consumer's distance between her brand choice and her ideal specification. Appendix 1 checks Assumption A for the standard Hotelling model of product differentiation.[18]

An analysis focused on the issuing side is incomplete. To understand the impact of the interchange fee, we must perform an equilibrium analysis. For, the interchange fee also impacts the merchant discount, and therefore the merchants' willingness to accept the card. In turn, the customers' willingness to purchase a card depends on the number of merchants accepting it. Last, prices charged by merchants to customers may depend on the interchange fee.

Acquirers: Acquirers face per transaction cost c_A and are competitive. Thus, for interchange fee a, they offer merchant discount m given by

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

Merchants: To study the impact of the interchange fee on final prices and social welfare, we use the standard Hotelling model of the "linear city"[19] (or cities: there may be an arbitrary number of such segments). Consumers are located uniformly along a segment of length equal to 1. Density is unitary along this segment. There are two stores selling the same physical good and located at the two extremes of the segment. Consumers wish to buy one unit and for this transaction must pick a store. They incur transportation cost t per unit of distance. As is usual, this transportation cost is meant to reflect the facts that products or services are differentiated and that different consumers prefer different products. Let d denote each firm's unit manufacturing/marketing cost (gross of the merchant discount). We normalize d so that it includes transaction costs associated with cash payments. Merchants enjoy benefit b_S per payment card transaction.[20] We assume that

$$\bar{b}_B + b_S > c_I + c_A. \tag{3}$$

[18]See Vives (1999, exercise VI.10) for an analysis of regularity conditions in a symmetric Bertrand oligopoly with differentiated products and CES demands.

[19]See, eg., Tirole (1988).

[20]At this stage we assume homogeneity of merchants. This assumption is relaxed in Section 6, where we allow for merchant observed and unobserved heterogeneity.

If condition (3) were violated, payment cards would generate no social surplus.

Merchants $i = 1, 2$ set their retail prices (p_1, p_2) noncooperatively as in Hotelling's model. They also decide noncooperatively whether to accept payment cards. We assume that the two decisions are sequential: card acceptance is followed by price setting (this is not crucial). Last, for the sake of conciseness, we will focus on "interior solutions". That is, a merchant never corners all consumers of a given type even if he is the only merchant to accept payment cards.[21]

Determination of the interchange fee: We will consider the two cases in which the issuers and a social planner maximizing total surplus, respectively, choose the interchange fee. Acquirers are indifferent as to the level of this fee.

Timing: The timing is as follows:

Stage 1: The interchange fee is set (either by the issuers or by a central planner).

Stage 2: Issuers set fees for their customers, who elect or not to have a card. Merchants decide whether to accept payment cards, and then set their retail prices.

Stage 3: Customers observe the retail prices and whether cards are accepted, and pick a store. If the selected store does not accept payment cards or if the consumer does not own a payment card, the consumer must incur his opportunity cost (b_B) of using the alternative payment method; and similarly the merchant incurs opportunity cost b_S.

4 Socially and privately optimal interchange fees

4.1 Merchant behavior

Let us now analyze the model described in section 3. Let us for the moment take the interchange fee as given. Because $f^*(c_I - a)$, the equilibrium customer fee in the oligopolistic issuing market, is a decreasing function of the interchange fee, the average benefit of a *cardholder*, $\beta[f^*(c_I - a)]$ is decreasing in a: the higher the interchange fee, the lower the customer fee; and so customers

[21] This assumption requires that $\bar{b}_B$ not be too large relative to t.

with lower willingnesses to pay for a card are induced to take a card when the interchange fee increases.

Let

$$m^n(a) \equiv m - b_S = c_A + a - b_S$$

denote the net cost (merchant discount minus merchant's benefit) for a merchant of selling to a cardholder rather than to a consumer using an alternative payment method. Note that this net cost does not embody possible strategic effects of accepting cards in the merchant's competitive environment. Finally, let $\bar{a}$ be uniquely[22] defined by

$$\beta[f^*(c_I - \bar{a})] = m^n(\bar{a}). \tag{4}$$

In words, $\bar{a}$ *is the level of the interchange fee at which the net cost to the merchants is equal to the average cardholder benefit* .

Proposition 1 *Under the no-surcharge rule, there exists an equilibrium in which all merchants accept the card if and only if* $a \leq \bar{a}$.

Proof of Proposition 1

(i) Suppose that consumers expect merchants to accept the card. Then issuers charge $f^*(c_I - a)$, and the demand for cards is $D(f^*(c_I - a))$. Is it indeed optimal for all merchants to accept the card? Suppose they do. Then a merchant's average cost per customer is $d + D(f^*(c_I - a))[c_A + a - b_S] = d + D(f^*(c_I - a))m^n(a)$. As is usual in the symmetric Hotelling model, the equilibrium price p^* is the same for both merchants and is equal to the merchants' marginal cost plus the transportation cost:

$$p^* = \Big[d + D(f^*(c_I - a))m^n(a)\Big] + t. \tag{5}$$

Each merchant's profit is equal to the margin times the market share:

$$\pi^* = \frac{t}{2}. \tag{6}$$

[22]The left-hand side of (4) is decreasing in $\bar{a}$, while the right-hand side is increasing; so there is at most one solution. To prove existence, note that the left-hand side of (4) is bounded, while the right-hand side can take arbitrarily small and arbitrarily large values.

To show this, note that, for given prices (p_i, p_j), merchant i's market share x_i among customers of type b_B is independent of b_B (since a customer pays either cash or with a payment card, independently of the merchant) and is given by

$$p_i + tx_i = p_j + t(1 - x_i),$$

yielding

$$x_i = \frac{1}{2} + \frac{p_j - p_i}{2t}. \tag{7}$$

So, merchant i solves

$$\max_{p_i} \Big\{ [p_i - (d + D(f^*(c_I - a))m^n(a))]x_i \Big\},$$

yielding, at equilibrium, equations (5) and (6).

Suppose now that merchant i deviates from this presumed equilibrium by not taking the card. Consumers with type $b_B < f = f^*(c_I - a)$ don't have a card, and are not affected by merchant i's decision. So, merchant i's market share among these customers is still given by (7). In contrast, merchant i's market share is reduced (for a given price) among cardholders. Among cardholders with benefit b_B, this market share is given by

$$p_i + tx_i = p_j + t(1 - x_i) - b_B,$$

or

$$x_i(b_B) = \frac{1}{2} + \frac{p_j - p_i - b_B}{2t}. \tag{8}$$

Aggregating over all customers (cardholders and non cardholders), merchant i's market share is

$$x_i = \frac{1}{2} + \frac{p_j - p_i - D\,(f)\,\beta\,(f)}{2t}.$$

On the other hand, merchant i's margin has increased to $(p_i - d)$. So, merchant i solves

$$\max_{p_i} \Big\{ (p_i - d)x_i \Big\},$$

yielding price

$$p_i = \frac{1}{2}\left[p_j + t + d - D\,(f)\,\beta\,(f)\right]. \tag{9}$$

On the other hand, the composition of merchant j's market share has changed, since the proportion of cardholders has increased. His profit function becomes:

$$\pi_j = [1 - D(f)](p_j - d)\left(\frac{1}{2} + \frac{p_i - p_j}{2t}\right) + D(f)(p_j - d - m^n)\left(\frac{1}{2} + \frac{p_i - p_j + \beta(f)}{2t}\right). \quad (10)$$

Merchant j's optimal price is therefore:

$$p_j = \frac{1}{2}[p_i + t + d + D(f)(\beta(f) + m^n)]. \quad (11)$$

The equilibrium prices are obtained by solving the system $\{(9), (11)\}$:

$$p_i = t + d - \frac{1}{3}D(f)[\beta(f) - m^n], \quad (12)$$
$$p_j = t + d + \frac{1}{3}D(f)[\beta(f) + 2m^n]. \quad (13)$$

Notice that it is no longer optimal for merchant i to offer the same price as merchant j, because merchant i offers a lower quality service. Equilibrium profits are given by:

$$2t\pi_i = \left[t - \frac{1}{3}D(f)(\beta(f) - m^n)\right]^2 \quad (14)$$
$$2t\pi_j = \left[t + \frac{1}{3}D(f)(\beta(f) - m^n)\right]^2 - m^n\beta(f)D(f)(1 - D(f)). \quad (15)$$

Comparing with (6), we see that the deviation is profitable ($\pi_i > t/2$) if and only if:

$$\beta(f) < m^n.$$

Since $\beta(f)$ is decreasing in a, while m^n is increasing in a, we see that universal acceptance is an equilibrium if and only if $a \leq \bar{a}$, which ends the proof of Proposition 1.

■

There is here less merchant resistance to an increase in the merchant discount than is described in Baxter's early contribution. Baxter assumes that merchants accept the card as long as $m^n(a) \geq 0$. That is, merchants believe that accepting the card does not help attracting consumers. This assumption is legitimate provided the consumers are unaware of which stores accept the card and furthermore still buy when they learn that the shop they patronise does not take the card. Our basic model makes the opposite assumption that consumers are perfectly informed. Section 6.1 analyzes a more general model that embodies the two models as special cases. Generically, though, Baxter's model overstates merchant resistance by ignoring that card acceptance is a competitive instrument.

The proof of Proposition 1 illustrates the externalities involved in the acceptance decision. When merchant i decides to refuse the card, this increases not only the market share of merchant j (for given prices p_i, p_j) but also the average cost of service of merchant j (see (10)), because the proportion of cardholders in his clientele increases. For both reasons, merchant j's equilibrium price increases (see (13)). Furthermore, merchant i's not taking the card makes merchant j's less willing to accept it. A consequence of this strategic complementarity between merchants is that multiple equilibria may exist:

Proposition 2 *For $a > \bar{a}$, card rejection is the unique equilibrium.*

For $a \leq \bar{a}$, either both merchants accept the card or both refuse it. For a equal to or slightly below $\bar{a}$, both merchants' refusing the card is also an equilibrium. The rejection equilibrium is less likely to exist when merchant differentiation increases (that is, if it exists for (a, t), then it also exists for (a, t') with $t' < t$). The set of interchange fees such that the rejection equilibrium exists is included in $(\underline{a}, \infty)$ where $\underline{a} \equiv b_S - c_A$ is the Baxter interchange fee.

Proof of Proposition 2:

When both merchants refuse the card, the Hotelling equilibrium is symmetric and the merchants' profit is

$$\pi^* = \frac{t}{2}.$$

Suppose now that merchant j deviates and accepts the card. His new equilibrium profit is given by (15). Therefore the deviation is unprofitable provided that:

$$t^2 \geq \left[t + \frac{1}{3}D(f)(\beta(f) - m^n)\right]^2 - m^n\beta(f)D(f)(1 - D(f)).$$

When $a = \bar{a}$, $\beta(f) = m^n$ and the above relation is satisfied with a strict inequality. By continuity, it is also satisfied for a close to $\bar{a}$. Last, for $a > \bar{a}$, refusing the card is a dominant strategy for each merchant. In Appendix 2 we show there is no hybrid equilibrium where only one merchant accepts the card.

■

Besides demonstrating the absence of hybrid equilibrium, Appendix 2 provides a much more general study of externalities between merchants. In particular, we derive general formulas that can be applied to the case where several systems coexist, in which consumers can hold several cards, and merchants have to decide which of these cards to accept. These formulas should be useful in future research.

4.2 Determination of the interchange fee

Section 4.1 showed that the "merchant acceptance subgame" admits either a unique equilibrium or two equilibria, depending on the level of the interchange fee. For $a > \bar{a}$, the card is always rejected. For $a \leq \bar{a}$, there exists a "low-resistance equilibrium" in which merchants accept the card; there may further exist a "high-resistance equilibrium" in which merchants reject the card. As we have seen this potential multiplicity arises because one merchant's decision to reject the card raises the other merchant's average cost of serving a customer and therefore makes the latter more reluctant to accept the card. When the two equilibria coexist, they yield the same profit to the merchants, while issuers prefer the low-resistance equilibrium. For expositional simplicity, we will focus on the low-resistance equilibrium described in Proposition 1 (which is essentially the unique equilibrium when merchants have substantial market power). But it will be clear that our welfare conclusions (presented in Proposition 3 below) hold *for any equilibrium selection* as long as $\bar{a}$ is replaced more generally by "the highest interchange fee $\hat{a}(\leq \bar{a})$ that induces merchant acceptance".

Because the issuers' profit is increasing in a, the optimal interchange fee for the issuers is the highest level that is consistent with the merchants' accepting the card, namely $a = \bar{a}$, corresponding to a customer fee

$$f = f^*(c_I - \bar{a}). \tag{16}$$

Suppose instead that a benevolent and omniscient social planner selects the interchange fee so as to maximize total welfare

$$W(f) = [\beta(f) + b_S - c_A - c_I]D(f) = \int_f^{\bar{b}_B} [b_B + b_S - c_A - c_I]dH(b_B).$$

Ignoring in a first step the constraint that the merchants must accept the card, at the socially optimal interchange fee a, the total cost and benefit of the marginal transaction are equal, or

$$f = f^*(c_I - a) = c_I + c_A - b_S.$$

We are thus led to consider two cases:

(i) $c_I + c_A - b_S \leq f^*(c_I - \bar{a})$.

In this case, the socially optimal provision of payment cards requires a low customer fee, which can be obtained only through an interchange fee that exceeds the level at which merchants accept the card. The socially optimal interchange fee is then equal to $\bar{a}$ and thus coincides with the issuers' preferred interchange fee.

(ii) $c_I + c_A - b_S > f^*(c_I - \bar{a})$.

In this case, the socially optimal interchange fee is smaller than the issuers' preferred interchange fee. This means that a payment card association controlled by issuers selects an interchange fee that leads to an *overprovision*[23] of payment card services.

Proposition 3 *Under the no-surcharge rule, the issuers' preferred interchange fee is equal to* $\bar{a}$.

(i) If $c_I + c_A - b_S \leq f^*(c_I - \bar{a})$, *then the socially optimal interchange fee is equal to the issuers' preferred interchange fee.*

(ii) If $c_I + c_A - b_S > f^*(c_I - \bar{a})$, *the interchange fee set by a payment card association controlled by issuers leads to an overprovision of payment card services.*

[23]This overprovision of card services is analogous to the overprovision of payment services by checks that occurs in some countries where regulation prevents banks from charging customers for the use of checks.

Appendix 3 fully solves a linear demand example and shows that each of the two cases envisioned in Proposition 3 may arise. Also, and as noted above, Proposition 3 would continue to hold if we presumed a higher degree of merchant resistance; the only difference is that overprovision of card services would become less likely: a low merchant resistance is the worst case scenario for the social optimality of an issuer-determined interchange fee.

5 Cash discounts and the no-surcharge rule

Let us now investigate the implications of lifting the no-surcharge rule. For concreteness, let "cash" be the alternative method of payment. We ignore the transaction costs associated with merchants' charging two different prices. Even so, allowing card surcharges has ambiguous welfare consequences. In essence, cash discounts raise the cost of payment cards and lead to a suboptimal diffusion of that means of payment.

When merchants are allowed to apply card surcharges, their accepting the card is no longer an issue, since they can charge a price for payment card transactions at least equal to the cash price plus their cost of payment card transactions.[24]

With cash discounts, merchants de facto compete on two segmented markets: that of consumers holding no card and that of cardholders. Let p^*_{cash} and p^*_{card} denote the two prices quoted by the merchants. These prices follow the Hotelling rule (price equals marginal cost plus the differentiation parameter):

$$\begin{aligned} p^*_{\text{cash}} &= d+t \\ p^*_{\text{card}} &= \Big[d+(m-b_S)\Big]+t. \end{aligned} \tag{17}$$

Note that, provided that $m^n(a) = m - b_S > 0$, $p^*_{\text{card}} > p^* > p^*_{\text{cash}}$, where p^* is the no-surcharge price given by (5). The no-surcharge rule leads, as one would expect, to a redistribution towards cardholders.

[24] In our model, merchants face no fixed cost of accepting payment cards. If they did (and that cost were not subsidized by payment card associations), then they might refuse payment cards for a high enough merchant discount.

For customer fee f, a consumer purchases a card if and only if

$$b_B \geq f + [p^*_{\text{card}} - p^*_{\text{cash}}] = f + a + c_A - b_S.$$

The key insight is that *the diffusion of payment cards can no longer be influenced by the interchange fee, since the interchange fee is entirely passed through by merchants to cardholders.* To see this, let

$$\tilde{f} \equiv f + a + c_A - b_S.$$

Then the issuers' margin $\tilde{f} + b_S - c_I - c_A$ and the demand for cards $D(\tilde{f})$ do not depend on a. Thus, in equilibrium, $\tilde{f}$ and market penetration, $D(\tilde{f})$, are independent of the interchange fee. Specifically, $\tilde{f} = f^*(c_I + c_A - b_S) > c_I + c_A - b_S$, which implies that lifting the no-surcharge rule systematically leads to an underprovision of cards.

Last, we compare payment card diffusion and social welfare under the no-surcharge rule and under cash discounts. There are more cardholders under the no-surcharge rule if and only if the net cost of a cardholder for an issuer is smaller under the no-surcharge rule:

$$c_I - a \leq c_I + c_A - b_S,$$

or

$$a \geq b_S - c_A = \underline{a}. \tag{18}$$

Condition (18) is satisfied for the privately optimal interchange fee $\bar{a}$, since $\bar{a} + c_A - b_S = m^n(\bar{a}) > 0$. It is also satisfied for the socially optimal interchange fee. This is obvious in case (i) of Proposition 3 since the privately and socially optimal interchange fees then coincide. In case (ii) of Proposition 3, $f = c_I + c_A - b_S$, and so the issuers' margin, $f + (a - c_I)$, is equal to $a + c_A - b_S$. If condition (18) were violated, then the issuers' margin and profit would be negative, which is impossible at equilibrium.

We thus conclude that card surcharges inhibit the diffusion of payment cards. Intuitively, price discrimination reduces the demand for payment cards. Issuers with market power then have an incentive to focus on the high end of the market rather than attract customers who anyway are not willing to pay much given that they will pay a second premium (beyond the annual fee) when they get to the store.

In terms of social welfare, the analysis is more complex:

• In case (i) of Proposition 3, lifting the no-surcharge rule unambiguously reduces social welfare. This is because in this case the no-surcharge rule leads to an efficient diffusion of cards, while lifting it leads to underprovision.

• In case (ii) however, the no-surcharge rule leads to overprovision ($f = f^*(c_I - \bar{a}) < c_I + c_A - b_S$) while lifting it leads again to underprovision ($\tilde{f} = f^*(c_I + c_A - b_S) > c_I + c_A - b_S$). Thus there is a trade-off between two sources of inefficiency: issuers' market power, which leads to underprovision of cards (when the no-surcharge rule is not imposed) and cross-subsidization between cardholders and non-cardholders, which leads merchants to refuse the card when their net cost is greater than the average (and not marginal, as efficiency would require) cardholder benefit. As a consequence, lifting the no-surcharge rule may increase social welfare when merchant resistance is weak and issuers have little market power.

We summarize the analysis of this section in

Proposition 4

(i) For a given interchange fee, allowing card surcharges raises the merchant price for cardholders and lowers it for noncardholders.

(ii) When the no-surcharge rule is lifted, the interchange fee is neutral and there is an underprovision of card services.

(iii) Lifting the no-surcharge rule reduces social welfare in case (i) of Proposition 3. In the absence of transaction costs associated with the merchants' charging different prices, lifting the no-surcharge rule may increase or reduce social welfare in case (ii) of Proposition 3.

6 Extensions and Discussion

6.1 Determinants of merchant resistance

A key factor in both the positive and the normative analyses is the degree of merchant resistance. Under the no-surcharge rule, lower merchant resistance is more conducive to overprovision, a higher resistance to social optimality of the association determined fee. The analysis of the low-resistance equilibrium described in Proposition 1 unveils the reason why overprovision may occur: Merchants, when deciding whether to take the card, consider the convenience benefit of the *average* cardholder (see equation (4)) rather than that of the *marginal* cardholder as the welfare analysis would command. We now discuss three factors impacting merchant resistance to payment cards. Section 6.3 will discuss yet another determinant of merchant resistance, system competition.

a) Reduced strategic edge of payment card acceptance. We have assumed that customers are always fully informed about individual merchants' acceptance policy. However, some consumers may not know which stores accept the card and may not leave the store once they patronize it and learn that it does not accept the card. Let us therefore consider the more plausible case where only a proportion $\alpha < 1$ of customers are informed of which merchants accept the card before they select a store (or equivalently customers are informed for only a fraction α of their purchases). For simplicity, let us further assume that the consumer either is informed of the acceptance policy of *both* merchants in the market for the good he considers buying (and this with probability α) or is uninformed of their acceptance policy (with probability $1 - \alpha$). The condition for universal acceptance to be an equilibrium becomes:

$$m^n(a) \leq \alpha\beta[f^*(c_I - a)],$$

which implies that the maximum possible interchange fee (given by this condition satisfied with equality) is smaller than $\bar{a}$.[25] Our analysis and Propositions 1 through 4 are otherwise

[25] While consumers may not be aware of whether the two merchants in a particular market take the cards, they have rational expectations (or are informed by the payment card association) as to the fraction (here 1) of merchants accepting the card.

unchanged. As we already noted, the case considered by Baxter (1983) corresponds implicitly to $\alpha = 0$; namely, Baxter assumes that merchants accept the card if and only if the net margin $m^n(a) = c_A + a - b_S$ is negative. In this case, the maximum possible interchange fee is $b_S - c_A = \underline{a}$. When $\alpha = 0$, *the merchants, who for* $\alpha = 1$ *internalized the convenience benefit of the average cardholder, do not even internalize that of the marginal cardholder.* Merchant behavior and issuer market power then both lead to an underprovision of cards.

Proposition 3' *Suppose that the association determined interchange fee induces an overprovision of card services under consumer full information about card acceptance* ($\alpha = 1$)*. Then, there exists* $\alpha^* \in (0, 1)$ *such that the association determined interchange fee is socially optimal if and only if* $\alpha \leq \alpha^*$*. If the association determined interchange fee is socially optimal for* $\alpha = 1$*, then it is also socially optimal for any* α*.*

b) Issuers' pricing stucture. We noted that, under our reasonable assumption of a fixed volume of transactions, it does not matter from the point of view of the issuer-customer relationship whether the issuer charges a fixed yearly fee or a volume proportional fee.[26] Interestingly, this "irrelevant" contractual choice turns out to make a big difference collectively, since it affects merchant resistance.

To see why, let us replace our assumption that the issuers offer yearly fees with no proportional payments by the opposite polar assumption[27] of a "perfect two-part tariff" with marginal cost pricing for the variable part; that is, for a number n (that we have normalized to one) of transactions and yearly fee F_i charged by issuer i, the total charge for the cardholder is:

$$f_i = F_i + (c_I - a)n = F_i + (c - m)n.$$

The analysis of competitive edge of card acceptance summarized in Proposition 1 carries over, as long as the convenience benefit b_B of using the card is replaced by the net benefit $b_B - (c - m)$ for the consumer of a card transaction. Competition among issuers for a given

[26]Recently, many Visa and MasterCard banks have introduced no-fee offerings. In contrast, Visa and MasterCard cards in Europe carry substantial yearly fees.

[27]More general models of price discrimination usually exhibit tariffs that are intermediate between a volume-insensitive fee and a perfect two-part tariff.

interchange fee a and universal card acceptance is unchanged: The equilibrium (now total) fee is still $f^*(c-m)$, and so the equilibrium (symmetric) yearly fee is defined as the residual

$$F^* = f^*(c-m) - (c-m).$$

The highest interchange fee $\hat{a}$, or equivalently the highest merchant discount $\hat{m} = \hat{a} + c_A$, that is consistent with universal card acceptance is therefore given by

$$\hat{m} = b_S + E\left[b_B - (c-\hat{m}) | b_B - (c-\hat{m}) \geq f^*(c-\hat{m})\right] \tag{4'}$$

instead of

$$\bar{m} = b_S + E\left[b_B | b_B \geq f^*(c-\bar{m})\right]. \tag{4}$$

We can now state:

Proposition 1' *Under the no-surcharge rule and when issuers use perfect two part tariffs, there exists an equilibrium in which all merchants accept the card if and only if* $a \leq \hat{a}$. *Furthermore* $\hat{a} > \bar{a}$ *if and only if*

$$\bar{m} > c.$$

Proof of Proposition 1': Only the last sentence requires a proof. To derive it, consider the two functions of m defined by the left-hand side minus the right-hand side of equations (4') and (4), respectively. These two functions are increasing in m, are equal for $m = c$, and that built from (4) is bigger for $m > c$ and smaller for $m < c$.

■

We therefore reach the very intuitive conclusion that variable payments reduce merchant resistance if and only if the interchange fee exceeds the issuer cost. This of course is more than a theoretical possibility. Many Visa and MasterCard banks as well as proprietary cards have introduced inducements for customers to use their card: cashback bonuses (Discover), discounts on products sold by affiliates, travel insurance, frequent flyer mileage, and so forth. In the case of associations, the noncooperative introduction of these volume related payments creates a positive externality among issuers.[28]

c) Intensity of issuer competition. Merchant resistance is indirectly influenced by the intensity of issuer competition. To see this, we parameterize the oligopolistic equilibrium fee by a number τ that increases when competition among issuers becomes more intense. That is, $f = f^*(c_I - a, \tau)$, with f^* decreasing in τ.

The maximum interchange fee that merchants accept is now a function $\bar{a}(\tau)$, implicitly determined by the relation:

$$\bar{a}(\tau) + c_A = b_S + \beta[f^*(c_I - \bar{a}(\tau), \tau)].$$

Since β is increasing, and f^* is also increasing in its first argument, $\bar{a}$ decreases in τ. *As competition among issuers becomes more intense, merchant resistance increases and the interchange fee $\bar{a}$ and the customer fee f decrease.*[29] Issuer competition makes the card available to a wider clientele and thereby lowers the average cardholder's benefit. Merchants are then less inclined to take the card.

d) Observed merchant heterogeneity. Our model can easily incorporate observable merchant heterogeneity: suppose indeed there are K categories of merchants (say supermarkets, grocery stores, gas stations ...) parametrized by $k = 1, \dots, K$. Each is characterized by (observable) merchant benefit b_S^k, and an exogenous transaction volume y^k (as earlier, we normalize total volume to 1: $\sum_k y^k = 1$) The average interchange fee, defined as

$$\bar{a} = \sum_{k=1}^{K} y^k \bar{a}^k,$$

[28]This externality is of course internalized in the case of a proprietary system.

[29]Indeed the implicit function theorem gives:

$$\frac{d\bar{a}}{d\tau} = \frac{\beta' f_2^*}{1 + \beta' f_1^*} < 0,$$

where lower indices denotes partial derivatives. Similarly we can compute the impact on the equilibrium customer fee:

$$\frac{d}{d\tau}[f^*(c_I - \bar{a}(\tau), \tau)] = \frac{f_2^*}{1 + \beta' f_1^*} < 0.$$

depends only on average merchant benefit

$$\bar{b}_S = \sum_{k=1}^{K} y^k b_S^k.$$

Indeed by multiplying each equation

$$\bar{a}^k + c_A = b_S^k + \beta(f^*(c_I - \bar{a})),$$

by y^k and summing over k we obtain:

$$\bar{a} + c_A = \bar{b}_S + \beta[f^*(c_I - \bar{a})]. \tag{4''}$$

Individual interchange fees $\bar{a}^k$ are then given by:

$$\bar{a}^k = \bar{a} + b_S^k - \bar{b}_S.$$

In particular, the net cost of the card is uniform across merchant categories:

$$m^{n,k} \equiv c_A + \bar{a}^k - b_S^k \equiv c_A + \bar{a} - \bar{b}_S.$$

This formula clarifies the nature of externalities between (observable) categories of merchants. Suppose for example that a new category of merchants with a low benefit (say supermarkets) participates in the system: the average benefit $\bar{b}_S$ decreases. Applying the implicit function theorem to formula (4″) shows that

$$0 < \frac{d\bar{a}}{d\bar{b}_S} < 1.$$

Therefore when supermarkets participate in the system, the average interchange fee $\bar{a}$ decreases, and the individual interchange fee of all other categories of merchants increases.

e) Unobservable merchant heterogeneity. We have not considered the possibility of unobservable heterogeneity among merchants since we have assumed that b_S is constant across merchants. If instead b_S is a random variable distributed according to a cumulative distribution function K, the acceptance decision by merchants becomes elastic. For example in the Baxter case ($\alpha = 0$) the proportion of merchants who accept the card becomes $1 - K(c_A + a)$. This modifies the potential surplus that a customer obtains by holding the card, since he will be able to use it only with probability $1 - K(c_A + a)$. Therefore he will hold the card if and only if:

$$b_B(1 - K(c_A + a)) \geq f.$$

The total profit of issuers becomes:

$$\pi_I = [f + (a - c_I)(1 - K(c_A + a))] \left[1 - H\left(\frac{f}{1 - K(c_A + a)}\right)\right],$$

or, denoting by $b_B^0 = \frac{f}{1-K(c_A+a)}$ the valuation of the marginal cardholder,

$$\pi_I = (1 - K(c_A + a))[b_B^0 - c_I + a][1 - H(b_B^0)].$$

This expression is proportional to the previous expression obtained in formula (1) in the case of homogenous merchants, provided that f is replaced by b_B^0. Let us assume that the statistical distribution of merchants is homogenous across issuers, so that this proportionality result also applies to individual profits of each issuer. Then previous formulas for prices and profits in the imperfect competition game between issuers are modified in a simple way: in particular the equilibrium customer fee in the case of heterogenous merchants is equal to the previous equilibrium fee $f^*(c_I - a)$ multiplied by the proportion of merchants $1 - K(c_A + a)$ who accept the card. Similarly, the total profit of issuers at equilibrium is equal to the previous one multiplied by $[1 - K(c_A + a)]$. The issuers' preferred interchange fee depends on the elasticity of the merchants' acceptance function $[1 - K(c_A + a)]$.

6.2 Comparison of business models: association vs for-profit

A for-profit system does not set an interchange fee properly speaking. However, it does have an implicit interchange fee through the level of the merchant discount. Let us therefore consider a for-profit system that either is vertically integrated (as is American Express today) or offers licenses to banks (as was the case for instance for Bank Americard Service Corporation before the creation of the NBI, now Visa). In our model, because issuers are symmetrical, a two-part-tariff license with a fixed payment and a per-transaction payment to the system is equivalent to vertical integration, provided the system offers a license to all issuers. So, we can just assume that the for-profit system is vertically integrated and sets the customer fee f and the merchant discount m directly.

If the merchants all have the same benefit b_S as in section 4, then the for-profit system sets the highest merchant discount, $m = \bar{a} + c_I$, that it can get away with. Hence, there is no difference in merchant discount with an association. The only difference is that the customer fee is in general higher because of the lack of competition on the issuing side (in the vertically integrated case) or the setting of a higher variable assessment (in the licensing case).

More interesting is the case of unobserved heterogeneity among merchants (see section 6.1e). A key difference between the for-profit and the cooperative paradigms is that the former has two separate instruments and optimizes over the merchant discount *and* the customer fee, while in the latter the customer fee is determined by issuer competition once the merchant discount/interchange fee is set. In particular, the cooperative must assess the extent to which an increase in the interchange fee is "competed away" through the competition among issuers. A high merchant discount reduces the issuers' marginal cost; if however this marginal cost saving is mostly passed through to the customers, then the issuers may not gain much from the reduction in their marginal cost and should rather choose a low merchant discount to ensure a wide acceptance of the card.

Let us analyze three standard models of oligopolistic competition among issuers to see whether this intuition is correct. [The formal results are derived in Appendix 4].

Example 1 (Hotelling).

In Hotelling's model of product differentiation, each consumer has a preferred brand and his surplus depends on the "distance" between his own preferred brand and the selected brand's characteristics. Consider (without loss of generality) an Hotelling duopoly in which the two issuers are located at the two extremes of a segment of length one and customers are located uniformly on the segment. If $y(m)$ is the proportion of merchants accepting the card,[30] then

[30] $y(m) = 1 - K(m)$ in Baxter's case; but we can allow for an arbitrary decreasing function of the merchant discount.

the net surplus of a consumer of issuer i located at distance x_i of the issuer is

$$y(m)[b_B - tx_i] - t'x_i - f_i,$$

where t and t' are the parameters of volume-related and fixed differentiation, respectively.

Assume that the elasticity of merchant acceptance $\frac{m|y'(m)|}{y(m)}$ is small for low values of m (technically, for merchant discounts such that the issuer industry is not in the Hotelling competitive region). This assumption is mild since one would expect that almost all merchants would accept the card ($y(m) \simeq 1$) for such low merchant discounts. Proposition 5 below establishes that for the Hotelling model, the association's optimal choice of an interchange fee leads to a smaller merchant discount than would be selected by a for-profit system.

Example 2 (linear demand systems).

Alternatively, suppose that the demand for cards confronted by each issuer is linear:

$$D_i(f_1, f_2, \dots, f_N) = \gamma y(m) - \alpha f_i + \beta \left(\sum_{j \neq i} f_j \right),$$

where α, β and γ are positive, and $y(m)$ reflects the proportion of merchants who accept the card. Proposition 5 shows that the elasticity of total issuers' profit to the merchant discount m is greater for a monopoly than for a Bertrand oligopoly. As a consequence, the merchant discount chosen by a proprietary system is bigger than the one selected by an association of competing issuers.

Example 3 (Cournot).

Proposition 5 shows that under Cournot competition among issuers and linear demand, the merchant discounts are the same for an association and for a for-profit system. The intuition is that a smaller fraction of the decrease in issuer marginal cost brought about by an increase in the interchange fee is passed through to consumers under Cournot then under differentiated Bertrand competition.

Let us summarize these results

Proposition 5 *Let m_P and m_C denote, respectively, the merchant discounts chosen (directly) by a proprietary system and (indirectly) by a cooperative of banks.*

- *When issuer competition is described by the Hotelling model, $m_C < m_P$, under the (weak) assumption that the elasticity of merchant acceptance is small when m is outside the competitive region.*
- *In the differentiated Bertrand model of issuer competition with linear demands, $m_C < m_P$.*
- *When issuers compete la Cournot and demands are linear, $m_C = m_P$.*

6.3 System competition

We consider now a situation in which two associations ($i = 1, 2$) compete for offering payment card services to customers and merchants. We denote by a_i ($i = 1, 2$) the interchange fees chosen by the associations, by $f_i (i = 1, 2)$ the customer fees and by $m_i (i = 1, 2)$ the merchant discounts.

We will not attempt to provide an in-depth analysis of system competition here. We however make two points that demonstrate that intuitions based on competition between for-profit corporations are misleading when applied to associations, and thereby stress the need for further research. We show, first, that competition between two associations need not result in a lower interchange fee, and, second, that even if it does lower the interchange fee, this reduction may lower welfare.

We maintain the assumption that acquirers are competitive: $m_i = c_A + a_i (i = 1, 2)$. Imperfect competition between issuers within each association in general becomes more complicated to model since it is in general influenced by the interchange fee charged by the competing association.[31] Let us therefore look at a few simple cases. Suppose first that each customer holds at most one card and that merchants are homogenous (b_S is constant). Then system competition has no impact on merchant resistance and the analysis of Section 4 is unchanged:

[31] Also, we abstract from the difficulties created by duality, i.e. the fact that issuers typically belong to both associations (see Hausman et al., 1999).

both associations choose the maximum interchange fee that is compatible with merchants' acceptance:

$$a_1 = a_2 = \bar{a}.$$

This is because the incentives of the associations and of the cardholders are perfectly aligned: both want to maximize the interchange fee under the constraint that the card is accepted by the merchants.

Second, keeping with the case in which merchants have a known b_S, $\alpha = 1$, and there is no unobserved heterogeneity among consumers, let us now assume that at least some consumers hold two cards not on the same system. While $(\bar{a}, \bar{a})$ is the equilibrium when consumers hold a single card, it in general is not an equilibrium here. Suppose that system 1 "undercuts" and chooses a slightly lower interchange fee. Then merchants, who for interchange fees $(\bar{a}, \bar{a})$ are indifferent (individually) between accepting a given card and rejecting it, now prefer to reject system 2's card, since the consumer may have the other card in her wallet and this card carries a lower merchant discount. In this case, system competition increases merchant resistance. Note last that from Proposition 3, $(\bar{a}, \bar{a})$ may lead to the socially optimal allocation. Thus, system competition may reduce social welfare by lowering the interchange fee. We of course do not want to draw general welfare implications from this special case, and only want to stimulate further research on this very interesting topic.

6.4 Robustness

Let us finally discuss the robustness of the results to our simplifying assumptions.

a) Elastic demand for retail goods. The assumption of an inelastic demand for retail goods, which may be reasonable in a first step analysis, implies that the interchange fee impacts only the diffusion of payment cards. While a higher interchange fee raises retail prices, when demand is inelastic this increase in retail prices only amounts to a redistribution of surplus between non-cardholders and cardholders, and is neutral from the point of view of aggregate surplus. On the other hand, with a downward sloping demand, changes in retail prices also affect final demand

and thus aggregate surplus. However, the global impact of a higher interchange fee on final demand is ambiguous, because a greater diffusion of cards has also a positive impact on the demand for retail goods (as new cardholders buy more) which may offset the negative impact due to the retail price increase.

b) Vertical integration. As we already noted, if the acquiring business is competitive, there is no strict incentive for an issuer to integrate with an acquirer. Suppose indeed that an issuer merges with an acquirer (or enters the acquiring business) and sets merchant discount m'. The per cardholder profit of the integrated bank corresponding to its cardholders' transactions is:

$$B = (1-\gamma)(f + a - c_I)[1 - H(f)] + \gamma(f + m' - c_I - c_A)[1 - H(f)],$$

where γ is A's share in the acquiring market. That is, a fraction γ of the bank's cardholders' transactions are "on us" transactions. Since the acquiring market is perfectly competitive, γ can be positive only if:

$$m' \leq m = a + c_A.$$

Then

$$B \leq (f + a - c_I)[1 - H(f)].$$

That $m' \leq a + c_A$ further implies that the bank makes no money or loses money on the transactions of cardholders of other banks who transact with the merchants it has signed up. Thus $(f + a - c_I)[1 - H(f)]$ is indeed an upper bound on the integrated issuer's profit. The issuer thus does not gain from operating in the acquiring business.[32]

c) Acquirer market power. The analysis needs to be modified if there is market power on the

[32] The reader may be concerned that the conclusion follows only in the case in which the issuers are (local) monopolies. In principle, there might be strategic effects that could induce the issuing bank to raise its cost of issuing cards by losing money on the acquiring side in order to soften competition in the issuing market. It can be checked this is not so in the Bertrand and Cournot illustrations given in this paper. In the differentiated Bertrand oligopoly model of Appendix 1, the demand for cards is fixed (as long as the equilibrium merchant discount does not exceed $\bar{a} + c_A$ and so there is a payment card market). Even though the issuer loses money on its acquiring transactions, it cannot reduce this loss by losing customers on the issuing side since customers then go to another issuer and still use a card. So, even though the issuer has a higher cost, its opportunity cost of issuing cards is unaffected and there is no strategic effect. In contrast, there is a strategic effect in the Cournot case; however, this effect goes the wrong way for the integrated issuer. In the Cournot model, the integrated issuer reduces its output if it loses money per transaction on the acquiring side. But this induces other issuers to increase their own output, resulting in a further loss for the integrated issuer. We thus conclude that in either model of strategic competition, vertical integration does not increase profit.

acquiring side as well. First, because acquirers now care about the interchange fee, one needs to consider, as Schmalensee (1999) does, the relative strength of issuers and acquirers within the payment card associations. To some extent, the two groups have conflicting interests with acquirers in favor of an interchange fee lower than the issuers' preferred level. Furthermore, there is now some incentive for vertical integration, so as to limit the double marginalization in the provision of payment card services. Second, and from a social point of view, the interchange fee must now reduce two distortions: it must be high in order to subsidize issuers and low so as to subsidize acquirers; a single instrument cannot achieve these two conflicting goals. Furthermore, as Schmalensee (1999) emphasizes, providing proper incentives to both sides in this "moral-hazard-in-teams" problem would require outside funding at the margin. The methodology developed in this paper could be used to analyze the welfare consequences of two-sided market power, but we leave this to future research.

d) Distortions in the provision of alternative payment methods. Our welfare analysis has implicitly assumed that the competing payment methods are efficiently supplied. As is usual, distortions in the provision of the alternative means of payment would lead to a second best situation, in which the interchange fee and the no-discrimination rule should be also assessed in the light of their impact on the alternative means of payment. Of particular interest here are the legal restrictions that exist in some countries on customer charges for checkwriting and the absence of interchange fees for checks, as well as the provision of cash through Automatic Teller Machines (ATMs) which involves a cooperative determination of interchange fees similar to that considered here. Our model can be used as a building block for this broader question, which, again, we leave for future research.

7 Summary and research agenda

To analyze the cooperative determination of the interchange fee, the paper has developed a framework in which banks and merchants may have market power and consumers and merchants decide rationally on whether to buy or accept a payment card. In the absence of unobserved heterogeneity among merchants and under the no-surcharge rule, an increase in the interchange fee increases the usage of payment cards, as long as the interchange fee does not exceed a threshold level at which merchants no longer accept payment cards. At this threshold level, the net cost for merchants of accepting the card is equal to the average cardholder benefit. The interchange fee selected by the payment card association either is socially optimal or leads to an overprovision of payment card services. If the no-surcharge rule is lifted, the interchange fee no longer impacts the level of payment card services. The merchant price for cardholders is increased and that for noncardholders decreased. Merchant price discrimination leads to an underprovision of card services.

A leitmotiv of our analysis has been the central role played by merchant resistance. A first insight is that, in the absence of unobservable heterogeneity, merchants accept the card even though the merchant discount exceeds the technological and payment guarantee benefit they derive from card acceptance. Payment card systems can exploit each merchant's eagerness to obtain a competitive edge over other merchants. Remarkably, though, the interchange fee need not be excessive. The exploitation of the merchants' "prisonner's dilemma" has two benefits from a social viewpoint: On the merchant side it forces merchants to internalize cardholders' convenience benefit, and on the customer side it offsets the underprovision of cards by issuers with market power. In some circumstances, though, the interchange fee may be too high since merchants' incentives are driven by the average cardholder's convenience benefit rather them the marginal cardholder's.

Merchant resistance is affected by several factors. Better consumer information (obtained through advertising or repeat purchases) about which stores accept the card, or an increased consumer willingness to quit the store when discovering it does not accept the card (due to the size of the payment or the proximity of a similar store) lower merchant resistance. Cash-back bonuses or other inducements offered by issuers for card usage also weaken merchant resistance. We would therefore expect associations not to mind when their members offer such inducements (and for-profit systems to make heavy use of these inducements), while in contrast being negatively affected by per-transaction payments charged by issuers.[33] Last, system competition increases merchant resistance when some cardholders have cards on several systems in their pocket.

Apart from our results on the comparison between the merchant discounts generated by cooperatives and for-profit systems, this paper has focused primarily on associations. However, several insights obtained in this paper carry over to for-profit systems. In particular, the analysis of the various factors impacting merchant resistance is unchanged. Still, it would be worth conducting an in-depth analysis of for-profit systems' strategies.

The payment card industry has received scant theoretical attention, and it won't come as a surprise to the reader that more research is warranted. We argued in section 6.4 that the framework developed here can be used as a building block to analyze more general situations with acquirer market power and distorted competing means of payments. The payment card industry offers many other fascinating topics for theoretical and empirical investigation, such as the impact of duality,[34] the governance of payment card associations, the competition between associations and proprietary systems, and the development of E-commerce. We leave these topics and others for future research.

[33] Associations however currently do not prohibit per-transaction payments (which sometimes exist for debit cards).

[34] Duality refers to the fact that banks can (and usually do) belong to both Visa and Mastercard. See Hausman et al (1999) for a start on the analysis of duality.

References

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

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

Carlton, D. W., and A. S. Frankel (1995) "The Antitrust Economics of Payment Card Networks," *Antitrust Law Journal*, 63 (2): 643-668.

Evans, D. S., and R. L. Schmalensee (1993) *The Economics of the Payment Card Industry*, National Economic Research Associates.

—— (1995) "Economic Aspects of Payment Card Systems and Antitrust Policy Toward Joint Ventures," *Antitrust Law Journal* , 63 (3): 861-901.

—— (1999) *Paying with Plastic: The Digital Revolution in Paying and Borrowing*, Cambridge, Ma: MIT Press.

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

Gilbert, R.J. (1990) "On the Delegation of Pricing Authority in Shared ATM Networks," mimeo, University of California Berkeley.

Hausman, J., Leonard, G., and J. Tirole (1999) "The Impact of Duality on Productive Efficiency and Innovation," mimeo, MIT.

Kim, J. (1998) "The Impact of Proprietary Positions and Equity Interest in the Pricing of Network ATM Services," mimeo, MIT.

Laffont, J.J., Rey, P. and J. Tirole (1998a) "Network Competition: I. Overview and Nondiscriminatory Pricing," *Rand Journal of Economics*, 29(1): 1-37.

—— (1998b) "Network Competition: II. Price Discrimination," *Rand Journal of Economics*, 29(1): 38-56.

Matutes, C. and J. Padilla (1994) "Shared ATM Networks and Banking Competition," *European Economic Review*, 38: 1113-1138.

McAndrews, J. (1997) "The Role of Direct Presentment in Non-Cash Payments", Research Paper, Federal Reserve Bank of New York.

McAndrews, J. and R. Rob (1996) "Shared Ownership and Pricing in a Network Switch," *International Journal of Industrial Organization*, 14: 727-745.

Schmalensee, R.(1999) "Payment Systems and Interchange Fees," mimeo, MIT, September.

Seade, J. (1987) "Profitable Cost Increases and the Shifting of Taxation: Equilibrium Responses of Markets in Oligopoly," mimeo.

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

Vives, X. (1999) *Oligopoly Pricing: Old Ideas and New Tools*, Cambridge, Ma: MIT Press.

Appendix 1: Modeling the issuing market as a differentiated Bertrand oligopoly

For the reader's convenience, we have not modeled explicitly in the text the imperfect competition between issuers. In this appendix, we give an example of a differentiated oligopoly model of the issuing market where we compute explicitly the demand for cards $D(f)$ and the average cardholder benefit $\beta(f)$.

Let us consider the case of a duopoly where two banks, $i = 1, 2$, issue each a different card. Consumers perceive these cards as differentiated products as in Hotelling's "transportation cost" model. The two banks are located at the two extremes of a line of length one, on which consumers are located uniformly. The "net" benefit b of a cardholder located in y equals the difference between a "gross" benefit $\tilde{b}$ and a "transportation cost" τy (if he goes to bank 1) or $\tau(1-y)$ (if he goes to bank 2). This "transportation cost" is associated to card usage.[35] For example, some consumers have more use than others for American Airlines frequent flyer miles.

For a given statistical distribution of gross benefits in the population, one can compute the (symmetric) equilibrium on the issuing market, the demand for cards and the average (net) benefit of cardholders. For example, if $\tilde{b}$ is uniformly distributed on $[0,1]$ and f represents the equilibrium customer fee, we easily obtain the (total) demand for cards:

$$D(f) = 2\int_0^{1/2} (1 - f - \tau y)dy = 1 - \frac{\tau}{4} - f,$$

and the average (net) benefit for cardholders

$$\begin{aligned} \beta(f) &= \frac{2}{D(f)} \int_0^{1/2} \int_{f+\tau y}^{1} (\tilde{b} - \tau y) d\tilde{b} dy \\ &= \frac{1}{2D(f)} \left[1 - \frac{\tau}{2} + \frac{\tau^2}{12} - f^2 \right]. \end{aligned}$$

$\beta(f)$ is increasing in f. Note also that the distribution $H(\cdot)$ referred to in the text corresponds to the convoluted distribution, namely the distribution of the *net* benefit.

[35]Our welfare analysis would be modified if the "transportation cost" were associated purely to holding the card. In this case, the welfare analysis would still depend on $\beta(a)$ (which is based on *net* benefits) but the merchants ' decision to accept cards or not would depend on the average *gross* benefit of cardholders.

Appendix 2: A more general formulation of externalities between merchants.

We consider here an extension of the Hotelling game studied in Section 4, where merchants have to make acceptance decisions, represented by an element σ of a finite set S. For example if there is a single system, like in Section 4, S is reduced to the couple $\{\sigma_0 =$ refuse the card, $\sigma_1 =$ accept the card$\}$. If there are two competing systems $\{1, 2\}$, S contains 4 elements each corresponding to a particular subset of $\{1, 2\}$. Each consumer is represented by his "type" τ, which contains all the relevant information about his preferences (parameter b_B) but also about which card(s) he holds. This allows to define two functions:

- $u(\sigma, \tau) =$ gross utility of a consumer of type τ when serviced by a merchant having chosen σ (for conciseness we will call σ the "type" of the merchant).

- $c(\sigma, \tau) =$ cost of servicing a consumer of type τ by a merchant of type σ.

In the example of Section 4, τ is reduced to b_B, since consumers hold the card if and only if $b_B \geq f$. Thus we have

$$\begin{cases} u(\sigma, \tau) & = b_B \, \mathbb{1}_{b_B \geq f} \\ c(\sigma, \tau) & = d + m^n \, \mathbb{1}_{b_B \geq f} \end{cases}$$

Suppose now that merchants 1 and 2 have chosen systems σ_1 and σ_2 and selected (non-discriminatory) prices p_1 and p_2. The market share of merchant 1 among consumers of type τ is:

$$\hat{x}(\tau) = \frac{1}{2} + \frac{1}{2t}[u(\sigma_1, \tau) - p_1 - u(\sigma_2, \tau) + p_2]. \tag{A.1}$$

Multiplying by the profit margins $p_i - c(\sigma_i, \tau)$ and integrating over τ we obtain the total profits

$$\pi_1 = E[\{p_1 - c(\sigma_1, \tau)\}\hat{x}(\tau)], \tag{A.2}$$

$$\pi_2 = E[\{p_2 - c(\sigma_2, \tau)\}\{1 - \hat{x}(\tau)\}], \tag{A.3}$$

where $E[\cdot]$ represents integration over τ.

Using formula (A.1) we get:

$$2t\pi_1 = E[\{p_1 - c(\sigma_1,\tau)\}\{t + u(\sigma_1,\tau) - p_1 - u(\sigma_2,\tau) + p_2\}].$$

This can be transformed by introducing a covariance term:

$$2t\pi_1 = \{p_1 - E(c(\sigma_1,\tau))\}\{t + E(u(\sigma_1,\tau) - u(\sigma_2,\tau)) - p_1 + p_2\}$$

$$-cov\ [c(\sigma_1,\tau), u(\sigma_1,\tau) - u(\sigma_2,\tau)].$$

Similarly,

$$2t\pi_2 = \{p_2 - E(c(\sigma_2,\tau))\}\{t + E(u(\sigma_2,\tau) - u(\sigma_1,\tau)) - p_2 + p_1\}$$

$$-cov\ [c(\sigma_2,\tau), u(\sigma_2,\tau) - u(\sigma_1,\tau)].$$

Solving for the price equilibrium (for given σ_1, σ_2) we obtain:

$$\begin{cases} p_1^*(\sigma) & = t + E(c(\sigma_1,\tau)) + \dfrac{1}{3}\delta(\sigma) \\ \\ p_2^*(\sigma) & = t + E(c(\sigma_2,\tau)) - \dfrac{1}{3}\delta(\sigma) \end{cases}$$

where $\delta(\sigma)$ measures the average competitive advantage of merchant 1:

$$\delta(\sigma) = E[u(\sigma_1,\tau) - c(\sigma_1,\tau) - u(\sigma_2,\tau) + c(\sigma_2,\tau)].$$

Finally the equilibrium profits are given by:

$$\begin{aligned} 2t\pi_1(\sigma) &= \left[t + \frac{\delta(\sigma)}{3}\right]^2 - cov\ (c(\sigma_1,\tau), u(\sigma_1,\tau) - u(\sigma_2,\tau)) \\ 2t\pi_2(\sigma) &= \left[t - \frac{\delta(\sigma)}{3}\right]^2 - cov\ (c(\sigma_2,\tau), u(\sigma_2,\tau) - u(\sigma_1,\tau)). \end{aligned}$$

In any symmetric situation ($\sigma_1 = \sigma_2$) both $\delta(\sigma)$ and the covariance terms vanish:

$$\pi_1(\sigma) = \pi_2(\sigma) = \frac{t}{2}.$$

The above formulas show clearly that the incentives for merchant 1 to deviate from such a symmetric situation depend on two things: the average competitive advantage $\delta(\sigma)$ that he will obtain by doing so, and the incremental cost that he will incur due to the modification of the structure of his clientele. Consider for example the single system model of Section 4, where σ can be A (acceptance) or R (refusal). The above formulas give

$$2t\pi_1(A,A) = 2t\pi_1(R,R) = t^2,$$

and more interestingly:

$$\begin{aligned} 2t\pi_1(A,R) &= \left[t + \frac{\delta(A,R)}{3}\right]^2 - m^n(a)\beta(f)D(f)(1-D(f)) \\ 2t\pi_1(R,A) &= \left[t - \frac{\delta(A,R)}{3}\right]^2, \end{aligned}$$

where $\delta(A,R) = D(f)[\beta(f) - m^n(a)]$.

Therefore it is immediate that (A,A) (i.e., universal acceptance) is an equilibrium if and only if $\pi_1(R,A) \leq \pi_1(A,A)$, which is equivalent to $\beta(f) \geq m^n(a)$. This establishes Proposition 1.

Similarly, (R,R) (universal refusal) is an equilibrium if and only if $\pi_1(A,R) \leq \pi_1(R,R)$, which is equivalent to:

$$\left[t + \frac{1}{3}D(f)(\beta(f) - m^n)\right]^2 \leq t^2 + m^n\beta(f)D(f)(1-D(f)).$$

When $a = \bar{a}$, $\beta(f) = m^n$ and this condition is satisfied with a strict inequality. By continuity, it is also satisfied for a close to $\bar{a}$, which establishes Proposition 2. Finally, (A,R) is an equilibrium iff

$$\left[t - \frac{\delta}{3}\right]^2 \geq t^2,$$

and

$$\left[t + \frac{\delta}{3}\right]^2 - m^n\beta D(1-D) \geq t^2,$$

which is clearly impossible.

Appendix 3: Illustration of Proposition 3.

Example: Let us consider a monopoly issuer and assume that the demand for payment cards is linear. That is, b_B is uniformly distributed on $[\underline{b}_B, \bar{b}_B]$:

$$1 - H(b_B) = \frac{\bar{b}_B - b_B}{\bar{b}_B - \underline{b}_B}.$$

Then

$$f^*(c_I - a) = \frac{1}{2}(\bar{b}_B + c_I - a),$$

and

$$E[b_B \mid b_B \geq f^*(c_I - a)] = \frac{1}{2}\left[f^*(c_I - a) + \bar{b}_B\right]$$

$$= \frac{1}{4}\left[3\bar{b}_B + c_I - a\right].$$

$\bar{a}$ is defined by

$$c_A + \bar{a} - b_S = \frac{1}{4}\left[3\bar{b}_B + c_I - \bar{a}\right],$$

or

$$\bar{a} = \frac{1}{5}\left[3\bar{b}_B + c_I - 4c_A + 4b_S\right].$$

Therefore,

$$f^*(c_I - \bar{a}) = \frac{1}{5}\left[\bar{b}_B + 2c_I + 2c_A - 2b_S\right].$$

The condition for overprovision,

$$f^*(c_I - \bar{a}) < c_I + c_A - b_S,$$

is thus equivalent to

$$c_I + c_A - b_S > \frac{\bar{b}_B}{3},$$

a condition that is compatible with the condition that payment cards generate social benefits:

$$c_I + c_A - b_S < \bar{b}_B.$$

Thus, for a monopoly facing a linear demand curve, the two cases envisioned in proposition 2 are possible.[36]

Appendix 4: Comparison of Business Models.

Example 1 (Hotelling).

By symmetry, a for-profit system will set the customer fee f and the merchant discount m

[36] In the alternative case of a demand with constant elasticity, it turns out that the first case is impossible: There is always overprovision of cards.

so as to maximize its total profit

$$2x(f-(c-m)y(m)),$$

where x is the location of the cut-off customer, defined implicitly by

$$(b_B-tx)y(m)-t'x-f=0.$$

Using x (instead of f) as a decision variable, the maximum profit of the system (for a fixed m) is thus given by

$$\pi_S(m)=\max_{x\leq 1/2} 2x\left[(b_B-tx-c+m)y(m)-t'x\right].$$

When t' is small enough, this maximum is attained for $x=1/2$, and the system will select the merchant discount m_S that maximizes

$$\pi_S(m)=\left(b_B-\frac{t}{2}-c+m\right)y(m)-\frac{t'}{2}.$$

Consider now an association of competing issuers. For low values of m, these issuers are in fact local monopolies, so that the total profit of the association, $\pi_A(m)$, is the same as that of the proprietary system derived above. However for m large enough, the issuers do actually compete. In this competitive region, and for fees $\{f_1,f_2\}$, the cut off customer is given by

$$y(m)[b_B-tx]-t'x-f_1=y(m)[b_B-t(1-x)]-t'(1-x)-f_2,$$

or

$$x=\frac{1}{2}+\frac{f_2-f_1}{2[ty(m)+t']},$$

and the profit of issuer 1 is

$$\pi_1=[f_1-(c-m)y(m)]x.$$

The equilibrium fee is symmetric and equal to:

$$f^*=[(c-m)y(m)]+[ty(m)+t'],$$

resulting in profit per issuer

$$\pi^* = \frac{1}{2}[ty(m) + t'],$$

which decreases in m.

This competitive region obtains as long as the cut off customer obtains a positive net surplus i.e.

$$y(m)\left[b_B - \frac{3t}{2} - (c-m)\right] \geq \frac{3t'}{2}.$$

When the elasticity of y is small (as we assume below) the left hand side of this inequality increases in m, and the condition is equivalent to $m \geq m_A$ where

$$y(m_A)\left[b_B - \frac{3t}{2} - (c-m_A)\right] = \frac{3t'}{2}.$$

The association will select a smaller merchant discount than the for-profit system if and only if $m_A \leq m_S$ (see Figure 2 below), which is satisfied (under classical concavity assumptions) whenever:

$$\pi'_S(m_A) = (b_B - t/2 - c + m_A)y'(m_A) + y(m_A) \geq 0,$$

which is equivalent to

$$\frac{|y'(m_A)|m_A}{y(m_A)} \leq \frac{m_A}{m_A + b_B - t/2 - c}.$$

Using the definition of m_A, one can see that this condition is satisfied whenever

$$\frac{|y'(m_A)|m_A}{y(m_A)} \leq \frac{m_A}{t}.$$

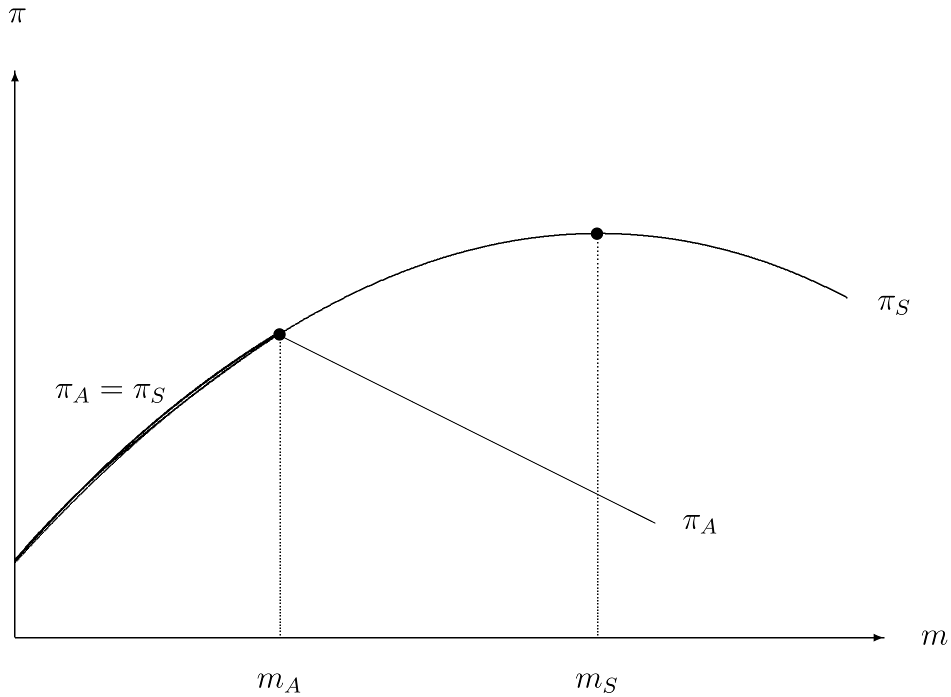

Figure 2: profits π_A of the association and π_S of the for-profit system as a function of the merchant discount m.

Example 2 (linear demand systems).

In the case of linear demands

$$D_i(f_1, \dots f_N) = \gamma y(m) - \alpha f_i + \beta \left(\sum_{j \neq i} f_j \right),$$

the profit of each issuer is given by

$$\pi_i(f_1, \dots, f_N) = \{f_i - (c - m)y(m)\} D_i(f_1, \dots, f_N).$$

We assume $2\alpha > \beta(N - 1)$, to guarantee that profits are concave.

Now π_i is maximum in f_i (for given f_j, $j \neq i$) when

$$2\alpha f_i - \beta \sum_{j \neq i} f_j = \{\gamma + \alpha(c - m)\} y(m).$$

The equilibrium fees are symmetric and given by:

$$f_1 = \dots = f_N = f^* = \frac{\{\gamma + \alpha(c - m)\} y(m)}{2\alpha - \beta(N - 1)}.$$

The total equilibrium profit of an association of issuers is

$$\pi_A(m) \equiv N\alpha \left[y(m) \left\{ \frac{\gamma - (c-m)(\alpha - \beta(N-1))}{2\alpha - \beta(N-1)} \right\} \right]^2.$$

The profit of a proprietary system (unique issuer) is obtained by taking $n = 1$ in the above formula:

$$\pi_S(m) \equiv \alpha \left[y(m) \left\{ \frac{\gamma - \alpha(c-m)}{2\alpha} \right\} \right]^2.$$

Given that $\beta > 0$, it is easy to see that π_A is less elastic to m than π_S:

$$\frac{m\pi_A'(m)}{\pi_A(m)} < \frac{m\pi_S'(m)}{\pi_S(m)}.$$

As a consequence, the value of m that maximizes π_A, which we denote again by m_A, is smaller than m_S, the corresponding value for π_S.

Example 3 (Cournot).

Consider now a Cournot oligopoly with N issuers and linear demand. After normalization, the inverse demand can be written

$$f(Q) = y(m)(1 - Q),$$

where $Q = q_1 + \ldots + q_N$ is the total number of cards issued. The profit of issuer i is

$$\pi_i(q_1, \ldots, q_N) = y(m)[1 - Q - (c - m)]q_i.$$

It is maximum (for fixed $q_j, j \neq i$) when

$$1 - Q - (c - m) - q_i = 0,$$

yielding a symmetric equilibrium:

$$q_1 = \ldots = q_N = \frac{Q}{N} = \frac{1 - (c-m)}{N+1}$$

and equilibrium total profit:

$$\pi_A(m) = Ny(m) \left[\frac{1 - (c-m)}{N+1} \right]^2.$$

Again the profit of a proprietary system is obtained by taking $N = 1$ in the above formula

$$\pi_S(m) = y(m) \left[\frac{1 - (c - m)}{2} \right]^2 .$$

It is immediate that π_A / π_S is independent of m so that in this case $m_A = m_S$.