A MODEL OF THE CARD PAYMENT SYSTEM AND THE INTERCHANGE FEE

PAUL DE GRAUWE
LAURA RINALDI

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the CESifo website: www.CESifo.de*

A Model of the Card Payment System and the Interchange Fee

Abstract

We analyse a card payment system to assess the economic impact of the interchange fee. This fee is paid by the bank of the merchant, the acquirer, to the bank of the consumer, the issuer. We build up a mode in order to explore whether the interchange fee can enhance the participation to the system and thus increase the benefits of the system as a whole. We also investigate what kind of incentives the interchange fee provides to the actors of the system. We show that the interchange fee selected by the card associations differs from the one that would be selected by a welfare maximising social planner. Moreover the interchange fee selected by the market is likely to produce a smaller market for cards than the socially optimal one.

JEL Classification: E42.

Paul De Grauwe
University of Leuven
Naamsestraat 69
3000 Leuven
Belgium
Paul.DeGrauwe@econ.kuleuven.ac.be

Laura Rinaldi
University of Leuven
Naamsestraat 69
3000 Leuven
Belgium

1   Introduction

Card use is expanding in all industrialised countries for both retail and remote transactions. Despite the wide and diverse range of card schemes active in the different countries, some common features can be identified in almost all schemes independently of their geographic dimension (local or international scheme) or specific functionality (credit card or electronic purse). One of these constant features, and probably also the most controversial one is the interchange fee (henceforth IF). In a card scheme there are three or four agents (three or four-party scheme). In the case of a three-party system there is a buyer, a seller and a card company. The latter signs up sellers and issues cards for buyers. In a four-party scheme buyer and seller draw a contract with two distinct financial institutions[1], respectively the issuer and the acquirer. The IF applies to four-party schemes and is a fee paid by the acquirer to the issuer[2]. Several motivations are put forward to justify the IF. Card associations claim that the IF is necessary to equilibrate the cost imbalance between acquirer and issuer, to internalise network externalities, to remunerate one agent for services supplied, to enhance

---

[1] Or at least not necessarily the same institution.

[2] This is the most common case. However there exist also schemes with 'reverse' or 'negative' interchange in which case the IF flows from acquirer to issuer.

1

network participation for the benefit of all users.... Because the IF appears to be simply a transfer between financial institutions, one might think that it has little economic impact.

This paper deals with the precise role played by the IF. In particular, we explore how the determination of the interchange fee affects the network's participation and efficiency. The motivation for studying such an issue stems from two facts. First, the interchange fee was recently questioned by several competition authorities[3]. Second, despite these concerns, the IF, first introduced for credit cards, is being extended to numerous other card schemes and banking services in general.

We provide a microeconomic model in which the IF is decided alternatively by a profit-seeking card association and by a benevolent social planner who maximises welfare. We then compare the two outcomes.

The paper is organised as follows. Section two deals with the characteristics of network industries and the particular features of card networks. In section three we provide a short review of the existing literature on card payment systems.   In section four we illustrate the main assumptions and set-up of the model. The model is then developed in section five assuming that the IF is chosen by a card association and we derive the private optimal interchange, the price level and the demand for payment services. In section six we perform the same analysis assuming that the IF is chosen so as to maximise public welfare. We then discuss the main findings of the model and we compare the public and private optimal situations. We also provide a numerical application of the model using data from the Belgian payment system. Finally, in section seven we conclude.

2   Cards as a network industry

2.1   Characteristics of network industries

Network industries are a very peculiar type of industry, with some characteristics distinguishing them from all the others. Examples of network goods and services are fax machines, computer software, e-mail, video players, some banking services.... What distinguishes these industries is the fact that they exhibit positive consumption externalities. Such externalities, also called network externalities, imply that the value of a unit of the good increases with the number of units sold, or to be more precise, the value increases with the expected number of units to be sold. It follows that no utility is derived from the consumption of the good in itself, while the larger the network, the larger the utility derived from the good.

Network externalities can have a strong impact on the development of the market because the adoption of a certain good depends on consumers' expectations with regard to the size

---

[3] For example see European Commission (2000) and Reserve Bank of Australia (2000).

of the network. Depending on expectations a network product could either develop very fast or never develop at all. In order for a network to develop, a critical mass of users must be reached. Only when the critical mass is reached all other potential consumers will be induced to join the network. As a consequence, other characteristics enter into play in network industries. These are complementarity and compatibility, switching costs and lock-in, and economies of scale in production (Shy 2001).

The complementarity refers to the fact that consumers buy systems of goods rather than individual goods. For example video players are useless without video tapes. Hence video players and video tapes are complements. This leads to the compatibility problem: complementary products must operate on the same standard, i.e. must be compatible. In this regard the question arises of whether different firms benefit from adopting a common standard. In case they indeed use the same standard, they must coordinate their decisions, while competing among each other.

When a product becomes widespread, it is costly for the consumer to switch from one product (technology) to another one, because of the learning effort involved or because of contract and search costs. Thus, switching costs are also typical of network goods. The size of switching costs determines to what extent consumers are locked-in. When consumers are locked-in "firms may raise prices knowing that consumers will not switch to a competing brand" (Shy, 2001, p.5).

Another recurrent feature of network industries is that production has very high sunk costs but almost zero marginal costs. For example a new software requires large development costs, but can be distributed via internet virtually without cost. Because of the large economies of scale involved, these markets are often characterised by a few dominant leaders, able to make large investments first and then exploit the economies of scale.

2.2   Peculiarities of the card industry

The card industry is also a network industry, i.e. it is subject to all the characteristics mentioned before. In fact, a card by itself provides no benefit to its holder: the benefit arises when the merchants are willing to accept it. Thus, potential consumers will increasingly be induced to buy a card as many merchants accept it. At the same time, the larger the number of cardholders, the more merchants will consider it worthwhile to accept cards. Thus, network externalities are crucial to the development of card payments. In fact a cardholder does not care directly about the total number of cardholders, but instead cares about the number of merchants accepting it. Nonetheless, indirectly, individual cardholders benefit from other consumers using the same card. This feature is what makes card networks peculiar. In card networks there are two different kinds of final users. This contrasts with

other network goods, like the telephone, that have only one type of user. Someone having a telephone can communicate with all other telephone-holders. As with other network industries, there exists a compatibility problem in the card network. This problem, however, will not be addressed in this paper[4].

With regard to switching costs, for cardholders these might take the form of loosing the benefit of loyalty programs or loosing access to a certain group of known merchants. More relevant are switching costs on the merchant's side. Apart from learning and training considerations, a merchant who decides to switch from cards to another payment instrument (or from one card to another one) faces an important cost in terms of reduced demand. In fact, by accepting a certain card, merchants can reach new customers. Thus a decision to stop acceptance would mean to take the risk of loosing these additional customers. It follows that once the merchant enters a card network he is locked-in.

Finally, the card industry is also subject to significant economies of scale. Large investments are needed in order to build up the infrastructure for the transmission of the information within the four-party network. However, once the infrastructure has been built, the cost of adding an additional user (being merchant or consumer) to the network is negligible.

2.3   The Interchange Fee

Another peculiarity of a card network is the interchange fee. To understand the interchange fee, it must be noted that there are two kinds of card schemes: three-party and four-party schemes. In three-party schemes there is only one player (the card association) dealing with the payer and the payee. The same institution signs up merchants, issues cards for consumers and decides the respective fees. This is the case for example of American Express and Diners Card. The more common four-party systems are characterised by two (possibly) different institutions taking care of payer and payee, respectively the issuer and the acquirer. A four-party system is often based on the interchange fee. This is a service fee paid by one institution to the other one. The IF finds its justification in the fact that the production of the card service is divided among two suppliers. As already mentioned, because of network externalities, a cardholder does not only care about the card he receives, but also about wide merchant acceptance. However, the latter does not belong to the responsibilities of the issuer (with whom the cardholder has a contract) but of the acquirer. Since on both

---

[4]In case some suppliers of card services decide to become compatible, i.e. to operate under the same standard, the level of externalities increases exponentially, producing a non negligible impact on the market for such cards. A good example of a similar strategy is represented by Visa and MasterCard. Their decision to adopt a common standard has resulted in a large expansion of their respective market and allowed them to become a leading worldwide system.

the issuer and the acquirer's side there are positive spill-overs, the interchange can take any direction or not being present at all. There is no economic rationale justifying a certain direction of this fee, nor is it clear how large it should be or whether it should be fixed or proportional to the transaction's value[5]. However, in practice usually the interchange flows from acquirer to issuer, as issuers seem to have higher costs than acquirers.

Recently, the IF was called into question by central banks and antitrust authorities[6]. This caused concerns about the precise role it plays. It should be noted that in network industries the success of a company in the market does not only depend on how efficient it is with respect to its competitors, but also on the timing and on the previously established market power.

The main purpose of this paper is to explore to what extent the IF can be efficient from an economic point of view. As one observer notes: "Interchange may be a tool that helps to build an efficient payment system. But it may also be a tool to keep an inefficient system alive. In the end the consumer would pay for such a system by higher prices. The trouble is that there is no easy way to decide which option is true. In fact, at the same time while the pricing policy of credit card companies is coming under fire some observers are arguing that cash is inefficient, keeping prices high and harming consumers" (Krueger 2001).

3   Related Literature

The first article in the economic literature dealing explicitly with card payment systems was Baxter (1983). Baxter argued that in order to maximise the benefit of the card payment system merchants and consumers should be priced at their marginal cost. According to Baxter the action of both issuers and acquirers determines the size of the payment system. The interchange fee does nothing more than reallocating their costs so that it cannot have any anti-competitive effect. However, Baxter's analysis does not take into consideration any incentive effect arising from the IF. For example in his model consumers do not know which merchant accepts cards. In addition card's acceptance has no impact on merchant's profits so that the merchant accepts cards only if he can afford their technological cost.

Schmalensee (2000) also investigates the card payment system, but focuses on the question of whether the multilateral fixation of the IF can be considered as a collective price fixing practice[7]. The author sets up a model with monopolistic issuer and acquirer where,

---

[5] Some card companies define the interchange as a demand management tool to maximise the usage of the card service, characterised by joint supply and joint demand. Others define it as a fee for precise services provided by one side to the other one. In sum, different card companies provide different explanations for the IF, but they all use it.

[6] To give an example, the Australian competition authority and the European commission have dealt with this issue.

[7] This is the main object of concern of the antitrust authorities.

like in Baxter (1983), the total size of the card network depends on the their joint effort. Network externalities are such that the value of the payment system to the issuer is affected by the behaviour of the acquirer and vice versa. The IF is determined through a two-stage game and plays the role of shifting costs from one side to the other. Schmalensee finds that increasing the IF also increases the size of the whole network. Moreover, Schmalensee shows that when demand functions are linear, the profit maximising IF is also welfare maximising. When the model is generalised to oligopolistic issuers and acquirers, the profit maximising IF departs from the output maximising one in the same direction as the welfare optimal IF. However, Schmalensee does not investigate whether issuers and acquirers have incentives to exploit the locked-in condition of merchants and the possible repercussions of this, in particular whether this can lead to distortions of retail pricing.

Rochet and Tirole (2000) develop a model of bank card payments were issuers have market power, while acquirers are characterised by perfect competition. Consumers and merchants decide on whether to enter the card network and the analysis concentrates on the strategic behaviour of merchants. More specifically, consumers take a card only if the card fee does not exceed the benefit from cards. The authors explore under what conditions a situation in which all merchants accept cards is an equilibrium. They find that there is a threshold level of the IF such that the acceptance of all merchants is an equilibrium[8]. If the IF is higher, no merchant entering the card network is also an equilibrium. Rochet and Tirole also find that in equilibrium under the Non Discrimination Rule (NDR)[9], the IF chosen by issuers will be either equal to the welfare maximising IF or higher, thus leading to over-provision of cards[10]. In contrast, without the NDR, the IF no longer affects the diffusion of cards. In this case the IF set by issuers can lead to under-provision of cards.

Wright (2000) and (2001) extend the model of Rochet and Tirole with Bertrand and monopolistic competition for merchants. Wright's model focuses on the efficiency of the NDR rule more than on the efficiency of the interchange fee and, as Rochet and Tirole, explores under what conditions all merchants accept cards. A new feature introduced by Wright is that he does not restrict the number of transactions, which allows him to distinguish between consumers' membership to the card network and their decision to use the card. This highlights the credit function of credit cards. He finds that in the case of Bertrand-type competition and if the NDR rule holds, merchants pass on to consumers all costs of the card

---

[8] Merchants compete according to the Hotelling model.

[9] This rule states that merchants cannot surcharge or offer discounts to consumers depending on the instrument they use to pay for the good.

[10] The realisation of one or another equilibrium depends on merchants' reaction. The authors write (p.21): "Low merchant resistance is more conducive to overprovision [of cards], and higher resistance to social optimality of the association determined fee". However in the real world, the achievement of a high merchant resistance equilibrium, where no merchant accepts cards, seems quite unlikely to happen, mainly because of coordination problems. Thus we must conclude that, in practise, card associations are very likely to select an IF higher than the socially optimal.

system by increasing prices. This has the consequence that nobody uses cards. In this case all benefits from card payments are eroded. The same occurs in the Hotelling model when there is no NDR. In general Wright finds that the NDR is beneficial in that it prevents merchants from extracting the additional surplus arising from card use. Without the NDR merchants would increase their prices and consumers would not be willing to sustain the cost of entering the card network, with the consequence that there would not be a card system.

4   A model of card payments

In the initial phase, the development of the network crucially depends on the attainment of a critical mass. This means that in the starting phase it is vital for the network to achieve a sufficiently large participation of merchants and cardholders. Merchants will decide to accept cards only if the profit they obtain from accepting both cards and currency is at least equal to the profit from cash transactions. In other words, in order to enter the network merchants must receive positive marginal profit from card acceptance, while consumers must have a positive utility from card usage. The card association and the financial institutions belonging to the network decide the interchange and the fees' level. The interaction between consumers and merchants and their banks allows the network to become sufficiently big and to move to a stable phase.

We assume that during this starting phase the number of cardholders and merchants participating to the network increases (and exceeds the critical mass) while the price level remains unchanged. This price level - and the resulting interchange and fee levels - that is compatible with a high level of card development is an equilibrium situation.

4.1   Assumptions

I. The so-called Non Discrimination Rule (NDR) holds: merchants are not allowed to surcharge cardholders or to offer discounts for cash payments.

II. Payments are by cash or card and there is only one payment card scheme.

III. Some consumers hold a card and they always prefer to use it when possible (i.e., when the seller accepts it).

IV. Consumers are fully informed about the acceptance attitude of merchants.

V. Issuers and acquirers have some degree of market power.

4.2 The agents

Issuers: Issuers bear some operating costs c^I and receive an interchange fee a, from acquirers for each transaction. Usually they also receive fees from cardholders[11]. We assume that cardholders pay a fee to issuers, $f^B = \beta$, where β can be positive in the event cardholders actually pay a fee, or zero. In the latter case issuers implicitly subsidise cardholders in order to stimulate their participation[12].

The profit function for issuers, π^I is as follows

$$\pi^I = \left[a + f^B - c^I \right] D_c$$

where D_c is the number of transactions made with all cards.

Acquirers: Acquirers also have operating costs, c^A assumed to be fixed, and pay a share of the transaction amount to issuers, the interchange fee, a[13]. They also collect fees from merchants: a fixed fee from every merchant f_0^S and a transaction-based fee *ad valorem* $f^S(a) = a + \phi$, with $\phi > 0$ being the transaction-based operating cost (and profit) of the acquirer[14]. Then, acquirers' profit π^A is:

$$\pi^A = \left(f_0^S - c^A \right) N^S + \phi D_c$$

with N^S being the number of all merchants belonging to the card scheme.

Consumers: Consumers use cash and cards.

For cash, consumers bear a cost mainly constituted by the risk of loss, theft, float and the time spent in trips to their bank. We call these costs m_0, and assume these to be fixed[15]. The utility value of money v_0, is the value consumers derive from paying by cash. So consumer's net utility from cash is

$$U_m^B = v_0 - m_0.$$

For cards we assume that cardholders pay a transaction-based fee, f^B to the issuer. This fee is sometimes equal to zero, because as card-issuers receive a revenue from acquirers, they can avoid charging cardholders. Card associations claim that this subsidization in favour of cardholders, made possible thanks to the interchange fee, is beneficial for the network as it

---

[11] However card issuers often claim that these fees are not sufficient to cover the costs incurred, so that in practise they provide cardholders with a positive transfer per transaction.

[12] The superscript "B" stands for buyer.

[13] To be more precise there is another entity in between card issuer and acquirer, that is the card company. The card company also receives a share of the interchange fee but that is omitted for sake of simplicity. In any case this does not affect the main conclusions.

[14] The superscript "S" stands for seller.

[15] With the exception of the float, these costs are clearly a fixed amount.

allows for a larger participation to the network and consequently increases the value of the network as a whole[16].

Consumers' valuation of cards depends on the network's expected size, q^e. Such positive valuation $v(q^e)$, is equivalent to a reduction in the marginal cost or an increase in the demand function. The cardholder's net utility takes the form

$$U_c^B = v(q^e) - f^B. \tag{1}$$

Merchants: We assume that merchants do not bear any costs for the goods they sell. They only have costs related to the payment instruments they accept. As a consequence, the final price that we derive is the amount just necessary to cover the transaction costs involved in the provision of payment services. For notes and coins merchants have a fixed cost, $T_m > 0$ which is related in the time spend in storing, counting, providing the change, etc.

For cards merchants pay a fixed cost (renting or purchase of terminals, telephone line etc.) f_0^S, and a proportional fee per card transaction $f^S(a)$.

Thus the cost for accepting cards by merchants is:

$$c_c^S = f_0^S N^S + f^S(a) D_c \tag{2}$$

where c_c^S is the merchants' total cost for card payments, D_c is the demand for such card transactions and N^S is the number of merchants accepting cards.

4.3 Two-phase development of the card network

In the first phase of the network development merchants decide to accept cards only if their activity based on card and currency acceptance is at least as profitable as their activity based only on cash transactions. Of course card's acceptance involves several benefits for merchants, but this is not enough to convince them to bear additional - and sometimes high - costs. In order to offer an additional payment solution, merchants have to gain additional sales (i.e. new customers). This idea is in line with Rochet and Tirole (2000, p.3) "A merchant's total benefit and thus its decision whether to accept a card depends not only on the merchant's technological benefit (fraud control, theft protection, speed of transaction, customer information collection,...), but also on the product of its increase in demand due to system membership and its retail markup". In other words, merchants must get non-negative marginal revenue from card transactions. This marginal revenue will be very large at the beginning, and decreases with the size of the network. Thus we measure merchants utility as the profit they derive from their activity.

---

[16]In reality cardholders usually pay a fixed cost (usually an annual fee), but we neglect it for sake of simplicity.

As the network becomes larger and reaches the critical mass of both consumers and merchants, the marginal revenue of card transactions for the merchant approaches zero. At this stage the merchant cannot refuse cards anymore because the loss from not accepting them would be much larger than the cost of cash (their opportunity cost of cards increases). Card associations know that merchants are, in a sense, "locked-in". Obviously they will now try to extract the maximum profit from merchants and therefore they have strong incentives to increase merchant fees. Since the merchants' elasticity with regard to the fees they pay tends to zero, card companies can increase merchant fees (and the IF) without loosing any merchants. Put differently, the monopoly power of card companies increases.

We assume that for some level of $a > a^*$ the acceptance condition holds till a certain level of a (a^{**}) that is too high. So for $a^* \leq a \leq a^{**}$ the merchant's profit from accepting only cash transactions, π^S, is smaller or equal to the profit derived from accepting both card and cash payments[17]. Let $p_c = p(a^*)$ be the common price for cards and cash payments set by a merchant that accepts cards and, D_m and D_c the demand for goods purchased by means of notes and coins and by card respectively

$$\pi^S(D_m) \leq \pi^S(D_m + D_c) \tag{3}$$

or, put differently, the profit from card transactions must be non-negative,

$$\pi^S(D_c) = p_c D_c - c_c^S \geq 0 \tag{4}$$

with c_c^S defined according to equation (2).

4.4   A linear city model

We consider a standard Hotelling model of a linear city with a continuum of risk-neutral buyers. The buyers are indexed by x and are uniformly distributed along the segment of length equal to 1. We assume that if they hold a card they always prefer to use their card rather than currency when possible. There are two merchants located at the extremes of the segment. They sell identical goods but are differentiated insofar as they offer different payment services. In particular one accepts only cash payments while the other also accepts cards. Potential consumers decide whether to take a card and choose a store incurring transportation costs t, reflecting the fact that the services offered by merchants are differentiated[18]. Merchants decide whether to accept cards and the price level non-cooperatively.

Note that we exclude corner solutions, i.e. we exclude the possibility for a single merchant to capture the whole market. Even if all buyers were cardholders it could be that

---

[17] Note that the utility of the merchant from a particular payment instrument corresponds to the profit he makes from the sales paid with that instrument.

[18] This is in addition to the costs and benefits mentioned in section 4.2.

a cardholder, even knowing that a merchant does not accept cards, still buys from him because it is closer to his location (in which case t is larger than the card's opportunity cost).

The actions of the agents take place in three stages:

Stage I: Issuers and acquirers (or alternatively a central planner) set the interchange, a, and consequently the fee level for merchants and consumers.

In the first case, the financial institutions, given their costs' structure, set the cardholder fee, f^B and the merchant fee, $f^S(a)$ so as to maximise their joint profit. In the alternative case a welfare maximiser sets interchange and user fees.

Stage II: Merchants observe the fee imposed on them and set the price level so as to satisfy their acceptance condition involving positive profits. Because of the no-surcharge rule[19], merchants must set the same price for cards and cash payments.

Stage III: Consumers decide whether to take a card, and what means of payment to use when buying a good. Consumers opt for a merchant given his price and knowing whether he accepts cards or not. If the merchant does not accept cards and the customer is a cardholder, the latter faces an opportunity cost equal to the card benefit.

5 Private determination of price and fees

We solve for the sub-game perfect equilibrium interchange fee set by the bank association and consequently the fee level imposed to merchants and cardholders.

At the last stage (stage III), buyers - who might hold a card or not - choose a particular merchant for their purchase, thus deciding what payment instrument to use.

The utility function of consumers indexed by x is

$$U_x^B = \begin{cases} v(q^e) - f^B - tx - p_c & \text{if they pay by card} \\ v_0 - m_0 - t(1-x) - p_m & \text{if they pay with cash} \end{cases} \qquad (5)$$

where t is the differentiation parameter for merchants and can be interpreted as spatial distance between them or, as different payment options available to their customers; we assume t to be bigger than zero. Finally p_c is the price for goods sold by merchant A accepting cash and cards and, p_m the price imposed by merchant B, accepting only notes and coins (see Figure 1).

As already mentioned in section 4.2, the utility function of consumers who pay by card exhibits network externalities since it increases with their expectations about the size of

---

[19]See Assumption I.

network[20]. Similarly to Schmalensee (2000) we assume that the value of transactions is $v(q^e) = f(x, y) = \alpha(xy)$, where x is the consumers' willingness to carry and use cards, y represents the merchants' willingness to accept cards and α is a positive measure of the network externality and is smaller than t[21]. In equilibrium expectations are fulfilled so that the expected size of the network equals its actual size.

Let $\widehat{x}$ denote a particular cardholder who is indifferent between buying from merchant A or B (see Figure 1). Thus from (5) we derive the consumers aggregate demand for card services,
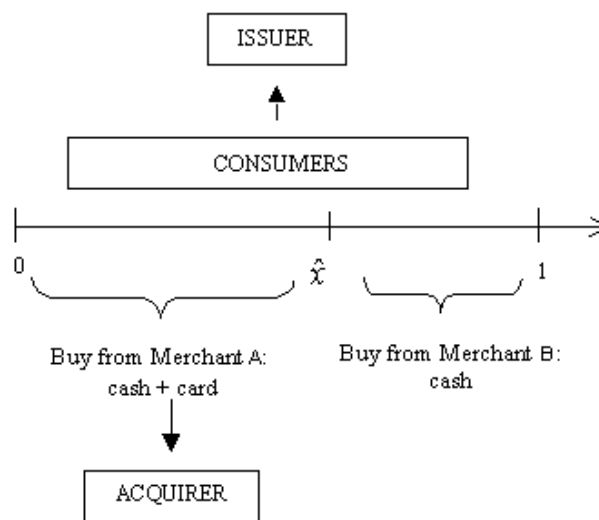
Figure 1: **The linear city model**

$$v(q^e) - f^B - t\widehat{x} - p_c = v_0 - m_0 - t(1 - \widehat{x}) - p_m$$

where $v(q^e) = \alpha x$, since $y = 1$ (in our linear city model only one merchant accepts cards), which implies:

$$D_c = \widehat{x} = \frac{t - v_0 + m_0 - f^B}{2t - \alpha} + \frac{p_m - p_c}{2t - \alpha} \tag{6a}$$

$$D_m = 1 - \widehat{x} = \frac{t - \alpha + v_0 - m_0 + f^B}{2t - \alpha} + \frac{p_c - p_m}{2t - \alpha}. \tag{6b}$$

---

[20] It should be noted that the cash-network also benefits from positive externalities. However this aspect is not considered here as in the model we do not take into consideration the whole circle and agents related to the cash network. Moreover, since notes and coins are legal tenders, no fees are imposed on them and are often perceived as for free by buyers and sellers. Finally the cash-network is not characterised by such a thing as the interchange fee in order to internalise the externalities.

[21] This assumption ensures us a downward sloping demand for card services.

Note that since one merchant accepts only cash and the other cash and cards and, since according to Assumption III, cardholders always prefer to use their card when possible, it follows that the demand for the two merchants corresponds respectively to the demand for card payments, D_c and, the demand for cash payments, D_m.

In stage II, the seller decides whether to accept cards and sets its price given the costs he incurs for cards and currency and given the demand he faces. From section 4.3 the card acceptance condition for merchant A implies the following problem,

$$\pi_A^S = (p_c - f^S(a))D_c - f_0^S \geq 0. \tag{7}$$

Thus merchant A will determine the price level for his goods paid both by cash or card, p_c, so to maximise his profit. Substituting (6a) into (7) we obtain

$$\max_{p_c} \pi_A^S = \left[p_c - f^S\right] \left(\frac{t - v_0 + m_0 - f^B}{2t - \alpha} + \frac{p_m - p_c}{2t - \alpha}\right) - f_0^S.$$

Similarly, merchant B chooses his price p_m to maximise his profit given by

$$\pi_B^S = (p_m - T_m)D_m, \tag{8}$$

substituting (6b) into (8) merchant B solves the following problem,

$$\max_{p_m} \pi_B^S = (p_m - T_m) \left(\frac{t - \alpha + v_0 - m_0 + f^B}{2t - \alpha} + \frac{p_c - p_m}{2t - \alpha}\right).$$

The prices derived from the merchants' problem are as follows:

$$p_c = t + \frac{m_0 - v_0 + T_m - \alpha - f^B + 2f^S}{3} \tag{9a}$$

$$p_m = t + \frac{v_0 - m_0 + 2T_m - 2\alpha + f^B + f^S}{3}. \tag{9b}$$

The last two equations deserve some attention. First, as expected both prices increase with the differentiation parameter, t. Second, from (9a) one can see that p_c decreases with the fee paid by consumers participating to the network, f^B. This means that if cardholders are subsidised (i.e. $f^B = \beta = 0$), the final price for goods will be higher. Third, α, the network externality effect, reduces optimal prices, while the interchange fee raises prices, with p_c rising twice as much as p_m (recall that $f^S = a + \phi$). Finally, prices increase with

the fixed costs related to cash, T_m and with acquiring costs, ϕ (again with p_c rising twice as much as p_m).

In sum, if the NDR is in place, a merchant accepting cards must set a common price for cards and currency. This price decreases with the cardholder fee, implying that if buyers are directly charged for card services, they pay a lower price. Conversely, if buyers receive incentives to participate to the system by receiving card services for free, the final price they pay for goods will be higher. The price level also increases with the interchange fee and with the merchant's differentiation parameter; the same holds also for the price set by the merchant accepting only cash. However the effect of network externalities decreases the price of both merchants.

5.1 Optimal interchange fee

In light of these results, we obtain the demand equations

$$D_c \;\; = \;\; \frac{3t - \omega - \alpha - f^B - f^S}{3\,(2t - \alpha)} \tag{10a}$$

$$D_m \;\; = \;\; \frac{3t + \omega - 2\alpha + f^B + f^S}{3(2t - \alpha)} \tag{10b}$$

where $\omega = v_0 - m_0 - T_m$.

In stage I we substitute the new demand functions found, (10a) and (10b), in the profit function of the two financial institutions supplying card services. These set fees for merchants and cardholders so as to maximise their joint profit Π[22],

$$\begin{aligned} \Pi \;\; &= \;\; aD_c - c^I D_c + f^B D_c + f_0^S N^S + f^S(a)D_c - c^A N^S - aD_c \tag{11} \\ &= \;\; f^B D_c + f^S(a)D_c - c^I D_c + f_0^S - c^A \end{aligned}$$

where N^S is the number of merchants accepting cards, which in our linear city is equal to 1. Substituting the value of f^B and $f^S(a)$ and maximising the joint profit with respect to the interchange fee we obtain the following first order condition,

$$- \frac{a + \phi + \beta - c^I}{3(2t - \alpha)} + \frac{3t - \omega - \alpha - \phi - a - \beta}{3(2t - \alpha)} = 0$$

from which we derive the profit maximising level of the interchange fee:

$$a^* = \frac{3t - \omega - \alpha + c^I - 2\phi - 2\beta}{2}. \tag{12}$$

From expression (12) it appears that the IF could be positive or negative depending on the parameters of the model, which is in line with the literature and with the existing schemes[23].

It is not clear from this expression what role the interchange plays in internalising externalities. On the one hand, the network externality effect α has a negative impact on interchange, that is to say that the larger the network, the smaller is the optimal interchange. On the other hand, the cardholder's fee β has a negative impact on the IF. This implies that if issuers decide to subsidise consumers' participation by setting $\beta = 0$, the optimal IF will increase. Similarly a higher transport cost and higher issuer's technical costs will also need a higher interchange fee. Thus, in order to be able to say something about the impact of the interchange fee we have to see in what way it affects merchants' market share.

Another consideration about equation (12) concerns its practical applicability. To be efficient the IF should be set on the basis of very detailed information about issuers and acquirers' cost structure, but also about consumers valuation of cash and cards. However, in the real world networks are sometimes transnational with a multitude of merchants and consumers with different tastes and many issuers and acquirers with different cost structure and profit preferences. Collecting such information is not an easy task for a card association. A wrong perception about individuals' preferences could easily lead to a non-optimal interchange fee.

Substituting the privately optimal interchange fee (12) in the demand functions (10a) and (10b), we obtain the equilibrium market share of each merchant

$$D_c^* = \frac{3t - \omega - \alpha - c^I}{6(2t - \alpha)} \tag{13a}$$

$$D_m^* = \frac{9t + \omega - 5\alpha + c^I}{6(2t - \alpha)}. \tag{13b}$$

Note that as the two merchants together serve the whole market their demand adds up to 1. The equilibrium profit level for acquirer and issuer together is obtained by substituting a back in (11),

$$\Pi = \frac{\left(3t - \omega - \alpha - c^I\right)^2}{12(2t - \alpha)} + f_0 - c^A. \tag{14}$$

---

[23] Usually the interchange fee is positive (flowing from issuer to acquirer), but is can also flow in the opposite direction. For instance, in Australia the EFTPOS debit card scheme is based on a reverse interchange fee bilaterally set between banks.

Finally, we substitute (12) into (9a) to find the equilibrium price,

$$p_c = \frac{6t - 2\omega - 2\alpha + c^I - 3\beta}{3}.$$ (15)

It is easy to see that a variation in the cardholders' fee β, implies an equal and opposite change in the price. Network externalities also have the effect of decreasing prices.

Finally,

$$p_m = \frac{9t + v_0 - m_0 + 5T_m - 5\alpha + c^I}{6}.$$

Note that while cardholder fees do not enter the equation for p_m, the network externality effect does. Surprisingly network externalities decrease the cash-price more than the card-price.

6   Welfare

We now analyse the same type of card network when the interchange is determined by a benevolent social planner maximising welfare, rather than by a card association. The social planner wants to achieve the largest possible network at the smallest cost.

Welfare is defined in the usual way, i.e. welfare is the sum of consumers' and producers' surplus. In particular, in our economy, welfare is given by the sum of consumers' utility, merchants' utility (U^S) and the profit of issuers (π^I) and acquirers (π^A).

$$W = \int_0^{\widehat{x}} U_x^B dx + \int_{\widehat{x}}^1 U_x^B dx + U^S + \pi^I + \pi^A$$
$$= \omega(1 - \widehat{x}) - c^I \widehat{x} - c^A - \frac{t}{2}(1 - \widehat{x})^2 - \frac{t}{2}\widehat{x}^2 + \frac{\alpha}{2}\widehat{x}^2$$ (16)

Since fees and the interchange fee are transfers among financial institutions, merchants and cardholders, they all cancel. As a result, social welfare is reduced to the net value to consumers and merchants from currency minus technological costs incurred by issuer and acquirer, net of aggregate differentiation cost. In addition there is also a last term due to the network externality effect, $\frac{\alpha}{2}\widehat{x}^2$, which has a positive effect on welfare. In other words, the effect of network externalities decreases the cost of aggregate differentiation of the payment services in the economy.

From equation (16) one can see that welfare is not directly affected by the interchange fee. Notwithstanding that, indirectly welfare still depends on IF as consumers' demands, $1 - \widehat{x}$ and $\widehat{x}$, are functions of a. It can be shown that an increase of the interchange fee reduces total welfare. We find the level of interchange that maximises welfare as

$$a^w = 3c^I + 2\omega - \alpha - \phi - \beta.$$ (17)

Based on this formulation of the IF we obtain the market share of merchant A and B respectively that are associated with the maximum welfare. In other words, we find the portion of consumers that makes purchases paying by card, D_c^w and the portion that pays with notes and coins, D_m^w

$$D_c^w = \frac{t - \omega - c^I}{2t - \alpha} \tag{18a}$$

$$D_m^w = \frac{t - \alpha + \omega + c^I}{2t - \alpha}. \tag{18b}$$

6.1 Optimal interchange fees and externalities

In the previous section we determined the functional forms of the interchange fee that would be chosen by a social planner maximising welfare and the associated demand for card services. In the present section we investigate how changes in the level of network externalities, α, affect the interchange and the market for cards. Moreover, we compare the results from the welfare set-up with the results of the profit maximising setting.

In order to do so we have to introduce some simplifying assumptions. We first assume that $\phi = 0$, that is to say that the acquirer charges the merchant only for the interchange and not for its operating cost or profit. In addition we assume $\beta = 0$, meaning that cardholders do not pay any per-transaction fee[24]. We also assume that $p_c = p_m$ and that ω is a positive value. This last assumption reflects the idea that the net utility of cash to its users is larger than zero[25]. With these assumptions, equations (13a) and (18a) remain unchanged, while (12) and (17) are simplified to

$$a^* = \frac{3t - \omega + c^I - \alpha}{2}, \quad \text{and} \quad D_c^* = \frac{3t - \omega - c^I - \alpha}{6(2t - \alpha)} \tag{19}$$

and,

$$a^w = 3c^I + 2\omega - \alpha, \quad \text{and} \quad D_c^w = \frac{t - \omega - c^I}{2t - \alpha}. \tag{20}$$

These equations are plotted in Figures 2, 3 and 4. In the upper chart, we set out the level of interchange fee on the vertical axis and the level of network externalities on the horizontal axis. In the lower chart we present the level of demand for card services on the vertical axis and the level of network externalities on the horizontal axis. The three figures show the same curves for different values of t. Remember that $\alpha < t$, thus the zone of interest is only the shaded area.

---

[24] If $\beta \neq 0$, i.e. if cardholders pay a fee for cards, the results of the comparison between the two optimal interchange fees does not change. In fact, β affects a^* and a^w exactly in the same way. The same is valid for ϕ.

[25] We believe that such assumption is not unrealistic, given the so widespread use of currency throughout the countries.

A first observation from these figures is that, for any t, the interchange fee decreases in α. This means that if a network is small - and as a consequence externalities are also small - the optimal interchange should be very high. The larger the externalities become, the smaller is the IF. This reflects the idea that the interchange is a device that should allow the network to reach the critical mass faster. However as the externalities become larger, the welfare optimal interchange decreases faster than the profit maximising one, which suggests that a^w can better internalise the positive externalities.

The situation is different for the demand for card services. If t is small (see Figure 2), meaning that consumers perceive cash and card services as very similar, then the market for cards decreases in α. In this case the externalities are too weak to make the card service valuable. In contrast, with a large t, i.e. when the service offered by merchants is perceived to be more differentiated (figures 3 and 4), the demand for cards increases with network externalities, producing the outcome typical of network goods: the larger is the network, the larger the utility and thus the demand for the service.

Let us now compare the results from the two set-ups[26]. Figure 2 shows what happens when t is small, where small is defined as $t < \frac{3\omega + 2c^I}{3}$. One can see that the welfare-optimal interchange, a^w lies above the private-optimal interchange a^* and both are positive. This means that when the differentiation parameter for merchant services is small and, consequently there are few network externalities[27], the benevolent social planner in order to internalise such externalities would set the IF at a level that is higher than the one set by a private card association. This higher interchange fee results in a demand for card services that is not only lower than the market produced by the privately optimal IF, but is even negative. In sum, when t is small and externalities are small too, if a social planner sets the IF there is no market for cards. While with the privately optimal IF there exists a market for cards for small enough levels of α. When α becomes large the market for cards also vanishes. It follows that if t is small, the benefits of cards do not become widespread and the network never develops.

Figure 3 shows the interchange fee and the demand when $\frac{3\omega + 2c^I}{3} \leq t \leq \frac{5}{3}(\omega + c^I)$. In this case for low levels of α the welfare optimal interchange fee is higher than the profit maximising one. However, as a^w decreases faster, it becomes smaller than a^* when externalities are large enough. The demand for card services follows the opposite trend: at the beginning the demand deriving from a^w is smaller, but as a^w goes down, the corresponding demand becomes larger and moves above the private optimal demand.

---

[26]We also derived the optimal interchange fee and demand for card services using alternative welfare definitions. These alternative definitions identified welfare only with users' surplus. However they produced results very similar to the one showed here. Thus the main conclusions are the same, whatever welfare definition we use.

[27]Since α has to be smaller than t.
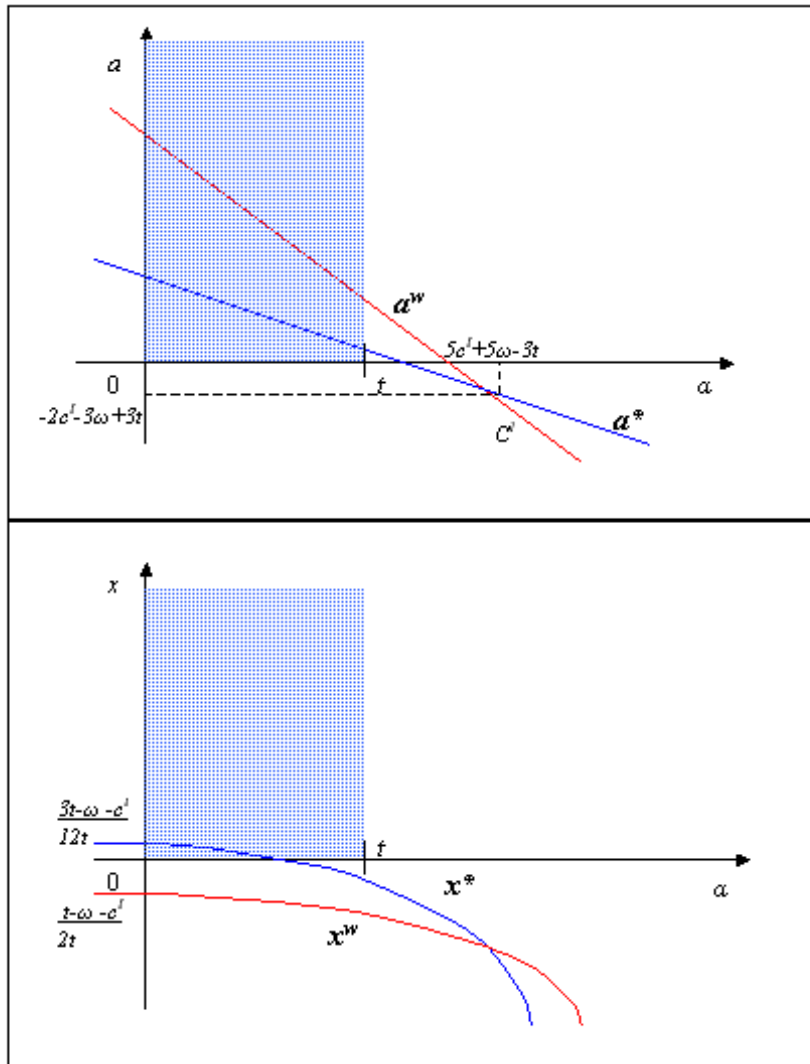
Figure 2: **Interchange fee and market for cards, small t**
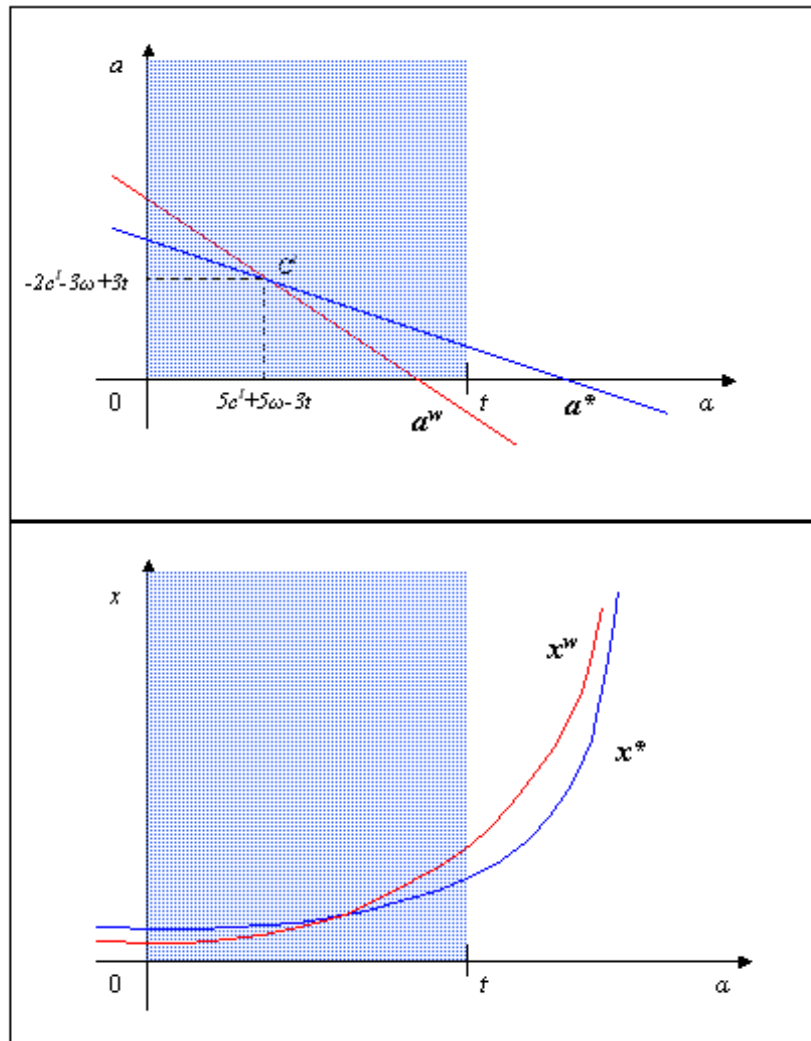
Figure 3: **Interchange fee and market for cards, intermediate t**

Finally, when t is very large ($t > \frac{5}{3}(\omega + c^I)$, see Figure 4), the private optimal IF is always larger than the social optimum IF and yields a demand for card services that is always smaller than the one derived from the welfare maximisation.
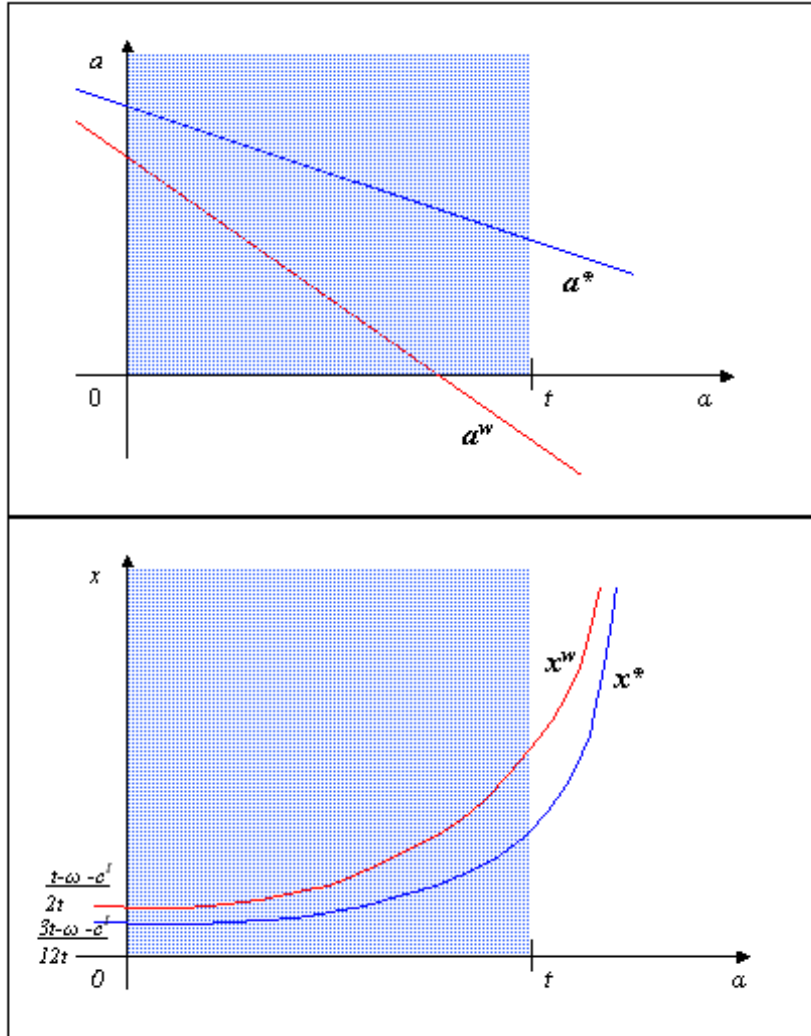
Figure 4: **Interchange fee and market for cards, large t**

6.2 An application of the model

In the previous section we compared the private and the socially optimal interchange fee by introducing some simplifying assumptions on the variables. In this section we compare again the two optimal IFs. However, this time we do so by using real data concerning the Belgian payment system. The figures we use come from the study by De Grauwe *et al.* (2000).

Again we consider equations (12) and (17) and we substitute the euro values into ϕ, c^I and β[28]. It was not possible to identify a proxy for ω, hence we simplify by assuming $\omega = 0$. Thus we obtain:

$$a^* = \frac{3t - \alpha - 0.74}{2} \tag{21}$$

and

$$a^w = 0.18 - \alpha \tag{22}$$

It follows that the private optimal interchange fee is likely to be positive for a wide range of values of t and α. By contrast, the welfare maximising IF is always negative unless externalities are very small (smaller than 0.18). In particular, a^* turns out to be larger than the welfare maximising interchange fee; a^* is smaller than a^w only in the case both α and t are particularly small. Therefore the Belgian example seems to suggest a situation similar to the case represented in Figure 4, where the privately optimal IF is larger than the welfare maximising one. Moreover a^w could be negative. Thus, according to this particular example a benevolent social planner would exploit the externalities typical of the card network by using a smaller interchange fee with respect to the one that would be chosen by profit seeking card associations.

7   Conclusion

In this paper we have investigated the functioning of card payment schemes paying a particular attention to the role played by the interchange fee. We developed a linear city model with two shops. The shops sell identical goods, but offer different payment possibilities. One merchant accepts only currency, while the other also offers card payment services but sells his goods at a unique price (according to the "non-discrimination rule"). We then derive the optimal interchange fee, as it would be set by a private card association and by a social planner maximising welfare. These two resulting interchange fees produce different consumer prices and demands' for cash and for card services.

We find that the optimal interchange fee always decreases as network externalities increase, independently of the way the interchange is determined. This finding suggests that the interchange fee, interpreted as transfer flowing from the acquirer to the card-issuer, is

---

[28] In Belgium there are three card schemes active for retail transactions: electronic purses, debit cards and credit cards. The figures we use in this section refer to the weighted average for the three types of card. Therefore we derive a value for the variables in question as if there was a single card scheme as opposed to cash. We find $\phi = 0.07$, $c^I = 0.22$ and $\beta = 0.42$.

particularly useful in the starting phase of the network development. However as the size of the network increases, the optimal interchange will go to zero or become negative, i.e. flowing from the issuer to the acquirer. As far as the optimal demand for card services is concerned, if the distance between the two shops and the level of externalities is small, the card network either never develops or dies out quickly. In this case the network size does not appear to be positively related to the level of network externalities. However, when the distance dividing the two shops becomes larger together with the level of externalities, the demand for card services grows with the externalities, at the beginning slowly, but then very fast. Comparing the two settings, we generally find that the welfare optimal interchange fee internalises the externalities better than the profit maximising one. Depending on the differentiation parameter for merchants, t, the private optimal interchange fee can be smaller or larger than the welfare optimal one. However, if t is large enough - and thus also the externalities -, the private optimal IF is larger, even if it shows a decreasing pattern. This IF is reflected in the pattern of demand for card services insofar as it produces a network that grows more slowly as externalities become stronger and, is smaller in size than the welfare optimal card network.

We also provide a numerical application of the model using data from the Belgian payment system. We find that the welfare optimal IF is negative for sufficiently high levels of externality. Conversely, the interchange fee that would be selected by a card association is very likely to be positive. Thus the Belgian payment system seems to reflect the last case of our theoretical analysis insofar as the IF selected by the market lies above the welfare optimal IF. This situation is very likely to produce a card payment system smaller than the optimal one.

One possible interpretation of our findings is that the IF may indeed create incentives for the creation and expansion of a card payment network, as reflected in a high welfare optimal IF for small levels of externalities. Nevertheless, as the installed base of merchants and cardholders increases (which is translated in large externalities), there is no longer a need to provide such incentives. As a consequence, the welfare optimal IF decreases faster and may become negative. However, card associations know that when the installed base of users is reached, merchants are very unlikely to drop out. At this stage, card associations have relatively little incentive to reduce the IF, even if the economies of scale would allow it. Thus, such a mechanism can explain why the privately optimal IF decreases more slowly than the welfare optimal one and is generally larger. As a result, merchants are induced to raise retail prices at the expense of final consumers and ultimately of the card system itself. Thus, there is indeed some ground for the concerns raised by the competition authorities. In any case we believe that this fascinating topic needs more investigation. It would be

interesting, for instance, to explore the same kind of model we used in this paper in the case surcharging is allowed. We leave this topic for further research.

References

[1] Balto D., 2000, The problem of Interchange fees: costs without benefits?, E.C.L.R., Issue 4.

[2] Baxter W., 1983, Bank Interchange of Transactional Paper: Legal and Economic Perspectives, Journal of Law and Economics, 26, 541-588.

[3] Bouckaert, J. and H. Degryse, 2000, Price competition between an expert and a non-expert, International Journal of Industrial Organization, 18, 901-923.

[4] Economides N., 1996, The Economics of Networks, International Journal of Industrial Organization, October.

[5] European Commission, 2000, Commission plan to clear certain Visa provisions, challenge other, Press release, DN: IP/00/1164, Brussels, October 16.

[6] De Grauwe P., Buyst E. and L. Rinaldi, 2000, The costs of cash and cards compared. The cases of Iceland and Belgium, mimeo, University of Leuven.

[7] Krueger M., 2001, Interchange fees in the land of fire, ePSO Newsletter, 4, January.

[8] Reserve Bank of Australia, 2000, Debit and credit card schemes in Australia, A study of interchange fees and access, October.

[9] Rochet J. and J. Tirole, 2000, Cooperation Among Competitors: The Economics of Payment Card Associations, mimeo, University of Toulouse.

[10] Schmalensee R., 2000, Payment Systems and Interchange Fees, December, mimeo, MIT.

[11] Shy O., 1995, Industrial Organization, Theory and applications, MIT Press: Cambridge.

[12] Shy O., 2001, The Economics of Network Industries, Cambridge University Press: Cambridge.

[13] Sutton, J., 1998, Technology and market structure, The MIT Press: Cambridge, Massachusetts.

[14] Wright J., 2000, An economic analysis of a card payment network, December, mimeo.

[15] Wright J., 2001, Optimal card payment systems, April, mimeo.

Appendix 1: The variables of the model

Table 1: List of variables	
a	interchange fee paid by issuer to acquirer
$D_c = \widehat{x}$	demand of the merch. A accepting both cards and cash
$D_m = 1\text{-}\widehat{x}$	demand of the merch. B accepting only cash
t	transport cost
α	network externality effect
$f^B = \beta$	fixed fee paid by cardholder to issuer
$f^S = a + \phi$	merchant fee paid to acquirer
c^A	acquirer's technological cost
c^I	issuer's technological cost
$\omega = v_0 - m_0 - T_m$	users' net utility from cash payments
$v_0 - m_0$	consumer's net utility from cash payments
T_m	merchant cost related to cash

CESifo Working Paper Series

(for full list see www.cesifo.de)

_____