

August 11, 2010



ESMT WORKING PAPER

ESMT No. 10-005

Pricing Payment Cards

Özlem Bedre-Defolie, ESMT
Emilio Calvano, Bocconi University

ISSN 1866-3494

Abstract

Pricing Payment Cards¹

Author(s):* Özlem Bedre-Defolie, ESMT
Emilio Calvano, Bocconi University

Payment card networks, such as Visa, require merchants' banks to pay substantial "interchange" fees to cardholders' banks, on a per transaction basis. This paper shows that a network's profit-maximizing fee induces an inefficient price structure, over-subsidizing card usage and over-taxing merchants. In contrast to the literature we show that this distortion is systematic and arises from the fact that consumers make two distinct decisions (membership and usage) whereas merchants make only one (membership). These findings are robust to competition for cardholders and/or for merchants, network competition, and strategic card acceptance to attract consumers.

Keywords: payment card networks, interchange fees, merchant fees

JEL Classification: G21, L11, L42, L31, L51, K21

¹ The authors would like to thank Bruno Jullien, Patrick Rey, Jean-Charles Rochet, seminar participants at MIT, U of Rochester, Brown, Purdue, HEC-Montreal, Toulouse School of Economics, U of College London, U of Pompeu Fabra, INSEAD-Paris, U of Maastricht, European School of Management and Technology, Norwegian Business School, Copenhagen Business School, Royal Holloway U of London, DIW, U of East Anglia (CCP), CEMFI-Madrid, U of Bocconi, Luiss-Rome, Collegio Carlo Alberto, European University Institute, U of Bogazici, U of Koc; European Central Bank, New York FED, Richmond FED, Central Bank of Norway, Bank of Italy, Ente Einaudi and participants at various conferences and workshops for helping us improve this paper.

* Contact: Özlem Bedre-Defolie, ESMT, Schlossplatz 1, 10178 Berlin, Germany;
Tel: +49 (0)30 212 31-1531, Email: ozlem.bedre@esmt.org.

1 Introduction

Spending at merchants on US credit, debit or prepaid cards topped almost a quarter of the US GDP in 2008¹, up 6.1 percent from 2007. Debit cards, e.g., Maestro and Visa debit, are expected to overtake cash as the primary method of payment by 2012. US merchants pay their banks on average 1.8 percent of every plastic card transaction to get the payment cleared. Such fees are the second-highest expense for many businesses after labor costs, exceeding the price of health care insurance for employees.² Cardholders in contrast are typically offered complementary benefits and services to use their cards at checkout counters. In some cases, up to 5 percent of the value of the transaction is returned to them under the form of “cash back” bonuses.³ This asymmetry is mainly due to the payment networks’ (e.g., MasterCard and Visa) practice of charging merchants’ banks a per transaction fee (called the “swipe” or interchange fee) and turning over the proceedings to the cardholders’ banks to increase the usage and issuance of cards.

The question that we address in this paper is whether the networks’ pricing policies promote an efficient use of these payment instruments. We study this issue, at least, for two reasons.

First, along with the growth of card transactions has come increased regulatory scrutiny of the industry practices worldwide. The concern is that these skewed pricing policies inflate retailers’ costs of card acceptance without enhancing the efficiency of the system. Many regulatory authorities and legislators, including the European Commission and the U.S. House of Representatives, either already intervened with specific interchange fee regulation or are considering the opportunity of doing so.⁴ In addition, various coalitions of trade associations, representing businesses that accept card payments, are seeking additional interventions and promoting class action lawsuits (interchange litigation), alleging, among other things, that the implicit price agreements that banks enforce through the networks’ pricing policies violate antitrust legislation.

Second, despite shedding a great deal of light on the workings of the industry, the existing literature delivers no straightforward normative implications when both consumer and merchant demands are assumed to be elastic. The relationship between socially and privately optimal card prices depend on *quantitative* considerations, e.g., surplus measures hinging on cost and preference attributes. Assessing distortions requires a significant amount of information, and in principle

¹That is more precisely 3.285 trillion dollars.

²Sources: Nilson Report (July 2009, issue 929, pp. 1,9); Nilson Report (April 2007, Issue 895, pp. 7); Hayashi (2009).

³For instance, as of December 2009 BoA rewards usage of its check cards with frequent flyer miles, cash back bonuses (Keep the Change) and insurance against theft of the merchandise. <http://www.bankofamerica.com/checkcard/>. The AMEX Blue Cash Card offers 5 percent cash back as of January 2010.

⁴After forcing MasterCard to cut its cross-border interchange fees to around zero, the European Commission (EC) is currently investigating Visa’s fees. (see MasterCard case: COMP/34.579, and Visa cases: COMP/29.373, COMP/39.398.). The U.S. congress issued “Credit Card Fair Fee Act of 2008” and is pondering further measures in two new pieces of legislation: the “Credit Card Interchange Fees Act of 2009” and the “Expedited CARD Reform for Consumers Act of 2009”. Price cap regulations on various fees (mainly interchange fees) to protect merchants from excess charges, have already been applied in Australia, Canada, Norway, Singapore, Switzerland, Mexico, Chile and Denmark. Other countries (e.g., U.K., Sweden, Brazil) have regulated card networks’ rules and agreements, or have outlawed them, aiming to reduce merchant fees.

interventions could go in either direction.

This paper shows that the profit-maximizing price structure over-subsidizes card usage at the expense of charging inefficiently high fees to merchants. This result is obtained in a very broad setup that nests several previous contributions. In contrast to existing formulations our model distinguishes consumers' card membership decision from card usage decision. This allows us to spell out the implications of an important asymmetry between consumers and merchants: consumers make two distinct choices (membership and usage) whereas merchants make only one (membership). This asymmetry is shown to be the ultimate root of the above distortion. We argue that this source of inefficiency is particularly important since it follows from a structural feature of the industry rather than from differences in idiosyncratic attributes between the two groups of users.

We develop a model of payment card pricing in which each bank is allowed to charge a two-part tariff to its customers, which could be either consumers or merchants, in return for providing access to one payment network. A two-part tariff consists of a fixed fee, which is paid when a customer signs up or renews a contract with her bank, and a marginal (per-transaction) fee, which is paid every time the card is used. We allow fixed and per-transaction fees to be negative to account for rewards.

To distinguish membership choices (and fees) from usage choices (and fees), the model allows for uncertainty on the consumers' benefit of paying by card rather than by other means. Cardholders learn their benefit and choose their payment instrument on a purchase-by-purchase basis. These (perhaps negative) benefits can be interpreted as stemming from the size and type of the transaction, the availability of foreign currency, the distance to the closest ATM and so on. At the membership stage individuals are willing to pay for the expected value of being able to pay by card in the future, which we refer to as "option value," and for the intrinsic benefits associated with membership (e.g., social prestige, insurance). Symmetrically merchants decide on card acceptance by comparing the value of membership benefits (e.g., easy accounting, safer transactions) and expected usage benefits (net of merchant fees) to the membership fee. Given that merchants make only membership choices, allowing for uncertainty on merchants' usage benefits has no impact on the outcome. We account for the fact that different end-users have different preferences for transactions and membership by allowing for four degrees of end-user heterogeneity and thereby elastic final demands.

We first consider a card network's pricing incentives in the simplest setting with one card issuer (i.e., the cardholders' bank) and many perfectly competitive card acquirers (i.e., merchants' banks). This analysis provides us with a useful benchmark. At the same time it sheds light on incentives within the so called "three party schemes" (such as American Express and Diners Club), in which a single company directly contracts with cardholders and merchants (the two setups are formally equivalent). In this context we show that profit maximization always implies allocating an inefficiently high fraction of the total per-transaction price to merchants. Intuitively, financing card usage perks through higher charges on merchants not only increases issuance of *new* cards but also fosters usage of *existing* cards. Such policy carries a double dividend as membership fees

make issuing banks the residual claimants of the change in the expected value of holding a card.

More precisely, issuers capture the average card *usage* surplus of cardholders through fixed card fees. On the other hand, the card network fails to account for merchants' card *usage* surplus, despite the availability of fixed merchant fees. This results from the fact that cardholders use their cards only if they realize positive usage surplus at a point of sale, whereas affiliated merchants cannot refuse cards (and thus cannot affect card usage volume), even when they realize that a card transaction incurs a cost to them. It follows that starting from the first best price structure, shifting price burden towards merchants is always profitable and hence desirable from the network's perspective. Such skewed card prices result in *over-usage* of payment cards in the sense that an inefficiently high fraction of sales are settled by cards at affiliated merchants.

Importantly, the above argument does not require membership fees to be positive. Zero or even negative membership fees can be simply explained by fierce competition for cardholders.⁵ What matters is that banks internalize the incremental surplus that stems from better card usage terms. Thus even in those cases where card membership is subsidized, by negotiating better usage terms with the networks, issuers are able to reduce the amount of subsidies required to reach their target membership level.

This result is robust to the introduction of imperfect competition for cardholders and for merchants as well as network competition (e.g., Visa versus MasterCard), provided that merchants accept different card brands (i.e., “multi-home”, as it is named by the literature).⁶ Finally, we argue that the result is also robust to many other factors affecting final demands, e.g., strategic card acceptance as a quality investment and/or to steal business from a rival.

The literature on payment cards has already identified several potential sources of inefficiencies.⁷ Wright (2001) and Schmalensee (2002) first emphasized the platform's role in “balancing” the demand of payment services by consumers and merchants. The sign and magnitude of the distortion is shown to depend on asymmetries in costs, in demand elasticities and in the relative intensity of competition for end users on the two sides of the market. Rochet and Tirole (2002) (later extended by Wright, 2004) recognized a further source of distortion by formalizing the idea that competing merchants may accept cost-increasing cards to steal customers from their rivals. The greater the competitive edge guaranteed by card acceptance, the more likely it is that card networks exploit the lower merchant “resistance” (to price increases) by setting an inefficiently high merchant fee. Rochet and Tirole (2003), Guthrie and Wright (2003), and Armstrong (2006) reach similar conclusions by studying the effect of competition among payment networks. If merchants accept the cards of multiple card networks (i.e., multi-home), competition increases the distortion

⁵Negative fixed fees (e.g., introductory “bonuses” for card membership) can be rationalized, for instance, when a card is either sold as a part of a bundle (i.e., checking account and other payment services) or provide complementary services (e.g., credit lines) which constitute alternative sources of profits (interest fees).

⁶According to the Nilsen Report the acceptance network of the two major players almost perfectly overlaps with 29 million shops for Visa and 28.5 million for MasterCard. Even if consumers adopt more than one payment card, Rysman (2007) finds empirical evidence that they mostly use only one of them.

⁷See Chakravorti and To (2003), Evans and Schmalensee (2005b), and Chakravorti (2010) for an overview of this debate.

even further, as networks try to woo cardholders back from their rivals by lowering their prices. Networks can then charge merchants the monopoly price to provide access to their exclusive turf of cardholders.

In general, the literature on two-sided markets⁸ shows that a monopoly platform distorts upwards the total level of the per-transaction prices due to its market power and also distorts the structure (allocation) of the total price between the two sides. Platform competition would correct the market power distortion on the total price level. However, the existence and sign of price-structure distortion is not straightforward when end users are heterogeneous on both sides.

We gain traction on the problem by considering an equilibrium model with a broader parameter space (allowing for ex-ante uncertainty on usage benefits) and a broader action space (allowing for non-linear prices). In contrast to the above papers, we find that the *sign* of the distortion does not depend on fundamental cost and/or preference attributes; only its *magnitude* does. Such a result is derived without imposing additional restrictions. In fact, we show that our formulation obtains the baseline characterizations of the equilibrium prices of the above contributions as special cases (that is, imposing, simultaneously, one or more restrictions on prices and/or information structure).

Regarding policy concerns, our model unambiguously predicts that regulations which cap interchange fees can improve social welfare. However, we do not find any support for widely used issuer-cost-based cap regulation. In line with the literature, we indeed find that the socially optimal fee structure reflects two considerations: relative demand elasticities (marginal users) and relative net surpluses (average users). We further show that regulating the interchange fee is not enough to achieve full efficiency in the industry. The interchange fee affects only the allocation of the total user price between consumers and merchants whereas the first-best efficiency requires also a lower total price level due to positive externalities between the two sides.

Section 2 introduces our framework. Section 3 derives the profit maximizing card fees and merchant fees and illustrates the distortion on card prices, contrasting the equilibrium outcome with the outcome that would arise if the network's choice were regulated. Section 4 goes further, characterizing the first-best prices. Section 5 considers alternative assumptions on market structure. In Section 6 we stretch out and discuss how our findings help us understand the link between market power and inefficiency for any two-sided market in which usage choices are delegated to one side. Section 7 relates our framework and findings to the existing literature. Section 8 concludes with some policy implications.

1.1 Background on the payment card industry

In a three-party card network, like AMEX and Discover, one financial institution (say, bank) provides card services to the two groups of users: consumers and merchants. On the other hand, in a four-party card network, like Visa or MasterCard, each member bank sets its prices for consumers

⁸These are businesses that have to attract at least two groups of users to create value from a transaction, e.g., payment card networks, software platforms, dating clubs, etc. See Caillaud and Jullien (2003), Rochet and Tirole (2003, 2005, 2006) and Armstrong (2006) for pioneering contributions. Weyl (2010) provides an important recent contribution.

and/or its prices for merchants in return for card services it provides. The card scheme sets an interchange fee to be paid by the merchant's bank (acquirer) to the cardholder's bank (issuer) for every card transaction.⁹ Through the interchange fee, the card network could control the structure of the total user prices.

The payment card industry is a two-sided market, in that a card transaction requires participation of two different groups of users - consumers and merchants - and the corresponding externalities between the two sides are not internalized.¹⁰ There are two-sided membership (network) externalities: the value of accepting (respectively holding) a card depends on how many consumers (respectively merchants) hold (respectively accept) that card. Moreover, there are one-sided usage externalities from cardholders to affiliated merchants: every time a cardholder pays by card, the merchant receives some benefits from the card transaction and pays a merchant fee to its bank. If the merchant and the cardholder could have efficient bargaining, they would internalize these externalities through appropriate monetary transfers. However, in practice, it is uncommon that merchants price discriminate between card users and cash users, either because this discrimination is prohibited by payment networks' rules (so called "no-surcharge rules"), or transaction costs make it unprofitable to discriminate.¹¹ Gans and King (2003b) show that when surcharging a payment card is costless, the interchange fee (and the structure of user prices) would become neutral for the volume of card transactions. However, the fact that merchants have been complaining about high merchant fees or interchange fees constitutes anecdotal evidence for non-neutrality of interchange fees, since merchants would not complain if the full pass-through of merchant fees was costless.¹² Hence, the allocation (structure) of the total user prices between merchants and consumers matters for the total volume of transactions, for bank profits, and for total surplus.¹³

2 Model

We analyze pricing incentives in a four-party card network (e.g., Visa) which provides card payment services to card users (cardholders and merchants) through issuers (cardholders' banks) and

⁹In practice an interchange fee is something between 0.5-2.5% of the transaction value. See the EC's Retail Banking Sector Inquiry (2007) and the Reserve Bank of Australia's (RBA) report (2007).

¹⁰See Rochet and Tirole (2003, 2006) and Rysman (2009) for a formal definition and different examples of multi-sided markets. Rochet and Tirole (2003) focus on markets where platforms charge both sides per-transaction fees, such as card platforms, trying to get cardholders and merchants on board. Armstrong (2006), instead, analyzes mainly two-sided markets where participation (membership) fees are more widely used, such as media platforms competing for viewers and advertisers. Weyl (2010) nests the previous contributions in a monopoly platform, charging general prices, and emphasizes the nature of user heterogeneity on both sides in assessing possible market failures. For the analyses of competition between platforms see Caillaud and Jullien (2003), Guthrie and Wright (2003, 2007), Anderson and Coate (2005), Hagiu (2006) as well as Rochet and Tirole (2003, 2006), and Armstrong (2006).

¹¹According to the EC's Sector Inquiry (2007), there is no widespread surcharging even when it is allowed, e.g., in the UK since 1989, in Sweden since 1995, and in Netherlands since 1997. In Australia, prohibition on surcharging has been lifted since 2003, however, according the RBA, in 2007 only 14% of very large merchants, and only 5% of very small merchants surcharge.

¹²The common view of the literature is that the price structure in the payment card industry is not neutral.

¹³When merchants cannot perfectly pass merchant fees on to consumers, Chakravorti and Bolt (2008) show that only a proportion of merchants accepts cards and the price structure affects the card usage volume.

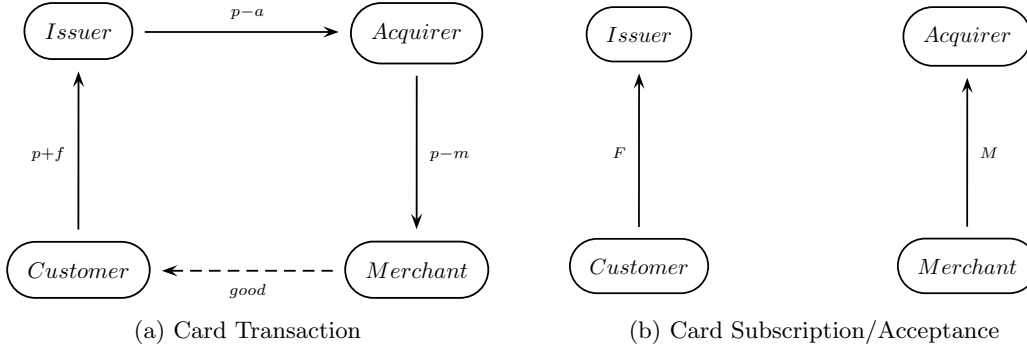


Fig. 1: Card Payments

acquirers (merchants' banks). We also assume that there is a price coherence, i.e., the price of a good is the same regardless if it is paid by cash or by card.

Consumption Surplus We consider a continuum (mass one) of consumers and a continuum (mass one) of local monopoly merchants.¹⁴ Consumers are willing to purchase one unit of a good from each merchant and the unit value from consumption is assumed to be the same across merchants. Let $v > 0$ denote the value of a good purchased by cash, that is the consumption value net of all cash-related transaction costs. A consumer gets $v - p$ from purchasing a unit good by cash at price p and the seller gets p from this purchase.¹⁵

Card Usage Surplus Consumers (or buyers) get an additional payoff of $b_B - f$ when they pay by card rather than cash. Let b_B denote the net per-transaction benefit¹⁶ and f denote the transaction fee to be paid to the issuer. Similarly, merchants (or sellers) get an additional payoff of $b_S - m$ when paid by card, where b_S denotes the net per-transaction benefit of a card payment¹⁷ and m denotes the merchant discount (or fee) to be paid to the acquirer. Note that we do not impose any sign restriction, potentially allowing for negative benefits, i.e., distaste for card transactions, and negative fees, e.g., reward schemes like cash-back bonuses or frequent-flyer miles. For each card transaction, the issuer (respectively the acquirer) incurs cost c_I (respectively c_A). Let c denote the total cost of a card transaction, so $c = c_I + c_A$. The card association requires the acquirer to pay an interchange fee a per transaction to the issuer. The issuer's (respectively the acquirer's) transaction cost is thus $c_I - a$ ($c_A + a$). Figure 1a summarizes the flow of fees triggered by a card transaction of amount p .

Card Membership Surplus Buyers and sellers are also subject to membership, i.e., transaction insensitive, fees (denoted respectively by F and M) and benefits (denoted respectively by B_B and

¹⁴In Section 5, we discuss the robustness of our results to merchant competition.

¹⁵Retailing costs play no role in the analysis and are wlog set to zero.

¹⁶Such as foregoing the transaction costs of withdrawing cash from an ATM or converting foreign currency.

¹⁷Such as convenience benefits from lower cash holdings, faster payments, easy accounting, saved trips to the bank, etc.

B_S) upon joining the card association (Figure 1b).¹⁸ To simplify the notation, we assume that the fixed costs of issuing an extra card and acquiring an extra merchant are zero.

In what follows, we assume that consumers and merchants are heterogeneous both in their *usage* and *membership* benefits from card payments. Specifically, benefits b_B, b_S, B_B and B_S are assumed to be distributed on some compact interval with smooth atomless cumulative distribution functions, satisfying the Increasing Hazard Rate Property (IHRP).¹⁹ Benefits b_B and b_S are i.i.d. across transactions.

Timing

Stage i: The payment card association (alternatively a regulator) sets the interchange fee a .

Stage ii: After observing a , each issuer sets its card fees and each acquirer sets its merchant fees.

Stage iii: Merchants and consumers realize their membership benefits B_S and B_B and decide, simultaneously, whether to accept and hold the payment card, respectively, and which bank to patronize.

Stage iv: Merchants set retail prices. Merchants and consumers realize their transaction benefits b_S and b_B , respectively. Consumers decide whether to purchase. Finally, cardholders decide whether to pay by card or cash.

Consumers and merchants maximize their expected payoff. We assume that the card association sets the interchange fee (IF) to maximize the sum of the profits earned by its issuers and acquirers. This assumption aims to represent real objectives of for-profit card associations.²⁰ In principle, for-profit card organizations could charge their members non-linear membership fees, and thus could internalize any incremental increase in their members' profits through fixed transfers.²¹ In the analysis, this means to define the profit of the association as the total fees collected from members, which could be proxied by the total profits of its member banks, allowing the association to charge fixed fees as well as transaction fees to its members. We are looking for a Subgame Perfect Nash Equilibrium (SPNE).

Consumption Surplus versus Card Usage Surplus Let buyer benefit b_B be distributed over interval $[b_B, \bar{b}_B]$ with cumulative distribution function $G(b_B)$ and probability density function

¹⁸E.g., cardholders enjoy the security of not carrying large amounts of cash, membership privileges (such as access to VIP), travel insurance, ATM services (such as account balance sheets, money transfers, etc.), social prestige (club effects); merchants benefit from safe transactions.

¹⁹The IHRP leads to *log-concavity* of demand functions (for cardholding, for card usage, and for card acceptance), which is sufficient for the second-order conditions of the optimization problems we solve.

²⁰Visa and MasterCard used to be non-profit organizations, but since 2003 Visa and since 2006 MasterCard are for-profit organizations and their shares are jointly owned by their member banks. See the EC's report (2007) and MasterCard decision (2007).

²¹Indeed, in virtually all countries where the card associations operate, a significant portion of the operating revenues are typically concentrated among a handful of large issuers.

$g(b_B)$. Similarly, let seller benefit b_S be distributed over some interval $[\underline{b}_S, \overline{b}_S]$ with CDF $K(b_S)$ and PDF $k(b_S)$. To simplify the benchmark analysis, we make the following assumption:

$$A1 : v \geq c - \underline{b}_B - \underline{b}_S + \frac{1 - G(\underline{b}_B)}{g(\underline{b}_B)}.$$

The assumption guarantees that v is sufficiently high so that merchants never find it profitable to exclude cash users by setting a price higher than v .²² In other words A1 rules out the case where merchants try to extract some of the surplus associated with card transactions (e.g., rewards) by increasing retail prices. Thus monopoly merchants set $p = v$ regardless of whether they accept card payments or not. After solving the benchmark model, we show that relaxing A1 reinforces our results.

2.1 Preliminary Observations

By A1, all merchants set $p = v$ and therefore all consumers purchase a unit good from each merchant. If a merchant accepts cards, a proportion, α_B , of its transactions (to be determined in equilibrium) is settled by card. The net payoff of type B_S merchant from accepting cards is:²³

$$B_S - M + E[b_S - m] \alpha_B, \tag{1}$$

which is the sum of the membership and expected transaction surpluses when merchant fees are (M, m) . The number of merchants that join the payment card network is thus:

$$\alpha_S \equiv \Pr(B_S - M + E[b_S - m] \alpha_B \geq 0).$$

Note that α_S depends only on the average merchant benefit and fee, which are defined respectively as:

$$\tilde{b}_S \equiv E[b_S] + \frac{B_S}{\alpha_B} \quad \text{and} \quad \tilde{m} \equiv m + \frac{M}{\alpha_B},$$

and thus $\alpha_S = \Pr(\tilde{b}_S \geq \tilde{m})$. There is, therefore, one degree of freedom in acquirers' pricing policy. Any $\hat{\alpha}_S$, resulting from some fees (\hat{M}, \hat{m}) , can also be implemented through a simple linear pricing scheme: $M = 0$, $m = \tilde{m}(\hat{m}, \hat{M})$.^{24,25} This observation is due to the fact that the card acceptance decision is sunk when the seller learns its benefit b_S , and therefore the card acceptance demand cannot be affected by the realization of b_S . Only the *average* benefit known *before* the acceptance

²²We refer the reader to Guthrie and Wright (2003, Appendix B) for a formal proof of this point in a related context.

²³Since we assume that b_S is i.i.d across transactions, even when a merchant realizes its benefit $b_S < m$ for one transaction, the merchant does not stop its card membership because it does not know the exact value of its benefit for future transactions.

²⁴This would still be the case if we assumed some market power on the acquiring side.

²⁵In fact, if merchants were risk-averse it would then be a dominant strategy to charge only for usage since payments are due only if a transaction occurs.

decision matters. For a given α_B , our framework is thus equivalent to a setup where merchants are heterogeneous in their *average* benefits prior to their card acceptance decisions.

Without loss of generality, in what follows, we focus on a model where $B_S = M = 0$ and merchants are heterogeneous in their average benefit (denoted by b_S), which they know before card acceptance decisions. We assume that b_S is continuously distributed on some interval $[\underline{b}_S, \overline{b}_S]$ with CDF $K(b_S)$, PDF $k(b_S)$ and increasing hazard rate $k/(1 - K)$.

Crucially, we cannot set $B_B = F = 0$ wlog since buyers make *two* decisions (card membership and usage) at different information sets. Cardholding depends on the average benefit and card fee, whereas card usage depends on the transaction benefit and fee. We assume that B_B is continuously distributed on some interval $[\underline{B}_B, \overline{B}_B]$ with CDF $H(B_B)$, PDF $h(B_B)$. Recall that benefits b_B , b_S and B_B are independently distributed and allowed to be negative.

3 Benchmark Analysis

In this section we study the benchmark case with one monopoly issuer and many perfectly competitive acquirers. In Section 5 we analyze alternative market structures. We elected this particular structure as our benchmark for two reasons. First, this setup is formally equivalent to a three-party scheme such as AMEX and, more generally, to any market in which platform coordinates two sets of users, one of which decides, exclusively, on usage (see Section 6). Second, the issuing side of the market is widely regarded as the dwelling of market power (for instance, see Evans and Schmalensee, 1999; Rochet and Tirole, 2002, 2005; and the EC's report, 2007) for a discussion of the cause and the extent of market power in the payment card industry).

□ Usage Decisions

Cardholders pay by card if and only if their transaction benefit exceeds the usage fee. Buyers' demand for card usage is

$$D_B(f) \equiv \Pr(b_B \geq f) = 1 - G(f),$$

which is the proportion of cardholders paying by card at transaction price f .

□ Membership Decisions

Merchant of type b_S accepts cards whenever $b_S \geq m$.²⁶ The proportion of merchants who accept payment cards is

$$D_S(m) \equiv \Pr(b_S \geq m) = 1 - K(m).$$

Define, respectively, buyers' and sellers' average surpluses from card usage as $v_B(f) \equiv E[b_B - f \mid b_B \geq f]$ and $v_S(m) \equiv E[b_S - m \mid b_S \geq m]$. The expected value of being able to pay by card at a point of

²⁶Card acceptance is not affected by card usage/membership, i.e., there is no externality imposed by consumers on merchant participation. We could restore this externality by allowing for fixed merchant fees, since the card usage demand then affects the average merchant fee, without changing our conclusions (see the discussion in the previous section).

sale, which we call *the option value* of the card, is denoted by Φ_B and equal to

$$\Phi_B(f, m) \equiv v_B(f)D_B(f)D_S(m),$$

where $D_B(f)D_S(m)$ is the volume of card transactions at card fee f and merchant fee m . Note that the option value increases with the expected card usage at affiliated merchants, D_B , and with merchant participation, D_S . Type B_B gets a card if and only if the total benefits from cardholding, i.e., the sum of fixed benefit B_B and the option value of the card, exceed the fixed card fee:

$$B_B + \Phi_B(f, m) \geq F.$$

The number of cardholders, which is denoted by Q , is then

$$\begin{aligned} Q(F - \Phi_B(f, m)) &= \Pr[B_B + \Phi_B(f, m) \geq F] \\ &= 1 - H(F - \Phi_B(f, m)), \end{aligned}$$

which is a continuous and differentiable function of card fees (F, f) and merchant discount m .

□ Behavior of the Issuer and Acquirers

Taking the IF as given, perfectly competitive acquirers simply pass-through their costs, charging $m^*(a) = a + c_A$ per transaction.

For a given IF, the issuer sets card fees by maximizing its profit, which is the sum of the card transaction profits and fixed card fees collected from cardholders:

$$\max_{F, f} [(f + a - c_I)D_B(f)D_S(m) + F] Q(F - \Phi_B(f, m)). \quad (2)$$

The usual optimality conditions bring the equilibrium fees:

$$f^*(a) = c_I - a, \quad F^*(a) = \frac{1 - H(F^*(a) - \Phi_B(a))}{h(F^*(a) - \Phi_B(a))}.^{27}$$

The fixed fee is characterized by a Lerner formula. The issuer introduces a monopoly markup on its fixed costs (for simplicity, here set to zero), inefficiently excluding some consumers from the market. The usage fee is set at the marginal cost of issuing since the issuer is the residual claimant of *changes* in the option value.

Privately and Socially Optimal Interchange Fees

Taking into account the equilibrium reactions (card fees and merchant fees) of banks to a given IF level, we proceed to define three critical levels of IF: the buyers-optimal IF, a^B , which maximizes the buyer surplus (gross of fixed fees), the sellers-optimal IF, a^S , maximizing the seller surplus,

²⁷To simplify the expressions, we write $\Phi_B(a)$ instead of $\Phi_B(c_I - a, c_A + a)$.

and a^V , which maximizes the volume of card transactions. These are, respectively, a solution to the following programs:

$$\max_a BS(a) = v_B(f^*)D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*) + \int_{F^* - \Phi_B(f^*, m^*)}^{\bar{B}_B} xh(x)dx \quad (3)$$

$$\max_a SS(a) = v_S(m^*)D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*) \quad (4)$$

$$\max_a V(a) = D_B(f^*)D_S(m^*)Q(F^*, f^*, m^*). \quad (5)$$

Lemma 1 *Problems (3), (4), (5) admit one and only one solution, denoted, respectively, a^B, a^S, a^V and satisfying $a^S < a^V < a^B$.*

Proof. *Appendix A.1.*

This lemma highlights the *tension* between buyers' and sellers' interests over the association's pricing policy. An increase in the interchange fee has three effects. On the one hand, it induces a higher merchant fee and thus lowers the number of shops where cards are welcome. On the other hand, it results in a lower card usage fee, and thus induces cardholders to settle more transactions by card at each affiliated store. Furthermore, a higher interchange fee changes buyers' expected surplus from card transactions (the option value of the card, Φ_B), and thus changes the net price of the card, $F - \Phi_B$. A unit increase in Φ_B increases the equilibrium fixed fee less than one, and therefore lowers the net price of the card resulting in a higher number of cardholders. Given that the number of cardholders, and thus total utility of buyers from cardholding, is increasing in the option value of the card, the IF maximizing the option value also maximizes the buyer surplus (gross of fixed fees). We show that the interchange fee maximizing the option value is higher than the volume maximizing IF, which is higher than the sellers-optimal IF, since the average buyer surplus from card transactions, v_B , is decreasing in card usage fee f , so increasing in IF, whereas the average seller surplus, v_S , is decreasing in merchant fee m , so decreasing in IF (due to the IHRP). Going above the volume-maximizing IF increases the buyer surplus (gross of fixed fees) at the expense of the seller surplus.

Indeed a^B is also the interchange fee maximizing the net buyer surplus (net of fixed fees). To see this suppose that this is not true and there exists another interchange fee, say $\bar{a} \neq a^B$, which maximizes the buyer surplus. Suppose that we change interchange fee, incrementally, starting from \bar{a} towards a^B . This change would increase the gross consumer surplus (ignoring fixed fee). This incremental increase would be *partly* captured by the issuer through a fixed fee. In other words, buyers would be better off by this change. This conflicts with the assumption at the beginning that \bar{a} maximizes the net consumer surplus.

□ Equilibrium Fees

Given the equilibrium reactions of banks,

$$f^*(a) = c_I - a \quad \text{and} \quad m^*(a) = c_A + a,$$

fixing the IF is formally equivalent to allocating the total cost of a transaction between the two sides of the market. Perfect competition on the acquiring side of the market implies that the association sets the IF that maximizes the issuer's profits subject to the equilibrium prices set by banks:

$$\max_{F,f,m} FQ(F - \Phi_B(f, m)) \quad st.: \quad i. f + m = c \quad ii. F = \frac{1 - H(F - \Phi_B(f, m))}{h(F - \Phi_B(f, m))}. \quad (6)$$

The issuer's profits are clearly increasing in the option value of the card Φ_B . To see this it suffices to apply the envelope theorem to the objective function. Hence, the privately optimal allocation is the allocation that maximizes the option value. It is such that the impact of a small variation of f on the option value is equal to the impact of a small variation of m .

From Lemma 1 we know that a^B maximizes Φ_B . We thus conclude that the privately optimal IF is equal to a_B , that is $a^* = a^B$.

□ Optimal Regulation

In this section we consider the problem of a regulator seeking to maximize the total surplus in the economy through an appropriate choice of a . Such a problem can also be stated as a price allocation problem similar to (6):

$$\max_{F,f,m} \{[v_B(f) + v_S(m)] D_B(f) D_S(m) + E[B_B \mid B_B \geq F - \Phi_B(f, m)]\} Q(F - \Phi_B), \quad (7)$$

subject to the same set of constraints.

The above formulation makes clear that the only difference between the regulator's problem and the association's problem is in the *allocation* of the total price c across the two sides of the market. As we shall see in the next section, full efficiency indeed requires a total price different than c .

To highlight the discrepancy between public and private incentives we shall restate problem (6) in terms of the indifferent cardholder, \tilde{B}_B :

$$\max_{\tilde{B}_B, f, m} (v_B(f) D_B(f) D_S(m) + \tilde{B}_B) Q(\tilde{B}_B) \quad st.: \quad i. \quad \text{and} \quad ii. \quad (3')$$

Comparing the association's objective with (7) highlights the two sources of welfare losses induced by the association's pricing policy. First, the association distorts the allocation of costs between card users and merchants, neglecting the impact of a marginal variation of the interchange fee on the merchant surplus. Starting from any IF between a_S and a_B , a marginal increase of a raises the buyer surplus (gross of fixed fees) at the expense of the merchant surplus (see Lemma 1). Through fixed card fees, the issuer, and thus the association, internalizes all *incremental* card usage surpluses of buyers due to this increase in IF. On the other hand, the lack of term $v_S D_B D_S Q$ in the association's objective reflects the seller surplus that the association fails to account for.

The second source of distortion is due to the monopoly markup of the issuer. Through setting a , the association determines, indirectly, equilibrium fixed fee, F^* , and the option value of the

card, Φ_B^* , which together determine the net price of the card, $F^* - \Phi_B^*$, and thus the equilibrium number of cardholders. Increasing membership on one side implies more surplus on both sides of the market since the number of interactions (i.e., card transactions) increases. For an additional cardholder, the association accounts for the fixed benefit of the marginal cardholder, \tilde{B}_B , whereas the social planner internalizes the fixed benefit of the average cardholder, $E[B_B | B_B \geq \tilde{B}_B]$. The fact that the association fails to capture the impact of an extra cardholder on the net benefit of an average cardholder, $E[B_B - \tilde{B}_B | B_B \geq \tilde{B}_B]$, results in an additional discrepancy between private and public interests.

We now compare the regulator's choice with the choice of the association:

Proposition 1 *The privately optimal IF is higher than the socially optimal IF. Hence, in equilibrium, cardholders pay too little and merchants pay too much per transaction.*

Proof. *Appendix A.2.*

For the special case where consumers get no fixed benefits from cardholding, $\underline{B}_B = \bar{B}_B = 0$, there is an intuitive characterization of the efficient fee:²⁸

$$\frac{f}{m} = \frac{\eta_B}{\eta_S} \div \frac{v_B}{v_S},$$

where $\eta_B = -\frac{f D'_B}{D_B}$ is the elasticity of buyers' card usage and $\eta_S = -\frac{m D'_S}{D_S}$ is the elasticity of sellers' card acceptance demand. The socially optimal allocation of the total price $f + m = c$ is achieved when relative user prices are equal to the ratio of the relative demand elasticities and the relative average surpluses of buyers and sellers.

3.1 Discussion

So far we have discussed how the discrepancy between private and public interests (respectively (6) and (7)) affects economic efficiency through the association's pricing policy. In the rest of this section we shall focus on the determinants of such discrepancy.

Observe that interchange fees, which constitute revenues for the issuing side, let the issuer extract (some of) the merchant surplus. It follows that by controlling the association's choice of a , the issuer acts effectively as a single platform owner. In fact one could think of the issuer as directly charging merchants for card services since competitive acquirers simply pass-through interchange fees to merchants. The benchmark framework is therefore formally equivalent to a monopoly platform, pricing both sides to maximize its profits (Fig. 2a).²⁹ The only asymmetry between the two sides of the market is that usage choices (i.e., the choice of the payment instrument)

²⁸An analogous property holds for the optimal access charge between backbone operators or between telecom operators where the access charge allocates the total cost between two groups of users (consumers and web sites in backbone networks, call receivers and call senders in telecommunication networks) (See Laffont et al. (2003)). This condition is analogous to Rochet and Tirole's (2003) characterization of Ramsey prices.

²⁹Indeed this observation extends our findings to so called proprietary (or 3-party) schemes such as AMEX.

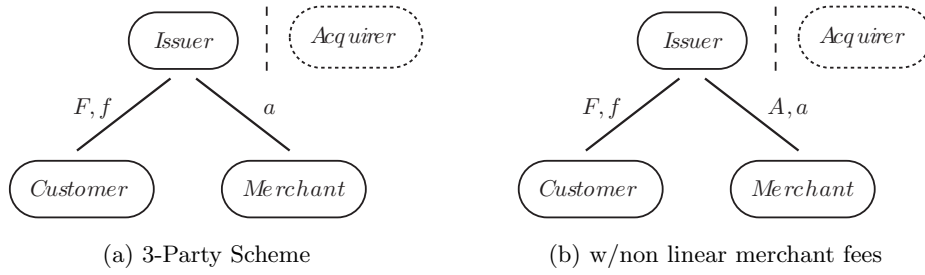


Fig. 2

are delegated to consumers. This structural feature of the payment card market is the ultimate foundation of the allocational distortion of Proposition 1. The intuition is as follows. Increasing the IF beyond the socially optimal level not only attracts new members through a higher option value but also fosters card usage among *existing* members. The incremental buyer surplus due to this extra, inefficient, usage can be extracted at the membership stage through higher fixed fees, while keeping the consumer participation fixed (i.e., keeping the average card fee fixed). The same is not true on the merchant side of the market. The association cannot fully internalize incremental losses in the merchant surplus due to this increase of the IF. As shown in section 2.1, considering non-linear charges on merchants (such as non-linear interchange fees or non-linear merchant fees (Fig 2b)) would not affect the result. Changing the marginal price, m , while keeping the average merchant price constant does not have any impact on the volume of card transactions. This is because merchants make only one decision, that is, whether to become a member of the card association, and the number of merchants accepting cards depends uniquely on the average merchant price and benefit from card acceptance. Hence, one of the two pricing instruments is redundant on the merchant side.

Rochet and Tirole (2003, 2006b) derive the optimal pricing structure for a monopoly platform setting linear prices to both sides. As opposed to theirs, our equilibrium fees do not maximize the total volume of transactions. We thus cannot conclude that in equilibrium there is *over-provision* of card services simply by noticing that the socially optimal IF is different (in our framework smaller) than the privately optimal one. Improving buyers' usage incentives through a higher IF (inducing, for instance, reward schemes and cash back bonuses) does not necessarily lead to a higher total volume of transactions, since some merchants abandon the platform in response to higher merchant fees. In our model there is *over-usage* in the sense that, in equilibrium, the proportion of buyers who choose to pay by card at an affiliated merchant is always inefficiently high.

4 Efficient Fees

In this section we characterize the first best (Lindahl) fees which are informative about the nature of the externalities in this market.

Consider the problem of a public monopoly running the industry in order to maximize the

total welfare, which is the total profits of issuer and acquirer banks plus the surplus of buyers and sellers:

$$\max_{F, f, m} W \equiv \{[f + m - c + v_B(f) + v_S(m)] D_B(f) D_S(m) + E[B_B \mid B_B \geq F - \Phi_B]\} Q(F - \Phi_B). \quad (8)$$

Proposition 2 *The first best total price (per transaction) is lower than the total cost of a transaction and equal to $c - v_B(f^{FB})$. The socially optimal allocation of such a price is achieved when*

$$v_B(f^{FB}) = v_S(m^{FB}),$$

that is, when the average buyer surplus is equal to the average seller surplus.

Proof. *Appendix B.1.*

Intuitively, each type of user is charged a price equal to the cost of a transaction minus a discount, reflecting its positive externality on the other segment of the industry. An extra card user (respectively merchant) attracts an additional merchant (respectively card user) which generates average surplus v_S (respectively v_B).³⁰ At the optimum, the two externalities must be equalized, the total price is thus

$$f^{FB} + m^{FB} = c - v_S(m^{FB}) = c - v_B(f^{FB}) < c$$

A Ramsey planner would solve (8) subject to an additional constraint: $\Pi_A, \Pi_I \geq 0$, where Π_A and Π_I denote, respectively, the acquirers' and the issuer's profits. The rationale for the latter comes from the problem of a regulator who can control end-user prices but cannot or does not want to run and/or subsidize operations, and therefore has to leave enough profits to keep the industry attractive for private investors. Using an argument analogous to that employed in the proof of Proposition 2 it is possible to show that

Proposition 3 *The second best (Ramsey) total price (per transaction) is higher than the first best, but still lower than the cost of a transaction.*

The Ramsey prices distort the total price level, but less than the card network. Intuitively, below-cost usage fees can be financed through fixed charges on the consumer side, and thus do not necessarily trigger budget imbalances.

5 Alternative Market Structures

5.1 Issuer Competition

In this section, we modify our benchmark setup by allowing for imperfect competition between two issuers, denoted by I_1 and I_2 , which provide differentiated payment card services within the same

³⁰Such a pricing rule was independently found by Weyl (2009).

card scheme and charge their customers two-part tariff card fees. Consumers have preferences both for payments made by card instead of other means and for the issuer itself, i.e., brand preferences. Brand preferences are due to, for instance, quantity discounts (family accounts), physical distance to a branch, or consumers' switching costs that derive from the level of informational and transaction costs of changing some banking products (current accounts).

Card i refers to the payment card issued by I_i , for $i = 1, 2$. We denote the net price of card i by t_i , which is defined as the difference between its fixed fee and the option value of holding card i : $t_i \equiv F_i - \Phi_B(f_i, m)$. The demand for holding card i is denoted by $Q(t_i, t_j)$ (or Q_i), for $i \neq j$, $i = 1, 2$. We make the following assumptions on Q_i :

$$\begin{aligned} A2 : \quad & \frac{\partial Q_i}{\partial t_i} < 0 & A3 : \quad & \frac{\partial Q_i}{\partial t_j} > 0 & A4 : \quad & \left| \frac{\partial Q_i}{\partial t_i} \right| > \frac{\partial Q_i}{\partial t_j} \\ A5 : \quad & \frac{\partial^2 \ln Q_i}{\partial t_i^2} < 0 & A6 : \quad & \left| \frac{\partial^2 \ln Q_i}{\partial t_i^2} \right| > \left| \frac{\partial^2 \ln Q_i}{\partial t_i \partial t_j} \right| \end{aligned}$$

A2 states that the demand for holding a card is decreasing in its net price. A3 ensures the substitutability between the card services provided by different issuers so that the demand for holding card i is increasing in the net price of card j . By A4, we further assume that this substitution is imperfect, and thus the own price effect is greater than the cross price effect. By assuming that Q_i is log-concave in net price t_i , A5 ensures the concavity of the optimization problems. A6 states that own price effect on the slope of the log-demand is higher than the cross price effect.

In Appendix C.1, we provide examples of classic demand functions for differentiated products, such as Dixit (1979), Singh and Vives (1984), Shubik and Levitan (1980) which satisfy all of our assumptions.³¹

□ Behavior of the Issuers and Acquirers

Perfectly competitive acquirers set $m^*(a) = c_A + a$. Taking the IF and card j 's fees as given, I_i 's problem is to set (F_i, f_i) to maximize its profit:

$$\max_{F_i, f_i} [(f_i + a - c_I) D_B(f_i) D_S(m) + F_i] Q(F_i - \Phi_B(f_i, m), F_j - \Phi_B(f_j, m)).$$

Like in the benchmark case, both issuers set $f_i^*(a) = c_I - a$ in order to maximize the option value of their card. The option value is therefore equal to $\Phi_B(c_I - a, c_A + a)$ (or compactly $\Phi_B(a)$) regardless of the identity of the issuer. The best reply fixed fee of I_i to its rival's fixed fee, F_j , is implicitly given by

$$\epsilon_i(F_i^*, F_j; a) = 1,^{32}$$

³¹ A simple illustration, here omitted for brevity, is a classic Hotelling one.

³² Observe that the optimality condition is indeed given by the Lerner formula:

$$markup_i = \frac{1}{\epsilon_i},$$

where the markup of each duopolist issuer is equal to 1 since there is no fixed cost in our setup. If instead each issuer

where $\epsilon_i \equiv -F_i \frac{\partial Q_i / \partial F_i}{Q_i}$ refers to the elasticity of I_i 's demand with respect to its fixed fee. Assumption A5 (log-concavity of the demand) guarantees that ϵ_i is increasing in F_i , and thus that F_i^* is well-defined. Whenever ϵ_i is greater (respectively less) than 1, I_i has incentives to lower (respectively raise) its fixed fee until $\epsilon_i = 1$. An equilibrium of issuer competition is any pair (F_i^*, F_j^*) such that $\epsilon_i(F_i^*, F_j^*; a) = \epsilon_j(F_j^*, F_i^*; a) = 1$.

□ Privately and Socially Optimal Interchange Fees

The association's problem is to set the IF, maximizing the sum of the issuers' profits $\Pi_1^* + \Pi_2^*$, where each issuer earns

$$\Pi_i^* = F_i^* Q(F_i^* - \Phi_B(a), F_j^* - \Phi_B(a)),$$

given that $\epsilon_i(F_i^*, F_j^*; a) = \epsilon_j(F_j^*, F_i^*; a) = 1$. Our claim is that the association sets $a^* = a^B$, maximizing the option value of the card, $\Phi_B(a)$. We prove the claim by showing that equilibrium profits increase with Φ_B . Applying the Envelope Theorem to the issuer profits, we derive

$$\frac{\partial \Pi_i^*}{\partial \Phi_B} = F_i^* \left[-\frac{\partial Q_i}{\partial t_i} - \frac{\partial Q_i}{\partial t_j} + \frac{\partial Q_i}{t_j} \frac{\partial F_j^*}{\partial \Phi_B} \right],$$

which helps us to identify two types of effects on I_i 's profit of a marginal increase in the option value:

1. Demand Effect: The direct effect of the net card prices on Q_i , which is composed of own and cross demand effects. The *own demand effect* (the first term in brackets) is positive because the demand decreases in the net price of the card (A2), increasing in the option value of the card. The *cross demand effect* (the second term in brackets) is negative because the demand increases in the net price of the rival's card (A3), decreasing in the option value. The overall demand effect is positive since the positive own demand effect dominates the negative cross demand effect (A4).

2. Strategic Effect: The last term in brackets accounts for the impact of a change in the option value on the rival's pricing policy.

Lemma 2 *Under A2 – A6, both equilibrium fees are increasing in Φ_B .*

Proof. *Appendix C.2.*

By Lemma 2 we find that the strategic effect is positive: Increasing the option value of the card softens price competition. As a result the profit of each issuer increases in the option value, Φ_B . A straightforward consequence is that:

Corollary 1 *Under A2 – A6, the issuers' incentives over the interchange fee are aligned.*

paid fixed cost C_I per card, the solution to I_i 's problem would be

$$markup_i \equiv \frac{F_i^* - C_I}{F_i^*} = \frac{1}{\epsilon_i},$$

whereas we simply assume that $C_I = 0$, so we have $markup_i = 1$.

To maximize the sum of the issuers' profits, the association sets $a^* = a^B$, which maximizes cardholders' surplus from card transactions.

We are now left to compare the profit maximizing interchange fee with the welfare maximizing fee. The regulator would set an IF, a^r , to maximize the total welfare, anticipating banks' pricing behavior in equilibrium:

$$\max_a \left\{ \begin{aligned} & [v_B(c_I - a) + v_S(c_A + a)] D_B(c_I - a) D_S(c_A + a) [Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)] \\ & + E[B_B \mid B_B \geq F_1^* - \Phi_B] Q(F_1^*, F_2^*, a) + E[B_B \mid B_B \geq F_2^* - \Phi_B] Q(F_2^*, F_1^*, a) \end{aligned} \right\}.$$

The solution to the usual optimality conditions characterizes a^r . By comparing the association's IF with the regulator's IF, we get our main result:

Proposition 4 *If A2-A6 hold, then the privately optimal IF is higher than the socially optimal IF. Hence, in equilibrium, cardholders pay too little and merchants pay too much per transaction.*

Proof. *Appendix C.3.*

Once we acknowledge the fact that the issuers' incentives are aligned to those of cardholders, the logic behind Proposition 4 is analogous to that of the previous section: The association sets the buyers' optimal IF, whereas the regulator would set a lower IF since it could internalize buyers' as well as sellers' surpluses and the seller surplus is decreasing in IF.

Finally, we explain the role of issuers' competition. Competition is effective in reducing membership fees and thus in reducing (even eliminating) the distortion due to the issuer market power (see the discussion before Proposition 1). This can easily be established, contrasting the equilibrium outcome with the outcome that would arise if the issuers were jointly owned. This observation coupled with the marginal cost pricing, $f_i = f_j = c_I - a$, implies that the total surplus is always higher under competition no matter what IF prevails in equilibrium. However, issuer competition fails to reduce the distortion originating from the inefficient allocation of transaction costs between consumers and merchants.

Interchange fees are relatively high in the United States, where membership fees are not often used. This observation might look contradictory to the mechanism explained above.³³ However, our analysis of issuer competition could indeed explain this: even when issuer competition lowers the equilibrium fixed fees to zero (or even makes them negative), the scheme sets a too high interchange fee and distorts the card price structure. To see this, consider an incremental change in interchange fee from the socially optimal level towards the buyers' optimal level. Since such a change would increase the expected card usage surplus, issuers could raise fixed card fees while keeping the number of cardholders fixed. They are thus better off by this change. This is true even when the equilibrium fixed fees are negative, since in this case issuers lose less from subsidizing membership.

³³We would like to thank Mark Armstrong for raising this point.

5.2 Acquirer Competition

We conjecture that our results are robust to the introduction of market power on the acquiring side of the market, however, we have not been able to extend our claims without adding further qualifiers. To see why this issue is subtle, consider the polar opposite case of a monopoly issuer and a monopoly acquirer. Optimal pricing on the acquirer's side involves a markup ($m^* > c_A + a$) which is characterized by a standard inverse elasticity rule over merchants' demand for card services. Such markup allows the acquiring bank to extract some of the surplus that merchants derive from card payments. Assuming that the association maximizes the total profits of its member banks, this creates a countervailing incentive to lower interchange charges. Such conflict between the issuer's and the acquirer's interests is due to the conflict between sellers' and buyers' interests. In particular, close to the issuer's optimal IF, the acquirer's profits decrease in a .

Following a reduction of the IF below the buyers' optimal level the acquirer internalizes only *a part* of the incremental surpluses that accrue to the merchant side of the market, since it has to leave some rent to heterogeneous merchants. Therefore, the privately optimal IF would likely be higher than the socially optimal one which fully takes into account both merchants' and consumers' incremental surpluses from card transactions.

5.3 Merchant Competition (Strategic Card Acceptance)

By assuming monopoly merchants, we abstract away from business-stealing effects of accepting payment cards. Rochet and Tirole (2002) are the first who analyze such effects in a model where merchants accept the card to attract customers from rival merchants who do not accept the card. For a given retail price, card acceptance increases the quality of merchant services associated with the option to pay by card. Consumers are ready to pay higher retail prices for the improved quality as long as they observe the quality.³⁴ Rochet and Tirole show that when merchants are competing à la Hotelling, they internalize the average surplus of consumers from card usage, $v_B(f)$, so merchants accept cards if and only if $b_S + v_B(f) \geq m$. In other words, merchants pay $m - b_S$ to accept cards since they could recoup v_B through charging higher retail prices for their improved quality of services.

It is important to note that we do not need merchant competition to make this argument. A monopoly merchant would also be willing to incur a cost per card transaction to offer a better quality of service to its customers (who value the option of paying by card), since it could then internalize *some*³⁵ of the average card usage surplus of buyers by charging higher retail prices.

We make assumption A1 to rule out card acceptance aiming to improve quality. Recall that A1 ensures a high enough consumption value by cash, v , so that merchants who accept cards do not want to exclude cash users by setting a price higher than v . In our setup, merchants accept

³⁴The authors assume that only a proportion, α , of consumers observe which store accepts cards before choosing a store to shop. Here, we consider simply their extreme case of $\alpha = 1$, which is sufficient to make our point.

³⁵Unlike Hotelling competition, total demand is decreasing in retail price. This is why the monopolist merchant could internalize *some* (not *all*) of the average card usage surplus.

cards only to enjoy convenience benefits from card payments, and thus they accept cards if and only if $b_S \geq m$. Once we relax A1, a merchant accepting cards might be willing to charge a price higher than v (exclude cash users, sell only to card users) since, by increasing its price, it could internalize some of the buyer surplus from card usage. Anticipating this extra revenue from card users, a merchant might accept cost-increasing cards. For instance, consider simply the case of homogeneous merchants and suppose that a merchant accepting cards prefers to set $p^* > v$, i.e., it gains more from setting $p = p^*$ than $p = v$. If the merchant sets p^* , only card users buy its product and the merchant gets³⁶

$$\Pi_S^* = (p^* + b_S - m)D_B(f + p^* - v), \quad (11)$$

If the merchant sets $p = v$, all consumers buy its product and the merchant gets

$$\Pi_S = v + (b_S - m)D_B(f) \quad (12)$$

We assume that $\Pi_S^* > \Pi_S$, and thus the merchant prefers to set $p^* > v$. Since $D_B(f) > D_B(f + p^* - v)$ for $p^* > v$, our assumption ($\Pi_S^* > \Pi_S$) implies also that

$$V(p^*, f) \equiv p^* - \frac{v}{D_B(f + p^* - v)} > 0,$$

where $V(p^*, f)$ is a positive function referring to the merchant's extra surplus from increasing its quality (so its retail price) through accepting cards. Putting it differently $V(p^*, f)$ refers to some of the average card usage surplus of buyers. The IHRP implies that p^* is decreasing in f (see the previous footnote). Using this together with the monotonicity of $D_B(\cdot)$, we get that $V(p^*, f)$ is decreasing in f .

If the merchant does not accept cards, it gets $\Pi_S = v$. A merchant thus accepts cards whenever

$$\begin{aligned} \Pi_S^* = (p^* + b_S - m)D_B(f + p^* - v) &\geq v \quad \text{or} \\ b_S + V(p^*, f) &\geq m \end{aligned} \quad (13)$$

Anticipating extra surplus $V(p^*, f)$ from card users, the merchant is willing to pay more than its convenience benefit to be able to accept cards, i.e., it resists less to an increase in m when it expects to get a higher surplus after accepting cards. Furthermore, $V(p^*, f)$, which measures the reduction in the merchant's resistance, decreases in card usage fee f , therefore increases in the IF.

³⁶A monopolist merchant accepting cards sets its price by

$$\max_p (p + b_S - m)D_B(f + p - v) \text{ st.: } p \geq v \quad (9)$$

The solution to the unconstrained problem is implicitly given by

$$p^* = m - b_S + \frac{D_B(f + p^* - v)}{D'_B(f + p^* - v)}. \quad (10)$$

The merchant's optimal price is p^* if it satisfies the constraint, i.e., $p^* > v$. Otherwise the merchant sets its price equal to v . We suppose here that the constraint is not binding in equilibrium.

When the association raises the IF, the merchant fee increases, which decreases the participation of merchants. Conversely, an increase in the IF decreases the card usage fee increasing $V(p^*, f)$. This in turn increases merchant participation. The latter effect does not exist in our original setup under A1. Hence, merchants would resist an increase in the IF less if we relaxed A1, in which case the privately optimal IF would be even higher than what we found. Hence, relaxing A1 would reinforce our results: cardholders would pay even less and merchants would pay even more. The same conclusions would hold if we allowed business-stealing effects, by introducing competition among merchants, since such a modification in our setup would again weaken the resistance of merchants to an increase in IF (see Rochet and Tirole, 2002, 2006a). For the case of heterogeneous merchants, we could make a similar argument for the marginal merchant: relaxing A1 would make the marginal merchant less resistant to an increase in the IF, and thus the association sets a higher IF.

5.4 Network Competition

Our setup does not incorporate the implications of competition between card schemes or other payment methods. The literature³⁷ shows that the implications of platform competition depend on which side of the market is more willing to use multiple platforms (“multi-homing”) rather than using a single platform (“single-homing”). Multi-homing on one side of the market intensifies price competition on the other side since platforms use low prices in an attempt to “steer” (undercut in order to encourage) end users on the latter side toward an exclusive relationship.

Typically merchants accept cards of several networks (multi-home), whereas consumers seem to multi-home less often than merchants and use one type of card (single-home in usage), even when they hold several cards.³⁸ This implies that card networks compete for cardholders (competitive bottlenecks), and thus ignore merchants’ surplus. In this case, platform competition would not decrease merchant fee, but it lowers fixed fees paid by cardholders. We, therefore, conjecture that our main result would be reinforced in this extension, i.e., cardholders are even more over-subsidized and the distortion on the price structure would be the same as in the case of a monopoly platform. On the other hand, multi-homing or single-homing decisions should be treated as endogenous variables. One interesting research agenda is to analyze which side of the market is going to use/accept one type of card in equilibrium of competing networks when networks (or their members) could charge fixed fees as well as per-transaction fees.

6 Implications for Other Two-sided Markets

It is important to note that in any two-sided market where one group of users make membership and usage decisions at different information sets, whereas the other group makes only membership

³⁷See Rochet and Tirole (2003), Guthrie and Wright (2003, 2007), and Armstrong (2006).

³⁸According to the Nilsen Report the acceptance network of the two major players almost perfectly overlaps with 29 million shops for Visa and 28.5 million for MasterCard. Even if consumers adopt more than one payment card, Rysman (2007) finds empirical evidence that they mostly use only one of them.

decisions, and transactions between the end-users are observed by the platform, our results suggest that the platform sets potentially inefficient user prices by over-subsidizing the side which decides on usage after membership decisions are made. Interesting examples are online search engines, such as Google, Yahoo!, and Bing, which try to attract viewers and advertisers; and video game platforms, such as Sony and Nintendo, which try to get game players as well as game developers on board. Regarding search engines, consumers decide whether to do an online search and then whether to click a title displayed on the search result, where online search services are free for consumers. However, advertisers decide only on membership, i.e., whether to purchase a place on the paid section of search results, and they pay per transaction, which is per click in this example. Our analysis predicts that, compared to the socially optimal price structure, a search engine will over-subsidize consumers and over-tax advertisers. Regarding video game platforms, consumers decide whether to purchase a game console and then decide whether to buy a game to play on their console, whereas game developers decide only whether to pay to a game platform to make their game compatible with the console, and then would pay rebates to the platform per game sold to consumers. In this context, our results predict that a video game platform will over-subsidize consumers and over-tax game developers.

7 Relationship with the Literature

The literature on two-sided markets³⁹ shows that a monopoly platform distorts upwards the total level of the per-transaction prices due to its market power, which is “market power distortion” as named by Weyl (2010). Besides this the platform distorts the structure (allocation) of the total price between the two sides, which Weyl refers to as “Spence distortion” since it arises from the platform’s inability to price discriminate across heterogeneous users on each side.

The common finding is that the existence and sign of “Spence distortion” is not straightforward since its sign depends on hardly measurable variables, e.g., the relative average surplus and the relative demand elasticity of the two sides when there is only usage heterogeneity (Rochet and Tirole, 2003); it does not exist when there is only membership heterogeneity (Armstrong, 2006); and, in general, it depends on the source of user heterogeneity (Weyl, 2010). Platform competition would correct the market power distortion on the total price level. However, it is not clear whether platform competition would reduce the distortion on the price structure, since it is difficult to identify the sign of *Spence* distortion.

As opposed to the literature, we show that *Spence* distortion (on the price structure) always over-subsidizes the side which determines the usage volume for given membership levels. In particular, in the payment card industry, cardholders are over-subsidized and merchants are over-taxed compared to the socially optimal price structure.

³⁹For general two-sided market frameworks see Rochet and Tirole (2003, 2006), Armstrong (2010), and Weyl (2010). See also Guthrie and Wright (2003, 2007) and Wright (2004) focusing on the payment card industry.

Table 1. Parameter values of our setup corresponding to the literature:

Literature	User Heterogeneity	Pricing Tools
Rochet and Tirole, 2003	two degrees: $B_B = B_S = 0$	$F = M = 0$
Armstrong, 2006	two degrees: $\underline{b}_B = \overline{b}_B$, $\underline{b}_S = \overline{b}_S$	$f = m = 0$

Table 1 shows how the pioneering literature corresponds to particular cases of our setup. Rochet and Tirole (2003) allow for population heterogeneity only on transaction benefits on the two sides, therefore two degrees of heterogeneity at the platform level, and considers only transaction prices. This corresponds to assuming $B_B = B_S = 0$ and $F = M = 0$ in our framework. Armstrong (2006) allows for population heterogeneity only on fixed benefits, therefore two degrees of heterogeneity at the platform level, and considers fixed (transaction-insensitive) prices. In our setup, this corresponds to assuming b_B is the same for all cardholders ($\underline{b}_B = \overline{b}_B$), b_S is the same for all merchants ($\underline{b}_S = \overline{b}_S$), and $f = m = 0$. The canonical model of Rochet and Tirole (2006) and Weyl (2010) allow for heterogeneity on both fixed and transaction benefits on both sides of the market, therefore four degrees of heterogeneity. Rochet and Tirole (2006) consider fixed fees as well as transaction prices, but they note that there is some redundancy in the pricing tools of the monopoly platform, since only per-transaction prices matter. Weyl (2010) does not specify any form of the tariff.

While allowing for four degrees of user heterogeneity, we boil down unclear normative implications of the literature to a unique equilibrium where the distortion on the price structure is always in one direction: too much cross-subsidization from merchants to cardholders. We obtain this clear result because, different from the previous work, we distinguish consumers' card membership decisions from card usage decisions by allowing for ex-ante (before membership decisions) uncertainty on transaction benefits.

Critical Distinction: Cardholding vs Card Usage Decisions and Fees The previous models assume that consumers are fully informed about their benefits before their membership (cardholding) decisions, so considering linear or non-linear card fees would give the same results in their analysis. This timing, in the payment card network, implies that consumers make only one decision, whether to hold the card or not, by comparing their average benefit with the average card fee. As a result, all cardholders pay by card at all merchants accepting cards.

In our setup, consumers get the card in order to secure the *option* (expected value) of paying by card in the future whenever this happens to be convenient for a particular transaction. This implies that only a fraction of the potential transactions takes place ex-post, i.e., cardholders might prefer to pay by another method if they realize lower benefits from card usage at the point of sale. Such formulation has mainly two advantages. First, it is able to rationalize frequent use of cash by many cardholders. Second, and most importantly, it distinguishes card membership from card

usage decisions (and fees), by assuming that these two decisions are made at different information sets. Such timing is also used by Guthrie and Wright (2003),⁴⁰ however, their paper restricts the analysis to linear fees, and thus does not recognize the implications of this timing.

Main Contribution: By distinguishing membership from transaction decisions, and, at the same time, by allowing firms to charge fixed fees in addition to transaction fees, we spell out the implications of an important asymmetry between consumers and merchants: Consumers make two distinct choices (membership and usage) whereas merchants make only one (membership). In our framework, it makes a difference whether tariffs to consumers are levied on a lump-sum or per-transaction basis or a combination of the two, since the platform is able to capture the average card usage surplus of cardholders through fixed card fees, but fails to account for the average card usage surplus of affiliated merchants, even if fixed merchant fees are in use. This is because cardholders use their cards only if they realize positive usage surplus at a point of sale, whereas affiliated merchants cannot refuse cards (and thus cannot affect card usage volume), even when they realize that a card transaction incurs a cost to them. As a result, the card network accounts for the card usage surplus of the average consumer and the card usage surplus of the marginal merchant, whereas a social planner also cares for the change in the surplus of inframarginal merchants.

Homogeneous Merchants: Allowing for population heterogeneity on both sides is critical to our results. If we assumed that all merchants have the same transaction benefits from card payments as any other payment method, $\underline{b}_S = \bar{b}_S$, all merchants would then accept cards if and only if $b_S \geq m$. Perfectly competitive acquirers set $m^*(a) = c_A + a$.

In this case, Baxter (1983) shows that setting an IF equal to $b_S - c_A$, which we call Baxter's IF, implements efficient card usage if issuers are also perfectly competitive, since then they set $f^*(a) = c_I + a$. Intuitively, the first best could be implemented through the usage fee that induces buyers to internalize the externality they impose on the rest of the economy while paying by card, i.e., $f^{fb} = c - b_S$. His analysis is restricted to be normative since perfectly competitive banks have no preferences over the level of IF. Going beyond Baxter, we assume imperfectly competitive issuers, and thus the privately optimal IF is well-defined in our analysis.

When issuers have market power, card fees are linear and fixed benefits from cardholding are zero (or the same for everyone), Guthrie and Wright (2003, Proposition 2) show that the socially optimal IF results in under-provision of card payment services. The reason is the following. The regulator would like to set an IF above Baxter's IF to induce the optimal card usage in the presence of an issuer markup. But then merchants would not participate (as $m > b_S$). At the second best, the regulator sets Baxter's IF, which is also the privately optimal IF and results in under-provision of card services. Hence, allowing for fixed card fees prevents inefficient provision of card services

⁴⁰As noted by Rochet and Tirole (2006, pp. 661), a similar timing assumption is used in several articles on two-sided markets. For instance, Anderson and Coate (2005), Bakos and Katsamakovs (2004), Caillaud and Jullien (2003), Hagiu (2006).

by eliminating issuer markups:⁴¹

Proposition 5 *When merchants are homogeneous, the privately and the socially optimal IFs always coincide. Furthermore,*

- i. If imperfectly competitive issuers can charge only linear usage fees, there is under-provision of card payment services.*
- ii. If membership (fixed) fees are also available, there is socially optimal provision of card payment services.*

Intuitively, since issuers could capture incremental card usage surpluses of buyers through fixed fees, they set the usage fees at their transaction costs, $c_I + a$. Baxter’s IF then implements the first best transaction volume.

8 Conclusion

This paper focuses on one payment card network, e.g., Visa or MasterCard, in which member banks provide card services to their customers: either consumers or merchants. The aim is to analyze whether the profit-maximizing user prices deviate from their welfare-maximizing levels. We illustrate a distortion in the structure of the user prices in the sense that, in equilibrium, cardholders pay too low card usage fees at the expense of too high merchant fees, compared to the socially optimal prices. This result implies that there is over-usage of cards at affiliated merchants; at point of sales accepting cards, an inefficiently high fraction of purchases are settled by card. We, therefore, predict that putting a cap on interchange fees, which are per-transaction fees paid by the merchant’s bank (*acquirer*) to the cardholder’s bank (*issuer*), could correct the distortion on the structure of user prices, and thereby improve the social welfare. Determining the optimal level of a cap goes beyond the scope of this paper because the optimal fee structure depends on barely measurable average net surpluses of cardholders and of merchants from card usage rather than using other payment methods.

The literature outlines conditions under which the privately optimal price structure is distorted (by favoring either cardholders or merchants too much). However, these conditions depend on unobservable consumer and merchant preferences, bank costs as well as the relative competitiveness between the market for consumers and the market for merchants. These conditions and therefore the direction of distortion on the price structure are very difficult to identify in practice.

Different from the literature, our results have clear policy implications since we show that the distortion on the card price structure is systematic, i.e., does not depend on the quantitative comparisons between the market for consumers and the market for merchants. We indeed find that merchants pay excessive transaction prices and cardholders pay inefficiently low card usage prices.

⁴¹A formal proof of the proposition is available upon request from the authors.

We demonstrate that the price structure distortion originates from an asymmetry between consumers and merchants: consumers make two distinct decisions, card membership and card usage, at different information sets, whereas merchants decide only on membership, i.e., whether to accept the scheme’s cards or not. To illustrate the implications of this asymmetry on the card price structure, we allow banks to charge fixed fees as well as transaction prices.

We develop a two-sided market framework of the industry where consumers and merchants are assumed to be heterogeneous in their card membership and card usage benefits. We thereby derive elastic demands of cardholding, card usage and card acceptance from first principals. In this setup, we show that the card scheme seeking to maximize its members’ profits sets an inefficiently high interchange fee since higher interchange fees induce lower card usage fees (or rewards) and thus encourage card usage, making the payment card more valuable at the membership stage. The additional surpluses of cardholders can be extracted by higher fixed fees. However, it is not possible to capture incremental card usage surpluses of merchants since they cannot affect card usage once they become a member of the card network. This inefficiency result is valid also in a third-party card scheme, like AMEX, since in this case one company has incentives to set inefficiently low card usage fees and high merchant fees.

By characterizing the first-best, “Lindahl” and the second-best, “Ramsey”, prices, we show that regulating the interchange fee corrects only the distortion in the price structure, but this is not enough to achieve full efficiency in the payment card industry, since efficiency requires each user fee be discounted by the positive externality of that user on the rest of the industry, and one tool (interchange fee) is not enough to achieve efficient usage on both sides. Intuitively, we suggest that if a card scheme charges its member banks fixed membership fees as well as transaction fees⁴², the platform could induce both consumers and merchants to internalize their externalities, and thus increase its profits as well as improve efficiency. We leave the characterization of an efficient interchange mechanism for future research.

The qualitative results are robust to issuer competition and merchant competition (or strategic card acceptance to attract consumers). We conjecture that our results could also be extended to the cases of imperfect acquirer competition and network competition.

This analysis provides significant insights for other two-sided markets. In any two-sided market where only one side decides on transaction after membership decisions are made and the platform could observe transactions between the two sides, our results predict that there will potentially be welfare distortions since the platform sets inefficiently low prices to the side, which determines the extent of trade at given number of participants, and over-taxes the other side. For instance, we anticipate that search engines over-subsidize viewers at the expense of advertisers, paying excessive fees per click and video game platforms over-subsidize game players and over-tax game developers.

Our setup inherits all the practical limitations of setting socially optimal prices that depend on barely observable characteristics of supply and demand. At this point we provide a theoretical framework which is hopefully rich enough to be used by an empirical analysis to characterize

⁴²In this case, different transaction fees could be set to cardholders’ banks versus merchants’ banks.

the socially optimal price structure, and thereby determine the optimal tool to correct welfare distortions in the payment card industry.

Appendix

A Benchmark Analysis

A.1 Proof of Lemma 1

We first show that $v'_B(f) < 0$ and $v'_S(m) < 0$ under the Increasing Hazard Rate Property (thereafter IHRP) of distribution functions respectively $G(f)$ and $K(m)$. Consider first $v_B(f)$. Using $D_B(f) = 1 - G(f)$ and integrating by parts, we get

$$v_B(f) \equiv \frac{\int_f^{\bar{b}_B} D_B(x) dx}{D_B(f)}. \quad (14)$$

Define $Y(f) \equiv \int_f^{\bar{b}_B} D_B(x) dx$. Notice that the IHRP is equivalent to say $D'_B/D_B \equiv Y''/Y'$ is a decreasing function. Given that Y''/Y' is decreasing and that $Y(\bar{b}_B) = 0$ and $Y(f)$ is strictly monotonic by definition, we have that Y'/Y is decreasing due to Bagnoli and Bergstrom (1989, Lemma 1).⁴³ Using (14), decreasing Y'/Y is equivalent to $v'_B(f) < 0$. Similarly, we can establish that $v'_S(m) < 0$. Since $v'_B \equiv -\frac{v_B D'_B + D_B}{D_B}$ and $v'_S \equiv -\frac{v_S D'_S + D_S}{D_S}$, inequalities $v'_B(f) < 0$ and $v'_S(m) < 0$ imply, respectively, that $v_B D'_B + D_B > 0$ and $v_S D'_S + D_S > 0$.

Define functional I as

$$I \equiv -\frac{(\text{HR}^{-1})'}{1 - (\text{HR}^{-1})'}$$

where HR^{-1} is the inverse of hazard rate, $\frac{1-H}{h}$, and thus decreasing by the IHRP. Note that $0 < I(\cdot) < 1$.

Given the best responses of the issuer ($f^*(a) = c_I - a$ and $F^*(a) = \frac{1-H(F^*(a)-\Phi_B(a))}{h(F^*(a)-\Phi_B(a))}$) and acquirers ($m^*(a) = c_A + a$), we now characterize interchange fees a^B , a^V , a^S which, respectively, maximize the buyer surplus (gross of fixed fees), the total transaction volume, and the seller surplus subject to the subgame perfection.

Existence and uniqueness of a^B :

First notice that the IHRP and $v'_B < 0$ imply respectively, the log-concavity of D_S and $v_B D_B$, and thus Φ_B is log-concave. An important property of continuous log-concave functions is that the first

⁴³The Generalized Mean Value Theorem of calculus ensures, for every x , the existence of a $\xi \in (x, \bar{b}_B)$ such that

$$\frac{Y'(x) - Y'(\bar{b}_B)}{Y(x) - Y(\bar{b}_B)} = \frac{Y''(\xi)}{Y'(\xi)}$$

If Y''/Y' is decreasing, for any $x < \xi$, it should then be the case that

$$\frac{Y'(x) - Y'(\bar{b}_B)}{Y(x) - Y(\bar{b}_B)} < \frac{Y''(x)}{Y'(x)}$$

Since Y is monotone and $Y(\bar{b}_B) = 0$, it must then be that $Y'(x)Y(x) < 0$ whenever $x < \bar{b}_B$. Multiplying both sides of the above inequality by $Y'(x)Y(x)$ gives $Y''(x)Y(x) < (Y')^2 - Y'(\bar{b}_B)Y'(x)$ and thus that $Y''(x)Y(x) - (Y')^2 < 0$, which is equivalent to Y'/Y decreasing.

order condition is both necessary and sufficient to have a local (and thus a global) maximum.⁴⁴

Hence, there exists a unique IF which maximizes the option value.

The buyers-optimal interchange fee, a^B , is a solution to:

$$\max_a BS(a) = \left[\int_{F^*(a) - \Phi_B(a)}^{\bar{B}_B} xh(x)dx + \Phi_B(a)Q(F^*(a) - \Phi_B(a)) \right],$$

$$\text{where } \Phi_B(a) = v_B(c_I - a)D_B(c_I - a)D_S(c_A + a).$$

This problem has an interior solution only if $f = c_I - a \leq \bar{b}_B$, which is equivalent to $a \geq c_I - \bar{b}_B$, because otherwise no one pays by card. The quasi-demand D_B is maximized and equal to 1 when $f = c_I - a \leq \underline{b}_B$, that is $a \geq c_I - \underline{b}_B$, and there is no gain from increasing a above $c_I - \underline{b}_B$. Without loss of generality, we thus restrict the domain of a to be $[c_I - \bar{b}_B, c_I - \underline{b}_B]$. By the Weierstrass Theorem, there exists a maximum of the continuous function $BS(a)$ on the compact interval $[c_I - \bar{b}_B, c_I - \underline{b}_B]$. By differentiating $F^*(a)$, we get

$$F^{*'}(a) = I(F^*(a) - \Phi_B(a))\Phi'_B(a), \quad (15)$$

which implies that $[F^* - \Phi_B]' = -(1 - I)\Phi'_B$. We therefore conclude that the IF, which maximizes $\Phi_B(a)$, minimizes $[F^*(a) - \Phi_B(a)]$, therefore maximizes $\int_{F^*(a) - \Phi_B(a)}^{\bar{B}_B} xh(x)dx$. Since cardholding demand $Q \equiv 1 - H$ is log-concave by the IHRP, the IF which maximizes $\Phi_B(a)$ also maximizes $\Phi_B(a)Q(F^*(a) - \Phi_B(a))$. We thus conclude that a^B is unique and equal to $\arg\max_a \Phi_B(a)$.

The *existence and uniqueness of a^S* : The sellers-optimal IF, a^S , is a solution to

$$\max_a SS(a) = v_S(c_A + a)D_S(c_A + a)D_B(c_I - a)Q(F^*(a) - \Phi_B(a)).$$

The Weierstrass Theorem guarantees the existence of a^S on $[c_I - \bar{b}_B, c_I - \underline{b}_B]$. Log-concavity of functions $v_S D_S$ (by $v'_S < 0$), D_B (by the IHRP), and Q (by the IHRP), implies that a^S is uniquely determined by the first-order optimality condition:

$$SS'(a^S) = -D_S(D_B + v_S D'_B)Q + (1 - I)\Phi'_B h v_S D_S D_B = 0 \quad (16)$$

The *existence and uniqueness of a^V* : The volume-maximizing IF, a^V , is a solution to

$$\max_a V(a) = D_B(c_I - a)D_S(c_A + a)Q(F^*(a) - \Phi_B(a)).$$

The Weierstrass Theorem guarantees the existence of a^V on $[c_I - \bar{b}_B, c_I - \underline{b}_B]$. Since quasi-demands D_B , D_S and cardholding demand Q are log-concave (implied by the IHRP), the volume of transactions $D_B D_S Q$ is log-concave. The unique interchange fee, a^V , is then implicitly given by the

⁴⁴To see this notice that by definition, a function $f(x)$ is log-concave if $\log(f(x))$ is concave, which is equivalent to f'/f decreasing or $f''f - (f')^2 < 0$. It follows that if f is log-concave, at any critical point the SOC must then be verified, i.e., for any x^* such that $f'(x^*) = 0$, we have $f''(x^*) < 0$

first-order optimality condition:

$$V'(a^V) = (-D'_B D_S + D'_S D_B) Q + (1 - I) \Phi'_B h D_B D_S = 0. \quad (17)$$

Now, our claim is $a^B > a^V$. By using the definition of a^B , i.e., $\Phi'_B(a^B) = D_B D_S + v_B D_B D'_S = 0$, we derive the volume of transactions at a^B :

$$V'(a^B) = -\frac{Q D_S}{v_B} (v_B D'_B + D_B).$$

We have $V'(a^B) < 0$ since $v_B D'_B + D_B > 0$ from $v'_B < 0$. Given that function $V(a)$ is concave (by the IHRP), condition (17) implies then that $a^B > a^V$.

Symmetrically, by using the IHRP and $v'_S < 0$, it can be shown that $a^S < a^V$. Hence, we prove that $a^S < a^V < a^B$.

A.2 Proof of Proposition 1

By definition a^B maximizes the surplus of buyers (gross of fixed fees) and a^S maximizes the surplus of sellers. Lemma 1 shows the existence and the uniqueness of a^B and a^S , and that $a^B > a^S$. By the revealed preference argument an interchange fee maximizing the sum of user surpluses necessarily lies in (a^S, a^B) .

B Efficient Fees

B.1 Proof of Proposition 2

We decompose the planner's problem of setting transaction prices f, m into a price allocation and a total price setting problem. We have already characterized in Proposition 1 the optimal allocation of total price $f + m = p = c$. We are thus left to generalize the optimal allocation of any total price p and characterize, then, the optimal p . Let $f(p)$ and $m(p)$ denote the respective fees which implement the optimal allocation of p between buyers and sellers.

The social planner first solves

$$\max_f [p - c + v_B(f) + v_S(p - f)] D_B(f) D_S(p - f) Q(F - \Phi_B(f, p - f)) + \int_{F - \Phi_B(f, p - f)}^{\bar{B}_B} x h(x) dx,$$

which characterizes implicitly $f^{FB}(p)$ and $m^{FB}(p) = p - f(p)$ as follows:

$$\begin{aligned} & [(p - c)(D'_B D_S - D_B D'_S) - v_B D_B D'_S + v_S D'_B D_S] Q - \\ & (p - c + v_B + v_S) D_B D_S Q' \partial_f \Phi_B + (F - \Phi_B) h(F - \Phi_B) \partial_f \Phi_B = 0, \end{aligned} \quad (18)$$

where $Q' < 0$ and $\partial_f \Phi_B$ denotes the derivative of the option value, $\Phi_B(f, p - f)$, with respect to f .

The planner next determines the socially optimal total price by

$$\max_p [p - c + v_B(f(p)) + v_S(p - f(p))] D_B(f(p)) D_S(p - f(p)) Q(F - \Phi_B) + \int_{F - \Phi_B}^{\bar{B}_B} x h(x) dx,$$

Using $[v_i D_i]' = -D_i$ and the Envelope Theorem, we get the first order condition:

$$(p - c + v_B) D_B D_S' Q - (p - c + v_B + v_S) D_B D_S Q' \partial_p \Phi_B + (F - \Phi_B) h(F - \Phi_B) \partial_p \Phi_B = 0. \quad (19)$$

Finally, the socially optimal membership fee $F^{FB}(p, f(p))$ is characterized by:

$$(p - c + v_B + v_S) D_B D_S Q' = (F - \Phi_B) h(F - \Phi_B). \quad (20)$$

Plugging (20) into 19 gives:

$$(p - c + v_B) D_B D_S' Q = 0, \quad (21)$$

which is verified if and only if $p^{FB} = c - v_B(f^{FB})$. Plugging (20) and p^{FB} into condition (18) we get:

$$(v_S - v_B) D_B' D_S Q = 0, \quad (22)$$

which implies that $v_S(p^{FB} - f^{FB}) = v_B(f^{FB})$.

C Competing Issuers

C.1 Examples of Demand Functions

The following examples of demand functions for differentiated products satisfy assumptions A2-A6.

(1) Linear symmetric demands of form, for $i = 1, 2, i \neq j$,

$$q_i = \frac{1}{1 + \sigma} - \frac{1}{1 - \sigma^2} p_i + \frac{\sigma}{1 - \sigma^2} p_j,$$

where q refers to demand, p refers to price, and σ measures the level of substitution between the firms (here, for imperfectly competitive issuers we have $\sigma \in (0, 1)$). These demands are driven from maximizing the following quasi-linear and quadratic utility function

$$U(q_i, q_j) = q_i + q_j - \sigma q_i q_j - \frac{1}{2} (q_i^2 + q_j^2),$$

subject to the budget balance condition, namely

$$p_i q_i + p_j q_j \leq I.$$

(2) Dixit (1979)'s and Singh and Vives (1984)'s linear demand specification, for $i = 1, 2, i \neq j$,

$$q_i = a - b p_i + c p_j,$$

where $a = \frac{\alpha(\beta-\gamma)}{\beta^2-\gamma^2}$, $b = \frac{\beta}{\beta^2-\gamma^2}$, $c = \frac{\gamma}{\beta^2-\gamma^2}$, and the substitution parameter is $\varphi = \frac{\gamma^2}{\beta^2}$, under the assumptions that $\beta > 0$, $\beta^2 > \gamma^2$, and $\varphi \in (0, 1)$ for imperfect substitutes.

(3) Shubik and Levitan (1980)'s demand functions of form, for $i = 1, 2, i \neq j$,

$$q_i = \frac{1}{2} \left[v - p_i(1 + \mu) + \frac{\mu}{2} p_j \right],$$

where $v > 0$, μ is the substitution parameter and $\mu \in (0, \infty)$ for imperfect substitutes.

Special case: Hotelling Demand, for $i = 1, 2, i \neq j$,

$$q_i = \frac{p_j - p_i}{2t} + \frac{1}{2}$$

satisfies the assumptions except for A4 and A6 since the own price effect is equal to the cross price effect, that is

$$\left| \frac{\partial q_i}{\partial p_i} \right| = \frac{\partial q_i}{\partial p_j}, \quad \left| \frac{\partial^2 \ln q_i}{\partial p_i^2} \right| = \frac{\partial^2 \ln q_i}{\partial p_i \partial p_j},$$

which imply that the equilibrium fixed fees are independent of the option value, and thus independent of the IF. In this case, the issuers would not have any preferences over IF. Hence, the privately optimal IF is not well defined.

C.2 Proof of Lemma 2

Consider the FOC of I_i 's problem:

$$FOC_i : Q(F_i - \Phi_B, F_j - \Phi_B) + F_i \frac{\partial Q_i}{\partial F_i} = 0.$$

Solving FOC_i and FOC_j together gives us the equilibrium fees as functions of the option value, i.e., $F_i^*(\Phi_B)$ and $F_j^*(\Phi_B)$. The second-order condition holds by A5:

$$SOC_i : 2 \frac{\partial Q_i}{\partial F_i} + F_i^* \frac{\partial^2 Q_i}{\partial F_i^2} < 0.$$

The solution of the issuers' problems gives us the symmetric equilibrium $F_i^* = F_j^*$. By taking the total derivative of the first-order conditions, we derive

$$\frac{\partial F_j^*}{\partial \Phi_B} = \frac{\partial F_i^*}{\partial \Phi_B} = 1 - \frac{\partial Q_i / \partial F_i}{SOC_i + \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j}}.$$

If $\frac{\partial^2 \ln Q_i}{\partial t_i \partial t_j} < 0$, we have

$$\frac{(\partial^2 Q_i / \partial F_i \partial F_j) Q_i - (\partial Q_i / \partial F_i)(\partial Q_i / \partial F_j)}{Q_i^2} < 0,$$

so that

$$\frac{\partial Q_i}{\partial F_j} - \frac{Q_i}{\partial Q_i / \partial F_i} \frac{\partial^2 Q_i}{\partial F_i \partial F_j} < 0.$$

From FOC_i we have, $F_i^* = -\frac{Q_i}{\partial Q_i / \partial F_i}$, so we get

$$\frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j} < 0.$$

Moreover, the log-concavity of Q_i (A5) implies that $SOC_i < \partial Q_i / \partial F_i$. Thus, we get $0 < \frac{\partial F_i^*}{\partial \Phi_B} < 1$.

If $\frac{\partial^2 \ln Q_i}{\partial t_i \partial t_j} > 0$, we have

$$\frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j} > 0.$$

Assumption A6 becomes $-\frac{\partial^2 \ln Q_i}{\partial t_i^2} > \frac{\partial^2 \ln Q_i}{\partial t_i \partial t_j}$, which implies that

$$-\left[\frac{\partial Q_i}{\partial F_i} + F_i^* \frac{\partial^2 Q_i}{\partial F_i^2} \right] > \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j},$$

Using SOC_i , we get

$$\partial Q_i / \partial F_i > SOC_i + \frac{\partial Q_i}{\partial F_j} + F_i^* \frac{\partial^2 Q_i}{\partial F_i \partial F_j},$$

proving that $0 < \frac{\partial F_i^*}{\partial \Phi_B} < 1$.

C.3 Proof of Proposition 3

Following the lines of our benchmark analysis, we first define three important IF levels: the buyers-optimal IF, the sellers-optimal IF, and the volume maximizing IF, which we denote, respectively, by a^{Bc} , a^{Sc} , and a^{Vc} , where superscript c refers to issuer competition:

$$\begin{aligned} a^{Bc} &\equiv \arg \max_a \left\{ v_B(c_I - a) D_B(c_I - a) D_S(c_A + a) [Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)] + \right. \\ &\quad \left. \int_{F_1^* - \Phi_B(a)}^{\bar{B}_B} x h(x) dx + \int_{F_2^* - \Phi_B(a)}^{\bar{B}_B} x h(x) dx \right\} \\ a^{Sc} &\equiv \arg \max_a v_S(c_A + a) D_B(c_I - a) D_S(c_A + a) [Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)] \\ a^{Vc} &\equiv \arg \max_a D_B(c_I - a) D_S(c_A + a) [Q(F_1^*, F_2^*, a) + Q(F_2^*, F_1^*, a)] \end{aligned}$$

From Lemma 2, we have $0 < \frac{\partial F_i^*}{\partial \Phi_B} = \frac{\partial F_j^*}{\partial \Phi_B} < 1$. Consider now the derivative of $Q(F_i^*, F_j^*, a)$ with respect to a :

$$Q'_i(a) = \left[\frac{\partial Q_i}{\partial F_i} \left(\frac{\partial F_i^*}{\partial \Phi_B} - 1 \right) + \frac{\partial Q_i}{\partial F_j} \left(\frac{\partial F_j^*}{\partial \Phi_B} - 1 \right) \right] \Phi'_B(a)$$

The first term inside the brackets represents the direct effect of the option value on Q_i , through changing the net price of card i , $F_i^* - \Phi_B$, and the second term represents the indirect effect of the option value on Q_i , through changing the net price of card j , $F_j^* - \Phi_B$. Imperfect issuer competition (A3 and A4) implies that the direct effect of the option value on Q_i dominates its indirect effect so

that the term inside the brackets is positive. We therefore conclude that when two differentiated issuers are competing with symmetric demands, the demand for holding card i is maximized at $a = a^B$, which is the interchange fee, maximizing the option value of the card, $\Phi_B = v_B D_B D_S$.

Following the lines of Lemma 1, we then conclude that the IF maximizing the option value of the card also maximizes the buyer surplus (gross of fixed fees) when the issuers are imperfect competitors, i.e., $a^{Bc} = a^B$. Recall that the association sets $a^* = a^B$ to maximize the issuers' payoffs. Hence, the privately optimal IF coincides with the buyers-optimal IF.

Since the average surplus of buyers and the average surplus of sellers are decreasing in their own usage fees, i.e., $v'_B(f), v'_S(m) < 0$ (see the proof of Lemma 1), we have $a^{Sc} < a^{Vc} < a^{Bc}$. The regulator wants to maximize the sum of buyers' and sellers' surpluses, the socially optimal IF is therefore lower than the privately optimal one.

The formal proof of Proposition 4 is available upon request.

References

- ARMSTRONG, M. (2002): “The Theory of Access Pricing and Interconnection,” *Handbook of Telecommunications Economics*, 1, 295–384.
- (2006): “Competition in Two-Sided Markets,” *Rand Journal of Economics*, 37(3), 668–681.
- ARMSTRONG, M., AND J. VICKERS (2006): “Competitive Nonlinear Pricing and Bundling,” Discussion paper.
- BAXTER, W. (1983): “Bank Interchange of Transactional Paper: Legal and Economic Perspectives,” *Journal of Law and Economics*, 26(3), 541–588.
- CHAKRAVORTI, S. (2003): “Theory of Credit Card Networks: A Survey of the Literature,” *Review of Network Economics*, 2(2), 50–68.
- (2010): “Externalities in Payment Card Networks,” *Review of Network Economics*, forthcoming.
- EUROPEAN COMMISSION (2007): “Final Report on the Retail Banking Sector,” Discussion paper.
- EVANS, D., AND R. SCHMALENSEE (2005a): *Paying With Plastic: The Digital Revolution In Buying And Borrowing*. MIT Press.
- EVANS, D., AND R. SCHMALENSEE (2005b): “The Economics of Interchange Fees and Their Regulation: An Overview,” *a Conference Interchange Fees in Credit and Debit Card Industries: What Role for Public Authorities*, pp. 4–6.
- GANS, J., AND S. KING (2003a): “Approaches to Regulating Interchange Fees in Payment Systems,” *Review of Network Economics*, 2(2), 125–145.
- GANS, J., AND S. KING (2003b): “The Neutrality of Interchange Fees in Payment Systems,” *Topics in Economic Analysis and Policy*, 3(1).
- GUTHRIE, G., AND J. WRIGHT (2003): “Competing Payment Schemes,” Departmental Working Papers wp0311, National University of Singapore, Department of Economics.
- (2007): “Competing Payment Schemes,” *Journal of Industrial Economics*, 55(1), 37–67.
- HAYASHI, F. (2009): “Do U.S. consumers really benefit from payment card rewards?,” *Economic Review*, (Q I), 37–63.
- LAFFONT, J., S. MARCUS, P. REY, AND J. TIROLE (2003): “Internet Interconnection and the Off-Net-Cost Pricing Principle,” *The RAND Journal of Economics*, 34(2), 370–390.

- RESERVE BANK OF AUSTRALIA (2007): “Reform of Australia’s Payments System: Issues for the 2007/08 Review,” Discussion paper.
- ROCHET, J. (2003): “The Theory of Interchange Fees: A Synthesis of Recent Contributions,” *Review of Network Economics*, 2(2), 97–124.
- ROCHET, J., AND J. TIROLE (2002): “Cooperation among Competitors: Some Economics of Payment Card Associations,” *The RAND Journal of Economics*, 33(4), 549–570.
- ROCHET, J., AND J. TIROLE (2003): “Platform Competition in Two-Sided Markets,” *Journal of the European Economic Association*, 1(4), 990–1029.
- ROCHET, J., AND J. TIROLE (2006a): “Must-Take Cards and the Tourist Test,” Discussion paper, Working Paper, IDEI.
- (2006b): “Two-Sided Markets: A Progress Report,” *The RAND Journal of Economics*, 37(3), 645–667.
- ROCHET, J., J. TIROLE, AND V. INTERNATIONAL (2002): *An Economic Analysis of the Determination of Interchange Fees in Payment Card Systems*. Visa International].
- RYSMAN, M. (2009): “The economics of two-sided markets,” *Journal of Economic Perspectives*, 23(3), 125–143.
- SCHMALENSEE, R. (2002): “Payment Systems and Interchange Fees,” *Journal of Industrial Economics*, 50(2), 103–122.
- WANG, Z. (2009): “Market structure and payment card pricing: What drives the interchange?,” *International Journal of Industrial Organization*.
- WEINER, S., AND J. WRIGHT (2005): “Interchange Fees in Various Countries: Developments and Determinants,” *Review of Network Economics*, 4(4), 290–323.
- WEYL, G. (2009): “Monopoly, Ramsey and Lindahl in Rochet and Tirole (2003),” *Economics Letters*.
- (2010): “A Price Theory of Multi-Sided Platforms,” *American Economic Review*, *forthcoming*.
- WRIGHT, J. (2001): “The Determinants of Optimal Interchange Fees,” Working paper no.220, Department of Economics, University of Auckland.
- (2004): “The Determinants of Optimal Interchange Fees in Payment Systems,” *The Journal of Industrial Economics*, 52(1), 1–26.

Recent ESMT Working Papers

	ESMT No.	Competence Center
Pricing Payment Cards Özlem Bedre-Defolie, ESMT Emilio Calvano, Bocconi University	10-005	European Competitiveness
Profiting from technological capabilities: Technology commercialization strategy in a dynamic context Simon Wakeman, ESMT	08-008R2	Management and Technology
Conditional Cooperation: Evidence for the Role of Self-Control Peter Martinsson, University of Gothenburg Kristian Ove R. Myrseth, ESMT Conny Wollbrant, University of Gothenburg	10-004	Management and Technology
Reconciling Pro-Social vs. Selfish Behavior: Evidence for the Role of Self Control Peter Martinsson, University of Gothenburg Kristian Ove R. Myrseth, ESMT Conny Wollbrant, University of Gothenburg	10-003R1	Management and Technology
Estimating Critical Mass in the Global Cellular Telephony Market Michał Grajek, ESMT Tobias Kretschmer, Ludwig-Maximilians-Universität München	08-004R1	Management and Technology
Marketing Social Responsibility Sumitro Banerjee, ESMT Luc Wathieu, Ferrero Chair in International Marketing, ESMT	10-002	Management and Technology
Quantification of Harm in Damages Actions for Antitrust Infringements: Insights from German Cartel Cases Hans W. Friederiszick, ESMT CA Lars-Hendrik Röller, ESMT	10-001	European Competitiveness
A Framework for Monitoring Relational Quality in B2B Technology Partnerships Francis Bidault, ESMT Manfred Lüth, University of Nice Olaf Plötner, ESMT	09-008	Management and Technology
Enjoy! Assertive Language and Consumer Compliance in (Non)Hedonic Contexts Ann Kronrod, Tel-Aviv University Amir Grinstein, Ben-Gurion University of the Negev Luc Wathieu, Ferrero Chair in International Marketing, ESMT	09-007	Management and Technology
ISO 9000: New Form of Protectionism or Common Language in International Trade? Joseph A. Clougherty, University of Illinois at Urbana-Champaign Michał Grajek, ESMT	09-006	European Competitiveness
Demography vs. Context: A Cross-Country Survey of the Willingness to Rely on Trust in Business Partnerships Francis Bidault, ESMT José R. de la Torre, Florida International University Stelios H. Zanakos, Florida International University	09-005	Management and Technology
Regulation and Investment in Network Industries: Evidence from European Telecoms Michał Grajek, ESMT Lars-Hendrik Röller, ESMT	09-004	European Competitiveness



ESMT
European School of Management
and Technology GmbH

Schlossplatz 1
10178 Berlin
Germany

Phone: +49 (0) 30 212 31-1279

www.esmt.org