Tying in Two-Sided Markets and The Impact of the Honor All Cards Rule

Jean-Charles Rochet*

and

Jean Tirole†

March 26, 2004

(Preliminary, Please do not cite without permission)

*IDEI, Toulouse University, and London School of Economics.
†IDEI, Toulouse University, and MIT.

Abstract

Payment card associations Visa and MasterCard offer both debit and credit cards and, until recently, used to engage in a tie-in on the merchant side through the so-called honor-all-cards (HAC) rule. This rule has come under attack on the grounds that the credit and debit card markets are separate markets and that the associations lever their market power in the "credit card market" to exclude on-line debit cards and thereby monopolize the "debit card market".

The objective of this article is to analyze the impact of the HAC rule. We construct a simple model of the payment card industry in which there are two types of transactions, debit and credit, and two platforms: one offering only a debit card, and the other offering both a debit card and a credit card.

In this benchmark model we show that in the absence of HAC, the interchange fee (IF, the transfer from the merchant's bank to the cardholder's bank) on debit is socially too low, and that on credit may be optimal or too high (depending on downstream members' market power). In either case, though, the HAC rule not only benefits the multi-card platform but also raises social welfare, due to a *rebalancing effect:* The HAC rule allows the multi-card platform to better perform the balancing act by raising the IF on debit and lowering it on credit, ultimately raising volume.

The paper then investigates a number of extensions of the benchmark model to allow for varying degrees of substitutability between debit and credit; different structures of cardholders' information about merchants' card acceptance policies; merchant heterogeneity with respect to their customers debit/credit mix; and platform differentiation. While the HAC rule may no longer raise social welfare under all values of the parameters, the basic and socially beneficial rebalancing effect unveiled in the benchmark model is robust.

1 Introduction

Buyers use both credit and debit cards. The credit facility brings about substantial benefits to some consumers, for some types of purchases or at specific moments of time. In other circumstances, credit is not needed.

Payment card associations Visa and MasterCard offer both debit and credit cards and, until recently, used to engage in a tie-in on the merchant side through the honor-all-cards (HAC) rule. In a class action initiated by WalMart (and involving more than five million U.S. merchants), this rule has come under attack on the grounds that the credit and debit card markets are separate markets and that the associations lever their market power in the "credit card market" to exclude on-line debit cards and thereby monopolize the "debit card market". Visa and Master Card have recently agreed to eliminate their HAC rules and pay over $ 3 billion in damages to the merchants.

The objective of this article is to analyze the impact of the HAC rule. We construct a simple model of the payment card industry in which there are two types of transactions, debit and credit. However, as a starting point, it is important to remind oneself of the economics of two-sided markets with a single type of card (credit or debit).

Relevant lessons from the literature

(1) With a single card, an association's choice of interchange fee is constrained in two ways:

- Even if the association faced no competition from another system, it would have to get both sides on board. The interchange fee must be high enough so as to induce consumers to use the card, but low enough so as not to meet merchant resistance.
- When competing with other payment systems, the association is further constrained, as each system tries to de-stabilize its rivals' balancing act, for example through steering strategies when consumers hold multiple cards (consisting in undercutting on the merchants' side, so as to incentivize them to turn down the rival card).

(2) Due to its not-for-profit status, an association cannot exercise its market power by inflating the overall price *level*. By contrast, the association has discretion in the allocation of cost between cardholders and merchants and, like ordinary firms, it may or may not get the price *structure* (i.e., the relative prices for different end users) "socially right". On the one hand, both a social planner and an association ought to design the price structure so as to account for the elasticities on both sides of the market and thereby get both sides on board. On the other hand, the literature has identified factors, such as downstream

(issuer, acquirer) market power or merchants' competition for market share, that may tilt an association's (or, for that matter, a proprietary system's) price structure away, in either direction, from the socially optimal one.

System competition is one such factor. Leaving aside the standard benefits of competition on managerial incentives (e.g., through the owners' ability to benchmark their management's performance), system competition's impact on prices has an ambiguous impact on welfare because competition influences only the price *structure* and not the price *level* (which, as we noted, must track cost, due to the not-for-profit status).

When consumers hold multiple cards, system competition tends to tilt the price structure toward lower merchant discounts and higher cardholder fees. The reason for this is that merchants have an incentive to turn down the card that is most expensive to them if consumers hold both cards. This is clearly the case if cardholders are unaware of the merchants' card acceptance policies before purchasing and so merchants obtain no competitive edge over rival merchants by accepting a card (the "Baxter[1] case").

Interestingly, this is still the case even when consumers are informed of the merchants' card acceptance policies and card acceptance buys merchants a strategic edge (as is assumed in the most of the treatment below). In the latter case, the card that is cheapest for the merchants is also less attractive for the cardholders (since the interchange fee is lower, cardholders benefits are lower — or their fees higher). Yet, we show below that competition reduces the interchange fee. Take the case in which merchants are homogenous (or more generally merchant heterogeneity is observable). In the polar case of system monopoly, the association, in order to maximize volume, chooses an interchange fee equal to the highest value that merchants will bear. In the other polar case of perfect system competition, the equilibrium interchange fee is the (lower) one that maximizes total user surplus. Intuitively, merchants prefer the card that gives them the highest sum of their own surplus (convenience benefit minus merchant discount) and of the cardholder's average surplus (convenience benefit minus cardholder fee), since they internalize the latter when trying to attract customers. Under monopoly, the system tries to please consumers in order to maximize volume; under competition, merchants have an important say on the usage of a specific card, because rejecting a card is much less costly to them; and so merchants receive a better deal (and the cardholders a worse deal).

System competition focuses an association's attention on the system's own elasticities rather than on the socially more relevant end-user elasticities. And so, whether competition improves social welfare depends on the initial price-structure bias of a monopoly

[1]This refers to Baxter (1983), the first formal analysis of the determination of interchange fees in payment card networks. Baxter assumes that merchants' acceptance decision are only driven by their convenience benefits from card payments and assume that card acceptance does not help them attract customers.

system . If the merchant discount (or the interchange fee) is initially too high, then competition forces it down and may improve welfare. By contrast, competition reduces welfare if the merchant discount (or the interchange fee) was initially too low.

Tying in two-sided markets.

The issue at stake, however, is not whether competition improves welfare, but whether, given competition, the HAC rule increases or decreases welfare. A first intuition might be that "bundling reduces competition, and so, if competition is socially desirable, bundling reduces welfare as well". This intuition turns out to be incorrect. We show that, *regardless of the desirability of system competition, the HAC rule always improves welfare.*[2]

Let us return to the two-card (debit, credit) context. A system issues both types of cards, and faces more intense competition on one segment. To simplify the exposition, suppose that the system is a monopoly on credit cards and faces an on-line competitor for debit cards that is a perfect substitute. Then in the absence of the HAC rule, the outcome is the monopoly outcome for credit and the competitive outcome for debit. From our previous analysis, the interchange fee is higher on credit than on debit. This interchange fee structure is not the one predicated by the demand specificities of the two-sided markets, but rather reflects the difference in the merchants' "bypass opportunities".

Suppose now that a merchant has to accept the system's two cards or none. The total user surplus (which is also the merchants' reservation utility from accepting the system's cards) is the same as in the absence of the HAC rule: It is equal to the highest total user surplus that can be offered by the on-line network. The new feature, though, is that the system gains *flexibility to rebalance its interchange fee structure* as the competitive constraint binds over the set of cards, rather than fully over the debit card. The system can therefore *increase volume* by raising the interchange fee on debit and lowering interchange fee on credit. Social welfare always increases.

Comparison with tying in one-sided markets.

There are a number of key differences with the literature on tying.

First, we analyze tying by an association. Hence, anticompetitive motives, like entry deterrence, cannot be associated with the standard purpose of raising the price level (an association can only effect the price structure).

Second, tying occurs in a two-sided market. There is then a natural benefit of tying in terms of a greater flexibility to rebalance charges between the two sides.

In Whinston (1990), tying is entirely motivated by entry deterrence: By lowering its

[2]This statement presumes that merchants are homogenous. When they are heterogeneous, the situation is more contrasted: this case is discussed in Section 6.

opportunity cost of selling in the competitive market (losing a sale in that market implies losing a sale in the monopoly market), the tying firm commits to be agressive and thereby may deter entry. Furthermore, tying must be technologically irreversible, since it reduces the tying firm's profit whether the rival enters or not. By contrast tying is time-consistent in our model.

In our model, like in Whinston's, tying may deter entry of a more efficient rival. But the consequences are different because a not-for-profit firm cannot exercise its market power by raising the price level. So, for example, deterrence of entry of a slightly more able rival through tying still raises welfare in our model.[3, 4]

2 The Model

We extend the model of the payment card industry developed in Rochet and Tirole (2002), by introducing two types of cards (debit cards and credit cards, respectively indexed by superscripts $k = d$ and c), and two competing networks (indexed by subscripts $i = 1$ and 2). Both networks are not-for-profit associations run by their members.[5] Network 1 only offers a debit card, while network 2 offers both a debit card and a credit card. The two debit cards are perfect substitutes for both cardholders and merchants.

Figure 1 shows the costs and benefits attached to a card transaction. The total cost of this service is the sum of the issuer's cost c_I and the acquirer's cost c_A. Suppose that the benefit accruing to the cardholder (or buyer) for the marginal use of a payment card is equal to b_B. Similarly, the benefit to the merchant (or seller) of this marginal use of a payment card is b_S. The benefits b_i and costs c_i referred to above are *net* benefits and costs. The cardholder and the merchant must compare the utilities they get by using payment cards with those associated with alternative payment methods (cash, checks, etc.). At the social optimum, the total benefit of the marginal transaction, $b_B + b_S$, is equal to its total cost, $c_I + c_A \equiv \gamma$. The left side of Figure 1 also features the payments from end users to intermediaries: cardholders pay f to issuers and merchants pay merchant discount m to acquiers. These two fees are market determined given the association's choice of interchange fee.

[3] We ignore "corporate governance" or "benchmarking" benefits of product market competition. But there are to a large extent internalized by the tying firm.

[4] Also, it is unclear whether tying really deters entry in practice. ATM cards and others after all have entered.

[5] We later examine the case where one (or both) network(s) is(are) for-profit.

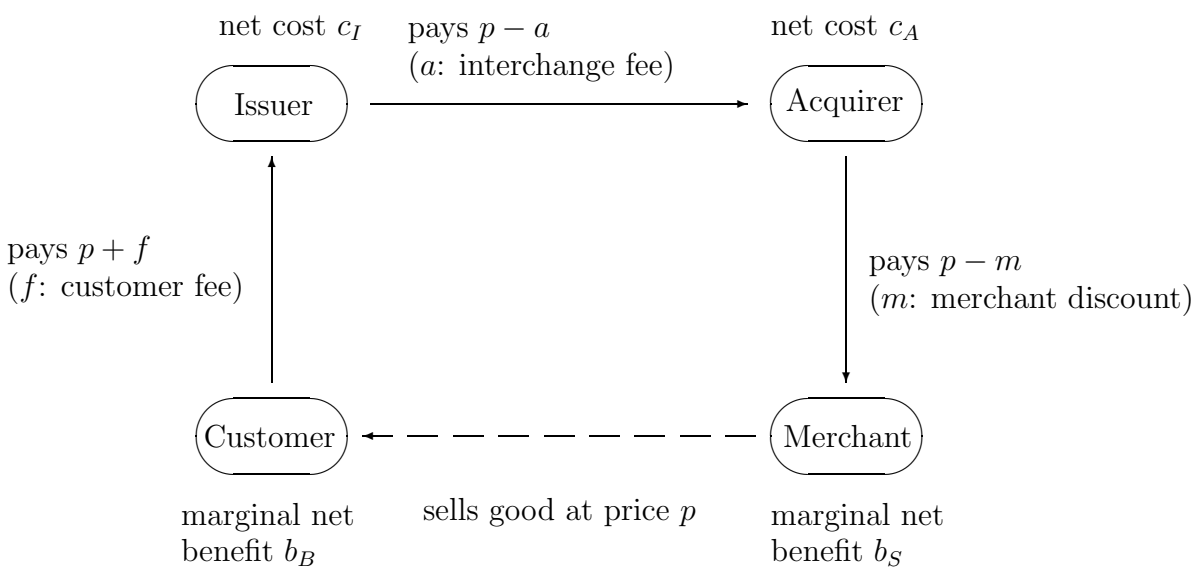

Figure 1

As in Rochet and Tirole (2002) we focus on the choice of interchange fees by the two networks. Final prices for both types of users (merchants and cardholders) are determined by the extent of competition in downstream (issuing and acquiring) markets. For simplicity we assume that margins on the issuing and acquiring sides are constant, that is users' prices react one for one to variations of issuers' and acquirers' net costs. Constant mark-up by issuers and acquirers offer the convenient simplification that all members are congruent (issuers and acquirers want to maximize network volume), and so the modeling of the governance structure is a no-brainer. Formally, for all three cards, cardholders' per transaction fees f_i^k (recall that $k = d, c$ represents the type of card, debit or credit, while $i = 1, 2$ represents the network offering the card) and merchant discounts m_i^k are related to interchange fees a_i^k by the following formulas:

$$f_i^k = f_0^k - a_i^k \tag{1}$$

$$m_i^k = m_0^k + a_i^k, \tag{2}$$

where f_0^k and m_0^k are given. Note that total user price is independent of the interchange fee:

$$f_i^k + m_i^k = f_0^k + m_0^k = \gamma^k + \pi^k, \tag{3}$$

where γ^k is the (total) unit cost of payment services with card k and π^k is the (total) margin of acquirers and issuers. These costs and margins are identical for the two debit cards.

Following Wright (2001), we assume that cardholders are ex-ante identical. However ex-post, i.e. once they have decided on a purchase in a given store, their transactional

benefit b_B^k from using card k ($k = d, c$) rather than cash (or check) is drawn from a distribution with a positive density $h^k(b_B^k)$.

In the simplest version of our model, we neglect any substitutability between credit and debit cards. In other words we assume for the moment that there are two distinct subsets of transactions: N^d transactions for which payment can be made by debit or cash, N^c other transactions for which payment can be made by credit or cash. The idea is that, for a given transaction, the cardholder may or may not need the credit facility.

Assuming that all buyers hold the two debit cards, buyers will want to use only the one(s) with the lowest cardholder usage fee (among the cards accepted by merchants).

3 Merchants' Acceptance Decisions

Rational merchants anticipate that their decisions whether to accept payment cards influence their competitive position. By accepting card k (with user prices f^k and m^k), a merchant indeed increases the expected utility of his future customers by

$$u^k = \int_{f^k}^{+\infty} (b_B^k - f^k) h^k(b_B^k) db_B^k.$$

However it also increases his net expected cost by

$$c^k = \int_{f^k}^{+\infty} (m^k - b_S^k) h^k(b_B^k) db_B^k,$$

where b_S^k denotes the convenience benefit derived by the seller from a payment by card k rather than a cash payment. If the price charged by the merchant (for the good or service he produces) is p, and c denotes the unit production cost, accepting card k amounts to charging an effective price $p - u^k$ and to incurring an effective cost $c + c^k$. It is as if the expected profit margin of the merchant were increased by an amount:

$$u^k - c^k = \phi^k(f^k) = \int_{f^k}^{+\infty} \left(b_B^k + b_S^k - f^k - m^k\right) h^k(b_B^k) db_B^k.$$

Notice that ϕ^k is equal to the expected *total users' surplus* derived from card k. Given relation (3), ϕ^k can also be written:

$$\phi^k(f^k) = \int_{f^k}^{+\infty} \left(b_B^k + b_S^k - \gamma^k - \pi^k\right) h^k(b_B^k) db_B^k.$$

Total users' surplus is a quasi-concave function of f^k (see Figure 2 below). It has a maximum for

$$f^k = f_c^k \equiv \gamma^k + \pi^k - b_S^k.$$

[For reasons that will become clear shortly, the subscript "c" refers to the competitive outcome.] This value is greater that the value f_*^k that maximizes social welfare:

$$f^k = f_*^k \equiv \gamma^k - b_S^k.$$

This is due to the facts that social welfare equals the sum of users' surplus and total profit of issuers and acquirers, and that issuers' and acquirers' profits increase with transaction volume, which itself decreases with cardholder usage fee f^k (this is because cardholders ultimately decide on the payment instrument).

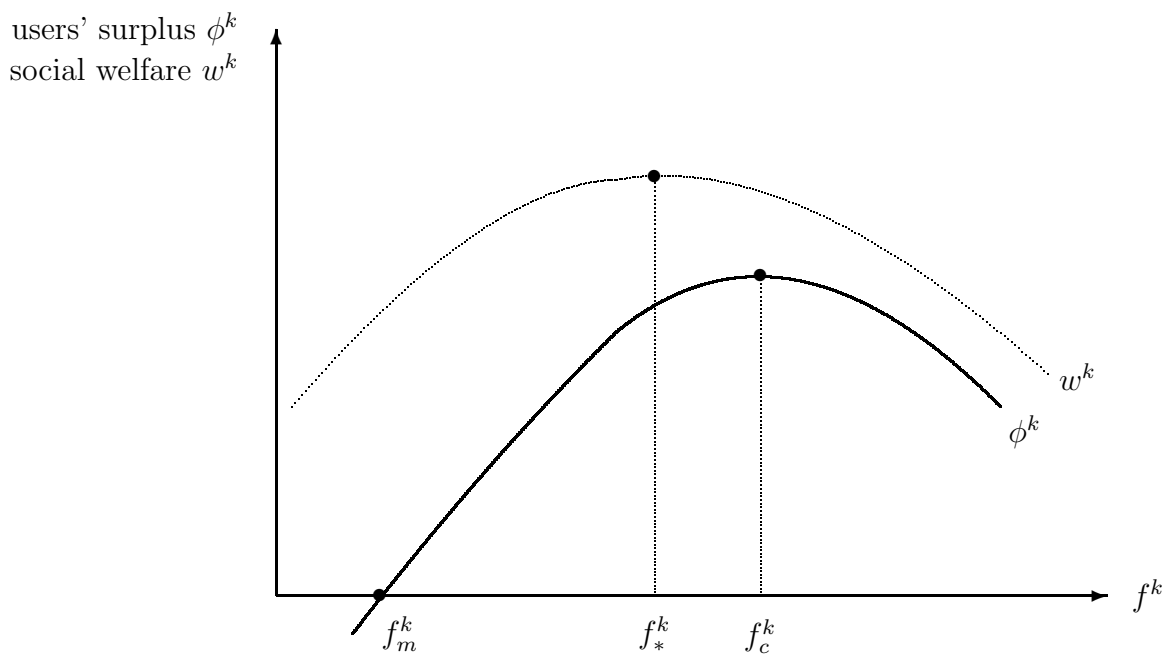

Figure 2: Users' surplus for card k, as a function of cardholders fee f^k. The dotted curve represents social welfare. f_m^k is the minimum fee such that users' surplus is non-negative.

Since volume decreases with f^k, it is a dominated strategy for networks to choose f^d above $f_c^d = \gamma^d + \pi^d - b_S^d$.

Making the innocuous assumption that the profit of each merchant is a decreasing function of its net cost $c - \phi$, the merchants' equilibrium behavior is to accept the set of cards that maximizes total users' surplus,[6] taking into account possible redundancies: Consider for example the case of debit cards alone. Total user surplus is $\phi^d(f_i^d)$ if the merchant accepts card i alone, and $\phi^d(\min(f_1^d, f_2^d))$ if the merchant accepts both. This is because in the latter case, only the least expensive card (from the cardholders' viewpoint) is used. Since networks never set fees above f_c^d, we can restrict attention to the interval[7] $[f_m^d, f_c^d]$, where ϕ^d is increasing and thus $\phi^d[\min(f_1^d, f_2^d)] \leq \phi^d(f_i^d)$. Thus merchants accept both cards if $f_1^d = f_2^d$ and only the one which gives the largest surplus (i.e. the one with the largest cardholder fee) if $f_1^d \neq f_2^d$.

To avoid technical problems[8] and without impact on the real allocation, we will assume that merchants accept card 1 only if accepting card 1 *strictly* increases their profit.

Proposition 1 *: If the Honor All Cards (HAC) rule is not imposed by network 2, merchants accept debit card 1 alone if* $\phi^d(f_1^d) > \max(0, \phi^d(f_2^d))$ *and debit card 2 alone if* $\phi^d(f_2^d) \geq \max(0, \phi^d(f_1^d))$*. The credit card is accepted whenever* $\phi^c(f^c) \geq 0$*.*

To see how the Honor All Cards rule modifies merchants' acceptance decisions, assume now that network 2 forces merchants to accept its debit card whenever they take its credit card. The choice confronted by merchants is summarized by the following table:

Cards accepted:	debit 1	debit 2 and credit	all cards
Users' surplus	$N^d\phi^d(f_1^d)$	$N^d\phi^d(f_2^d) + N^c\phi^c(f^c)$	$N^d\phi^d(\min(f_1^d, f_2^d)) + N^c\phi^c(f^c)$

Table 1: Merchants' options under the HAC rule.

[Recall that $N^k(k = d, c)$ represents the total number of transactions for which card k can be used.]

Merchants compare the users' surpluses associated with their three acceptance options. It is still the case that inducing f_k^d above f_c^d is a dominated strategy. Given that networks

[6] This result is derived in Wright (2002) in the context of a Hotelling model and in Wright (2003) in a Cournot model, but the reasoning is clearly fully general.

[7] Recall that $f_m^d = \min\{f/\phi^d(f) \geq 0\}$.

[8] Technically, this assumption is made in order to avoid an "openess problem" (similar to the standard openess problem encountered under Bertrand competition with unequal costs) in the case of bundling.

charge interchange fees such that $f_i^d \leq f_c^d$, if merchants accept network 2's cards, they cannot increase their profit by accepting network 1's debit card: Either $f_1^d < f_2^d$ and then card 1 yields a lower total users' surplus, so that merchants strictly prefer their customers not to use card 1 (which they would do if merchants took card 1); or $f_2^d \leq f_1^d$ and card 1 is not used by buyers, and so accepting or refusing card 1 is a matter of indifference to merchants. Thus:

Proposition 2 *: Under the HAC rule, merchants accept debit card 1 alone if*

$$N^d \phi^d(f_1^d) > \max\left[0, N^d \phi^d(f_2^d) + N^c \phi^c(f^c)\right]$$

and both cards of network 2 if

$$N^d \phi^d(f_2^d) + N^c \phi^c(f^c) \geq \max\left[0, N^d \phi^d(f_1^d)\right].$$

4 The Outcome of Network Competition

Networks select interchange fees (or equivalently cardholder fees) so as to maximize their members' profit.[9] In order to attract merchants, network 2 has to offer a total users' surplus at least equal to the one offered by network 1. The equilibrium strategy of network 1 is to choose f_1^d that maximizes ϕ^d, namely $f_1^d = \gamma^d + \pi^d - b_S^d$. When the HAC rule is not enforced, this leaves no choice for network 2 other than also "setting" $f_2^d = \gamma^d + \pi^d - b_S^d$. As for credit cards, network 2 maximizes its members' profit by choosing $f_m^c \equiv \{\min f \mid \phi^c(f) \geq 0\}$.

Proposition 3 *: In the absence of HAC rule, network competition results in identical fees for debit cards*

$$f_1^d = f_2^d = \gamma^d + \pi^d - b_S^d \qquad \textit{(competitive fee).}$$

The credit card fee is equal to the minimum cardholder fee that is compatible with merchants' acceptance:

$$f_m^c = \min\{f | \phi^c(f) \geq 0\} \qquad \textit{(monopoly fee).}$$

The outcome of network competition in the absence of HAC rule is represented by Figure 3 below. For simplicity, this figure depicts the "symmetric" case, that is, the case where total users' surpluses ϕ^c and ϕ^d and welfares w^c and w^d are identical for credit and debit cards.

[9] This is one simplification afforded by the assumption that issuers and acquires levy a constant margin. Their interests (the maximization of volume) are then congruent. Otherwise, we would need to investigate the bargaining power of members on either side, as in Schmalensee (2002).

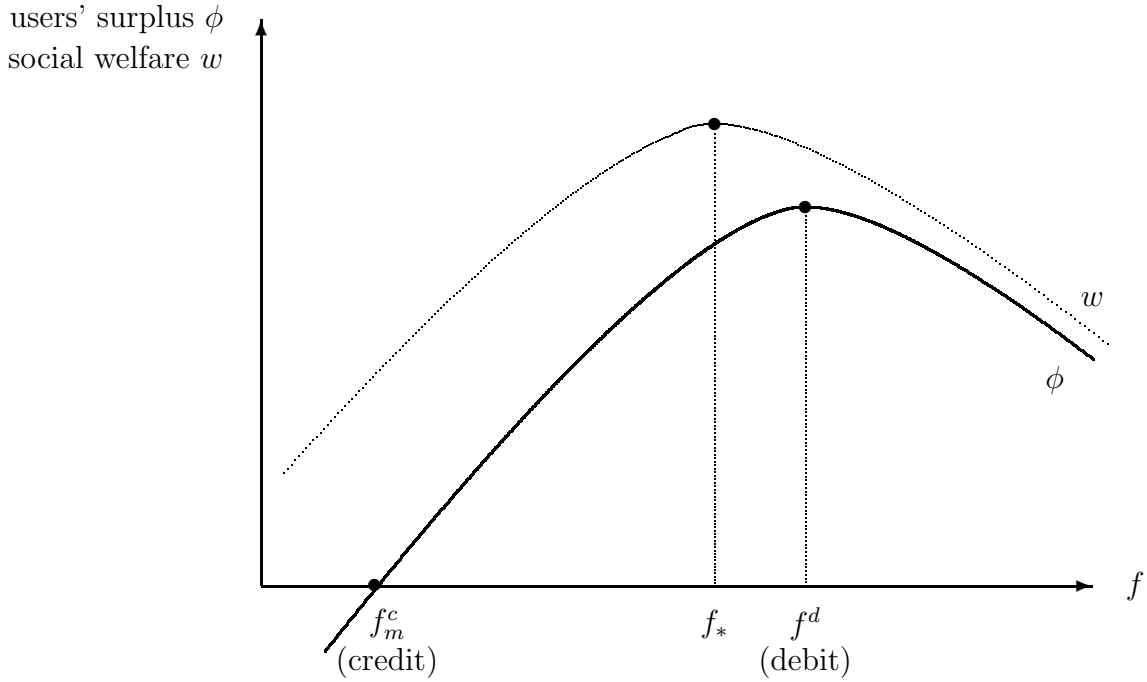

Figure 3: The outcome of network competition in the absence of HAC rule. The dotted curve represents social welfare.

More generally, since cardholder fees are decreasing functions of interchange fees, Proposition 3 implies that, in the absence of HAC rule, network competition leads to interchange fees that can be too high or too low for credit ($f_m^c \gtrless f_*^c$) but are always too low for debit (since $f_c^d > f_*^d$).

When the HAC rule is enforced, network 2 has one more degree of freedom. It can choose any combination of cardholder fees that provides users with at least the surplus created by network 1:

$$N^d \phi^d(f_2^d) + N^c \phi^c(f^c) \geq N^d \phi^d(f_1^d). \tag{4}$$

Network 2 chooses (f_2^d, f^c) so as to maximize its members' profit:

$$\Pi = \pi^d D^d(f_2^d) + \pi^c D^c(f_2^c), \tag{5}$$

where

$$D^k(f^k) \stackrel{\text{def}}{=} N^k \int_{f^k}^{+\infty} h^k(b_B^k) db_B^k$$

represents the demand functions of cardholders ($k = c, d$).

Proposition 4 *The HAC rule raises the interchange fee for debit and lowers the interchange fee for credit.*

Proof: Without the HAC rule, network competition forces network 2 to choose a high cardholder fee $f_c^d = \arg\max \phi^d$ for debit. On the other hand, cardholder fees for credit are as low as possible (i.e. the minimum fees that are compatible with merchants acceptance). With the HAC rule, the same combination of cardholder fees is still feasible. However by slightly increasing cardholder fees for credit cards, the network can substantially decrease cardholder fees for debit cards (while maintaining total users surplus constant). This is because the marginal user surplus is zero at f_c^d. Such a change increases total volume and is thus profitable for network 2. ■

The outcome of network competition with the HAC rule is represented in Figure 4 below, again in the "symmetric" case. The interchange fee for debit card 1 is unchanged (Recall that the customer fee of equals a constant minus the interchange fee). On the other hand, the HAC rule allows network 2 to increase interchange fees for debit card 2, and thus to increase volume. This is compensated by a decrease in the interchange fee for credit. Total user surplus remains unchanged, but total volume (and thus social welfare) increases.

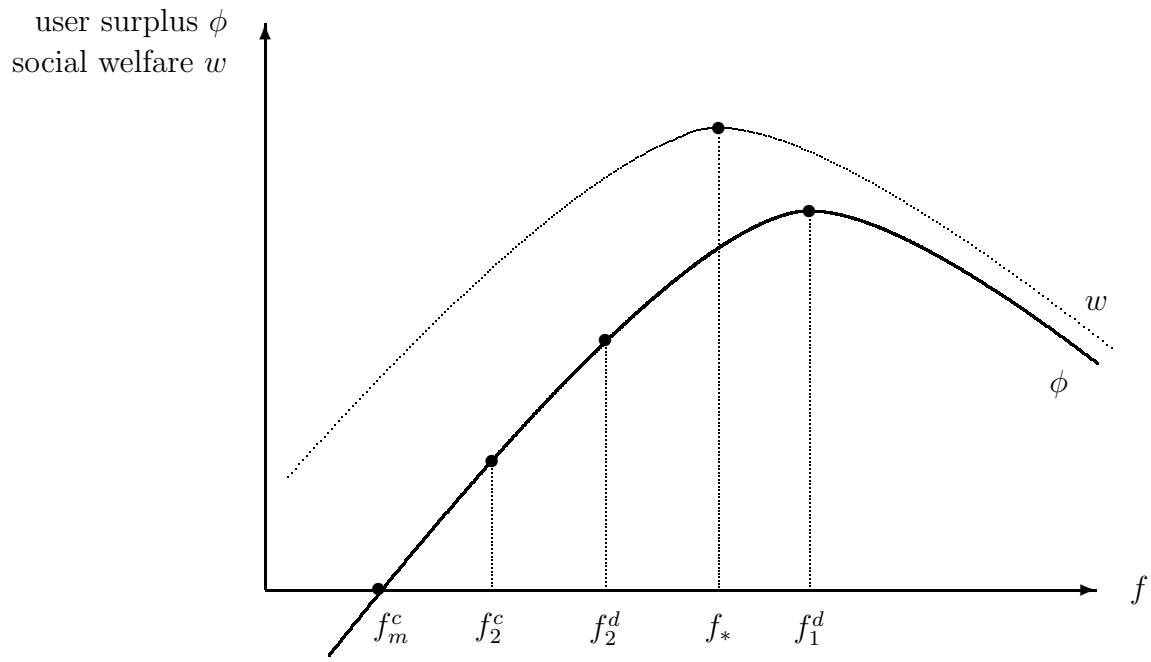

Figure 4: The outcome of network competition with the HAC rule.

5 Analyzing the Impact of the HAC Rule

We have thus been able to determine the outcome of network competition with and without the HAC rule. We have seen (Proposition 4) that the HAC rule allows network 2 to lower cardholder fees for debit and increase cardholder fees for credit. Total users'

surplus is identical. In both cases it is equal to the maximum surplus provided by network 1 alone. However total volume (and total profit) is higher with the HAC rule, since network 2 has more degrees of freedom. Therefore social welfare, which is the sum of total users' surplus and members' profits is enhanced by the HAC rule.

Proposition 5 *Under our assumptions (competing platforms, informed consumers, no substitution between debit and credit, and homogenous merchants), social welfare is higher under the HAC rule.*

6 Robustness

For simplicity, we have worked so far under several specific assumptions. The purpose of this section is to explore the robustness of our results to alternative assumptions.

6.1 Monopoly system

Our results do not rely on the assumption of competing systems. With a single platform, tying again improves social welfare. Indeed, in the absence of HAC rule, a monopoly platform would separately choose the maximum interchange fees (or equivalently minimum cardholder fees) that are acceptable to merchants. Namely: $f_m^d = (\phi^d)^{-1}(0)$ and $f_m^c = (\phi^c)^{-1}(0)$. With the HAC rule, the monopoly would select the combination of cardholder fees (f^d, f^c) that maximizes the members' total profit $\pi^d N^d D^d(f^d) + \pi^c N^c D^c(f^c)$ under the (single) constraint that merchants accept cards:

$$N^d \phi^d(f^d) + N^c \phi^c(f^c) \geq 0. \tag{6}$$

The previous combination of fees (f_m^d, f_m^c) satisfies this constraint because it has to secure acceptance of each card individually. However, the network has now more flexibility and generically will choose a different combination that gives the same total users' surplus (constraint (6) is binding) but a larger volume of transactions. Once again, social welfare is increased by the HAC rule.

Proposition 6 *: When there is a unique platform, social welfare is higher under the HAC rule.*

6.2 Uninformed consumers (Baxter's case)

One may object to our assumption that all consumers are aware of merchants' acceptance of cards before they choose which store to patronize. In order to check the robustness of

our results to this assumption, we now consider the polar case of Baxter (1983) where consumers are supposed to be totally uninformed. In this case, merchants only consider their net expected convenience benefit $B^k(f^k) = (b_S^k + f^k - \gamma^k - \pi^k)D^k(f^k)$ when deciding whether to accept card k. Propositions 1 through 4 carry through by replacing users' surpluses $\phi^k(f^k)$ by this convenience benefit $B^k(f^k)$. When the HAC rule is not enforced, competition forces both networks to choose the cardholder fee that maximizes merchants' convenience benefit for debit:

$$f_1^d = f_2^d = \arg\max B^d(f),$$

while the cardholder fee for credit is the minimum acceptable to merchants:

$$f^c = (B^c)^{-1}(0) = \gamma^c + \pi^c - b_S^c.$$

When the HAC rule is enforced, network 1 does not change its interchange fee, while network 2 selects (f_2^d, f^c) so as to maximize total profit under the (global) acceptance decision of merchants:

$$\begin{cases} \max \pi^d N^d D^d(f^d) + \pi^c N^c D^c(f^c) \\ \text{s.t.} \\ N^d B^d(f^d) + N^c B^c(f^c) \geq N^d \max_f B^d(f). \end{cases}$$

Once again the interchange fee increases for debit and decreases for credit and total volume is increased. Moreover the (total) convenience benefit of merchants remains the same. But the impact on social welfare is a priori ambiguous since cardholder usage fees for debit and credit move in opposite directions, and therefore the net surplus of cardholders may decrease or increase. However in the symmetric case (identical costs, identical margins and identical demand functions for debit and credit) tying still improves social welfare under a mild technical assumption (log concave demand function).

Proposition 7 *: In the symmetric Baxter case, tying improves social welfare when the demand function of cardholders is log-concave.*

Proof: See the appendix.

In the rest of the paper, we return to the case in which consumers are informed of card-acceptance policies before shopping.

6.3 Debit-credit Substitution

A simple way to introduce substitution is to assume that credit cards can also be used for debit transactions (while the opposite is not true). However this substitution is imperfect:

the utility obtained by the cardholder is only $b_B^d - \Delta$ (with $\Delta \geq 0$), when using a credit card for a debit transaction.[10] The possibility of substitution constrains the price differential:

$$f^d - f^c \leq \Delta. \tag{7}$$

For simplicity we stick to the symmetric case and therefore drop the index $k = d$ or c. When the HAC rule is enforced, the equilibrium is not altered by the possibility of substitution, since $f^c = f^d$, and thus constraint (7) does not bind. The same is not true when the HAC rule is not enforced: when $\Delta < f_c - f_m$, the previous equilibrium no longer obtains.

Proposition 8 *: Take the symmetric case and introduce the possibility of using credit cards for debit transactions (for a utility loss $\Delta < f_c - f_m$). Then:*

a) When the HAC rule is enforced, the equilibrium fees remain the same as in the absence of potential substitution.

b) When the HAC rule is not enforced, the interchange fee for credit is unchanged, the interchange fee for debit increases, and users' surplus decreases.

c) The HAC rule may decrease welfare when π is large and Δ is small. Otherwise it increases welfare.

6.4 Heterogenous Merchants

Suppose that merchants differ in the ratio $x = \frac{N^c}{N^d}$ of (potential) credit and debit transactions. x is private information of the merchant, and distributed on $[x_0, x_1]$ according to a c.d.f. G (and a density g). It is easy to see that the equilibrium without the HAC rule is not altered: One still has $f_1^d = f_2^d = f_c^d$ and $f^c = (\phi^c)^{-1}(0)$. Since $\phi^c(f^c) = 0$, merchants' acceptance decisions are the same, independently of x. The same is not true when the HAC rule is enforced, since in this case $\phi^c(f^c) > 0$ and $\phi^d(f_2^d) < \phi^d(f_1^d)$ at equilibrium. As can be seen from Table 1(p. 9), there exists a critical threshold $\hat{x}$ such that:

- merchants with $x < \hat{x}$ accept only debit card 1,
- merchants with $x > \hat{x}$ accept only the two cards of network 2.

[10] Δ may be strictly positive for several reasons: the cardholder may fail to pay the credit balance on time and thus may have to pay penalties for delay and high interest rates. Or else he may suffer from time inconsistent behavior as in the work of Angeletos et al. (2001).

Given that network 1 has now a positive volume of transactions, it will select a lower cardholder fee than in the previous equilibrium (by optimally trading off volume and probability of acceptance):

$$f_1^d < \arg\max \phi^d.$$

This implies that users' surplus generated by card 1 is smaller than when the HAC rule is not enforced:

$$\phi^d(f_1^d) < \max \phi^d.$$

On the other hand, for a given threshold $\hat{x}$, network 2 will select the combination (f^c, f_2^d) that maximizes volume under the constraint

$$\phi^d(f_2^d) + \hat{x}\phi^c(f^c) \geq \phi^d(f_1^d).$$

Since this constraint binds at equilibrium, the marginal merchant $\hat{x}$ obtains a lower users' surplus when the HAC rule is enforced (this is because $\phi^d(f_1^d) < \max \phi^d$).

Therefore the consequences of enforcing the HAC rule are less clearcut when merchants are heterogenous:

- Merchants with $x < \hat{x}$ refuse the two cards of network 2, and in particular the credit card. This generates a welfare loss due to foregone credit transactions.
- Merchants with intermediate x (above $\hat{x}$ but close to it) incur a loss of users' surplus due to the fact that f_1^d decreases.
- Merchants with large x benefit from the HAC rule since it induces network 2 to decrease interchange fee on credit.

Even though we cannot exclude that the HAC rule decreases social welfare when merchants' heterogeneity is substantial, a simple continuity[11] argument shows that this is not true when $(x_1 - x_0)$ is small (or when the distribution of x is fairly concentrated around a given value): In this case the equilibrium under HAC rule is very close to the previous one (when merchants are identical) and the HAC rule unambiguously increases welfare.

6.5 Competition between Differentiated Platforms

Our basic model considers the extreme case of a single credit card and two undifferentiated debit cards. Suppose, more generally, that the multi-card platform faces competition by

[11] More precisely, we use the fact that f_1^d converges to $\arg\max \phi^d$ when $(x_1 - x_0)$ converges to zero.

differentiated single-card platforms on each segment $k = c, d$. A simple way to introduce such a differentiation is to assume that the competing card on segment k generates a net transaction benefit $b_B^k + b_S^k - \gamma^k - \pi^k - \delta^k$ (instead of $b_B^k + b_S^k - \gamma^k - \pi^k$ for the k-card offered by the multi-card platform). Our basic model corresponds to the extreme case $\delta^c = +\infty$ (no competitor on credit), and $\delta^d = 0$ (perfect competition on debit). We now consider general values of (δ^c, δ^d). δ^k can be associated with differences in marginal cost or in convenience benefit. It can be positive (in which case the multi-card platform is more efficient on segment k) or negative (in which case the competing network is more efficient).

Without the HAC rule, only the more efficient card is accepted on each segment. When $|\delta^d|$ and $|\delta^c|$ are small, fees are then determined by "limit pricing" conditions on each segment: the most efficient card is priced at the minimum level that induces merchants to reject the competing card. The cardholder usage fee is thus such that it generates a users' surplus that could be provided by the alternative, less efficient card (provided of course that the latter be positive, so competition has some bite).

By contrast, when the HAC rule is enforced, competition between platforms is globalized. Consider for instance the case where $\delta^c > 0$ (the multi-card platform is more efficient for credit) and δ^d is either positive or slightly negative. In this case, the multi-card platform is in a position to exclude the two rival cards. It chooses a combination of fees that maximizes volume under the constraint that aggregate users' surplus (on both cards) is higher than the sum of maximum users' surplus that can provided by competing systems. Aggregate users' surplus (on both cards) is thus unchanged if $\delta^d \geq 0$ (i.e. when the multi-card platform is more efficient on both segments) and slightly decreased if δ^d is slightly negative (if δ^d is largely negative, entry deterrence is impossible).

In the first case ($\delta^d \geq 0$), social welfare is clearly increased by the HAC rule. Our usual argument remains valid: the HAC rule does not alter users' surplus, but increases volume and thus social welfare. When δ^d is negative, the HAC rule has two opposing effects on welfare:

- users' surplus is decreased, since the efficient debit card platform is excluded,
- price structure is "rebalanced" and total volume is increased.

Again the net effect of the HAC rule on social welfare becomes ambiguous, but by continuity our conclusion still holds: for δ^d negative but small, the rebalancing effect dominates the inefficiency effect, and social welfare is increased by the HAC rule.

APPENDIX:

Proof of Proposition 7:
Let us define the auxiliary function

$$v(b) = w(B^{-1}(b)), \quad \text{for} \quad b \leq \max_f B(f).$$

With the HAC rule, we have seen that

$$f^d = f^c = B^{-1}\left(\frac{N^d b^*}{N^d + N^c}\right), \quad \text{with} \quad b^* = \max_f B(f).$$

Welfare is thus equal to $(N^d + N^c)v\left(\frac{N^d b^*}{N^d+N^c}\right)$.

By contrast, when the HAC is not enforced

$$f^d = B^{-1}(b^*) \quad \text{and} \quad f^c = B^{-1}(0).$$

Welfare is thus equal to $N^d v(b^*) + N^c v(0)$. Thus social welfare is increased by tying if and only if

$$v\left(\frac{N^d b^*}{N^d + N^c}\right) \geq \frac{N^d v(b^*) + N^c v(0)}{N^d + N^c},$$

which is true whenever v is concave.

Proposition 7 then results from the following lemma:

Lemma 1 *: If* $\log D$ *is concave then* v *is concave.*

Proof:

$$\dot{v}\,(b) = \frac{\dot{w}\,[B^{-1}(b)]}{\dot{B}\,[B^{-1}(b)]}.$$

Since B^{-1} is increasing in the relevant range, we have to establish that $\frac{\dot{B}}{\dot{w}}(f)$ increases with f.

Now

$$B(f) = (f - \gamma + b_S - \pi)D(f)$$

$$\Rightarrow \dot{B}\,(f) = (f - \gamma + b_S - \pi)\,\dot{D}\,(f) + D(f).$$

Moreover,

$$w(f) = -\int_f^{+\infty} (b_B - \gamma + b_S)\,\dot{D}\,(b_B)db_B.$$

$$\Rightarrow \dot{w}\,(f) = (f - \gamma + b_S)\,\dot{D}\,(f).$$

Thus

$$\frac{\dot{B}}{\dot{w}}(f) = 1 - \frac{\left(\pi - \frac{D}{\dot{D}}\right)}{f - \gamma + b_S}.$$

By assumption $-\frac{D}{\dot{D}}$ is decreasing function of f (and it is > 0) therefore $-\frac{\dot{B}}{\dot{w}}$ increases with f and the lemma is established. ■

Proof of Proposition 8:

a) Since $f^d = f^c$ at equilibrium, constraint (7) is not binding. Network could set $f_1^d < f^c + \Delta$ to trigger substitution, but debit card 1 would then be refused by merchants. So the equilibrium is unchanged.

b) If $\Delta < f_c - f_m$ the previous equilibrium (when the HAC rule is not enforced) no longer obtains. We now establish that

$$f_1^d = f_2^d = \phi^{-1}(0) + \Delta, \quad f^c = \phi^{-1}(0)$$

is the new equilibrium, by looking at possible deviations of the two networks.

Network 1:

- $f_1^d < \phi^{-1}(0) + \Delta$ is not accepted by merchants.
- $f_1^d > \phi^{-1}(0) + \Delta$ generates no transactions for network 1, since cardholders prefer to use their credit card, "or network 2's debit card".

Network 2:

- $f^c = \phi^{-1}(0)$ is optimal given $f_2^d = \phi^{-1}(0) + \Delta$. Similarly $f_2^d = \phi^{-1}(0) + \Delta$ is optimal given f^c (for the same reason as for network 1).
- So, to benefit, network 2 must move both interchange fees, and in such a way that $f^c = f_2^d - \Delta$.
- If $f_2^d > f_1^d$, volume is reduced on debit and also on credit.
- If $f_2 < f_1^d$ (and thus $f^c < \phi^{-1}(0)$) merchants no longer accept the credit card (nor do they accept network 2's debit card).

So either way, a deviation by network 2 is not profitable. This completes the proof that $f^c = \phi^{-1}(0) + \Delta$, $f^c = \phi^{-1}(0)$ is the new equilibrium. f^c is thus unchanged, while f^d is lower (which implies that interchange fee for debit increases and users' surplus decreases).

c) To see that the HAC rule may decrease welfare, consider the case where π is so large that $\gamma - b_S < \phi^{-1}(0)$, and Δ is so small that $\phi^{-1}(0) + \Delta < \phi^{-1}\left(\frac{N^d}{N^d+N^c}\max\phi\right)$. In this case social welfare is a decreasing function of f on the interval $[\phi^{-1}(0), +\infty[$ and both $\phi^{-1}(0)$ and $\phi^{-1}(0)+\Delta$ are smaller than the equilibrium fee $\phi^{-1}\left(\frac{N^d}{N^d+N^c}\max\phi\right)$ under the HAC rule. In this case the HAC rule decreases welfare. ■

REFERENCES

Angeletos, G.M., D. Laibson, A. Repetto, J. Tobacman and S. Weinberg (2001), "The Hyperbolic Consumption Model: Calibration, Simulation and Empirical Evaluation", *Journal of Economic Perspective*, August, 47-68.

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

Guthrie, G. and J. Wright (2003), "Competing Payment Schemes", University of Auckland, Working Paper.

Rochet, J.C. and J. Tirole (2002), "Cooperation among Competitors: The Economics of Payment Card Associations", *Rand Journal of Economics*, 33(4), 1-22.

Schmalensee, R. (2002) "Payment Systems and Interchange Fees", the *Journal of Industrial Economics*, 50(2), 103-122.

Whinston, M.D. (1990), "Tying, Foreclosure and Exclusion", *American Economic Review*, 80(Sept.), 837-859.

Wright, J. (2001), "Determinants of Optimal Interchange Fees in Payment Systems", University of Auckland, Working Paper.

Wright, J. (2003), "Why Do Firms Accept Credit Cards?", University of Auckland, Working Paper.