Must-Take Cards and the Tourist Test*

Jean-Charles Rochet† and Jean Tirole‡

May 9, 2006

Abstract

Antitrust authorities often argue that merchants cannot reasonably turn down payment cards and are therefore forced to accept unacceptably high merchant discounts. The paper attempts to shed light on this "must-take cards" view from two angles.

It first identifies four key sources of potential social biases in the payment card associations' determination of interchange fees: internalization by merchants of a fraction of cardholder surplus, issuers' per-transaction markup, merchant heterogeneity, and extent of cardholder multi-homing. It compares the industry and social optima both in the short term (fixed number of issuers) and the long term (in which issuer offerings and entry respond to profitability).

Second, the paper gives some operational content to the notion of "must-take card" by introducing the "tourist test" (would the merchant want to refuse a card payment when a non-repeat customer with enough cash in her pocket is about to pay at the cash register?) and analyzes its relevance as an indicator of excessive interchange fees.

Keywords: Card payment systems, interchange fee, internalization, multi-homing, tourist test.

*We are grateful to Luis Cabral for useful comments.
†IDEI and GREMAQ (UMR 5604 CNRS), Toulouse University, France.
‡IDEI and GREMAQ (UMR 5604 CNRS), Toulouse, and MIT.

1 Introduction

In payment cards systems such as Visa or MasterCard, the interchange fee (IF) paid by the merchant's bank to the cardholder's bank allocates the total cost of the payment service between the two categories of users, cardholders and merchants. In several regions of the world, courts of justice, competition authorities, and banking regulators have claimed that these IFs are set at unacceptably high levels. Merchants, the argument goes, accept to pay the resulting high merchant discount because they are concerned that turning down cards would impair their ability to attract customers; that is, cards are "must-take cards" (Vickers 2005).[1]

In the UK, the Office of Fair Trading, following a multi-year investigation of MasterCard's credit card IFs, has announced its intention to regulate down these IFs, as well as plans to investigate the IFs set by Visa for credit card transactions. Similarly, under the pressure of the European Commission, Visa International agreed in 2002 to reduce its cross-border interchange fees on credit and debit transactions within the European Union. In Australia, after the publication of an extensive study of debit and credit card schemes in 2000, the Reserve Bank of Australia mandated a sizeable reduction of credit IFs, and is considering doing the same (or perhaps even mandating a zero IF) for debit transactions. Other countries where similar decisions have been made (or are seriously considered) by courts of justice, competition authorities or banking regulators include Israel, Spain, Portugal, Belgium, and the Netherlands.

This paper offers a model of the payment card industry[2] that is sufficiently rich to account for the complex effects of IFs on volumes of card payments, banks' profits, consumer

[1]The potential anticompetitive effects of IFs have been discussed in a sizeable antitrust literature, in particular Carlton and Frankel (1995), Evans and Schmalensee (1995), Frankel (1998), Chang and Evans (2000) and Balto (2000). This literature is surveyed in Schmalensee (2003).

[2]This model elaborates on previous theoretical analyses of the impact of IFs, in particular Rochet and Tirole (2002), Schmalensee (2002), Wright (2003a, 2003b, 2004). This literature is surveyed in Rochet (2003).

welfare, retailers' profits and retail prices, yet simple enough to assess their regulation.

It first identifies four key sources of potential social biases in the payment card associations' determination of interchange fees: internalization by merchants of a fraction of cardholder surplus, issuers' per-transaction markup, merchant heterogeneity, and extent of cardholder multi-homing. It compares the industry and social optima both in the short term (fixed number of issuers) and the long term (in which issuer offerings and entry respond to profitability).

Second, the paper tries to give some operational content to the notion of "must-take card". It introduces the "tourist test" (would the merchant want to refuse a card payment when a non-repeat customer with enough cash in her pocket is about to pay at the cash register?) and analyzes its relevance as an indicator of excessive interchange fees.

The paper is organized as follows: It first models the retail sector and assesses the impact of the pricing of payment cards services on card acceptance decisions by merchants, card usage decisions by consumers and the level of retail prices (Section 2). It then looks at the impact of interchange fees on the pricing of payment cards services, first in the case of a monopoly platform (Section 3), then when several platforms compete (Section 4). The following three sections then develop the core contributions of the paper. Section 5 revisits consumer surplus when issuer entry and offerings respond to industry profitability. Section 6 discusses the tourist test and derives policy implications under merchant homogeneity. Section 7 shows that retailer heterogeneity makes the tourist test likely to produce false positives. Finally, Section 8 summarizes the key insights.

2 Modeling the impact of payment cards on the retail sector: A benchmark case

2.1 The model

The benchmark model has a single card payment system. For the moment, the pricing of payment card services is exogenous: every time a transaction between a consumer (buyer) and a retailer (seller) is settled by card, the buyer pays[3] transaction fee p_B and the seller pays merchant discount p_S. There are no annual fees and all consumers have a card.

There is a continuum of consumers (of total mass normalized to one) with quasi-linear preferences. They spend their income on a composite good or "cash good" taken as a numeraire and on one unit of a "card good" sold by R retailers (being a "card good" means that consumers can pay by card as long as merchants accept it. "Cash goods" include leisure/work). The utility from purchasing the card good can differ across consumers, but is large enough, so that the aggregate demand for the card good is constant and equal to one.[4] To capture the intensity of competition in the (card good) retail sector, we use the Lerner-Salop model of product differentiation: Retailers and consumers are located uniformly on a circle of length normalized to one. The timing is as follows:

- First, each consumer learns his preference across brands of the card good offered by the retailers, as well as the prices chosen by the retailers. Furthermore, he learns all stores' card acceptance policies with probability α, and does not learn any with probability $1-\alpha$. The consumer then chooses which store to patronize. The optimal choice minimizes the sum of three terms: the retail price p_R^j, the transportation cost $t\Delta^j$ incurred for going to the store (where Δ^j is the distance to the store and $t > 0$ is a given parameter), and the expected transaction cost associated with the payment mode (this term is detailed below).

- Second, after choosing a store, the consumer learns his convenience benefit of using a card rather than cash in the particular instance,[5] and chooses the payment mode

[3] We allow p_B to be negative, in which case the cardholder receives a payment from his bank, in the form of cash back bonuses or air miles awarded to the buyer every time he uses his card.

[4] The analysis of the variant where the demand for the card good is elastic is more complex but gives similar results. It is available from the authors upon request.

[5] This assumption, borrowed from Wright (2004), simplifies the analysis of merchants' card acceptance decision. In Rochet and Tirole (2002) by contrast, we assumed that $\tilde{b}_B$ was drawn ex-ante. In this case, merchants' acceptance decisions become complementary, multiplicity of equilibria may arise and the timing of merchants' decisions matters.

among the ones accepted by the retailer. For simplicity, we restrict the analysis to two payment modes: card (if it is accepted by the retailer) and an alternative payment mode (cash or check). The relative cost $\widetilde{b}_B$ of this alternative payment mode for the consumer (also equal to his convenience benefit for a card payment) is random, and drawn from another continuous distribution with c.d.f. H:

$$H(b_B) = \Pr\left(\widetilde{b}_B \leq b_B\right). \tag{1}$$

We adopt the convention that $\widetilde{b}_B$ is the convenience cost of a cash/check payment and 0 is that for a card payment.

As we noted, this convenience benefit $\widetilde{b}_B$ is observed by the consumer only once he is in the store. The net benefit of paying by card is thus equal to the difference $\widetilde{b}_B - p_B$. A card payment is optimal for the consumer whenever this net benefit is positive. The proportion of card payments is denoted $D_B(p_B)$:

$$D_B(p_B) = \Pr\left(\widetilde{b}_B > p_B\right) = 1 - H(p_B). \tag{2}$$

Retailers $j = 1, \cdots, R$ compete in two stages:[6]

- First, they simultaneously decide whether to accept the card. We denote the decision of retailer j by a variable x^j equal to one if retailer j accepts the card, and zero if he does not.

- Second, they simultaneously set their retail prices: p_R^j is chosen by retailer j so as to maximize his profit:

$$\pi^j = \left[p_R^j - \gamma - b_S - x^j (p_S - b_S) D_B(p_B)\right] y^j, \tag{3}$$

where γ is the cost of producing the card good, and b_S is the cost of the alternative payment mode for the seller (assumed for the moment to be constant across sellers).

[6]The timing here is irrelevant: the equilibrium would be the same if the first and second stages were simultaneous. This is because we assume that consumers' transactional benefits are drawn ex post.

We adopt a convention similar to that for cardholders: b_S is the retailer's cost of a cash/card payment, while that for a card payment is normalized at 0. Thus $(p_S - b_S) D_B (p_B)$ represents the expected net cost of card payments for the seller (incurred only when $x^j = 1$). Finally, y^j represents the market share of retailer j.

2.2 Card acceptance policies and consumer welfare

Retailer j's market share is a function of the retail prices set by retailer j and his neighbors $j-1$ and $j+1$, as well as the card acceptance decisions x^j, x^{j-1} and x^{j+1}:

$$y^j = \frac{1}{R} + \frac{1}{2t}\left[p_R^{j-1} + p_R^{j+1} - 2p_R^j - \alpha\left(x^{j-1} + x^{j+1} - 2x^j\right) s_B (p_B)\right] \tag{4}$$

where

$$s_B (p_B) \equiv \int_{p_B}^{\infty} (b_B - p_B)\, dH (b_B)$$

denotes the expected surplus that a buyer derives from the option of paying by card, and $\alpha \in [0,1]$ represents the proportion of consumers who are informed about the card acceptance decisions of retailers. Our first proposition relates the retailers' acceptance decision to prices p_B and p_S and to the internalization parameter α.[7]

Proposition 1. *The retail sector equilibrium is unique and symmetric.*

- *retailers (all) accept the card if and only if the "weighted total user surplus" is positive:*

$$\phi_\alpha \equiv (b_S - p_S) D_B (p_B) + \alpha \int_{p_B}^{\infty} (b_B - p_B)\, dH (b_B) \geq 0 \tag{5}$$

(otherwise none accepts the card). If this condition is satisfied, then:

- *retailer pass through card transaction costs (or benefits) into the retail price:*

$$p_R^* = \gamma + b_S + \frac{t}{R} - (b_S - p_S) D_B (p_B) \tag{6}$$

[7] Proposition 1 can easily be extended to the case where two card networks offer identical cards and set identical IFs.

(where $\gamma + b_S$ is the marginal cost of a cash transaction and $\frac{t}{R}$ the Hotelling mark-up), and the total profit of the retail sector is constant:

$$\pi = \frac{t}{R}, \tag{7}$$

- *consumers' total purchase cost is given by:*

$$\left[\gamma + E\left(\widetilde{b}_B\right) + b_S\right] + \frac{5t}{4R} - \phi_1, \tag{8}$$

where ϕ_1 is obtained by taking $\alpha = 1$ in formula (5).

Proof of Proposition 1

Suppose, first, that all retailers accept the card. In this case, formulas (3) and (4) show that the profit of retailer j is maximized when

$$0 = \frac{\partial \pi^j}{\partial p_R^j} = y^j - \left[p_R^j - \gamma - b_S - (p_S - b_S)\, D_B(p_B)\right]/t.$$

The equilibrium market shares are all equal $\left(y^j \equiv \frac{1}{R}\right)$, and so are retail prices:

$$p_R^j \equiv p_R^* = \left[\gamma + b_S + \frac{t}{R}\right] - (b_S - p_S)\, D_B(p_B)$$

which establishes formula (6). Consumer total purchase cost is then equal to the sum of the retail price p_R^*, the average transportation cost $\frac{t}{4R}$ and the expected transaction cost for the cardholder $E\left(\widetilde{b}_B\right) + \int_{p_B}^{\infty} (p_B - b_B)\, dH(b_B)$. Formulas (8) and (7) are then immediate.

Suppose now that retailer j considers rejecting the card. A new price equilibrium arises where all retailers except j increase their price and market share, whereas retailer j decreases his price and market share but also his cost (assuming $b_S < p_S$). It is easy to check that the net effect is to decrease retailer j's profit if and only if

$$(b_S - p_S)\, D_B(p_B) + \alpha \int_{p_B}^{\infty} (b_B - p_B)\, dH(b_B) \geq 0.$$

Intuitively, with a linear demand (stemming from the uniform distribution of consumers), retailer j can lower his price by $\alpha \int_{p_B}^{\infty}(b_B - p_B)dH(b_B)$ and keep the same market share as when he accepts the card. The first term in the latter inequality is (minus) the cost saving associated with rejecting the card. The condition is thus equivalent to the assertion that accepting the card maximizes the perceived or weighted total user surplus, where only a fraction α of the buyer surplus from using the card is internalized. This condition ends the proof of formula (5) and, thus, of Proposition 1. The equilibrium is unique. ■

2.3 Discussion

Proposition 1 has a certain number of interesting implications:

• First, formula (5) shows that in general (that is, if $\alpha > 0$), retailers are willing to accept cards even if they lose money on card transactions (i.e., $p_S > b_S$). They are willing to incur a cost $p_S - b_S$ (providing it is not too large) on each card transaction, in order to offer a better quality of service to their customers (who value the option of paying by card). The intensity of this phenomenon is proportional to the probability α that card acceptance makes their store more attractive to the consumer.

This internalization of consumer surplus is unrelated to competition among retailers. Indeed, the same formula (5) would apply to a retail monopolist and has much broader generality than the Lerner-Salop model of Hotelling competition would lead us to believe. Card acceptance increases both the retailers' cost (if $p_S > b_S$) and quality of service to the consumer. Provided that consumers attach the same value (as they do here) to the increase in the quality of service, regardless of their willingness to pay for the good sold by the retailer, the retailer's card acceptance decision depends only on the sum of the merchant net convenience benefit and the quality increase brought about by card acceptance, weighted by the proportion α of informed consumers.

• Second, equilibrium retail prices reflect the expected cost of card transactions for merchants. In particular, formula (6) shows that if the merchant discount p_S increases, retailers pass through this increase into retail prices.[8]

• However, retail prices are not a good measure of consumer surplus since they don't take transaction costs into account. Formula (8) shows that the relevant index of social efficiency of the payment card network is

$$\begin{aligned}\phi_1 &\equiv (b_S - p_S)\, D_B(p_B) + \int_{p_B}^{\infty} (b_B - p_B)\, dH(b_B) \\ &\equiv \int_{p_B}^{\infty} (b_B + b_S - p_B - p_S)\, dH(b_B).\end{aligned}$$

ϕ_1, which we call *total user surplus*, represents the expectation of the total surplus (total benefit b_B+b_S minus total price p_B+p_S) derived from card payments by the two categories of users.

Finally, let us discuss our choice of the Lerner-Salop model for the description of retail demand. This model's linear demand allows a convenient aggregation of demands by consumers who are informed and uninformed about card acceptance policies. The retailers' card acceptance policy rule ($\phi_\alpha \geq 0$) holds for arbitrary demand functions when $\alpha = 0$ or 1. To see this, introduce the *perceived* and *real* hedonic prices:

$$\widehat{p}^j \equiv p^j - x^j \alpha s_B(p_B)$$

and

$$\widetilde{p}^j \equiv p^j - x^j s_B(p_B).$$

For arbitrary demand functions and $\alpha \in \{0, 1\}$, retailer j's profit is:

$$\left[\widehat{p}^j - \left[\gamma + b_S - x^j \phi_\alpha\right]\right] y^j \left(\widehat{p}^j, \widehat{p}^{-j}\right).$$

[8]Merchants may still want to sue payment card associations for high IFs for a variety of reasons. First, under current law, they are entitled to recoup the damages even if the latter are ultimately borne by cash users. Second, and as we will see in Section 7, merchants in general are heterogeneous and those with the lowest demand for card services may object to a policy that is tailored to the average merchant.

Thus when $\alpha \in \{0,1\}$, retailer j accepts the card iff $\phi_\alpha \geq 0$, independently of the shape of the demand function y^j. That $\phi_\alpha \geq 0$ is also the exact criterion for acceptance when $0 < \alpha < 1$ by contrast hinges on the linearity of demand.

We now model the payment card industry and investigate the impact of interchange fees on prices p_B and p_S, and ultimately on consumer surplus.

2.4 The tourist test

Retailers often complain that they are "forced" to accept card transactions that increase their net costs. To understand this "must-take card" argument, one must distinguish between ex post and ex ante considerations. Once the customer has decided to buy from the retailer, it is in the latter's interest to "steer" the former to pay by cash or check instead of by card whenever $p_S > b_S$ or equivalently $p_B < c - b_S + m(p_B)$. But from an ex ante point of view, the retailer must also take into account the increase in store attractiveness brought about by the option of paying by card. Because retailers can always ex ante turn down cards, the "must-take card" argument refers to the ex post perspective.

Let us accordingly introduce the "tourist test": suppose the buyer in question is a tourist, who will never patronize the store again in the future. The buyer shows up at the cash register with ostensibly enough cash to pay the wares. It is then in the interest of the seller to reject the card if and only if

$$a > a^T \equiv b_S - c_S \quad \Leftrightarrow \quad p_B < p_B^T = c - b_S + m\left(p_B^T\right). \tag{9}$$

3 Impact of the interchange fee on user surplus

Recall that, in a payment card association, the interchange fee (IF) a represents the amount paid[9] by the seller's bank (the acquirer) to the buyer's bank (the issuer) for each

[9]Nothing prevents, both in our model and in reality, a to be negative. In that case the IF flows from the issuer to the acquirer.

card transaction. It reallocates the total cost[10] $c = c_B + c_S$ of processing the transaction between the two banks. The acquirers' net marginal cost becomes $c_S + a$ and the issuers' becomes $c_I \equiv c_B - a$. We simplify the analysis by assuming that acquirers are perfectly competitive:

$$p_S = c_S + a. \tag{10}$$

By contrast, issuers may have market power. We model competition between them in reduced form, denoting by $p_B(c_I)$ and $\pi_B(c_I)$ the issuers' price and profit as functions of c_I. We assume that p_B increases and that π_B decreases with c_I.[11] Issuers' margin m is a function of p_B defined implicitly by:[12]

$$p_B - c_I + a = m\,(p_B)\,. \tag{11}$$

For convenience, we take p_B (instead of a) as the variable of interest. By assumption, p_B is increasing in $c_I = c_B - a$, which implies that p_B is decreasing in a. We can thus reason on p_B, keeping in mind that an increase in the IF results in a decrease in p_B.

The total profit of the members of the association (that is of the issuers, since acquirers make no profit in our model) is thus:

$$\pi_B = m\,(p_B)\, D_B\,(p_B)\,. \tag{12}$$

By assumption, π_B is decreasing in p_B.

[10] As in the rest of the paper, indices B refer to the buyer side, and indices S refer to the seller side. Thus c_B represents the marginal cost of the issuer (the buyer's bank) and c_S that of the acquirer (the seller's bank).

[11] Revealed preference implies that these conditions are always satisfied for a monopoly issuer.

[12] This is more general than Rochet and Tirole (2003) where we assumed $p_B = f\,(c_I)$ with $0 \leq f' < 1$. $m(\cdot)$ is derived from $f(\cdot)$ by a simple change of variable: $m[f(c_I)] = f(c_I) - c_I$. Here we maintain the assumption that $f' \geq 0$ (p_B increases with c_I), but we do not require $f' < 1$. The assumption that $f' < 1$ implies that $m' < 0$ (margins are decreasing). We call this case the "cost absorption" case. The case $f' = 1$ corresponds to that of a constant margin. We also consider here the case where $f' > 1$ (e.g. Cournot oligopoly with isoelastic demand) which we call the "cost amplification" case. In this case m increases with p_B. Note however that $m' = \dfrac{f' - 1}{f'} < 1$.

3.1 Total user surplus

When issuers are perfectly competitive ($m(p_B) \equiv 0$), total user surplus

$$\phi_1(p_B) = \int_{p_B}^{\infty} (b_B + b_S - c - m(p_B))\, dH(b_B)$$

is a single-peaked function of p_B, reaching its maximum at

$$p_B^0 \equiv c - b_S. \tag{13}$$

With perfectly competitive issuers, the price p_B^0 perfectly internalizes the externality associated with the decision of paying by card, which is made by the consumer. Indeed, the social cost of such a decision is not just c_B (the marginal cost of the buyer's bank) as it incorporates $c_S - b_S$, the externality exerted on the seller's side. The associated interchange fee corresponds to the threshold given by the "tourist test" defined in Section 2:

$$a^T = b_S - c_S.$$

It was first put forward by Baxter (1983). The corresponding merchant discount ($p_S = c_S + a^T = b_S$) makes the retailer ex-post indifferent about the choice of the payment instrument by the buyer (Farrell, 2006).

When issuers have market power ($m > 0$), p_B^0 is still the value of p_B that maximizes social welfare. Indeed, retailers' profit is constant, and social welfare is equal (up to a constant) to the sum of total user surplus and banks' profit:

$$W \equiv \phi_1 + \pi_B = \int_{p_B}^{\infty} (b_B + b_S - c)\, dH(b_B).$$

W clearly is maximal when $p_B = c - b_S = p_B^0$. The corresponding value of the interchange fee is then:

$$a^0 = c_B - p_B^0 + m(p_B^0) = b_S - c_S + m(p_B^0) > a^T.$$

By contrast, total user surplus ϕ_1 is maximized for a larger value p_B^{TUS}, obtained by equating to zero the derivative of ϕ_1:

$$\phi_1'(p_B) = [p_B + b_S - c - m(p_B)] D_B' - m' D_B = 0.$$

Thus

$$p_B^{TUS} = c - b_S + m\left(p_B^{TUS}\right) + \frac{m' D_B}{D_B'}\left(p_B^{TUS}\right). \tag{14}$$

p_B^{TUS} exceeds p_B^0 because user surplus does not include the issuers' profit. Since issuers' expected profit $m(p_B)D_B(p_B)$ decreases with p_B, a higher p_B (and thus a lower interchange fee) implies a lower expected profit for issuers and thus, around the social welfare optimum, a higher expected total user surplus.

The corresponding interchange fee a^{TUS} is given by:

$$a^{TUS} = c_B - p_B^{TUS} + m\left(p_B^{TUS}\right) = b_S - c_S - \frac{m' D_B}{D_B'}\left(p_B^{TUS}\right). \tag{15}$$

Proposition 2. *When issuers have market power ($m > 0$):*

i) The interchange fee a^0 that maximizes social welfare is higher than Baxter's interchange fee a^T, because it offsets issuers' margin:

$$a^0 = a^T + m(p_B^0).$$

ii) The interchange fee a^{TUS} that maximizes total user surplus is lower than a^0. It is higher than a^T in the cost amplification case ($m' > 0$) and lower than a^T in the cost absorption case ($m' < 0$).

The behavior of functions W and ϕ_1 is represented in the following diagram:

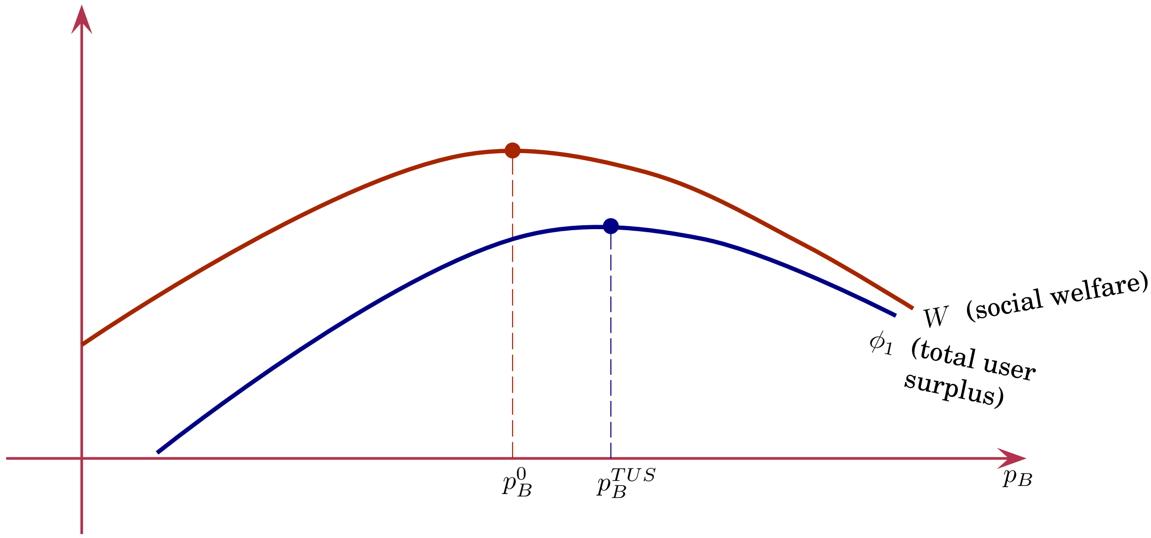

Figure 1: *Total user surplus ϕ_1, and social welfare SW. The vertical difference between these functions represents the expected profit of issuers. It decreases with p_B. This explains why p_B^{TUS}, which maximizes ϕ_1, is to the right of p_B^0, which maximizes W.*

We now examine the privately optimal interchange fee, i.e., the one set by the association in the absence of regulation.

3.2 The privately optimal interchange fee

Because issuers' profit $m(p_B)D_B(p_B)$ decreases with p_B, and in the absence of competition with another network, the card association sets the IF at the maximum value that retailers accept.

Thanks to Proposition 1, we can characterize retailers' acceptance decisions by looking at the behavior of function ϕ_α:

$$\phi_\alpha(p_B) \equiv (b_S - p_S) D_B(p_B) + \alpha \int_{p_B}^{\infty} (b_B - p_B)\, dH(b_B),$$

Noting that $\partial^2\phi_\alpha(p_B)/\partial\alpha\partial p_B < 0$, merchant resistance to an increase in the interchange fee is smaller when merchants' internalization coefficient α increases. The price p_B^m chosen by the monopoly association is given implicitly by:

$$\phi_\alpha(p_B^m) = 0.$$

Because $p_S = c - p_B + m(p_B)$ from formulas (10) and (11), we can rewrite ϕ_α as:

$$\phi_\alpha(p_B) = [b_S - c + p_B - m(p_B) + \alpha v_B(p_B)]\, D_B(p_B), \tag{16}$$

where $v_B(p_B)$ denotes the average net cardholder benefit per card payment:

$$\begin{aligned} v_B(p_B) &\equiv E[b_B - p_B | b_B \geq p_B] \\ &= \frac{\int_{p_B}^{\infty} (b_B - p_B)\, dH(b_B)}{\int_{p_B}^{\infty} dH(b_B)} > 0. \end{aligned}$$

Using formula (16), the association's optimal buyer price can be rewritten as:

$$p_B^m = c - b_S + m(p_B^m) - \alpha v_B(p_B^m). \tag{17}$$

Comparing (17) with formula (14) we see that p_B^m may be bigger or smaller than p_B^0, depending on issuer market power and on the value of α. The interchange fee chosen by the association is thus:

$$a^m = c_B - p_B^m + m(p_B^m) = b_S - c_S + \alpha v_B(p_B^m). \tag{18}$$

Proposition 3. *i) A monopoly association selects the maximum interchange fee a^m that is accepted by retailers.*

ii) When $m(p_B^0) < \alpha v_B(p_B^0)$ (a condition that is more likely to be satisfied when issuers' margin is small, merchant internalization is large and the net average cardholder benefit is large), a^m is larger than the socially optimal IF.

iii) When $m(p_B^0) \geq \alpha v_B(p_B^0)$, the interchange fee a^m chosen by the association coincides with the (second best) socially optimal IF.

Proof of Proposition 3

Part i) has already been noted.

To establish parts ii) and iii), let us recall that social welfare is equal (up to a constant) to the sum of total user surplus and banks' profit:

$$W = \phi_1 + mD_B = \int_{p_B}^{\infty} (b_B + b_S - c)\, dH(b_B),$$

which is maximum for $p_B = p_B^0$. Since W is quasi-concave in p_B, the socially optimal buyer price is equal to p_B^0 when this is compatible with merchant acceptance, i.e., when $\phi_\alpha(p_B^0) > 0$, and to p_B^m otherwise. Now

$$\phi_\alpha\left(p_B^0\right) = \left[-m\left(p_B^0\right) + \alpha v_B\left(p_B^0\right)\right] D_B\left(p_B^0\right),$$

which establishes ii) and iii). ■

Proposition 3 extends an earlier result of Rochet and Tirole (2002) to the case of an arbitrary internalization coefficient α. It shows that when there is a single association, when acquiring is perfectly competitive and when there is no unobservable heterogeneity among retailers, the association sets the highest possible IF a^m that retailers accept. a^m is always larger than the level a^{TUS} that maximizes total user surplus (and thus consumer surplus). However it is not necessarily larger than the socially optimal IF. If issuers' margin is large, or if retailers' acceptance of cards has a limited impact on their competitive position (for example if α is close to zero and/or the average benefit v_B of cardholders per card payment is small) the interchange fee that maximizes social welfare is too large to be acceptable by retailers. The (second best) socially optimal IF then coincides with the privately optimal one.

4 The impact of platform competition

We now extend our analysis to the competition between two card associations (indexed by $k = 1, 2$). For simplicity we assume that the two cards are perfect substitutes for both buyers and sellers: However the two associations may set different IFs a_1 and a_2, in which case user prices (denoted p_B^k and p_S^k, $k = 1, 2$) also differ. To fix ideas we assume (without loss of generality) that $a_1 \leq a_2$ so that $p_B^1 \geq p_B^2$. With perfectly competitive acquirers, we have also:

$$p_S^1 = c_S + a_1 \leq p_S^2 = c_S + a_2. \tag{19}$$

Since cards are perfect substitutes and card 2 is more expensive, retailers would be inclined to accept only card 1 but, like in Section 2, they must take into account the impact of their acceptance decisions on consumers' patronage. The retailers' acceptance decision is analyzed in the next subsection.

Note that when retailers accept both cards, then a consumer who holds both cards (we call such a consumer a multi-homer) only uses card 2 (since $p_B^2 \leq p_B^1$). Also, it is a dominated strategy for the associations to choose p_B^k in the decreasing part of ϕ_α. Thus we can assume without loss of generality that $\phi_\alpha(p_B^1) \geq \phi_\alpha(p_B^2)$. In turn, this implies that it is a dominated strategy for merchants to accept only card 2. If they only accept only one card, it will be card 1.

A complete analysis of platform competition lies outside the scope of this paper. We content ourselves with the analysis of two polar cases: Subsection 4.1 looks at the case of complete multi-homing, and Subsection 4.2 studies complete single-homing. Appendix 1 analyzes the retailers' acceptance decisions under partial multi-homing.

4.1 Complete multi-homing

We stick to the convention that $a_1 \leq a_2$ and therefore $p_B^1 \geq p_B^2$. Now, if all consumers have both cards, retailers accept both cards if and only if

$$\phi_\alpha\left(p_B^2\right) \geq \max\left(0, \phi_\alpha\left(p_B^1\right)\right).$$

Since issuers' profit in network 2 is a decreasing function of p_B^2, network 2 wants to choose p_B^2 as small as possible, but it is constrained by the condition $\phi_\alpha(p_B^2) \geq \phi_\alpha(p_B^1)$. By symmetry the competition between networks results in (equal) prices, set to maximize

$\phi_\alpha(p_B)$.[13] The consequences on merchant discounts and retail prices are immediate and are summarized in the next proposition.

Proposition 4. *In the case of complete multi-homing (all consumers have the two cards), both associations set the same interchange fee:*

$$a^{MH} = b_S - c_S - \frac{D_B}{D'_B}(p_B^{MH})\left[m'(p_B^{MH}) - (1-\alpha)\right]$$

where $p_B^{MH} \equiv \arg\max \phi_\alpha(p_B)$. *The merchant discount is then given by*

$$p_S^{MH} = c - p_B^{MH} + m(p_B^{MH}).$$

Proposition 4 implies that in the case of complete multi-homing, the price p_B^{MH} paid by cardholders exceeds the value p_B^{TUS} that maximizes consumer surplus. Equivalently the interchange fee set by competing networks is too low with respect to the value that maximizes consumer surplus.

We now examine the polar case of complete single-homing.

4.2 Complete single-homing

If all consumers have a single card, card i is accepted if and only if $\phi_\alpha(p_B^i) \geq 0$. This implies that, like in the case where there is a single network, card associations select the highest IF a^{SH} that retailers accept. The outcome of network competition is the same as the monopoly outcome characterized in Proposition 3:

- The price p_B^{SH} paid by cardholders is characterized implicitly by $\phi_\alpha\left(p_B^{SH}\right) = 0$, which gives:

$$p_B^{SH} = c - b_S + m\left(p_B^{SH}\right) - \alpha v_B\left(p_B^{SH}\right) = p_B^m.$$

- As illustrated by Figure 2, this price is lower than the value p_B^{TUS} that maximizes consumer surplus (or equivalently total user surplus) when issuers' margin decreases

[13] This is the two-sided version of Bertrand's undercutting argument.

with p_B (cost absorption case) but the reverse may hold when issuers' margin increases with p_B (cost amplification case).

Proposition 5. *When there is complete single-homing, the interchange fee a^{SH} set by competing networks is equal to the monopoly interchange fee a^m, and higher than the value a^{TUS} that maximizes consumer surplus.*

However a^{SH} may be smaller or larger than the socially optimal IF.

The following figure synthesizes our results:

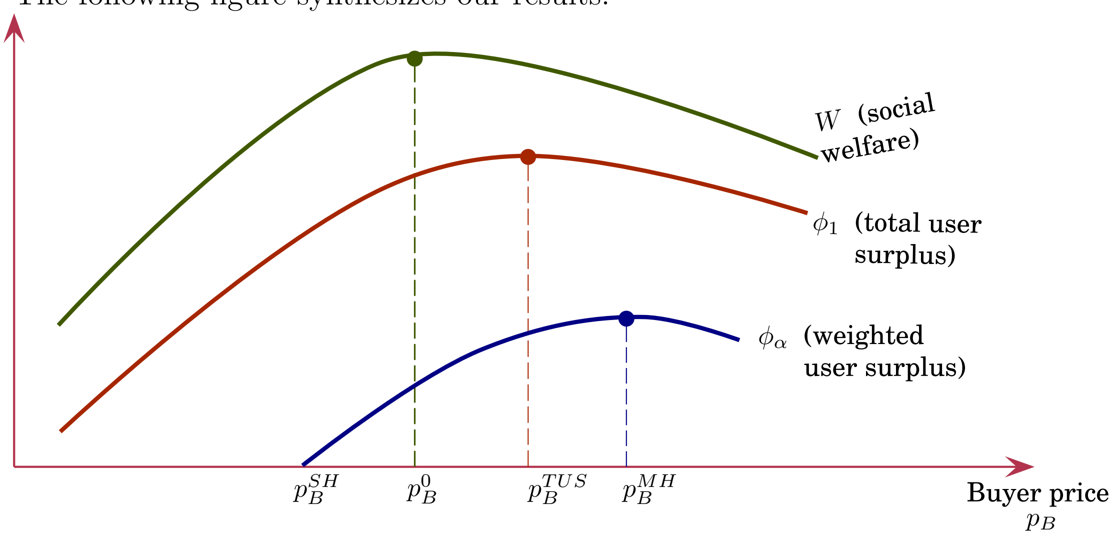

Figure 2: *Comparison of buyer prices (in the cost absorption case) when two card associations compete:*

- *when consumers single-home, the outcome is p_B^{SH} (i.e., the same as in the case of a single network),*
- *when consumers multi-home, the outcome is p_B^{MH}. It is larger than the price p_B^{TUS} that maximizes total user surplus in the cost absorption case (as represented here) but may be lower in the cost amplification case.*

5 Entry: revisiting the notion of total user surplus

The previous section analyzed the interchange fees associated with two benchmarks, corresponding to the maximization of social welfare and to that of consumer surplus (TUS). Focusing on the narrow notion of consumer surplus is legitimate for a short-term analysis as long as the welfare of shareholders is weighted much less heavily than that of consumers. In the medium and long term, though, issuers respond to increased profitability by offering a wider variety of products or by reducing prices.

To illustrate the impact of entry, this section computes the "long-term total user surplus" first in the context of an homogenous issuing industry in which issuers do not compete perfectly (so entry reduces price but does not increase variety), then in a context of monopolistic competition between differentiated issuers.

5.1 Homogenous issuing industry

Let $c_I \equiv c_B - a = c - p_S$ denote the issuers' marginal cost, N the number of issuers, and $P(Q)$ the inverse demand function. The fixed cost of being in the issuing industry is $F > 0$.

Adapting our previous notation to account for the number N of issuers, we denote by $p_B = p_B(c_I, N)$ and $m = m(p_B, N)$ the equilibrium price and margin for a fixed number N of issuers. We assume that $\frac{\partial m}{\partial N} < 0$ (more issuers imply smaller margins). The number of issuers is now endogenous, and given by the unique solution $N = N(p_B)$ to the zero-profit equation:

$$m(p_B, N)D_B(p_B) = NF. \tag{20}$$

Note that, as a consequence of our assumption that issuers' profit $m(p_B, N)D(p_B)$ decreases with p_B, the number of issuers $N(p_B)$ also decreases with p_B.

The long-term total user surplus (also equal to social welfare since issuers make no supra-normal profit) is equal to the sum of cardholder and merchant surpluses:

$$TUS^{LT} = \int_{p_B}^{\infty} \left[b_B + b_S - c - m(p_B, N(p_B))\right] dH(b_B).$$

The only difference with short-term total user surplus ϕ_1 (see formula (13)) is that the margin $m(p_B)$ depends on p_B also through the number N of issuers. TUS^{LT} is maximized for

$$p_B^* = c - b_S + m + \frac{D_B}{D_B'}\left[\frac{\partial m}{\partial p_B} + \frac{\partial m}{\partial N} N'\right]. \tag{21}$$

Comparing with p_B^{TUS}, that maximizes short-term total user surplus ϕ_1, we see that p_B^* contains an additional term $\frac{D_B}{D_B'}\frac{\partial m}{\partial N}N'$, corresponding to the impact of p_B on the number of issuers, and thus indirectly on the level of issuers' margin. This additional term is negative since D_B', $\frac{\partial m}{\partial N}$ and N' are all negative. Thus

$$p_B^* < p_B^{TUS}.$$

The intuition for this result is that a lower cardholder fee p_B leads to the entry of more issuers and therefore increases competition and, ultimately, long-term total user surplus. Note that the term between brackets in (21) is negative (this is because $m\left(p_B, N\left(p_B\right)\right) = \dfrac{N\left(p_B\right)F}{D_B\left(p_B\right)}$ is decreasing in p_B). Thus $p_B^* \leq c - b_S + m = p_B^T$.

The comparison between p_B^* and $p_B^0 = c - b_S$ (that maximizes short term welfare) is also easy. By differentiating the zero-profit condition that defines $N(p_B)$:

$$m(p_B, N(p_B))D(p_B) = N(p_B)F,$$

we get:

$$\left(\frac{\partial m}{\partial p_B} + \frac{\partial m}{\partial N}N'\right)D_B + mD_B' = N'F. \tag{22}$$

By using formula (21) defining p_B^*, we get:

$$p_B^* - p_B^0 = \frac{1}{D_B'}\left[mD_B' + D_B\left(\frac{\partial m}{\partial p_B} + \frac{\partial m}{\partial N}N'\right)\right].$$

where the right-hand side is computed at p_B^*. Thus

$$p_B^* - p_B^0 = \frac{N'}{D_B'}F > 0.$$

We can now state our results in terms of the associated interchange fees a^*, a^{TUS} and a^0.

Proposition 6. *For an homogenous issuing industry with free entry, the total user surplus maximizing interchange fee a^* lies in between a^{TUS}, that maximizes short term total-user-surplus, and a^0, the short-term first best:*

$$a^{TUS} < a^* < a^0.$$

We also have that $a^* \geq a^T$.

Cournot example: Consider Cournot competition with linear demand: $D_B(p_B) = 1 - p_B$. The industry is viable if the monopoly profit $\left(\frac{1-c_I}{2}\right)^2$ exceeds entry cost F. In this case the short-term equilibrium price and the short-term margin are:

$$p_B(c_I, N) = \frac{1+Nc_I}{1+N} \text{ and } m(p_B, N) = p_B - c_I = p_B - \frac{(1+N)p_B - 1}{N} = \frac{1-p_B}{N}. \quad (23)$$

Thus there is short-term cost absorption $\left(\frac{\partial m}{\partial p_B} < 0\right)$, implying, by formula (15):

$$a^{TUS} = a^T - \left(\frac{\frac{\partial m}{\partial p_B} \cdot D_B}{D'_B}\right)\left(p_B^{TUS}\right) < a^T.$$

By contrast, there is long-term cost passthrough. Indeed, the zero-profit condition,

$$m(p_B, N)(1 - p_B) = NF,$$

yields :

$$m(p_B, N(p_B)) = \sqrt{F}.$$

Thus long-term total user surplus is maximized for $a^* = a^T = b_S - c_S < a^0 = b_S - c_S + \sqrt{F}$.

5.2 Pure product variety

The homogenous-good case confers limited benefits on the entry mechanism: While entry benefits consumers though lower prices, the incentive to enter comes in large part from business stealing. Indeed, Mankiw and Whinston (1986) show that (with at least two firms) there are always too many firms in an homogenous-good, free-entry industry.

To illustrate the product-diversity argument, we build a stylized example which is by contrast probably biased towards high consumer benefits from entry since it does not embody any business stealing effect. Suppose a continuum of niche markets for cards indexed by the fixed cost of entry F. All markets are identical but for the fixed cost of entry. Let $K(F)$ denote the c.d.f. of i, with density $k(F)$.

Each market is contestable (is an "auction market"). In equilibrium of a contestable market, there is a single firm, and this firm makes no profit; the markup $m(c_I, F)$ is the smallest solution of:

$$mD_B(c_I + m) = F.$$

As $\frac{d}{dm}(mD_B(c_I + m)) > 0$ in the relevant range, the contestable market example exhibits long-term cost amplification $\left(\dfrac{\partial m}{\partial c_I} > 0\right)$.

Let $F^*(c_I)$, a decreasing function, be defined by:

$$\max_m\{mD_B(c_I + m)\} = F^*(c_I).$$

It corresponds to the maximum fixed cost that the card issuing industry can sustain. $K[F^*(c_I)]$ represents the mass of active issuers.

Then

$$\begin{aligned} TUS^{LT} &= \int_0^{F^*(c_I)} [v_B(c_I + m(c_I, F)) + (b_S - p_S)] D_B(c_I + m(c_I, F)) dK(F) \\ &= \int_0^{F^*(c_I)} \left[v_B(c_I + m(c_I, F)) + \left(c_I - c_I^T\right)\right] D_B(c_I + m(c_I, F)) dK(F), \end{aligned}$$

where $c_I^T \equiv c_B - a^T = c - b^S$.

Using

$$v_B D_B = \int_{c_I + m(c_I, F)}^{\infty} [b_B - [c_I + m(c_I, F)]] dH(b_B),$$

we see that

$$TUS^{LT} = \int_0^{F^*(c_I)} \int_{c_I + m(c_I, F)}^{\infty} \left[b_B - m(c_I, F) - c_I^T\right] dH(b_B) dK(F).$$

Then at $c_I = c_I^T$:

$$\frac{dTUS^{LT}}{dc_I} = \frac{dF^*}{dc_I} v_B D_B k(F^*) - \int_0^{F^*(c_I)} D_B \frac{\partial m}{\partial c_I} dK(F) < 0$$

(where $F^* \equiv F^*(c_I)$).

Proposition 7. *The total-user-surplus maximizing interchange fee in the pure-product-variety model) always exceeds the level given by the tourist test.*

6 Is the tourist test a good test?

The attraction of the tourist test resides in the fact that the merchant pays no more than his convenience benefit from card payments. Capping the merchant discount at the merchant's convenience benefit prevents card payment systems from exploiting the internalization effect to force merchants to accept card payment that they do not want and that might therefore be inefficient. Whether the cap implied by the tourist test is reasonable, though, depends on whether cardholders are provided with the proper social incentives: The social optimum is reached only when the cardholders make the efficient decision with regards to the choice of payment method. An efficiency analysis however hints at two shortcomings of the tourist test:

(i) *cardholders' incentives are already distorted:* As is usual, a "first-best rule" usually is no longer adequate when the rest of the economy is already distorted. Here, if merchants pay their convenience benefit, the cardholder pays more than the net social cost of the card transaction (equal to the total cost of card payments, $c_B + c_S$, minus the merchant's benefit, b_S) whenever issuers (or acquirers for that matter) levy markups about cost. This suggests that cardholders may underconsume card payments. As our analysis has shown, the welfare analysis is however complicated by the fact that markups are not constant, and so the interchange fee may also be used also to reduce the markup. In the rest of this section, we will sum up the insights obtained so far in this respect.

(ii) *merchants are heterogenous:* When merchants differ in their convenience benefits (b_S), inframarginal merchants derive more benefits from card payments than marginal ones. The tourist test can be applied to each merchant (or at least to merchants who end up accepting the card), but it is clear that efficiency cannot require that it be net even by all participating merchants; for, when internalize the welfare of the *average* merchant and not of the marginal one, who, recall, values card payments less than the average merchant. Capping merchant discounts at the convenience benefit of the most reluctant merchants provides the cardholder with an incentive for underconsumption of card payments. We will come back to this point with more detail in Section 7.

Interchange fee	a^{TUS}	a^{T}	a^{0}	
Tourist test	PASSES	PASSES	FAILS	FAILS
Social welfare	IF too low			IF too high
Ex post user surplus	IF too low	IF too high		

Cost absorption case

Interchange fee	a^{T}	a^{TUS}	a^{0}	
Tourist test	PASSES	FAILS	FAILS	FAILS
Social welfare	IF too low			IF too high
Ex post user surplus	IF too low		IF too high	

Cost amplification case

Table 3: Different thresholds for the interchange fee. (consumer surplus is total user surplus, and, unlike social welfare, does not account for issuers' markup).

Table 3 shows that the tourist test is not a good test for judging whether interchange fees are too high or too low. Indeed, from either point of view of social welfare or ex post consumer surplus, the optimal IF does not in general correspond to the threshold given by the tourist test.

In the case of perfect competition among issuers, the three thresholds coincide, as shown in Figure 3.

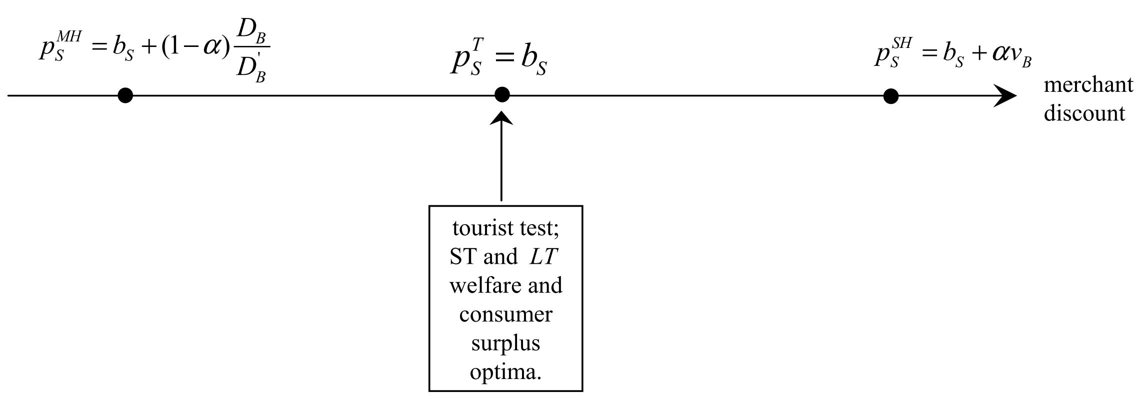

Figure 3: The perfect competition case.

However, when issuers have some market power (which seems to be the belief of Competition Authorities in many regions of the world) the tourist test threshold can be too high or too low, according to whether cost absorption ($m' < 0$) or cost amplification prevails. This is represented in Figure 4.

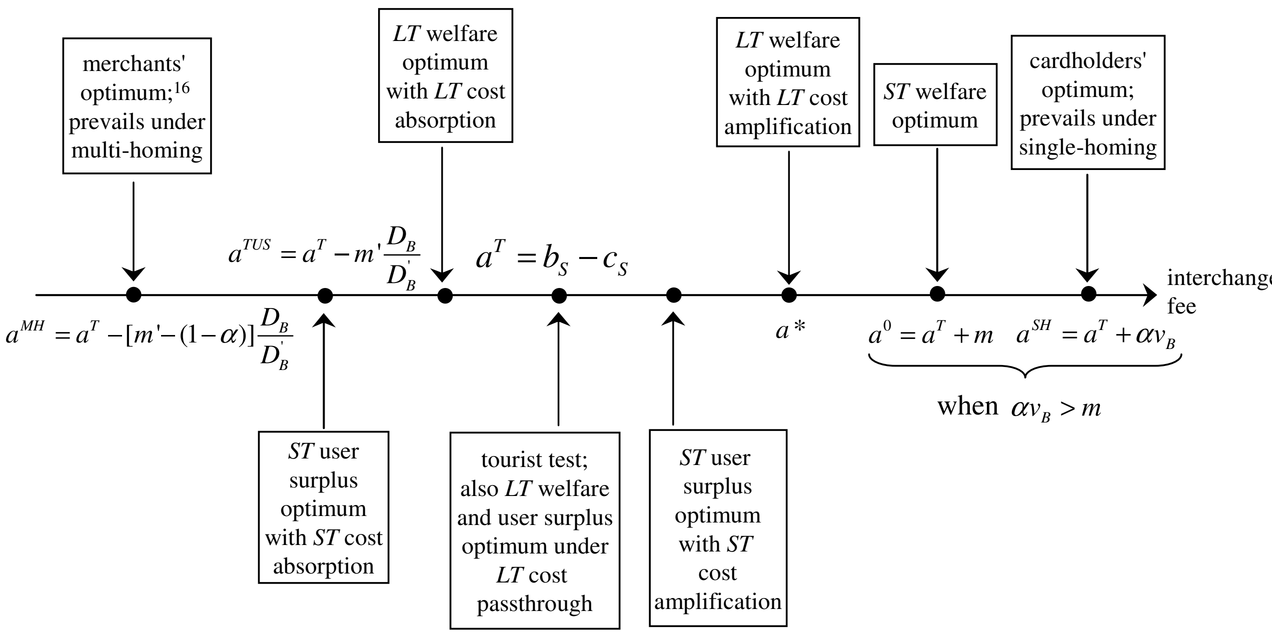

Figure 4: Optimal values of the interchange from the points of view of merchants, cardholders and social welfare.

7 Heterogeneous retailers

To bring our benchmark model more in line with reality, we need to introduce unobservable heterogeneity either on the internalization parameter α or on the convenience benefit[17] b_S of retailers for card payments.

[16] Strictly speaking, merchants are indifferent about the level of the interchange fee in our model. This is because in our simple Hotelling model with linear demand they pass through their cost one for one into retail prices. However as soon as their profit decreases, even slightly, when their cost increases,their preferred interchange fee is a^{MH}.

[17] Recall that this convenience benefit is equal to the convenience cost for retailers of using the alternative payment mode (say checks).

7.1 Heterogeneity in convenience benefit

The first type of heterogeneity among retailers has α constant and b_S heterogenous. Following Schmalensee (2002), we assume that sellers' convenience benefit is drawn from a continuous distribution with c.d.f. G:

$$G(b_S) = \Pr\left(\widetilde{b}_S \leq b_S\right).$$

Following Wright (2004), we assume that the retail sector consists in a continuum of Lerner-Salop circles or markets,[18] each corresponding to a value of $\widetilde{b}_S$. The buyers' distribution of convenience benefits, $H(b_B)$, is independent of the market in which they buy. Consumers buy one good in each of these markets and patronize the store that offers the best combination of retail price, transportation cost and quality of service (determined here by the retailer's decision of whether to accept cards). For given prices for card services (p_B and p_S), the equilibrium behavior of retailers is characterized by the same conditions as in Proposition 1 but now this behavior is conditional on the realization of b_S:

- A retailer in "sector" b_S accepts card payments if and only if:

$$b_S \geq \widehat{b}_S \equiv p_S - \alpha v_B(p_B), \tag{24}$$

where $v_B(p_B) \equiv E\left[\widetilde{b}_B - p_B | \widetilde{b}_B \geq p_B\right]$.

- Retail prices in "sector" b_S are given by:

$$p_R^*(b_S) = \left[\gamma + b_S + \frac{t}{R}\right] - (b_S - p_S) D_B(p_B) \quad \text{if} \quad b_S \geq \widehat{b}_S \quad \text{and}$$
$$p_R^*(b_S) = \left[\gamma + b_S + \frac{t}{R}\right] \quad \text{otherwise.}$$

[18] Wright (2004) builds a model of a payment card association with heterogeneous merchants. He shows how the privately and socially optimal IFs depend on the elasticities on the two sides (merchants, cardholders) and argues that there is no systematic bias between the IF chosen by the association and the socially optimal IF.

- Finally, consumers' total purchase cost is:

$$\left[\gamma + E\left(\widetilde{b}_B\right) + E\left(\widetilde{b}_S\right) + \frac{5t}{4R}\right] - \int_{\widehat{b}_S}^{\infty}\int_{p_B}^{\infty}\left(b_B + b_S - c\right) dH\left(b_B\right) dG\left(b_S\right). \tag{25}$$

The volume of card transactions is also easily computed:

$$V = D_B\left(p_B\right) D_S\left(\widehat{b}_S\right), \tag{26}$$

where $D_S\left(\widehat{b}_S\right) = 1 - G\left(\widehat{b}_S\right)$ represents the "demand" for card transactions by retailers.[19]

Social welfare is maximized when two symmetric "Samuelson conditions" are satisfied:

$$p_B^* = c - E\left[b_S | b_S \geq \widehat{b}_S\right],$$

and

$$\widehat{b}_S = c - E\left[b_B | b_B \geq p_B^*\right].$$

These conditions imply that

$$v_B(p_B^*) = v_S(\widehat{b}_S) \equiv E\left[b_S - \widehat{b}_S | b_S \geq \widehat{b}_S\right],$$

where $v_S(\widehat{b}_S)$ denotes the average retailer surplus per card payment.

When merchants fully internalize cardholder surplus ($\alpha = 1$ and so the second Samuelson condition is met) and $m = 0$ (perfect competition among issuers), this social optimum can be implemented by setting the interchange fee at the following level:

$$a^* = c_B - p_B^* = E\left[b_S | b_S \geq \widehat{b}_S\right] - c_S.$$

The corresponding merchant discount is

$$p_S = E\left[b_S | b_S \geq \widehat{b}_S\right].$$

In this case the *average* merchant (among those who accept cards) is ex post indifferent about the means of payment chosen by the consumer. This means that unless all retailers are identical, some of them would like to reject cards ex post.

[19] A similar multiplicative formula for the volume of card transactions (with $\alpha = 0$ and thus $\widehat{b}_S = p_S$) was first proposed by Schmalensee (2002) and later used in a more general context by Rochet and Tirole (2003).

Proposition 8. *Assume that merchants differ in the net convenience benefit that they derive from card transactions, that all consumers are informed about merchants' card acceptance, and that issuers' margin is nil (perfect competition). Then the social optimum can be implemented by selecting a merchant discount equal to the average net convenience benefit among the merchants who accept cards.*

We now can state the section's main result (whose proof can be found in the appendix). We assume for simplicity that the issuers' margin m is constant, so that $p_B + p_S$ is itself a constant, and equal to $c + m$.

Proposition 9. *Assume that merchants differ in the net convenience benefit b_S that they derive from card transactions and that issuers' margin is constant (cost passthrough). Provided that social welfare is quasi-concave in the buyer price, the privately optimal IF is lower than the socially optimal value if and only if merchant per card payment surplus exceeds cardholders':*

$$v_S\left(\widehat{b}_S^m\right) > v_B\left(p_B^m\right). \tag{27}$$

Proposition 9 shows that merchant heterogeneity has interesting consequences on the price structure chosen by a monopoly platform. To prove Proposition 10, one first demonstrates the following:

i) In the absence of IF regulation, a monopoly network chooses the price structure (p_B^m, p_S^m) that maximizes volume. It is characterized by:

$$\frac{D_B'}{D_B}\left(p_B^m\right) = \frac{D_S'}{D_S}\left(\widehat{b}_S^m\right)\left(1 + \alpha v_B'\left(p_B^m\right)\right) \tag{28}$$

ii) The socially optimal price structure $\left(p_B^W, p_S^W\right)$ is characterized:

$$\frac{D_B'}{D_B}\left(p_B^W\right)\left[1 + \frac{v_S\left(\widehat{b}_S^W\right) - v_B\left(p_B^W\right)}{m + (1-\alpha)v_B\left(p_B^W\right)}\right] = \frac{D_S'}{D_S}\left(\widehat{b}_S^W\right)\left(1 + \alpha v_B'\left(p_B^W\right)\right) \tag{29}$$

Since formulas (28) and (29) are complex, it is useful to consider first Baxter's case ($\alpha = 0$) in which merchants do not internalize cardholder surplus. In this case $\widehat{b}_S^m = p_S^m$ and formula (28) coincides with the formula characterizing the price structure of a monopoly association in Rochet and Tirole (2003). Prices for buyers and sellers are chosen in such a way that semi-elasticities of demand $\frac{D_B'}{D_B}$ and $\frac{D_S'}{D_S}$ are equal on both sides. When $\alpha > 0$ however, merchants care about the surplus of cardholders and the price structure also depends on the marginal impact of changes in p_B^m on the expected surplus of cardholders $v_B(p_B^m)$.

The buyer price p_B^{TUS} that maximizes total user surplus satisfies a formula very close to (29):

$$\frac{D_B'}{D_B}\left[1 + \frac{v_S - v_B}{(1-\alpha)v_B}\right] = \frac{D_S'}{D_S}\,(1 + \alpha v_B')\,. \tag{30}$$

Thus if total user surplus is quasi-concave in the buyer price, condition (27) (i.e., $v_S\left(\widehat{b}_S^m\right) > v_B\left(p_B^m\right)$) is also necessary and sufficient for the privately optimal IF to be lower than the value that maximizes total user surplus.

Condition (27) shows that the price structure chosen by a monopoly platform, in the absence of a regulation, is not systematically biased in favor of cardholders. When the average net benefit of retailers from card payments v_S is greater than the average net benefit of consumers v_B, the IF chosen by a monopoly platform is too low, from both viewpoints of social welfare and total user surplus.

7.2 Heterogeneity in internalization

Let us next investigate heterogeneity in α (b_S being the same for all merchants). Merchants accept cards if and only if their internalization parameter α exceeds a critical value $\alpha^*(p_B)$ defined implicitly by the relation:

$$b_S = p_S - \alpha^* v_B(p_B) = c + m(p_B) - p_B - \alpha^* v_B(p_B). \tag{31}$$

(We assume that there is cost absorption or cost passthrough ($m' \leq 0$). In this case, α^* is a decreasing function[20] of p_B. Thus increasing p_B gets more merchants on board. Social welfare has a simple expression:

$$W = \left[\int_{p_B}^{\infty}(b_B + b_S - c)dH(b_B)\right] \cdot \Pr(\alpha \geq \alpha^*(p_B)).$$

At the (first best) social optimum all merchants accept cards ($\alpha^* = \alpha_{\min}$, where $\alpha_{\min}$ is the minimum value of α among merchants) and:

$$p_B^0 = c - b_S.$$

This can be implemented by setting the interchange fee at

$$a = b_S - c_S + m = a^0,$$

provided this is compatible with universal merchant acceptance:

$$\alpha_{\min} \geq \frac{m}{v_B(p_B^0)}.$$

If this condition is not satisfied, let $\psi(p_B) = \Pr(\alpha \geq \alpha^*(p_B))$ denote the proportion of merchants who accept cards, as a function of the cardholder price p_B. The socially optimal level of p_B, which we denote p_B^0, satisfies:

$$\frac{W'}{W}(p_B) = \frac{(p_B + b_S - c)D_B'}{\int_{p_B}^{\infty}(b_B + b_S - c)dH(b_B)} + \frac{\psi'}{\psi}(p_B) = 0.$$

By contrast, a monopoly network selects a cardholder price p_B that maximizes the profit of issuers:

$$\Pi = m(p_B)D_B(p_B)\psi(p_B).$$

Therefore p_B^m satisfies:

$$\frac{\Pi'}{\Pi} = \frac{m'}{m}(p_B) + \frac{D_B'}{D_B}(p_B) + \frac{\psi'}{\psi}(p_B) = 0.$$

[20] Since $\alpha^* \leq 1$ and $v_B' \geq -1$, the assumption $m' \leq 0$ implies that the right-hand side of (31) decreases in p_B.

Under monotone comparative statics (that is, if Π is log concave) we can compare p_B^m and p_B^0. Indeed if $m' \leq 0$ then, given that $D_B' \leq 0$, and

$$\frac{p_B + b_S - c}{\int_{p_B}^{\infty}(b_B + b_S - c)dH(b_B)} < \frac{1}{D_B},$$

we have:

$$\frac{\Pi'}{\Pi}(p_B^m) = 0 = \frac{W'}{W}(p_B^0) \geq \frac{D_B'}{D_B}(p_B^0) + \frac{\psi'}{\psi}(p_B^0) \geq \frac{\Pi'}{\Pi}(p_B^0),$$

which implies, by log concavity of Π, that

$$p_B^0 \geq p_B^m.$$

Proposition 10. *When the internalization parameter α is heterogenous across merchants and there is cost absorption or pass through ($m' \leq 0$), the interchange fee selected by a monopoly association is (weakly) bigger than the socially optimal buyer price.*[21]

8 Summary

Merchants may accept the card even when their net benefit from doing so is negative. Accordingly, we introduced the "tourist test" (would the merchant want to refuse a card payment when a non-repeat customer with enough cash in her pocket is about to pay at the cash register?) and analyzed its relevance as an indicator of "excessive" interchange fees. The relevant welfare benchmarks for this analysis are (a) short-term total welfare, (b) short-term consumer welfare, and (c) consumer welfare when issuer entry and offerings respond to industry profitability.

In the absence of platform competition or under cardholder single-homing and merchant homogeneity, the interchange fee chosen by issuers exceeds the short-term socially optimal level if and only if the fraction of cardholder benefits internalized by merchants

[21] For analogous reasons, the buyer price p_B^{SH} that results from competition between identical platforms when cardholders single-home is greater than p_B^m. This contrasts with the case of homogenous merchants, where $p_B^{SH} = p_B^m$.

(which depends on how knowledgeable about merchants' acceptance policies cardholders are) exceeds the issuers' per transaction markup. Under platform competition and multi-homing, the IF is smaller than the value that maximizes consumer surplus (cardholders' plus merchants' surplus), and a fortiori the value that maximizes social welfare.

The paper's second contribution has been to assess whether the tourist test is a proper test for detecting excessive interchange fees. The attraction of the tourist test is that merchants are not forced into transactions that they would not wish individually. We unveiled a number of reasons why this test may yield false positives even if we focus on total user surplus (TUS), which does not account for issuers' markups. First, in the short run, the TUS-maximizing IF fails the tourist test if the issuing industry's prices exhibit cost amplification (conversely, cost absorption leads to false negatives). Second, in the long term, issuer markups translate into entry and thereby lower prices and increased variety, and so the short-term analysis yields TUS-maximizing IFs that are smaller than their long-term counterpart. Third, merchants are heterogeneous, and an interchange fee that properly guides cardholders' decisions must reflect the average, not the marginal merchant benefit. This implies that the merchants that benefit least from the card, say the large retailers, are likely to fail the tourist test at the social optimum.

References

[1] Balto, D. A.(2000) "The Problem of Interchange Fees: Costs without Benefits?" *European Competition Law Review*, 21(4): 215–224.

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

[3] Carlton, D. W., and A.S. Frankel (1995) "The Antitrust Economics of Credit Card Networks," *Antitrust Law Journal*, 63(2): 643–668.

[4] Chang, H. H., and D.S. Evans (2000) "The Competitive Effects of the Collective Setting of Interchange Fees by Payment Card Systems," *Antitrust Bulletin*, 45(3): 641–677.

[5] Evans, D. S., and R. Schmalensee (1995) "Economic Aspects of Payment Card Systems and Antitrust Policy toward Joint Ventures," *Antitrust Law Journal*, 63(3): 861–901.

[6] Farrell, J. (2006) "Efficiency and Competition between Payment Instruments", *Review of Network Economics*, 5(1): 26–44.

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

[8] Mankiw, G., and M. Whinston (1986) "Free Entry and Social Inefficiency," *Rand Journal of Economics*, 17(1): 48–58.

[9] Rochet, J.-C. (2003) "The Theory of Interchange Fees: A Synthesis of Recent Contributions," *the Review of Network Economics*, 2(2): 97–124.

[10] Rochet, J.-C., and J. Tirole (2002) "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, 33(4): 549–570.

[11] —— (2003) "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1(4): 990–1029.

[12] Schmalensee, R. (2002) "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50(2): 103–122.

[13] —— (2003) "Interchange Fees: A Review of the Literature," in *The Payment Card Economics Review*, Vol. 1, Cambridge: payingwithplastic.org/National Economic Research Associates, 2003: 25–44.

[14] Vickers, J. (2005) "Public Policy and the Invisible Price: Competition Law, Regulation and the Interchange Fee", Proceedings of a conference on "Interchange Fees in Credit and Debit Card Industries" (Federal Reserve Bank of Kansas-City, Santa Fe, New Mexico, May 4-6, 2005) 231–247.

[15] Wright, J. (2003a) "Optimal Card Payment Systems," *European Economic Review*, 47: 587–612.

[16] —— (2003b) "Pricing in Debit and Credit Card Schemes," *Economics Letters*, 80: 305–309.

[17] —— (2004) "The Determinants of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, 52(1): 1–26.

Appendix 1: Retailers' acceptance decisions under partial multi-homing

For simplicity, this appendix takes as given the proportions of buyers who own the two cards. Specifically, let β_k $(k = 1, 2)$ denote the proportion of buyers who own only card k (the single-homers) and β_{12} denote the proportion of buyers who own both cards (the multi-homers). Like before, we assume for simplicity that all consumers have a card[22] (i.e., $\beta_1 + \beta_2 + \beta_{12} = 1$) and also that only a proportion $\alpha \in [0, 1]$ of consumers are aware of retailers' card acceptance policy before they select which store to patronize. The next proposition characterizes the equilibrium of the retail sector (both in terms of card acceptance and retail prices) as a function of payment card prices and parameters α, β_1, β_2 and β_{12}.

Proposition 11. *At the equilibrium of the retail sector:*

- *retailers accept both cards if and only if:*

$$\beta_1 \phi_\alpha \left(p_B^1\right) + \left(\beta_2 + \beta_{12}\right) \phi_\alpha \left(p_B^2\right) \geq \max \left[0, \left(\beta_1 + \beta_{12}\right) \phi_\alpha \left(p_B^1\right), \left(\beta_2 + \beta_{12}\right) \phi_\alpha \left(p_B^2\right)\right] \tag{32}$$

If this condition is satisfied then:

- *retail prices*[23] *are given by:*

$$p_R^* = \left[\gamma + b_S + \frac{t}{R}\right] - \beta_1 \left(b_S - p_S^1\right) D_B \left(p_B^1\right) - \left(\beta_2 + \beta_{12}\right) \left(b_S - p_S^2\right) D_B \left(p_B^2\right),$$

- *aggregate demand for the card good equals*[24] $\beta_1 D\left(u_1^*\right) + \left(\beta_2 + \beta_{12}\right) D\left(u_2^*\right)$*, where*

$$u_k^* = p_R^* + \frac{t}{4R} + E(\widetilde{b}_B) - s_B\left(p_B^k\right), \qquad k = 1, 2.$$

[22] This is not inconsistent with our model, which assumes that issuers do not charge fixed fees to cardholders.

[23] We assume that retailers find it too costly or are not allowed to charge different prices for transactions settled through different cards.

[24] For simplicity, we assume that the determinants of the choice of cards by consumers are independent from the parameters that determine the gross utility $\tilde{u}$ obtained by consumers when they consume the card good.

- *consumer surplus equals* $\beta_1 S(u_1^*) + (\beta_2 + \beta_{12}) S(u_2^*)$,
- *finally, the total profit of the retail sector is*

$$\pi_R = \frac{t}{R}\left[\beta_1 D(u_1^*) + (\beta_2 + \beta_{12}) D(u_2^*)\right]$$

Proof of Proposition 11:

It proceeds similarly to that of Proposition 1. In equilibrium, retailers accept the (set of) cards that maximize the expectation of weighted user surplus ϕ_α over all buyers. Accepting only card $k(k = 1, 2)$ allows a fraction $(\beta_k + \beta_{12})$ of buyers to pay by card generating weighted user surplus $\phi_\alpha(p_B^k)$. The right-hand side of condition (32) corresponds to the maximum of three outcomes: accepting no card, accepting card 1 alone, and accepting card 2 alone. The left-hand side of condition (32) corresponds to the expectation of weighted user surplus when the merchant accepts both cards. In this case multi-homers prefer to use card 2 (since $p_B^2 \le p_B^1$), which explains the fraction $(\beta_2 + \beta_{12})$ of buyers who use card 2. This establishes the first bullet point in Proposition 11. Retail prices (second bullet point) are given at equilibrium by the average unit cost faced by merchants (including the net cost of card payments) plus a constant margin $\frac{t}{R}$. The other bullet points are immediate. ■

Appendix 2: Proof of Proposition 9

Proof of Proposition 9:

i) Given our assumption that issuers' margin is constant, their profit is maximized for a value p_B^M of buyers' price that maximizes the volume of card transactions, given by (26):

$$V(p_B) = D_B(p_B) D_S(\widehat{b}_S(p_B)),$$

where $\widehat{b}_S(p_B) = c + m - p_B - \alpha v_B(p_B)$.

Now p_B^{SH} is given by the first-order condition:

$$\frac{V'}{V}(p_B^{SH}) = \frac{D'_B}{D_B}(p_B^{SH}) - \frac{D'_S}{D_S}(\widehat{b}_S^{SH})(1 + \alpha v'_B(p_B^{SH})) = 0,$$

which gives (28).

ii) Social welfare is maximized for a value p_B^W of buyers' price such that:

$$\frac{\partial W}{\partial p_B}(p_B^W, \widehat{b}_S^W) = \frac{\partial W}{\partial \widehat{b}_S}(p_S^W, \widehat{b}_S^W)(1 + \alpha v_B'(p_B^W)). \tag{33}$$

Using formula (25) we see that

$$\frac{\partial W}{\partial p_B}(p_B^W, \widehat{b}_S^W) = D_B'(p_B^W) \int_{\widehat{b}_S^W}^{\infty} (p_B^W + b_S - c) dG(b_S).$$

Now by definition:

$$v_S(\widehat{b}_S^W) D_S(\widehat{b}_S^W) = \int_{\widehat{b}_S^W}^{\infty} (b_S - \widehat{b}_S^W) dG(b_S).$$

Thus we can write:

$$\frac{\partial W}{\partial p_B}(p_B^W, \widehat{b}_S^W) = D_B'(p_B^W) D_S(\widehat{b}_S^W) \left[v_S(\widehat{b}_S^W) + m - \alpha v_B(p_B^W) \right].$$

Similarly

$$\frac{\partial W}{\partial \widehat{b}_S}(p_B^W, \widehat{b}_S^W) = D_S'(\widehat{b}_S^W) D_B(p_B^W) \left[(1-\alpha) v_B(p_B^W) + m \right].$$

Thus the first-order condition for welfare maximization (condition (33)) can be rewritten as:

$$\frac{D_B'(p_B^W)}{D_B(p_B^W)} \left[v_S(\widehat{b}_S^W) + m - \alpha v_B(p_B^W) \right] = \frac{D_S'(\widehat{b}_S^W)}{D_S(\widehat{b}_S^W)} [(1-\alpha) v_B(p_B^W) + m](1 + \alpha v_B'(p_B^W)).$$

This is equivalent to formula (29).

iii) When social welfare is quasi-concave with respect to the buyer price, the privately optimal IF (associated with buyer price p_B^{SH}) is excessively low whenever

$$\Delta \equiv \frac{\partial W}{\partial p_B}(p_B^{SH}, \widehat{b}_S^{SH}) - \left[1 + \alpha v_B'(p_B^{SH})\right] \frac{\partial W}{\partial \widehat{b}_S}(p_B^{SH}, \widehat{b}_S^{SH}) < 0.$$

Adapting the formulas obtained above, we see that

$$\begin{aligned} \Delta &= D_B'(p_B^{SH}) D_S(\widehat{b}_S^{SH}) \left[v_S(\widehat{b}_S^{SH}) + m - \alpha v_B(p_B^{SH}) \right] \\ &\quad - [1 + \alpha v_B'(p_B^{SH})] D_S'(\widehat{b}_S^{SH}) D_B(p_B^{SH}) [(1-\alpha) v_B(p_B^{SH}) + m]. \end{aligned}$$

Applying condition (28) allows us to simplify this expression:

$$\begin{aligned}\Delta &= D'_B(p_B^{SH})D_S(\widehat{b}_S^{SH})\left[v_S(\widehat{b}_S^{SH}) + m - \alpha v_B(p_B^{SH}) - (1-\alpha)v_B(p_B^{SH}) - m\right], \\ &= D'_B(p_B^{SH})D_S(\widehat{b}_S^{SH})\left[v_S(\widehat{b}_S^{SH}) - v_B(p_B^{SH})\right].\end{aligned}$$

Thus $\Delta < 0 \Leftrightarrow v_S(\widehat{b}_S^{SH}) > v_B(p_B^{SH})$, and the proof of Proposition 9 is complete. ■