Carl Christian von Weizsäcker

Antitrust Problems of Four Party Credit Card Systems

February 2009

weizsaecker@coll.mpg.de

Abstract: Visa and MasterCard are highly successful payment systems. Until recently they were bank cooperatives. The cooperative agreements contained clauses which led to collectively decided "interchange fees" paid from the acquiring bank (the bank serving the merchant) to the issuing bank (the bank serving the cardholder). Antitrust authorities in many countries considered these interchange fees as price fixing. Is this the case? The answer is: no. In particular I show that the system, if left alone, sets interchange fees so as to maximize business volume, whereas a restriction of competition would tend to restrict output.

A Payment Systems

Payment systems are institutions, which allow buyers and sellers to execute a payment from the buyer to the seller as a compensation for the good delivered from the seller to the buyer. Examples of payment systems are cash, debit cards, credit cards, direct transfer from one bank account to the other, and cheques. The demand for the services of a payment system is always joint demand of a buyer and a seller. Thus payment systems are a prime example of two-sided demand. If we measure the annual quantity of output of a payment system in terms of the total Euro sum of payments performed by the payment system within a year, then we understand that the service provided to the seller (we call him the "merchant" from now on) and the service provided to the buyer (frequently called the cardholder from now on) is supplied and obtained in strict complementarity: Within a payment system the total amount paid by buyers to merchants and the total amount received by merchants from buyers is the same.

Not the least because of this structure of payment systems, but also because of similar structures in the field of telecommunications and media including internet based media economists have developed a new theoretical field, the field of "two-sided demand" and of platforms supplying the services which are in "two-sided demand". This field has caught considerable interest also because of many antitrust proceedings related to this two-sided demand structure. Thus payment systems like the credit card organizations Visa and MasterCard are now for many years under considerable scrutiny of antitrust authorities. Some

of the important insights of economic theory on two-sided demand have been summarized by Rochet and Tirole (2006 A) and Armstrong (2006). Payment systems in particular have drawn substantial academic attention. See Rochet and Tirole (2006 B) and Guthrie and Wright (2007) as well as the literature cited there.

Being an example of joint demand, buyer and merchant have to agree which payment system to use for any particular transaction. If there were no transaction costs then, using the Coase theorem, we could predict that they would decide to use that payment system which, among all payment systems, maximizes the sum of their net benefits derived from the transaction. Here the net benefit is the gross benefit of having this payment performed minus the cost of the payment system. Thus, if p is the price paid by the buyer and if q is the price paid by the merchant, demand for the services of the payment system only would depend on the total price $t = p + q$. Indeed, in this case without transaction costs between buyer and merchant, the consumer surplus of this payment system can be read in the usual way from the demand curve for the payment system as a function of its price t: it is the area below the demand curve and above the horizontal line which corresponds to the actual total price t.

Things become more complicated as soon as transaction costs are involved before buyer and seller can agree on the choice of a payment system.

Joint demand for, or joint choice of the payment service may be determined by the customer and the merchant within the negotiation of a purchase agreement of some item of some value so that it is worth negotiating about the price of the item. There is then no real distinction between the transaction cost of negotiating about the price of the item and the transaction cost of choosing the best payment system. We then can expect that that one payment system is jointly chosen that has the highest joint net benefit for merchant and customer, like in the hypothetical case without transaction costs

Or this joint demand for a payment system is standardised to avoid the need at each instance for negotiation about the choice of payment system. This is the case in routine purchases of generally lower value, say in a grocery store or at a petrol station or in a hotel or in a restaurant or with purchases over the Internet. These routine purchases have the following features: the merchant sets the price of the item, the customer decides about the quantity. The standardisation of the joint demand for the services of a payment system then is this: 1.the

merchant previously has decided which payment systems generally to accept; 2. for any given purchase transaction, within the payment systems accepted by the merchant, the customer has the choice which payment system to use; 3. but the merchant can influence the choice of the customer by offering discounts depending on the choice of payment system or by limiting the choice of the customer further, for example as a function of the total purchase volume (say, no credit cards for purchases below 30 €) or by simply talking to the customer and telling her which payment system he, the merchant, prefers.

.

Generally speaking, the competition authorities have misunderstood the character of choice of payment system for routine purchases: they see this as a one-sided choice of the buyer. They say, merchants are forced to (have no choice but to) accept MasterCard and/or Visa or American Express for competitive reasons, so step 1. is not a free decision; then the customer in step 2. can choose as she wants without taking account of the preference of the merchant for one payment system over the other. They basically deny the realism of step 3.

But this is a fundamentally mistaken view. The standardisation of the joint choice of a payment system has the purpose to save transaction costs. Thus, any deviation from a negotiated joint choice, i.e. from a choice that takes account equally of the interest of both parties, is only stable as long as it does not hurt one party more than it would be hurt by bearing the transaction costs of a negotiated joint choice. But, most purchases are repeat purchases between the same merchant and the same customer. Transaction costs to negotiate a mutually beneficial choice of payment system then can be spread over the expected sum of purchases over an extended period of time. Thus they are low in comparison with the total expected volume of purchases covered by the negotiated agreement on the joint demand for payment services. Hence the sacrifice of the merchant when allowing the customer to choose the payment system is limited by the low level of transaction costs in negotiating the joint choice of payment system.

The services of the payment system to the merchant and to the buyer are provided in joint supply. There is then strict complementarity not only on the demand side (joint demand), but also strict complementarity on the supply side. Now we know that we cannot disentangle the separate costs of two goods which are produced in strict complementarity. Only their joint average or marginal cost can be obtained. Let g be the marginal cost of the joint supply of the payment service. Assume (only for simplicity of presentation) that the marginal cost g is

independent of the output level X. We assume that demand X depends on the price for buyers as well as on the price for merchants. Suppressing the influence of prices of other payment systems on the demand for this payment system, we can write $X = X(p,q)$. Profit then is

$$\pi = (p + q - g)X(p,q)$$

We can further assume that the profit maximum is characterised by the inequality $p + q - g > 0$. It is then easily seen that the following holds: for any given total price $t = p + q$, profit is maximised by setting p and q so as to maximise demand X. This proposition is so obvious that I like to call this proposition the "Obvious Theorem".

B Four Party Payment Systems

It is usual to distinguish between three party payment systems and four party payment systems. At three party payment systems any given payment directly involves three parties: 1. the buyer, who pays, 2. the merchant, who receives the payment, and 3. the payment system, which makes the payment happen. The prototypical three party payment system is "cash". The manager or "owner" of this system is the central bank. Another three party payment system is the credit card system of American Express. In a three party system the payment system directly (or via its agents) deals with the buyer, who pays, and with the merchant, who receives payment.

Four party payment systems have a different structure. Here one bank, the "acquirer" deals with the merchant; another bank, the "issuer", deals with the buyer or cardholder. Thus here the payment goes from the cardholder to the issuing bank, then to the acquiring bank and from there to the merchant. Four parties are involved. The reason that, despite their more complicated structure, four party payment systems are successful is that banks have their "natural" customers: the issuing bank in all likelihood is the bank where the (potential) cardholder has her account; the acquiring bank tends to be the bank of the merchant that provides credit to him and serves other functions for the merchant. Thus, then the "selling cost" of the payment system comprising all these banks is low in comparison with a three party system, like American Express, which must put in much more effort to obtain customers who originally are not close to the payment system.

On the other hand these efficiencies of four party payment systems only come, if certain "rules of the game" are observed by its member banks. A credit card like MasterCard or Visa, or a debit card like Maestro is the more useful for the cardholder the more merchants accept it as a means of payment. Similarly, a payment system is the more useful to the merchant the more customers use it or prefer to use it. These are the well known network effects which have been studied in the economics literature. For a four party payment system to be able to take advantage of these network effects it must require its merchant customers to comply with what is called the "honour all cards rule" (HACR). If the merchant accepts a particular four party payment system, like MasterCard, then he is obliged by the contractual terms to accept all cards of the system. He cannot selectively accept MasterCard cards from certain issuers without accepting MasterCard cards from certain other issuers. The four party payment system can develop its advantages only if the honour all cards rule prevails. Thus the merchant, or his bank, the acquirer, does not individually negotiate with each of thousands of issuing banks about acceptance of their cards. It is a zero-one decision for the merchant whether to accept MasterCard or not. Also it is impracticable to let the price paid by the merchant to the acquirer depend on the issuer. It is the same for all MasterCard cards. Without here going into the details, it can easily be shown that a four party payment system cannot survive, unless it works with the HACR.

But the HACR has another consequence. How do the issuing and acquiring banks involved in the scheme decide about the terms and conditions concerning the transactions between them? On the occasion of a payment by a cardholder to a merchant using the particular four party payment system, say MasterCard, the issuing bank of that cardholder provides a service to the cardholder and at the same time provides a service to the acquiring bank of the merchant, who is to receive the payment. On that same occasion the acquiring bank of the merchant provides a service to the merchant and at the same time provides a service to the issuer of the cardholder. Remember that a payment service is always a joint service to the payer (the cardholder) and to the payee (the merchant). Thus the service provided to the cardholder by her bank, the issuer, (= making an effective payment) requires as an input the service of the acquiring bank, the bank of the merchant. At the same time, the service provided to the merchant by his bank, the acquiring bank, (= effectively receiving a payment) requires an input from the issuing bank. Thus, we might expect a price paid by the issuer to the acquirer and a price of the acquirer paid to the issuer. The difference between these two prices is, what

is actually being paid. This difference of two prices has been given a particular name: the interchange fee. The convention is to say that the interchange fee is positive, if the net payment goes from the acquirer to the issuer and to say it is negative if the net payment goes from issuer to acquirer.

But how is the interchange fee determined? Does each acquirer negotiate with each issuer about the level of the interchange fee? Given that thousands of banks all over the world are issuers and a (smaller, but still) substantial number of acquirers are members of the system, this would imply perhaps millions of bilateral agreements. But the sheer quantity is not the most important obstacle to bilateral agreements. The additional problem is that acquirers do not have any negotiating power vis à vis issuers: because of the HACR acquirers had already agreed that they accept each issuer's cards. Thus, even before a bilateral agreement between any issuer and any acquirer exists, the acquirer has already accepted that it has to provide its service to the issuer. Moreover, what happens if a buyer presents a card from an issuer without a bilateral agreement with the acquirer of the merchant? The money flows from the buyer to the issuer and then to the acquirer and then to the merchant. The issuer then is free to subtract an interchange fee according to his own taste, obviously a very uncomfortable situation for the acquirer. Indeed the acquirer in this situation is in the position of the passive part in a dictator game, whereas the issuer is the "dictator" of that dictator game. We can expect that a four party payment system which would rely on bilateral agreements between each issuer and each acquirer simply cannot work. No acquirer in his right mind would agree to join such a payment system.

The answer then is that the four party payment system does not only require an honour all cards rule (HACR). To protect the interests of acquirers it also needs a collective rule on the interchange fee. To be exact, it needs a fallback rule about the interchange fee in case bilateral agreements between two banks do not exist in a situation in which they have to deal with each other on the occasion of a transaction between their customers, a cardholder and a merchant. If, independent of the precise form of transaction, the same interchange fee would apply we would expect it to be the same as the fall-back interchange fee collectively determined. If there is scope for a differentiated interchange fee depending on the precise form of transaction then bilateral agreements may come about which deviate from the collectively set fall-back interchange fee. Thus for example, costs of executing the payment by the system may be different if they are performed electronically than if they are performed in a more traditional

manner. Then it may be attractive for issuer and acquirer to differentiate the interchange fee according to the technology used so that the acquirer can offer the merchant a differential merchant service charge depending on the technology used; thereby the acquirer can provide incentives to the merchant to buy the equipment needed to switch to the electronic technology. But even if bilateral agreements lead to interchange fees which are different from the one established collectively we still would expect the average interchange fee to be near the collectively decided fall-back interchange fee.

C Restriction of Competition?

What is a restriction of competition? The different cases throughout the world involving MasterCard and/or Visa throw up rather fundamental questions about the concept of a restriction of competition. In the following I shall refer to the legal framework of the European Community – pars pro toto. But similar law exist throughout the world. Article 81, 1 of the Treaty prohibits agreements among undertakings which have the intent or the effect to restrict competition. Article 81, 3 makes a restriction of competition legal, if it contributes to economic progress, if it in addition provides benefits to consumers, if that restriction is indispensable for this beneficial effect, and if by this restriction competition is not completely eliminated.

One point of contention then is: is the collectively determined interchange fee an Article 81, 1 case or is it an Article 81, 3 case. This distinction is important, among other things, because in an Article 81, 1 case the authority (here the European Commission) has the burden of proof that a restriction of competition prevails. If it is established that a restriction of competition is there then seeking an exemption under Article 81, 3 from its prohibition puts the burden of proof on the parties.

The fact that some collectively determined interchange fee is needed to keep the four party payment system going indicates that not all possible levels of interchange fees can be a restriction of competition. For, if each level of a collectively determined interchange fee were a restriction of competition among participants in the four party scheme, then the counterfactual without a restriction of competition would be a four party payment system without a collectively determined interchange fee. But then, as we have shown in the

preceding section, the scheme would not exist, i.e. there would be no competition to restrict anyway.

We may now ask: given that some collectively determined interchange fee exists, which is not a restriction of competition, which one is it? Or, are there more than one interchange fees which are not a restriction of competition?

In addition the following question arises: provided that there are certain levels of the collectively determined interchange fee which are not restrictive of competition, has the four party payment system an incentive to decide on such a non-restrictive interchange fee or not?

Before we can obtain answers to these questions, we have to find out what are criteria by which we appraise the competitiveness of a situation? Economists generally have emphasised the efficiency criterion. There is the general tenet among economists that there are market failures, if the market outcome is not efficient and that then the government should step in to correct that market failure. Here I do not want to quarrel with this general statement, except to add that government correction of market outcomes only makes sense, if there are no countervailing "government failure" phenomena, which prevent the government from improving the situation by its intervention. My main point is a different one: history has developed a legal structure of government action which is characterised by a significant division of labour between different government departments. There is specialisation not only in the private sector of the economy and society; there is also specialisation within the government sector. Thus, the different government departments do not pursue the goal of efficiency per se, as defined by the economist; they pursue specific goals: the police pursues the prevention of crime, schools pursue certain educational goals as defined by the legislators, as for example the ability of people to read, write and do elementary forms of arithmetic. The transport department pursues the goal of efficient forms of mobility. And the anti-trust authorities have as their goal the protection of competition among private undertakings. It is true that each government department is required to pursue its goals in an efficient manner. This it does so as to achieve a maximum of its specific goal (i.e. the prevention of crime or literacy of the people) with a given budget. But it is not the function of the police to consider a trade-off between the prevention of crime and the level of literacy according to an overriding efficiency criterion. And there are good reasons for this: policemen are trained to prevent crime efficiently; they are not particularly good at dealing with literacy. Thus they are

not well equipped about trade-offs between crime prevention and literacy. Similarly teachers may be experts in achieving a maximum of literacy within given resources available for the school system; but they are not trained to do a good job about the trade-off between literacy and crime- prevention.

D The Goal of Competition

Competition is the goal to be achieved by competition policy and thus by antitrust laws. Human ingenuity has not found another effective principle to organise markets which generally is superior to competition for the goal of enhancing consumer welfare. But, competition is a concept which is distinct from consumer welfare. Thus, for example, it is not always the case that density of retail outlets and prices of retail outlets in a competitive equilibrium are set at levels which maximise consumer welfare under the constraint that the retail outlets break even. It could well be that slightly less price competition and thus somewhat higher prices, which then afford a higher density of retail outlets, provides more consumer welfare than the competitive outcome. Or to take another example, there is no guarantee that opening hours of shops under competition are such as to maximise consumer welfare. It could well be that (cost saving) shorter shopping hours with lower retail margins and thus lower consumer prices provide more consumer welfare than is achieved in the competitive equilibrium.

Nevertheless, the lawmakers in most countries instruct their antitrust agencies to pursue the goal of competition and not the goal of welfare maximisation. The reason is clear: finding the welfare maximising deviation from the competitive equilibrium and implementing it by government order is an impossible task for anti-trust authorities. It would completely overwhelm them and would generate one mistake after the other with the result that things are much worse than with the more modest goal of promoting competition.

In the real world there is almost never “perfect competition”, i.e. there is almost never a situation, in which the goods of the different suppliers are perfectly interchangeable. Generally there is “imperfect competition”: there is a high, but not infinite elasticity of substitution between the goods supplied by the competing suppliers. And, indeed, product

differentiation is one of the important strategies of competitors in their attempt to obtain a competitive advantage over their rivals. Thus, for example, suppliers may have a choice of the quality level of their products. They could decide to offer high quality and thereby to attract consumers even though they charge higher prices than some or most of their competitors, or they could decide to save costs and to offer lower quality and to attempt to attract consumers by charging a lower price than some or most of their competitors. If different suppliers choose different quality strategies consumers have a choice between high quality at a high price and lower quality at a low price. Thus product differentiation strategies may enhance consumer welfare relative to a situation, in which every supplier offers the same quality. On the other hand, it could well be that, from a consumer welfare point of view, competition leads to "too much choice" of consumers relative to a hypothetical situation, in which suppliers still charge cost covering prices, but incur lower costs due to a lower degree of product differentiation. In real life it is simply asking too much from competition authorities to find out whether, from a consumer welfare point of view, there is too little or too much product differentiation in any particular market.

As a rule, consumers are not very well informed about the total array of offerings in a market. It would take a substantial effort for the consumers to obtain full information about available products, before they decide which one to buy. So they buy without incurring the cost of getting fully informed. Thus, suppliers have to act in a market with imperfect information on the demand side. They then, as a rule, try to change the state of information of consumers by their marketing strategies. For example, a supplier advertises his brand to make consumers aware of his product and, indeed, to convince consumers that his products are "better" than those of the competitors. Brand advertising then is an important competitive instrument in the struggle of suppliers to survive under competitive conditions. Experience tells us that in a majority of markets for consumer goods brands go under and their suppliers go out of business, if they do not advertise. Yet, economists have been unable to show, indeed are very sceptical about the possibility to show that the level of advertising in a competitive market is at the level which maximises consumer welfare. Nevertheless, unless a supplier is dominant in the sense of antitrust law, no competition authority so far has challenged his right to advertise as he wishes. If, for specific reasons, society comes to the conclusion that advertising for a product is harmful, then it is not antitrust law which is invoked to prevent this advertising activity. There are then specific regulations like, for example, regulations against TV advertising for cigarettes.

Generally, there is an intuition in most people to think that efforts to sell a product, say, by advertising it, is distinct from efforts to improve the quality of the product in terms of, say, performance on any specific scale (weight of a notebook, for example) or durability or serviceability or probability of breakdown or side-effects (of pharmaceutical products, for example) or instant availability nearby or convenient availability of expert help for using the product or aesthetic attractiveness or improved taste (for food products) or ….. As said above, this distinction in the effort to raise demand for the product by improving quality or by raising the selling effort is not really relevant in anti-trust matters. Moreover, in the case of two-sided services, the platform advertising addressed to side 1 and thereby capable of raising side 1´s demand for the platform thereby simultaneously raises the quality of the service in the view of side 2. Thus MasterCard's advertising telling potential cardholders that so many merchants accept MasterCard cards may raise the number of actual MasterCard cardholders and thereby raises the quality of the MasterCard payment system in the eyes of merchants.

How do we then decide about the competitiveness of a market? A fairly robust characteristic of competition in a market is that it maximises output in the market subject to the constraint that suppliers cover their costs. Moreover, suppliers – in their attempt to maximise profits – try to minimise costs, subject to the constraint that they can deliver the goods which customers want to buy. Thus, given quality levels and advertising levels and other selling effort levels of suppliers in the market, a lower price by one supplier will raise demand for the product of this supplier, but at the same time will reduce the demand for the products of competing suppliers. It will also raise total market demand: that is, the price reduction raises the demand for the product of this supplier by more than it reduces the demand for the products of his competitors. Thus, given quality and selling effort, price competition among suppliers leads to higher market demand than would a restriction of price competition, say, due to a price fixing agreement.

The same applies for quality competition. Given prices and given the level of selling effort of suppliers in the market, raising the quality of the product by one supplier will raise his sales and will reduce the sales of the other suppliers, but it will raise market demand. In other words, it will raise demand for the product of that supplier by more than it reduces the demand for the products of his competitors. Thus, competition in terms of quality will raise the market demand in comparison with a situation, in which – by whatever mechanism -

quality is kept below the level which it obtains under competition, given prices and selling effort of all suppliers.

And again the same is true for selling effort, including advertising. Given prices and given quality levels of products, an increased selling effort by a supplier will raise demand for his product at the expense of the demand for the products of his competitors. But, market demand will be up, that is, the higher demand for his product outweighs the loss in demand for his competitors´ products. Other things equal, higher advertising budgets of suppliers will raise market demand. Thus, with given prices and quality levels the level of advertising under competition will raise market demand as compared to a restriction of the advertising levels, say, due to a restriction of competition by artificially limiting advertising budgets below their competitive levels.

Without going into the specifics of any given market we cannot say very much about the interplay of the three competitive instruments so far mentioned: price, quality, selling effort. But this is also not necessary for our purposes. It is enough to state that there are markets, in which price competition dominates, whereas there is a high level of standardisation of the quality of the products. Consumers know the product well and ask for the standardised quality level, so that a quality deviation does not pay for the supplier. The selling effort may also be limited, for the very reason that everybody sells the same quality. Consumers here tend to be well informed about potential sources of supply. (Indeed, typically these are markets for intermediate products, so that the demand side does not consist so much of end-consumers, but of undertakings producing a product). Then there are markets, in which product characteristics change all the time, due to product innovation. Consumers do not have perfect knowledge of the sources of supply and of the quality of the products (the “lemons” problem, as economists usually call this phenomenon). Branding and building up a reputation for good quality is important. The quality level itself becomes an important strategic variable to be determined. It is a “high margin business” in the sense that the price even under competition is much higher than the marginal cost of production; but there are substantial outlays needed in terms of product development, marketing, advertising, quality control etc so that despite the high margin of price over marginal cost it is not easy to obtain a profit.

Let us now observe that life in a market is characterised by large uncertainties. Thus, it is not the case that suppliers have automatically found the profit maximising combination of price,

quality and selling effort. They may change their decisions. For example, a manufacturer may decide that he must engage in heavier advertising and that it is possible to "finance" this additional outlay by raising the price of the product. If he is right in this, it may mean that demand for his product increases despite the fact that he has raised the price. This decision may have the result that market demand also increases despite one price having gone up – due to the heavier advertising of the product. Indeed, if one supplier has been successful with this strategy change, other suppliers in the market may follow, so that we reach a new competitive equilibrium with a higher price level, with more advertising - and with higher market demand. Or a new equilibrium is reached with higher prices, higher advertising, but with lower market demand. Any given supplier's advertising effort then has the main effect of cannibalising the sales of his competitors. Whether this shift from a lower price level and lower advertising level to a higher price level and a higher advertising level raises or reduces market demand, it is not the business of competition policy authorities to prevent this shift. Both equilibria are competitive equilibria; and it may simply be accidental which of the two equilibria ensues, given that the suppliers themselves do not know with certainty which is their profit maximising strategy: higher prices and higher advertising or lower prices and lower advertising.

We may define the concept of competitive pressure imposed by any one supplier on his competitors. We may not have a scale by which to measure competitive pressure like, say, we can measure the temperature at some place (the economist says: we may not have a "cardinal" measure of competitive pressure. But there has been a tradition in economics to measure competitive pressure by the percentage equilibrium mark-up of price over marginal cost: the Lerner index, or rather its inverse. Yet, this in my view is a mistake, because it focuses exclusively on price competition for given quality levels and given selling effort levels. If we make quality and selling effort endogenous the inverse of the Lerner index becomes a poor measure of competitive pressure). But we may be able to say that in situation 1 competitive pressure is higher than in situation 2.

Keeping quality levels and selling effort levels the same a reduction in price by supplier A puts additional competitive pressure on his competitors B, C, etc. This is seen in the fact that this price reduction in A-price reduces demand for B-product, C-product, etc., other things being equal. On the other hand it raises demand for A-product. The same with quality levels: keeping the price and selling effort the same a rise in the quality level of supplier A puts

additional competitive pressure on his competitors B, C, etc. This is seen in the fact that the higher quality of A-product reduces demand for B-product, C-product etc., other things being equal. On the other hand, it raises demand for A-product. The same again with selling effort: keeping the price and the quality level the same a rise in selling effort (say, by raising the advertising budget) by supplier A puts additional competitive pressure on his competitors B,C, etc. This is seen in the fact that the higher selling effort of A-product reduces demand for B-product, C-product etc., other things being equal. On the other hand, it raises demand for A-product.

We then can conclude: an action by supplier A, which raises demand for his product, imposes additional competitive pressure on suppliers B, C, etc. in the same market. This is seen in the fact that this action tends to reduce demand for B-product, C-product, etc. Now, price, quality and selling effort all are instruments which work on the demand side for the product. They presuppose that we look – realistically – at a market where sales are demand constrained and at which suppliers are "transaction hungry". But instruments which are there to influence the demand constraint are – in game theoretic language – strategic complements. So a price reduction by A will be followed by a price reduction by B, C, etc. A quality improvement by A will be followed by a quality improvement by B, C etc. A higher selling effort by A will be followed by a higher selling effort by B, C. etc. Thus increased competitive pressure by A leads to lower prices, higher quality, higher selling effort of B, C, etc. And we can measure increased competitive pressure from A on B, C, etc. by looking at the change in demand for the product of A, before B, C, etc. have adapted their price, quality, selling effort. If a change of policy by A raises demand for A-product competitive pressure on B, C, etc has been raised. If a change of policy by A reduces demand for A-product, competitive pressure on B, C, etc has been reduced.

I conclude that unrestricted competition in a market leads to higher market output than a situation of restricted competition. As I have emphasised above most of the time a welfare analysis of a market is much too complicated to be handled by a competition authority. But there is one quasi welfare statement: for the quantitative composition of output prevailing under unrestricted competition the price index of the goods sold on the market is lower than it would be under a situation of restricted competition. Thus, if – at a given price - supplier A raises demand for his product by raising its quality then, due to increased competitive pressure on B, C etc they will reduce their prices, other things equal and/or they will raise their quality.

Hence, for given quality levels and taking the quantities sold under competition as weights, the price index of the goods sold on this market will be lower under competition than in a situation of restricted competition.

I then propose that competitiveness of a market can be measured in ordinal terms (not cardinally) by the output volume of that market: we expect a restriction of competition to reduce output relative to the counterfactual without the restriction of competition.

E The Competitive Counterfactual in Terms of the Interchange Fee

Following this thinking I then suggest that the interchange fee of a four party payment system should be considered not to be restriction of competition, if it maximises the business volume of the scheme.

This is in contrast to the view taken by many anti-trust authorities. They see a zero interchange fee as the competitive counterfactual. In their view this counterfactual is a regime without a collectively determined interchange fee. For example the European Commission calls this counterfactual a regime without a collectively determined interchange fee, but with a rule for the members which prohibits "ex post pricing". By this they mean a collectively determined rule that prohibits issuers to deduct any fee from a payment going to any acquirer.

It is obvious that the counterfactual favoured by the authorities simply is a regime with a collectively determined interchange fee of zero. Now we have seen in Section B above that the interchange fee is the difference between two prices for two different services, one service of the issuer to the acquirer and another service of the acquirer to the issuer. These services are provided in strict complementarity on the occasion of the execution of any particular payment from a cardholder customer of the scheme to a merchant customer of the scheme. But they are different services, involving different activities and different costs for issuer and acquirer. (Generally, in the case of credit cards, the issuer incurs higher costs than does the acquirer). There is no particular reason why the prices of two complementary goods should be the same, so that their difference would be zero. There is no particular reason why the annual leasing price for a car should be equal to the annual cost of the petrol used for driving the car. There is no particular reason why the annual leasing cost for a piano should be the same as the annual outlay for the lessons taken from a piano teacher. There is no particular reason why the

cream going with a cup of coffee should cost the same as the number of coffee beans going with a cup of coffee.

A zero interchange fee thus is a completely unconvincing competitive counterfactual for assessing the competitiveness of any given collectively determined interchange fee. This analysis is not changed by giving the zero- interchange fee regime a different name: a "regime without an interchange fee, but with a rule prohibiting ex-post pricing." In a sense the competition authorities are in a state of enlightenment which in arithmetic corresponds to the state before the number "zero" was invented by an Indian mathematical genius almost two thousand years ago. Indeed, if you work with Roman numbers you may have difficulty distinguishing between a situation without an item present and a situation in which the particular item is present, but happens to have the value zero. Remember that the Romans did not yet measure the weather temperature by "degrees Celsius". Today we make a clear distinction between the statement: "we do not know the outside temperature" and the statement "the outside temperature is zero degrees Celsius". The authorities seem to have a difficulty to understand that a collectively decided rule "which prohibits ex-post pricing" is exactly the same as a collectively determined interchange fee of value zero.

Also, to come back to the efficiency criterion, there exists no particular reason to assume that a zero interchange fee is more efficient than another level of the interchange fee. Look at a hypothetical world ("World No 1"), in which the handling costs of payments by the MasterCard system are equally distributed between issuers and acquirers, i.e. they are, say, 50 cents per € 100 payment for the issuer and 50 Cents per € 100 payment for the acquirer. Assume now that in this hypothetical world a zero interchange fee would be maximizing consumer welfare.

Consider now a different hypothetical world ("World No 2"), in which the demand functions of cardholders and cash payers, as well as of merchants are the same as in World No 1, but issuer handling costs are 80 cents per € 100 payment and acquirer handling costs are 20 cents per € 100 payment. Now, by using an interchange fee of 30 cents per € 100 payment, the cost situation in World 2 is transformed into the cost situation of World 1. Thus, evidently, the consumer welfare maximizing interchange fee in World 2 is equal to 30 Cents per € 100 payment.

More generally: the maximum consumer welfare achievable by the existence of the MasterCard system only is a function of the sum of the issuer and the acquirer unit costs, not of the distribution of these costs between issuer and acquirer. There is one potential distribution (call it "distribution zero") of the sum such that the welfare maximizing interchange fee is zero. All other potential distributions of the costs (for a given sum) can be transformed into the "distribution zero" by means of a positive or negative interchange fee. It would be purely accidental, if the distribution of costs between issuers and acquirers in the real world would be the "distribution zero". Thus, in all likelihood, the consumer welfare maximizing interchange fee is different from zero.

F Without Government Intervention the Four Party Payment System Tends to Maximise Volume

I now turn to the question: what happens to the competitiveness of the situation, if authorities leave the four party payment system alone? As I will show in this section we can expect the system to set the collectively determined interchange fee so as to maximise business volume. For this I assume that the system is interested to set the collectively determined interchange fee so as to maximise the joint profits of its participating members, the issuing banks and the acquiring banks. I do this, for ease of presentation, by means of a simplified "class room model". But, as will be seen in the next section, the proposition is much more general.

In a three-party system – given the prices of competing systems – profits Π are maximised by selecting the appropriate price level on the issuing side, p, and the appropriate price level on the acquiring side, q. We equivalently can also work with the total price $t = p + q$ and the price difference $\Delta = q - p$. Obviously, we have $q = \frac{t + \Delta}{2}$ and $p = \frac{t - \Delta}{2}$.

In a t, Δ-diagram we can draw "iso-profit-curves". They are elliptic closed curves, one inside the other. The more "inside" the curve is, the higher is the profit. In the centre of this collection of iso-profit curves there is a point at which profit reaches its maximum. Let us then observe certain properties of this "profit- landscape" in terms of volume of business, X. Let g be marginal cost of issuing, let h be marginal cost of acquiring. Let $X = X(t, \Delta)$ be the demand curve for the services of the payment system. Profit then is $\Pi = (t - g - h)X = (t - g - h)X(t, \Delta)$. Provided that $t - g - h$ is positive, it is obvious that the

profit maximising value of Δ is the volume maximising Δ. Keeping total price constant, a three party payment system sets the difference between the prices on the two market sides so as to maximise business volume. This is, what I called in Section A, the "obvious theorem".

In a four party payment system it is generally difficult to reach the profit maximising combination of t and Δ, or equivalently of p and q. The three party system can choose t and Δ independently. The four party payment system has one parameter, the collectively determined interchange fee, which influences both the total price and the price difference. In a two- dimensional diagram with a $t-$ axis and a $\Delta-$axis, the points attainable by setting the interchange fee form a curve. But there is no guarantee that this curve passes through the point, where under a free choice of total price and price difference profit is maximised. Nevertheless MasterCard, or any other four party payment system is likely to set the interchange fee so as to find the restricted profit maximum - restricted, because it has to be on the curve of $t,\Delta-$combinations representing different interchange fees.

The main impact, and the one easy to understand, of the interchange fee is on the price difference $\Delta = q - p$. The higher the interchange fee the larger is this difference. The interchange fee also may have an impact on the total price t. But the sign of this impact is not a priori clear. If it were the case that a change in the interchange fee has no impact on the total price then the interchange fee would be set so as to find that price difference Δ, which maximises volume: for the given total price profit is maximised if volume is maximised. Whether a rise in the interchange fee raises or reduces the total price depends on the pass-on ratios of the two market sides. If the pass-on ratio of a change in the interchange fee is higher on the acquiring side than on the issuing side then a rise in the interchange fee will raise the total price. If the pass-on ratio of a change in the interchange fee is lower on the acquiring side than on the issuing side then a rise in the interchange fee will reduce the total price.

The European Commission and other anti-trust authorities claim that pass-on ratios of the interchange fee are lower on the issuing side than on the acquiring side. Therefore, they say, a higher interchange fee implies a higher total price. They then conclude that it is in the profit interest of the four party payment system to set the interchange fee at a level which is higher than the volume maximising level. In other words, they predict that at the going interchange fee a small reduction in the interchange fee raises the business volume. It is also the case that so far in the academic literature the pass-on ratios on the issuing side and on the acquiring

side have not been variables to be determined by the system. Rather they have been taken as exogenous parameters. For example Rochet and Tirole assume that the pass-on ratio is 100 % on the acquiring side and less than 100 % on the issuing side. They do not derive this pass-on structure from other model assumptions. They only say in words that the acquiring business is more competitive than the issuing business.

In the following I show that the anti-trust authorities and the academic literature are mistaken in this respect. Indeed, we can show that the interchange fee which maximises volume also maximises total joint profits. The basic issue is a correct understanding of the forces which determine the level of pass-on ratios. For the anti-trust authorities and the academic literature it is mainly market structure which determines pass-on ratios. They consider the issuing markets as being local oligopolistic markets and they consider the acquiring market to be highly competitive. From there they derive the conclusion that the pass-on of the interchange fee is low on the issuing side and high on the acquiring side. I do not deny that market structure may have an influence on pass-on ratios. But the anti-trust authorities (as well as the academic literature, so far) ignore the specific effects on pass-on ratios which are due to the two-sidedness of the service provided by the payment platform. As we will see, the market structure effect on pass-on ratios is less important in the context of two-sided markets than in the context of one-sided markets.

The class-room model is a model with linear demand curves, where the slope of the demand curve (or the slope of the inverse demand curve) is independent of the prices coming into the picture. Further, for simplicity of presentation I assume that there are n issuers, who are alike, and that there are m acquirers, who are alike. I start with the issuing side of the four party payment system. For issuer i there is an inverse demand function

$$p_i = a_i - bx_i$$

Here the parameter b is a constant, the same for all issuers, whereas a_i reflects the influence of prices of other players (competing issuers of the same system, acquirers of the same system, prices in other payment systems). Marginal cost of issuer i is assumed to be a constant g, the same for all issuers Then there is the interchange fee r. For simplicity of presentation I assume that there is just one interchange fee, the same for all forms of transactions. Profits of issuer i then are

$$\pi_i = (p_i - g + r)x_i = (a_i - bx_i - g + r)x_i$$

The profit maximum is achieved at a price

$$p_i = \frac{a_i + g - r}{2}$$

Assuming Bertrand- price competition from this follows the well known equation (*) (which also holds for more general cases)

$$p_i - g + r = bx_i \qquad (*)$$

In the profit maximum the margin as a percentage of price equals the negative of the inverse of the price elasticity of demand, or, after multiplying both sides by p_i, we obtain the equation (*). The quantity sold at the profit maximum is

$$x_i = \frac{a_i - g + r}{2b}$$

Given that all issuers are alike, we know that in Nash-Bertrand price equilibrium among issuers all issuer prices are the same. That is, also all quantities x_i are the same. Thus, in this equilibrium all values a_i are the same.

Let p be the price charged by all issuers in equilibrium. Let x be total quantity demanded from all n issuers. Thus, we can write the equation (*) $p_i - g + r = bx_i$ in the following form

$$p(r) - g + r = \frac{b}{n} x(r)$$

Differentiation of this expression with respect to the interchange fee r yields

$$p'(r) + 1 = \frac{b}{n} x'(r)$$

The pass-on ratio of the interchange fee among issuers is of course equal to $-p'(r)$. We then see that its value is equal to $1 - \frac{b}{n} x'(r)$. As long as a rising interchange fee induces a rising MasterCard business volume x, the pass-on ratio is less than one. At the point where the volume maximising interchange fee is located the volume derivative is zero. Then the pass-on ratio of the interchange fee is exactly 100%. If a rise in the interchange fee reduces business volume then the pass-on ratio on the issuing side is above 100%.

In a similar way we can proceed at the acquiring side of the business. Again, I assume that the m acquirers are all alike. Here we look at a particular acquirer j and its inverse demand function

$$q_j = d_j - cy_j$$

q_j is the price of acquirer j; y_j is the business volume of acquirer j; c is a constant representing the inverse price sensitivity of demand for the services of acquirer j; and d_j represents the influence of other prices on demand for the services of acquirer j. In a similar way as with the issuer we find the profit maximising price of acquirer j:

$$q_j = \frac{d_j + h + r}{2}$$

Here h denotes the marginal cost of the acquirer other than the interchange fee, which is the same for all acquirers; and r again is the interchange fee.

Again, as on the issuer side, in Bertrand- Nash equilibrium all acquirer prices are the same; moreover business volume y_j of acquirer j is equal to $\frac{y}{m}$, where $y = x$ is the business volume of all acquirers, which of course is equal to the business volume of all issuers. In a similar way as on the issuing side we can write the market equilibrium equation as

$$q(r) - h - r = \frac{c}{m} y(r) = \frac{c}{m} x(r)$$

Differentiation with respect to r yields

$$q'(r) - 1 = \frac{c}{m} x'(r)$$

The pass-on ratio on the acquiring side then is

$$q'(r) = 1 + \frac{c}{m} x'(r)$$

It is greater than 100 %, if a rise in the interchange fee raises business volume. It is smaller than 100%, if a rise in the interchange fee reduces business volume.

We may then ask the question: what happens to total price as a function of r? We have

$$t(r) - g - h = p(r) - g + r + q(r) - h - r = (\frac{b}{n} + \frac{c}{m}) x(r)$$

Upon differentiation with respect to r we obtain

$$t'(r) = (\frac{b}{n} + \frac{c}{m}) x'(r)$$

As long as a rising interchange fee goes with a rising business volume the total price rises with a rising interchange fee. At the volume maximising interchange fee the total price reaches its maximum. If a rising interchange fee reduces business volume then the total price

declines with a rising interchange fee. Thus the total price is a "single peaked" function of the interchange fee with its maximum at that interchange fee which maximises business volume. It is obvious that the four party payment system maximises profit of the system $\Pi = (t - g - h)x$ at the interchange fee at which the highest business volume and at the same time the highest total price is reached.

The economics of this result is much more general than just presented for the "classroom model". In the next section I show the mathematics of this generalisation. Even though the precise result in more complicated models may deviate somewhat from the precise coincidence of maximum total price with maximum business volume, I show there, why this deviation in all likelihood is too small to distract the four party payment system from its goal to maximise business volume.

Here I now explain the result in economic terms. I thereby show that the result of the "class-room model" is not a curiosity, but gives us a very general structure of two-sided markets. In one-sided markets it may indeed be the case that the pass-on ratio of costs depends on market structure. In a linear monopoly model it is known that this pass-on ratio is 50 %. The corresponding linear model of monopolistic competition shows a pass-on ratio of 100 %. In linear oligopoly models we obtain pass-on ratios which are above 50 %, but which may fall short of 100 %. If we go to models with non-linear demand curves the results are somewhat different, but it remains true that pass- on ratios rise as the number of competing suppliers rises. The fact that pass-on ratios are lower in oligopoly or monopoly than under perfect competition or monopolistic competition can be understood by looking again at a simplified version of equation (*)

$$p - g = bx \quad (*).$$

For a given slope b of the inverse demand curve of the individual supplier the profit maximising price cost margin $p - g$ is proportional to the supplier's business volume. Under competitive conditions the price is driven down nearly to marginal costs. There the market demand sensitivity is low, relative to the individual demand sensitivity 1/b. The price mainly is determined by marginal cost. Thus, whenever marginal cost rises the price rises by the same amount, whereas the business volume, due to inelastic market demand does not change much. So we have a pass-on ratio close to 100 %. In a monopoly the market price sensitivity of

demand is equal to the price sensitivity for the supplier. Thus, a rise in costs and thus price has an appreciable negative effect on business volume. This then by equation (*) means that the mark-up of price over cost is reduced by an appreciable amount. The pass-on ratio is appreciably lower than 100%. Oligopoly then is an intermediate case between monopolistic competition and monopoly.

In the case of four party payment systems a rise in cost due to a change in the interchange fee on the one market side goes together with a fall in cost on the other market side. Thus, even under oligopoly or monopoly, the appreciable reduction in demand due to the cost rise and own price rise is tied with an increase in demand on the other side of the two-sided market, because there cost and thus price have gone down. It is then no longer clear that the cost rise on the own side induces an appreciable reduction in demand. In cases, where the demand effect of the cost increase on one side is fully compensated by the demand effect of the cost reduction on the other side, the value of x in the relevant equation (*) has not declined, so that the mark-up of price over cost has not declined and thus the cost increase has been passed on to customers by 100%.

The full compensation of the demand effect of a marginal change in cost due to a marginal change in the interchange fee by the demand effect on the other market side is what precisely happens at the volume maximising interchange fee. There a marginal change in the interchange fee has no effect on business volume. Thus it is precisely there that a 100 % pass-on ratio prevails. And this result is independent of the market structures on both market sides, as long as we still assume the Bertrand-Nash mode of price formation, which is the basis for equation (*).

It is interesting to find an analogy for our finding in one-sided markets. Take the case of a monopoly, where the pass-on ratio of an exogenous cost change is likely to be appreciably lower than 100%. But we can imagine a case in which the cost increase does not fall from heaven, but is induced by the monopolist himself: he may decide to upgrade the quality of his product, thereby raising the marginal cost of production, say, by one Euro. But, at the same time the quality has become better; and thus demand does not go down, if the monopolist raises his price by one Euro. Thus this cost increase can be passed on by 100 % or more, because demand has at least remained stable, if not even increased due to the quality change. Indeed, the whole quality upgrading only makes sense, if the corresponding cost increase can

be passed on by 100 % or more. Otherwise the quality upgrading would have eaten into the monopolist's profits.

We then can understand the passing on effect of the interchange fee in analogy to the quality effect just mentioned. If a product is in higher demand on one market side then in a two-sided market this is like a quality improvement of the product for the other market side. If more merchants accept MasterCard or if fewer accepting merchants discourage the use of MasterCard then – in the eyes of cardholders – the quality of a MasterCard card has increased: the heavily discussed network externality of payment systems. If more people use the MasterCard card as a means of payment then it becomes a more attractive proposition for merchants to accept the card. A change in the interchange fee then always means a cost reduction and thereupon a price reduction on one market side, thus enhancing demand on that side, which thereby means a quality improvement for the other market side that goes together with price increase that this other market side faces, due to the same change in the interchange fee. Whether this quality improvement falls short of the corresponding price increase or compensates it or even over-compensates it is then a question for the specific situation. To the extent that a higher interchange fee implies a rise in MasterCard business volume the cost increase for merchants has been overcompensated by the "quality" increase and thus the pass-on ratio of acquirers is above 100 %. In the same situation the price reduction for cardholders has not been fully eaten away by a "quality" reduction (due to less acceptance by merchants, due to a higher proportion of discouraging the use of the card among merchants) so that demand by cardholders increases. Then the pass-on ratio of the lower cost of issuers is below 100 %, because they now provide a "better" product price-plus-quality-wise. But then, the interchange fee goes with a higher total price, the pass-on ratio on the cost rising side being higher than the pass-on ratio on the cost reducing side. The basic reason for this price rise, which also goes with a rise in demand, is that overall the "quality" of the product has risen. It has gone up by more than the price rise on the merchant side. And it has come down by less than the price reduction on the cardholder side. So part of the overall improvement in quality is absorbed by issuers and acquirers in the form of a higher total price.

We can see this in analogy to an outward shift in the demand curve in a market, be it a monopoly, be it an oligopoly, be it a market with monopolistic competition. This outward shift may be interpreted as a quality improvement of the product. Whatever its precise cause, we generally expect this outward shift to have the effect that volume as well as price

increases. So it is here. Think of the demand for the services of MasterCard as a function of total price, given any interchange fee. If the change in the interchange fee induces an outward shift in the demand curve then we expect volume as well as price to increase. This then makes it clear why the total price reaches its maximum at the interchange fee which maximises business volume.

We then can understand what the basic mistake of the competition authorities and the academic literature so far has been: they consider pass-on ratios on the issuing and the acquiring side as determined by the market structure of these two sides of the business, and thus treat these pass-on ratios as exogenously given. But the pass-on ratios are endogenous and depend on where we are in terms of the interchange fee. As long as a marginal rise in the interchange fee raises business demand for the system the pass-on ratio is higher on the acquiring side than on the issuing side, and thus the total price rises. If the interchange fee is too high to be volume maximising the pass-on ratio is higher on the issuing side than on the acquiring side, and a further rise in the interchange fee reduces price as well as volume. Only at the volume maximising interchange fee, which is also the price maximising interchange fee and thus the profit maximising interchange fee are the pass-on ratios the same on both sides of the business.

G The General Case of the Pass-On-Ratio

For issuer i we have a demand function which expresses the quantity demanded as a function of his own price, but also of the prices of his competitors inside and outside of the MasterCard system and of the prices of MasterCard acquirers and other acquirers. These other prices are influenced by the MasterCard interchange fee r; and in the following the influence of "other prices" is represented by the value of the MasterCard interchange fee. We then transform the direct demand function into the inverse demand function, which expresses issuer i`s price as a function of his quantity demanded and as a function of the interchange fee r, which represents the influence of "other prices" on the inverse demand function. Let then

$$p_i = f_i(x_i, r)$$

be the inverse demand function of issuer i.

The equation (*), indicative of the Bertrand-Nash equilibrium, then reads

$$f_i(x_i, r) - g_i + r = -x_i f_i'(x_i, r) \quad (*)$$

Here g_i is the marginal cost issuer i and $f_i'(x_i, r) = \frac{\partial f_i}{\partial x_i}$ is the partial derivative of the inverse demand function with respect to quantity demanded. In the "class-room model" of a linear demand function the latter has the constant value $-b$. Thus $f_i'(x_i, r)$ also is the inverse of the price sensitivity of demand.

Using this equation (*) we now can look at two different concepts of a pass-on ratio. First I discuss the individual cost pass-on ratio (ICPOR). Assume that only g_i changes, whereas all other prices and also the interchange fee remain the same. The individual cost-pass-on ratio ICPOR can be seen by differentiating equation (*) with respect to g_i. We then obtain

$$f_i'(x_i, r)\frac{dx_i}{dg_i} - 1 = \left[-f_i'(x_i, r) - x_i f_i''(x_i, r)\right]\frac{dx_i}{dg_i}$$

where $f_i''(x_i, r) = \frac{\partial^2 f}{\partial x_i^{\,2}}$ is the second partial derivative of the inverse demand function with respect to the quantity x_i. The individual-cost-pass-on-ratio ICPOR is

$$ICPOR = f'(x_i, r)\frac{dx_i}{dg_i}$$

And thus we obtain

$$ICPOR = 1 + \left[-f_i'(x_i, r) - x_i f_i''(x_i, r)\right]\frac{dx_i}{dg_i} \qquad (1)$$

$f_i'(x_i, r)$ is negative: the inverse demand function is negatively sloped. Thus ICPOR only can be positive, if $\frac{dx_i}{dg_i}$ is negative. So we assume this to be the case, for it is reasonable to assume that the individual pass-on ratio is positive: a higher marginal cost of production will raise the profit maximising price. We generally expect the individual-cost-pass-on-ratio, ICPOR, to be less than 100%, because costs and prices of competitors have not changed, so that the supplier under consideration cannot fully pass on his cost rise. So we assume this to be the case: $ICPOR < 1$. But then it follows that the expression in square brackets is positive.
Hence we have the inequality

$$-f_i'(x_i, r) - x_i f_i''(x_i, r) > 0 \qquad (2)$$

I now turn to the other pass-on ratio, which I call the interchange fee pass-on ratio IFPOR. To evaluate it we now differentiate equation (*) with respect to r.

$$f_i'(x_i,r)x_i'(r)+\frac{\partial f_i}{\partial r}+1=-x_i\left[f_i''(x_i,r)x_i'(r)+\frac{\partial f_i'}{\partial r}\right]-f_i'(x_i,r)x_i'(r)$$

The interchange fee pass-on-ratio IFPOR of issuer i is

$$IFPOR=-f_i'(x_i,r)x_i'(r)-\frac{\partial f_i}{\partial r}$$

Thus

$$IFPOR=1+\left[x_i f_i''(x_i,r)+f_i'(x_i,r)\right]x_i'(r)+x_i\frac{\partial f_i'(x_i,r)}{\partial r} \qquad (3)$$

Due to inequality (2) the square bracket in this expression (3) is negative. If it were not for the term $x_i\frac{\partial f_i'(x_i,r)}{\partial r}$ the sign of $IFPOR-1$ would be the same as in the "class-room model". It would be negative, as long as $x_i'(r)$ would be positive. It would be zero at the volume maximising interchange fee, and it would be positive, whenever $x_i'(r)$ would be negative.

In a similar way we can derive the IFPOR for any acquirer j of the MasterCard system. Let $\varphi_i(j,r)$ be the inverse demand function faced by acquirer j, conditioned on the interchange fee r. The IFPOR then is given by

$$IFPOR=1-y_i'(r)\left[y_j\varphi_i''(y_j,r)+\varphi_j'(y_j,r)\right]-y_j\frac{\partial\varphi_j'(y_j,r)}{\partial r}$$

In analogy to inequality (2) we also can consider the square brackets to be negative. Thus, if it were not for the term $y_j\frac{\partial\varphi_i'(y_j,r)}{\partial r}$ the sign of IFPOR would be the same as in the "class-room model". If $y_i'(r)>0$ then IFPOR-1 would be positive; if $y_i'(r)<0$ then IFPOR would be negative. Thus at $y_i'(r)=0$, IFPOR would be exactly equal to unity; this is the value of r which maximises volume $y_i(r)$.

In this general model it is no longer necessarily the case that the same r maximises business volume for the different issuers and acquirers. But it makes sense to ask which value of r maximises total MasterCard volume. For this purpose I introduce an issuer price index and an acquirer price index. Let x_i * be the business volume of issuer i at the interchange fee that

maximises MasterCard business volume. Let $x^* = \sum_{i=1}^{n} x_i^*$ be that maximum MasterCard business volume. I now define $\gamma_i = \frac{x_i^*}{x^*}$ to be the weight of issuer i in the issuer price index. The issuer price index then is $p(r) = \sum_{i=1}^{n} \gamma_i p_i(r)$

Similarly let y_j^* be the business volume acquirer j at the interchange fee that maximises MasterCard business volume. Let $y^* = x^*$ be that MasterCard business volume. I then define $\beta_j = \frac{y_j^*}{y^*}$ to be the weight of acquirer j in the acquirer price index. That price index then is $q(r) = \sum_{j=1}^{m} \beta_j q_j(r)$.

I then define the average interchange fee pass-on ratio on the issuing side as

$$IFPOR(I) = \sum_{i=1}^{n} \gamma_i IFPOR_i = 1 + \sum_{i=1}^{n} \gamma_i x_i'(r)\left[x_i f_i''(x_i, r) + f_i'(x_i, r)\right] + \sum_{i=1}^{n} \gamma_i x_i(r) \frac{\partial f_i'}{\partial r}$$

Similarly we define the interchange fee pass-on ratio on the acquiring side as

$$IFPOR(A) = \sum_{j=1}^{m} \beta_j IFPOR_j = 1 - \sum_{j=1}^{n} \beta_j y_j'(r)\left[y_j \varphi_j''(y_j, r) + \varphi_j'(y_j, r)\right] - \sum_{j=1}^{m} \beta_j y_j \frac{\partial \varphi_j'}{\partial r}$$

Let then t be the index for the total price. It is defined as $t = p + q$. Thus, we obtain for the rate of change of the total price (index)

$$\frac{dt}{dr} = IFPOR(A) - IFPOR(I) =$$
$$-\sum_{j=1}^{m} \beta_j y_j'(r)\left[y_j \varphi_j''(y_j, r) + \varphi_j'\right] - \sum_{j=1}^{m} \beta_j y_j \frac{\partial \varphi_j'}{\partial r} - \sum_{i=1}^{n} \gamma_i x'(r)_i \left[x_i f_i''(x_i, r) + f_i'\right] - \sum_{i=1}^{n} \gamma_i x_i \frac{\partial f_i'}{\partial r}$$

Note the following: the main impact of a change in the interchange fee on the inverse demand function of an issuer is a downward shift: competitor's prices have fallen, prices for acquirers, whose activity is a complement to the business of issuers, have risen. Thus, it is a reasonable approximation to assume that a change in r means a parallel downward shift of the issuer's inverse demand function. But this then means that at any given level of x_i the slope of the inverse demand function remains unaffected by a change in r. Thus we obtain

$$\sum_{i=1}^{n} \gamma_i x_i \frac{\partial f_i'}{\partial r} = 0 \text{ and similarly } \sum_{j=1}^{m} \beta_j y_j \frac{\partial \varphi_j'}{\partial r} = 0$$

In the case of a parallel shift of the demand curve and thus the inverse demand curve upon a change in the interchange fee we then obtain the equation

$$\frac{dt}{dr} = -\sum_{i=1}^{n} \gamma_i x_i'(r)\left[x_i f_i''(x_i, r) + f_i'\right] - \sum_{j=1}^{m} \beta_j y_j'(r)\left[y_j \varphi_j''(y_j, r + \varphi_j'\right]$$

It shows the following. As long as a rise in the interchange fee predominantly leads to rising business volumes the total price t rises with the rising interchange fee. If a rise in the interchange fee predominantly leads to a fall in business volumes the rise in the interchange fee induces a reduction in the total price. The total price thus reaches a maximum as a function of the interchange fee where the total sum of business volumes is near its maximum. This obviously implies that the total profit maximum is at an interchange fee, which also maximises volume.

H Conclusion

I have derived the result that it is in the profit interest of the members of a four party payment system to set the collectively determined interchange fee at a level which maximises volume. Thus, given the criterion that more competition leads to higher output in a market, there is no need for the anti-trust authorities to enforce a change of the freely determined collectively determined interchange fee.

References

Armstrong, Mark, Competition in Two-Sided-Markets, Rand Journal of Economics, 37, (2006), p. 668-691

Guthrie, Graeme and Julian Wright, Competing Payment Schemes, Journal of Industrial Economics, 55, 2007, p.37-67

Rochet, Jean Charles and Jean Tirole, (2006A) Two-Sided Markets: A Progress Report, Rand Journal of Economics, 37, (2006), p. 645-667.

Rochet, Jean Charles and Jean Tirole,(2006B) Externalities and Regulation in Card Payment Systems, Review of Network Economics, 5, 2006, p. 1-14