

# ■ Payment system efficiency and pro-competitive regulation

BY MATS A. BERGMAN

Associate professor, Department of Economics, Göteborg University and Uppsala University (mats.bergman@nek.uu.se)

*Effective competition guarantees that prices are set at such a level that efficiency is promoted (resulting in so-called allocative efficiency), but in markets characterised by large returns to scale and strong network effects, it may be too costly to have more than one firm operating. In such markets, known as natural monopolies, technical efficiency is best achieved by a single firm. More generally, a given market may be too small to support enough firms to guarantee effective competition. This creates a dilemma: if the market is monopolised, market power and monopoly pricing will result in allocative inefficiencies, while if several firms are active, there will be technical inefficiencies (or positive network effects will not be fully exploited). In many natural monopolies, regulators have sought to resolve this dilemma by introducing pro-competitive regulation. This article discusses the merits of introducing regulations in the payment-system markets.*

The article is based on a paper presented at a workshop on Central Bank Efficiency organized by Sveriges Riksbank on May 23–24, 2003.

I am grateful to Gabriela Guibourg for valuable suggestions and comments and to Sveriges Riksbank for financial support.

## Why is telecom infrastructure regulated, but not payment infrastructure?

The payment-system industry and the telecom industry are two examples of industries characterised by large returns to scale and strong network effects. In the telecom industry, the above dilemma has been resolved by regulation. But why do we see so little pro-competitive regulation of bank-owned payment systems? In particular, although Sveriges Riksbank is responsible for the efficiency of the payment system, this responsibility is not reflected in Swedish law. Would it not be possible to promote efficiency and to achieve consumer benefits by instigating a legal requirement that the owners of payment infrastructure must provide access to competitors and new entrants – just as the incumbent telecom operators are required to provide non-discriminatory access to their networks?

**Two examples of industries characterised by large returns to scale and strong network effects are the payment-system industry and the telecom industry.**

Here, three possible answers to these questions are considered. Firstly, competition in the provision of payment infrastructure services is arguably greater than in the provision of telecom infrastructure services. While telecom services typically have been national monopolies, payment services have been provided by a competitive industry. Secondly, payment infrastructure is often jointly owned by several banks, while important parts of the telecom infrastructure are still owned by the former telecom monopolies. Thirdly, it is possible that the general competition law is sufficient to guarantee fair competition in the payment-system market.

**Competition is a means of achieving efficiency – within a payment system, or in any market.**

Before going deeper into these three issues, however, it is necessary to consider why pro-competitive regulation of infrastructural services is sometimes required. Competition is a means of achieving efficiency – within a payment system, or in any market. Competition gives the producers incentives to be efficient and it ensures the buyers in the market a price that is not too high. If competition is lacking, production will often be less efficient and prices will tend to be higher. In industries where access to infrastructural services is important, a lack of competition in the provision of infrastructural services can propagate into subsequent stages of the production chain. This is known as the bottleneck problem.

**There is sometimes a trade-off between, on the one hand, cooperation and returns to scale, and, on the other, competition.**

However, as noted above, technical efficiency sometimes requires large-scale production and standardisation. Hence, there is sometimes a trade-off between, on the one hand, cooperation and returns to scale and, on the other, competition. In addition, a conflict can exist between the incentives for short-run and long-run competition. Maximum short-run competition sometimes comes at the expense of reduced incentives for long-run competition. This can also be phrased as a choice between competition *in* the market and competition *for* the market. Short-run competition (competition in the market) can for example be maximized by implementing favourable conditions for access to other parties' infrastructures or by denying patent rights. However, this reduces the incentives for investments and, hence, reduces long-run competition (competition for the market).

**Historically, there have often been competing providers of payment system services as well as competing infrastructures.**

Historically, there have often been competing providers of payment system services as well as competing infrastructures. The Swedish postal giro system (Postgirot), for instance, has competed with the banks' giro system (Bankgirot) and there are a number of competing international payment card systems, such as Visa, Mastercard and American Express. Presently, however, there appears to be a trend towards larger scale in payment systems and integration of the systems.<sup>1</sup> Some examples are the consolidation of ATM (Automated Teller Machine) networks, the integra-

<sup>1</sup> Guibourg (2001).

tion of previously non-compatible card payment systems and domestic and international integration of giro systems. Furthermore, there is likely to be a process of standardisation of internet-based payment systems.

Possibly, these trends will eventually result in national or even regional monopolies. This suggests that an analysis of the possible benefits of pro-competitive regulation of payment systems is warranted. Experience, as well as theory, tells us that the regulations should be in place *before* private firms undertake significant investments in new infrastructure and *before* government-owned infrastructure is privatised.

Furthermore, there have already been controversies over access to existing payment infrastructures. In many cases, the concerns have centred on the level of the interchange fee (see below for a discussion of this concept). This has been the case for card payments and for ATM transactions. There have also been controversies over the level of the fixed costs and the entry costs for small banks and new entrants in systems for card payment and ATM transactions, for national giro transactions and for clearing and settlement institutions.<sup>2</sup> There is a clear parallel, at least superficially, with the telecom industry, where many observers have pointed to high interconnection charges<sup>3</sup> as a cause of the lack of genuine competition.

**An analysis of the possible benefits of pro-competitive regulation of payment systems is warranted.**

**There have already been controversies over access to existing payment infrastructures.**

#### Giro payments, card payments and ATM transactions

Giro payments can be made without the payer and the payee meeting, but the payer needs to know the payee's giro number. A card payment is typically made in situations when the payer and the payee meet in person, but it can also be made at a distance, if the payee knows the payer's card number. In both cases, the payer's bank makes a withdrawal from the payer's account and sends information to the payee's bank, which, in turn, credits the payee's account. Both types of transaction require an agreement between the banks on a standard for the exchange of information. In principle, each bank could communicate directly with all other banks. In practice, central clearing houses typically process the transaction information and, i.e., calculate the net amounts that each bank owes every other bank. The net amounts are settled once or a few times every day. Settlements are made by transferring the net amounts between the banks' accounts in the central bank. The clearing houses and other critical assets, such as trade marks, are often jointly owned by the banks. If it is not regulated, the access price will be determined by the owners of the systems (i.e., the owners of the clearing houses and the trade marks).

<sup>2</sup> See, e.g., the EU Commission's press release IP/01/462, March 31, "Commission Raises Competition Concerns about Behaviour of Clearstream Banking AG". Clearstream is the German clearing and settlement institution. Preliminary, it has been found to have abused its dominant position by discriminating Euroclear.

<sup>3</sup> The fee a telecom operator charges other operators for terminating or originating a call in its own network.

ATM (Automated Teller Machine) transactions have much in common with card payments, but instead of crediting a merchant's account, the money is made available in cash. After a cash withdrawal, the cardholder's bank makes a withdrawal from the account that corresponds to the card and pays to the ATM-owning bank, through a clearing institution. In "foreign" ATM transactions, the banks' customers use other banks' terminals; typically the cardholder's bank has to compensate the bank that owns the terminal (in addition to transferring the amount withdrawn).

Clearing and settlement institutions, e.g., VPC in Sweden, are also used for financial assets, such as equities and fixed-income instruments. Trade in electronically registered financial assets results in two transactions that must be cleared and settled: one for the payment and one for the asset that change hands. While the (net) payments are eventually settled in the central bank, the transfer of asset ownership is typically both cleared and settled in a clearing and settlement institution dedicated to financial assets.

The next section discusses scale economies and competition as sources of efficiency in payment services and in general. This is followed by an introduction of the bottleneck problem pertinent to infrastructural industries, as well as the traditional solutions to this problem. After that I focus on possible anti-competitive concerns in payment services, and continue by discussing the role of the general competition law in this industry.

The tentative conclusion of this article is that there *is* a role for pro-competitive (access) regulation of payment services, just as for telecom services. However, there are important differences between the telecom and the payment industries. A regulatory framework that is appropriate for the former cannot be expected to be appropriate for the latter without significant modifications. Such modifications, in turn, require a careful analysis of the particularities of the payment-system industry – and its interaction with the banking industry at large.

**The tentative conclusion of this article is that there is a role for pro-competitive regulation of payment services, just as for telecom services.**

## Scale economies versus benefit of competition

When there are increasing returns to scale, one sometimes has to compromise between two means for achieving efficiency: large-scale production and competition. This is likely to be true for payment services, as there is strong evidence of this industry being characterised by significant returns to scale, both from the supply (or cost) side and from the demand side. The latter is due to the network effects that exist in payment services. Furthermore, as discussed above, there are indications that the scale of operations is increasing in payment systems, suggesting that the balance of benefits from competition and from large scale may be shifting.

**When there are increasing returns to scale, one sometimes has to compromise between two means for achieving efficiency: large-scale production and competition.**

## THE BENEFIT OF SCALE

There are a number of reasons why there should, in general, be increasing returns to scale in production. Some fixed costs (e.g., management and R&D) may not need to follow the scale of production. Increased scale may allow a shift towards more efficient technology (typically a more automated technology, with relatively higher fixed costs and lower variable costs).<sup>4</sup> A higher level of production will allow employees to become more specialised and will allow individuals and firms to move along the so-called learning curve. Finally, so-called economies of massed reserves will allow firms to economise on production equipment, as random breakdowns or idiosyncratic fluctuations in demand and supply will have less impact. However, these sources of scale economies will eventually peter out, and diseconomies will set in, such as increasing managerial costs due to the complexity of the operation, agency problems and, in many industries, transportation costs.<sup>5</sup>

**There are a number of reasons why there should be increasing returns to scale in production.**

Most studies have found that returns to scale in general banking are relatively modest. Wheelock & Wilson (2001), for example, in a study on US banks, found significant returns to scale only for banks with total assets below US \$ 300–500 million (approximately equivalent to the assets of some new entrant in the Swedish banking market, such as IKANO bank and ICA Banken, or a medium-sized savings bank, such as Sparbanken Skaraborg). In addition, statistically insignificant point estimates suggested that positive returns to scale existed up to perhaps US \$1 billion in assets.<sup>6</sup> Lindquist (2003) criticises the methodology used in most previous studies and finds positive returns to scale also for large banks, in particular for deposits. She also finds that returns to scale increase when electronic payment services increase.

**Most studies have found that returns to scale in general banking are relatively modest.**

Returns to scale appear to be stronger in payment systems than in other banking services. Humphrey et al. (2003) find scale economies of 0.2 in a European cross-country study, implying that costs increase by 2 per cent when volumes rise 10 per cent. Similarly, although not directly relevant for Sweden, Bauer reports average scale economies of 0.7–0.8 for cheque-processing offices, which, however, appear to be exhausted for the largest US offices.<sup>7</sup> Bauer & Ferrier (1996) report scale economies of approximately 0.5 for automated clearing houses. In contrast, Felgran (1986) found that scale economies were insignificant for ATM networks with more than 1,000 service points.

**Returns to scale appear to be stronger in payment systems than in other banking services.**

In conclusion, there appear to be strong economies of scale in the

<sup>4</sup> Cf. the so-called two-thirds rule, shown to apply for many chemical and metallurgical processes.

<sup>5</sup> See Tirole (1988) section 1.2 and Scherer & Ross (1990) chapter 4.

<sup>6</sup> See Wheelock & Wilson (2001) for further references.

<sup>7</sup> Quoted from Bergendahl et al. (2002).

production of payment services, suggesting that cooperation between banks should be allowed.

#### NETWORK BENEFITS

In addition to returns to scale from the supply (or cost) side, payment systems are characterised by returns to scale from the demand side.

In addition to returns to scale from the supply (or cost) side, payment systems are characterised by returns to scale from the demand side. These are referred to as network effects.<sup>8</sup> Network effects represent a special case of positive externalities. When an additional user connects to the network, this increases the utility for other connected users. The network effects can be direct, as in telephone networks or in giro-payment systems and other types of ACHs (Automated Clearing Houses).<sup>9</sup> They can also be indirect, as in payment-card systems and in the markets for computer software and hardware. If more consumers choose a particular payment card (e.g., Visa or a certain ATM card) or a particular type of office computer, there will be an incentive for more merchants to accept the payment card, more ATMs will be installed or there will be a larger supply of compatible software, respectively.<sup>10</sup>

In network markets, competition between networks must be distinguished from competition within systems.

In network markets, competition *between* networks must be distinguished from competition *within* systems. Examples of *inter*-network competition (competition between firms using different networks) are PC computers versus Apple computers and American Express credit cards versus Visa versus Mastercard. Examples of *intra*-network competition (competition between firms that use the same network) are competition between various PC producers; competition between banks offering to process Visa transactions for merchants; and competition between commercial banks that all provide giro solutions based on the Bankgiro. Note also that the distinction between intra-network competition and inter-network competition is not absolute, as there is often a degree of compatibility even between supposedly non-compatible systems. For example, the same EFTPOS (Electronic Funds Transfer Point Of Sale) terminals, i.e., terminals used for card payments in ordinary stores, can be used for cards from several networks. Similarly, customers can make giro payments from a bank account through the Bankgiro to a Postgiro account.

Guibourg (2001) surveys the existing literature on network effects in ATM and ACH markets and provides an empirical study of network effects in the EFTPOS market. The general conclusion is that strong net-

<sup>8</sup> See also Guibourg (1998) for a general discussion of network effects.

<sup>9</sup> Electronic interbank networks used to process transactions.

<sup>10</sup> With a higher level of demand, producers can benefit from returns to scale in production, offer more variety and offer higher service levels or a denser distribution network. A prerequisite for this to occur is that consumers choose compatible products.

work effects and large returns to scale dominate competition effects, in the sense that adoption rates are higher in markets with few competing networks – in practice often a single network – although there is normally *intra*-network competition. Guibourg reports that in 1999, the number of transactions per capita was ten times higher in countries with a single compatible EFTPOS network than in countries with two or more incompatible networks. In addition, growth in the number of transactions per capita increased dramatically in countries where incompatible systems merged into a single compatible system. There appears to be less support for the existence of strong economies of scale in production, as the average number of proprietary (although often compatible) systems increased from three in 1988 to over seven in 1999. In addition, the three smallest countries in the sample are among the four countries with the highest number of transactions per capita, while the three largest countries are among the four with the lowest number of transactions per capita.

The above discussion suggests that network effects may be an even more important motive for allowing cooperation between banks in payment systems.

#### THE BENEFIT OF COMPETITION

Just as fundamental as economies of scale, are the benefits of competition. When competition is lacking, one or a few firms will possess market power, which has four main adverse consequences. Firstly, it will transfer welfare from consumers to producers.<sup>11</sup> Secondly, as the price rises above the competitive level, demand will fall below the optimal level – i.e., there will be allocative inefficiencies. Thirdly, low competitive pressure is generally believed to result in sub-optimal effort and X-inefficiencies (i.e., weak cost control will result in unduly high costs). Fourthly, the existence of a monopoly profit may trigger socially costly lobbying for the favoured position, as well as other types of rent-seeking behaviour. Although regulation can mitigate problems of the first and second type, there is a substantial risk that it will not properly address problems of the third and fourth types. Moreover, regulation introduces new problems, such as regulatory risks (the risk that investment incentives et cetera will be reduced because the regulator may be tempted to exploit the regulated firm once the latter has taken the investment cost) and the direct costs of regulation.<sup>12</sup>

**The general conclusion is that strong network effects and large returns to scale dominate competition effects.**

**When competition is lacking, one or a few firms will possess market power, which has four main adverse consequences.**

<sup>11</sup> Although, strictly speaking, this will only reduce welfare if consumer surplus is valued more highly than producer profit, welfare transfers from consumers to producers are normally considered to be negative.

<sup>12</sup> See Bergman (2002).

**Firms have a strategic interest in overstating economies of scale and downplaying the benefits of competition.**

Firms have a strategic interest in overstating economies of scale and downplaying the benefits of competition. This is because having fewer competitors is typically beneficial for the industry and negative for consumers, while economies of scale will tend to benefit both categories. Hence, it is tempting to appeal to economies of scale even in situations where the true rationale is a desire to reduce competition.

Some observers argue that the introduction of competition normally gives rise to cost savings and price reductions in the 25–75 per cent range (Winston (1998) and a number of OECD studies, referred to in Gonce & Nicoletti (2000)). However, based on an extensive review of the empirical literature on deregulation, Bergman (2002) concludes that savings in the 5–10 per cent range are more realistic.<sup>13</sup> He draws a similar conclusion in the section where studies of banking deregulation are surveyed.

## CONCLUSIONS

**Possibly, returns to scale from the demand side are more important than those from the supply side; thus standardisation and consolidation of payment systems should be encouraged.**

It appears that returns to scale are important in the payment industry. Possibly, returns to scale from the demand side (network effects) are more important than returns to scale from the supply side (traditional scale economies). This suggests that standardisation and consolidation of payment systems should be encouraged. On the other hand, we are not eager to relinquish the benefits of competition. Possibly, a policy that encourages consolidation and standardisation of payment systems and simultaneously stimulates competition between service providers, is an optimal compromise. Note, however, that such a development would make the payment industry (and the banking industry at large) more similar to industries such as telecom and electricity, relative to what has previously been the case. This, in turn, suggests that bottleneck problems in the payment and retail banking industry may become a more pressing concern in the future. The bottleneck problem will be discussed more thoroughly in the next section.

Returning to the motivating questions, raised in the beginning of this article, and the first of the proposed answers, it appears that although there are grounds for claiming that up to now competition in the provision of infrastructural services has been stronger for payments than for telecom, this will not necessarily be true in the future. While a trend towards consolidation in payment infrastructure may be discernable, it is clearly the case that competition in the provision of telecom infrastructure

<sup>13</sup> The large effect of deregulation found in some studies appears to stem from two types of shortcomings in the empirical research design. One is that a falling trend in costs and prices that existed even before the deregulation is not accounted for (e.g., in telecom and rail freight). The other is that short observation periods are used to estimate changes in the rate of productivity growth – and that these estimates are used to extrapolate deregulation gains far beyond the period of observation.



is becoming more intense, as the number of parallel mobile telephony networks increases.

## The bottleneck problem – and possible remedies

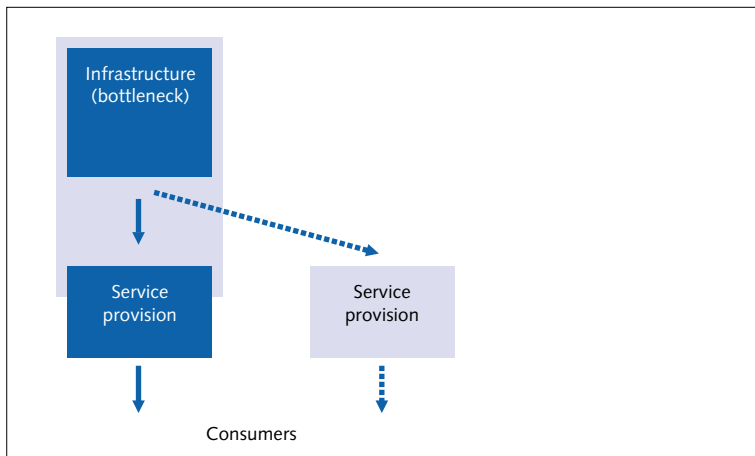
Many industries, in particular network industries, have a vertical production structure. In one (or a few) stages, competition would be non-viable – or at least inefficient. In other stages, competition *is* viable. For example, although it would be inefficient to build more than one electricity network, it is entirely reasonable to allow numerous firms to generate electricity.<sup>14</sup> Various regulatory methods are used to exploit returns to scale in the network stage, while preventing the network owner from exploiting its market power in the potentially competitive stage. Before reviewing these methods, however, the bottleneck problem as such will be addressed.

### THE FUNDAMENTAL BOTTLENECK PROBLEM

The typical situation in a bottleneck industry (or network industry) is illustrated in Figure 1. To emphasise the point, it is assumed that only one firm can be active in the upstream infrastructure market, the bottleneck, while several firms can be active in the downstream market, the market for service provision. For example, the main activity in the upstream market can be to establish and maintain a local telecom network or a payment system. Here, these activities are assumed to be natural monopolies. The

**Only one firm can be active in the upstream infrastructure market – the bottleneck – while several firms can be active in the downstream market – the market for service provision.**

Figure 1. The bottleneck problem



<sup>14</sup> The electricity industry is an example of an industry with two distinct vertical stages where competition is non-viable: (local) distribution and (long-distance) transmission of electric energy. At the same time, electricity generation and retail electricity sales are (potentially) competitive.

downstream market can then be the telecom services market or retail banking, respectively.

Control over the bottleneck stage will be associated with market power in that stage of production, which, in itself, gives rise to negative consequences of the type discussed above. In addition, control over the bottleneck can give rise to market power and lack of competition also in the potentially competitive downstream market.

**The non-viability of competition in the bottleneck stage can result from economies of scale in production or in consumption.**

The non-viability of competition in the bottleneck stage can be a result of economies of scale in production (which is typically the case for physical networks, like telecom, electricity and rail) or in consumption, i.e., network effects. In the latter case, network effects may be so strong that “tipping” will occur: as soon as one network gains an upper hand, all users will find it optimal to adopt the technology or products associated with that network.<sup>15</sup> Examples of such tipping are the processes that led the video system VHS to dominate over BetaMax and almost has made Microsoft the monopoly provider of operating systems for personal computers.

**The firm that controls the bottleneck is in a good position to extract the industry-monopoly profit.**

Although duplication of the bottleneck is not desirable because large scale economies or strong network effects, the creation of a monopoly will, if unchecked, normally result in welfare losses due to market power. The firm that controls the bottleneck is in a good position to extract the industry-monopoly profit. This can be achieved either by completely excluding competitors from access to the essential infrastructure, or by charging such a high price for infrastructural services that the monopoly profit for the industry as a whole accrues in the infrastructural stage alone.

In order to overcome these problems, various policies have been used in network industries. These policies can be divided into two broad categories – regulatory measures and structural measures – which can be employed in combination or separately.<sup>16</sup>

## REGULATORY MEASURES

**The traditional Swedish and European response to the bottleneck problem has been government ownership.**

The traditional Swedish and European response to the bottleneck problem has been government ownership. Postal services, telecom and rail services, for example, have been national monopolies in most EU countries. In banking, this method has been used less frequently. Note, however, that central banks perform some functions that can be seen as bottleneck services. One example is the settlement function, which enables banks to

<sup>15</sup> Besen & Farrell (1994).

<sup>16</sup> Bergman (2002, 2003).

make net payments to one another.<sup>17</sup> Another example is control of the physical supply of notes and coins and the right to seigniorage.<sup>18</sup>

In the US, the preferred response to the bottleneck problem used to be regulation of consumer prices. In contrast to the situation in Europe, production was mainly in the hands of private firms. This method has been used for some industries also in Europe – among them the taxi market. The interest rate regulation previously used in banking was also a form of consumer-price regulation, although it was not introduced in response to a bottleneck problem.

Instead of regulating consumer prices, it is possible to regulate the bottleneck price only (the price of the infrastructural service), i.e., access (price) regulation. If competition is viable in the non-bottleneck stages of production, access regulation should suffice to ensure effective competition.

Internationally and across many industries, there has been a move towards “deregulation”. Typically, market entry has been liberalised. In the US, firms have been given greater freedom over prices, as the consumer-price regulations have been lifted or eased. In Europe, however, price regulations have often been *introduced* during deregulation. This is because the government monopolies were often unregulated in a formal sense. In the early deregulations in Britain, privatisation was often accompanied by the introduction of consumer-price regulations. Later on and in other countries, there has been a tendency towards increased reliance on access regulation. This tendency is particularly pronounced in telecom. In banking, on the other hand, deregulation of entry and of consumer prices has seldom been followed by access regulation.

This general pattern is reflected in Sweden, where Sveriges Riksbank is responsible for system efficiency (as well as stability) in the financial markets, but there is no sector-specific regulation that corresponds to this responsibility. Instead, the Riksbank relies on moral suasion and on its ability to influence the legislator and the Swedish Financial Supervisory Authority, which is responsible for the soundness of individual institution. This is in contrast to the situation in, e.g., Australia and Norway, where the central banks have more explicit regulatory powers (Andersson et al. (2001)).

The central banks of the developed countries, including the Riksbank, have agreed on a set of guidelines that aim to foster safety and efficiency in payment systems, the so-called Core Principles for Systemically

**In the US, the preferred response to the bottleneck problem used to be regulation of consumer prices.**

**Internationally and across many industries, there has been a move towards “deregulation”.**

**In Sweden, Sveriges Riksbank is responsible for system efficiency in the financial markets but there is no sector-specific regulation that corresponds to this responsibility.**

<sup>17</sup> The Swedish settlement system is called RIX.

<sup>18</sup> Sveriges Riksbank handles notes and coins through Pengar i Sverige AB, a wholly-owned subsidiary.

Important Payment Systems.<sup>19</sup> In particular, Principle IX calls for fair and non-discriminatory access to such systems. However, unless they are enacted by a national legislator, these recommendations have no formal legal status.

## STRUCTURAL MEASURES

**Vertical and horizontal separations have been proposed as structural measures for addressing the bottleneck problem.**

**Vertical separation disconnects ownership of the bottleneck from ownership of downstream activities.**

Vertical and horizontal separations have been proposed as structural measures for addressing the bottleneck problem. Vertical separation disconnects ownership of the bottleneck from ownership of downstream (competitive) activities. The advantage is that the owner will have no incentives to favour one of the downstream operators. In contrast, an entity that is vertically integrated and faces competition downstream will often have an incentive to provide inferior infrastructural services, or price the services excessively high. The disadvantages of vertical separation are that vertical synergies cannot be fully exploited and that the problem of market power is not resolved. The latter implies that vertical separation must typically be combined with price regulation or government ownership.

**Horizontal separation means that one large bottleneck monopoly is divided into several smaller bottleneck monopolies.**

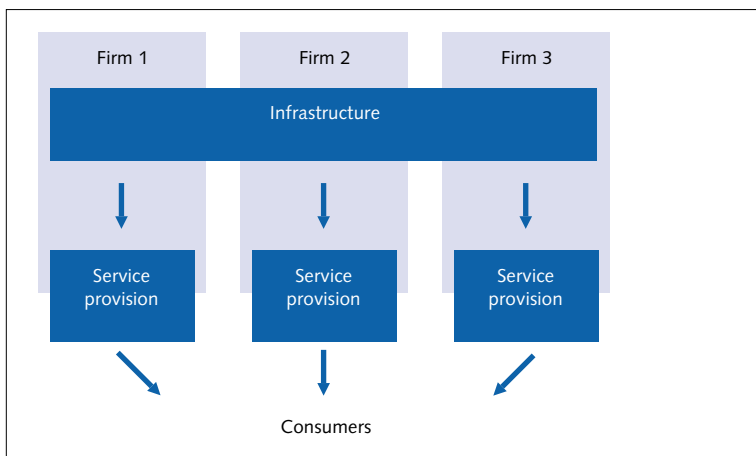
Horizontal separation means that one large bottleneck monopoly is divided into several smaller bottleneck monopolies. Typically, these small monopolies operate in distinct geographical areas. A horizontal separation will not in itself resolve the bottleneck problem. However, dividing one national bottleneck between several regional bottleneck operators can provide more information and may allow benchmarking between the operators. Hence, horizontal separation, as well as vertical separation, may facilitate regulation.

**In banking infrastructural clubs appear to be the preferred institution.**

In banking, neither vertical nor horizontal separation has been used very often (again, the central banks and their functions are the main exceptions). Instead, a third structural measure – infrastructural clubs – appears to be the preferred institution. An infrastructural club is an arrangement wherein firms that compete horizontally own the essential infrastructure jointly (see Figure 2). Examples are the Visa and Mastercard systems, national giro systems like Bankgirot, and ATM networks. Examples from other industries are the airlines' computer reservation systems (CRSs) for air tickets, taxi switches and, sometimes, joint ownership of mobile-telephony infrastructure.

<sup>19</sup> BIS (2001). Two core-principles evaluations of Swedish payment systems are presently available at the Riksbank's website, [www.riksbank.se](http://www.riksbank.se), under "Financial stability" and "Payment systems".

Figure 2. Infrastructural clubs



Infrastructural clubs provide some prospects of self-regulation. The owners/customers have a common interest in holding costs and prices down, while the firms may have strong incentives to compete for customers in the downstream market (e.g., retail-banking services). On the other hand, the common ownership of the infrastructure can conceivably be used to coordinate pricing in the downstream market and there is a risk that large (incumbent) firms will not allow small (entrant) firms to join the clubs. In addition, conflicting views between the owners may increase transaction costs, which in turn may reduce efficiency and give rise to excessive inertia (lack of innovations). The possible anti-competitive consequences of infrastructural clubs will be explored in the next section.

**Infrastructural clubs provide some prospects of self-regulation.**

## CONCLUSIONS

The second proposed reason, mentioned in the introduction, for regulating the banking and the payment industries less strictly than the telecom industry is that infrastructural clubs are common in the former, but less common in the latter. The above discussion suggests that this argument has some merits. However, joint ownership of a single infrastructure does not resolve all problems. Potentially, the bottleneck structure of most payment systems may still have anti-competitive consequences.

## Anti-competitive concerns in payment systems

Four different types of anti-competitive concerns may be relevant for infrastructural clubs and for jointly-owned payment systems.

Above, it was argued that infrastructural clubs to some extent can be expected to achieve efficiency without regulatory intervention, but it was also pointed out that self-regulation will not eliminate all the risks of anti-competitive effects. This section will discuss four different types of anti-competitive concerns (or mechanisms) that may be relevant for infrastructural clubs in general and for jointly-owned payment systems in particular. The four mechanisms are:

- Coordination of downstream behaviour through a high vertical fee for infrastructural services. Monopoly profits can be generated in the bottleneck and redistributed to the owners.
- Coordination of downstream behaviour through high horizontal (multilateral) fees. This can generate monopoly profits in potentially competitive market segments (service provision).
- Discrimination against small and joining members of the infrastructural clubs. This may be in the interest of large firms that are in control of the clubs, but will be detrimental to efficiency.
- Facilitation of collusion through joint operation of assets. Information sharing and the creation of legal “links” (jointly-owned assets) between the firms may align their interests.

### COORDINATION THROUGH VERTICAL FEES

Firms that jointly own an infrastructural club can use the club's fee structure strategically, in order to achieve super-competitive profits.

A number of firms that jointly own an infrastructural club can use the club's fee structure strategically, in order to achieve super-competitive profits. In particular, if the owners agree to pay monopoly prices for the services provided by the club, the outcome will be the same as if the infrastructure were controlled by an independent monopolist. This is so, since the owners, who relies on the services provided by the club, would experience higher marginal costs while a monopoly profit would accumulate in the club. This profit could subsequently be distributed to the owners. However, if the profit is distributed in proportion to each owner's production, the firms will realise that their true marginal cost of using the infrastructure is not the nominal fee, but the nominal fee minus the average profit margin. Hence, if the firms wish to use this mechanism to restrict competition, the profit that accrues in the club must be distributed between the firms in fixed proportions.<sup>20</sup>

<sup>20</sup> An analogous effect has been identified in the context of patent pools; see Priest (1977).

As an example, assume that the fees paid to Bankgirot were to be set at such an elevated level that the banks would earn no profit in the downstream market. This would, potentially, give Bankgirot a handsome profit, which could then be divided between the owners, i.e., the banks themselves. In effect, this would give the banks a mechanism that allows them to earn monopoly profits, even though they compete in the downstream market.

In practice, Bankgirot's fee structure follows the cost-plus principle and it is unlikely that the fees could be raised to the monopoly level, for a number of reasons. A drastic price increase, to the benefit of the owner banks, would probably violate competition law (see below). In addition, individual banks are not solely dependent on Bankgirot for giro services and they have some ability to shift the consumers' demand towards other payment solutions. Hence, an individual bank would have incentives to buy fewer services from Bankgirot, than the collective of owner banks would prefer. In fact, although the possibility cannot be ruled out, there are no strong indications that huge profits are generated within the banking market's infrastructural clubs as such.

**There are no strong indications that huge profits are generated within the banking market's infrastructural clubs as such.**

#### DISCRIMINATION AND THE RISK OF COLLUSION

The main concern in Sweden has been that fees are set in such a way that new entrants and small firms are discriminated against. This could, for example, be achieved by giving steep discounts to banks that use large quantities of the infrastructural services, by having new banks pay high entry fees or by having all banks pay high annual fixed fees.

**However, there have been concerns that fees may be set so that new entrant and other small firms are discriminated.**

Joint ownership of infrastructure necessitates contacts between competing banks. It is often believed that if such contacts become too intimate, there is a risk of this facilitating collusion.

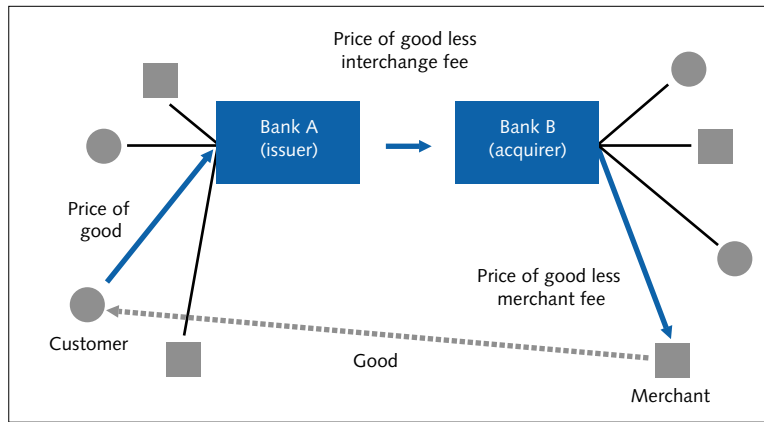
These two mechanisms are not explored further here, but we will return to the issue of discrimination in the next section, when Swedish applications of competition law to payment systems are discussed.

#### MULTILATERAL INTERCHANGE FEES

At the international level, the competition authorities' main concerns with regard to payment systems have centred on the level and structure of the multilateral interchange fees. The concern has been that *horizontal fees* between banks that compete in providing payment services have been fixed at unduly high levels. Such horizontal fees are known as interchange fees. Figure 3 illustrates the role of these fees in the context of a card payment. (Interchange fees are also used for, e.g., ATM transactions and auto-giro payments.)

**Internationally, the main concerns have centred on the level and structure of the multilateral interchange fees.**

Figure 3. Interchange fees



In the figure, the holder of a card from Bank A buys from a merchant whose card transactions are acquired (collected and processed) by Bank B. Thus, for this transaction, Bank A is the “issuer” and Bank B is the “acquirer”. The issuer deducts an amount corresponding to the purchase price from the cardholder’s account (assuming a debit-card transaction). This amount, less the interchange fee, is transferred from the issuer to the acquirer. The acquirer, in turn, transfers the purchase price less the merchant fee to the merchant’s account. These payments are illustrated by the bold unbroken arrows.

In most countries, the bulk of the cost of card payment systems is borne by the merchants, through the merchant fee paid to the acquiring bank. The merchant fee is paid in the opposite direction to the right-most arrow in Figure 3. In practice, the fee is not paid separately. Instead, a fraction of the price of the good sold by the merchant is withheld by the acquiring bank. The interchange fee redistributes parts of the merchant-fee payment to the issuing bank. Again, the fee is paid in the opposite direction to one of the bold arrows in the figure – in this case the arrow between the two banks. The issuer charges the customer the full price of the good, withholds a fraction of that price and pays the rest to the acquirer (which, in turn, pays the merchant).

Often, banks are both issuers and acquirers. Depending on which customer buys from which merchant, either of the banks can be the acquirer, the issuer, both acquirer and issuer – or not involved at all.

The merchant fees are set by the banks individually, but the interchange fee is determined multilaterally. Since the interchange fee constitutes a relatively large fraction of the marginal cost of an acquisition – approximately 80 per cent (see below) – the level of the interchange fee influences the level of the merchant fee. It follows that the interchange

**The merchant fees are set by the banks individually, but the interchange fee is determined multilaterally.**



fee can be used to coordinate the behaviour of acquirers, by giving the banks incentives to set higher merchant fees. The result is similar to what can be achieved by setting a high *vertical fee* for infrastructural services, as discussed above. Furthermore, there is no need to think about how the resulting profits should be redistributed, since the payments go directly to other banks.

It may be thought that a multilateral interchange fee is an effective instrument for coordination and for achieving super-competitive profits. However, a high interchange fee will create incentives to compete for incoming fees, i.e., to compete for cardholding customers. This competition is likely to result in lower annual fees and lower per-transaction fees charged to cardholders than would otherwise be the case.<sup>21</sup> In other words, the super-normal profit that results from higher merchant fees will be competed away on the issuing side of the market.

In an interesting recent paper, Rochet & Tirole (2002) analyse the welfare consequences of a multilateral interchange fee (a MIF). The basic intuition of their model is the one presented above: even if the MIF can be used to achieve super-competitive profits on one side of the market, (most of) this profit will be competed away on the other side. However, this phenomenon does not guarantee that a MIF is innocuous – even if the banks do not earn higher-than-normal profits, it may still be the case that the MIF distorts the pricing away from the socially optimal price structure and, hence, reduces welfare.

The most important result of Rochet & Tirole's paper is that in a two-sided network market, such as the payment-card market, it is in general *not* optimal to set prices on each side of the market equal to the marginal costs incurred on that side. On the contrary, it is normally optimal to let one side of the market subsidize the other side, in order to reap the full benefits of positive network externalities. However, the authors also show that the banks *may* set the MIF too high, if the authorities do not impose a ceiling.<sup>22</sup> The EU Commission's policy vis-à-vis MIFs will be discussed below.

**Even if the MIF can be used to achieve super-competitive profits on one side of the market, this profit will be competed away on the other side of the market.**

**In a two-sided network market, such as the payment-card market, it is in general not optimal to set prices on each side of the market equal to the marginal costs incurred on that side.**

<sup>21</sup> It follows that high multilateral interchange fees will be particularly attractive for firms that cannot compete for customers that generate incoming fees. Assuming that the firms only sell to domestic customers, this will be the case for international agreements on multilateral interchange fees (or corresponding fees in other industries, which go under other names). Examples are (fees corresponding to) interchange fees for international payments, international telephone calls and international mail delivery. For the latter, see the EU Commission's cases concerning the REIMS I and II agreements. The Commission's cases concerning international card payments will be discussed below.

<sup>22</sup> Similar results have been shown to hold in the context of telecom infrastructure. High interconnection fees (corresponding to interchange fees) raise the marginal cost of providing a telephone call, which in turn gives the operators incentives to increase the per-minute price of phone calls. On the other hand, high interconnection fees raise the profit of the terminating operator, giving them incentives to compete for customers who receive incoming calls. For example, the operators can set low monthly fees, subsidise handsets or offer subsidies for receiving incoming calls. See Laffont et al. (1988a, b) and Laffont & Tirole (2000). See Rochet & Tirole (2003) for a general analysis of two-sided markets. Schmalensee (2002) provides a complementary analysis of card-payment systems.

## The role of competition law in payment-system markets

**In most jurisdictions competition law is based on three main prohibitions.**

In most jurisdictions, including those of the EU, USA, Sweden and the other Nordic countries, competition law is based on three main prohibitions. Firstly, agreements between firms that reduce competition (notably cartels) are prohibited. Secondly, firms are not allowed to merge, if the merger will confer sufficient market power to significantly impede competition. Thirdly, a “dominant” firm (a firm with significant market power) is not allowed to abuse its dominant position.<sup>23</sup>

**The first two can be seen as a system for weighing the benefits of competition against the benefits of large scale and of cooperation.**

The first two of these prohibitions can be seen as a system for weighing (and protecting) the benefits of competition against the benefits of large scale and of cooperation. Firms are often allowed to cooperate, but are not allowed to enter into agreements that restrain competition unreasonably. In principle, the benefits of cooperation are weighed against cooperation’s anti-competitive effects, although the weighing is not always done on a case-by-case basis.<sup>24</sup> Similarly, firms can merge unhindered by the law – unless the post-merger market share is too large – as low-market-share mergers are presumed to be beneficial, or at least harmless.

**The third limits the freedom of large firms in some respects, in order to preserve competition on the market or in order to protect the consumers.**

The third main prohibition of competition law, that against abuse of dominance, limits the freedom of large firms in some respects, in order to preserve competition on the market or in order to protect the consumers. For example, dominant firms are not allowed to discriminate other firms and there are restrictions on their pricing and on their freedom to use rebate schemes. A special aspect of this prohibition is the “essential facilities doctrine”. Under this doctrine, a dominant firm may sometimes be required to provide access to its infrastructure (or other “essential” facilities) at non-discriminatory prices.<sup>25</sup>

**Sweden has voluntarily chosen to make the domestic competition law almost identical to EU’s rules.**

EU’s competition rules are applicable within the member states in the event of an appreciable effect, actual or potential, on cross-border trade. This means that domestic and EU-wide competition rules are often applicable simultaneously. Sweden has voluntarily chosen to make the domestic competition law almost identical to EU’s rules. In addition, European legal practice has confirmed that the member states’ courts and competition authorities cannot take decisions that directly conflict with EU’s competition rules. As of 1 May 2004, the member states’ freedom to uphold

<sup>23</sup> The closest US correspondence is “monopolization”.

<sup>24</sup> To simplify application of the law, there are “white lists”, “black lists” and “group exemptions”. In the US, black-listed behaviour is said to be prohibited “per se”.

<sup>25</sup> See Guibourg (1998) and the Swedish Competition Authority (1999) for analyses of this doctrine, with applications to payment systems.

national competition rules that are not aligned with EU's rules will be further reduced.<sup>26</sup>

In Sweden and elsewhere, competition law and its prohibition against anti-competitive agreements have been applied to payment systems. Within the financial industry, the merger regulation has most often been applied to bank mergers, rather than to mergers directly between firms providing infrastructural services.<sup>27</sup> In the following, however, I will focus on cases that have concerned the prohibition against anti-competitive agreements.

#### SWEDISH CASE LAW

In a sequence of cases, the Swedish Competition Authority evaluated the banks' cooperations concerning infrastructure for payment cards, ATMs and giro payments.<sup>28</sup> The Authority's analysis in these cases focused on the possible discrimination of small or entrant banks. There was a concern that the payment systems employed pricing schemes that were handicapping the smaller players in the market. For example, relatively high discounts were given to banks with a large number of transactions per year. In other instances, banks with few transactions had to pay surcharges; part of the payment systems' costs were covered through fixed annual fees, there were entry fees and in a series of cases concerning the jointly owned credit-information agency *Upplysningscentralen*, there was mention of the possibility that large profits might be accumulated and subsequently distributed to the owners of *Upplysningscentralen*.<sup>29</sup> Effectively, this gave non-owners a higher price.

One specific case dealt with the CEKAB, a jointly-owned processor of electronic card-based payment transactions; ATM transactions as well as EFTPOS transactions.<sup>30</sup> CEKAB is 97 per cent owned by three of the four largest Swedish banks. The Competition Authority maintained that CEKAB's fee structure was discriminatory vis-à-vis smaller banks, which to a large extent were dependent on it. However, the Market Court found that on three grounds, the fee structure was acceptable.<sup>31</sup> Firstly, the fees were cost based. Secondly, the fees did not have an appreciable effect.

**Systems for card payments, ATM transactions and giro transactions have been evaluated under the Swedish competition law.**

<sup>26</sup> The merger rules do not have to be harmonised, except in the sense that all mergers above a certain threshold are handled by the EU Commission.

<sup>27</sup> An interesting merger case, which concerned payments systems, is the Swedish case *Svenska Giro*, case No. 159/2001, the Swedish Competition Authority.

<sup>28</sup> See, e.g., case No. 1128/97, *FöreningsSparbanken AB (publ.) et al.* (1999-10-29), concerning the Bankgiro and case No. 12/1999, *ABN Amro Bank N.V. et al.* (1999-05-19), concerning "Dataclearingen", a jointly-owned clearing institution.

<sup>29</sup> See cases No. 1124/93, 386/96, 861/97 and 851/2002, which granted UC individual exemptions.

<sup>30</sup> Swedish Competition Authority, case No. 605/1998.

<sup>31</sup> Case No. 1999:12, A 16/98, 1999-05-04.

Thirdly, commercially motivated discounting to a firm's largest customers may be acceptable, even in the absence of a cost justification.

**Competition law requires jointly-owned infrastructural enterprises to grant small rivals access at non-discriminatory conditions.**

After the Competition Authority lost the *Cekab* case in court, competition law has not been used very actively in order to facilitate small and entrant banks' access to payment systems in Sweden. However, it appears safe to conclude that in principle, competition law requires jointly-owned infrastructural enterprises to grant small rivals access at non-discriminatory conditions.<sup>32</sup> Similarly, it is likely that if the competition law had not existed the small banks would have been offered worse access terms.

#### EU CASE LAW

**The prohibition against anti-competitive agreements has been used to challenge the rules of Visa and Mastercard.**

On the European level, the prohibition against anti-competitive agreements has been used to challenge the rules of the international bank cooperatives Visa and Mastercard. In particular, Visa's so-called no-discrimination and honour-all-card clauses have been challenged, as has the level of Visa's interchange fee as such.

**Visa International has adopted a no-discrimination clause that prevents merchants from surcharging card users.**

Visa International has adopted a rule that require individual member banks to respect the no-discrimination clause in the banks' agreements with individual merchants. The no-discrimination clause, in turn, ensures that customers paying with cards are not surcharged, relative to customers paying cash. The European version of the clause prohibits any price differentials between cash and card payments. The US version of the clause only prohibits *surcharges* to card-paying customers; rebates to customers paying cash are accepted.<sup>33</sup>

**In a decision by the EU Commission in 2001, the no-discrimination clause was cleared on the European level.**

Prior to the EU Commission's decision concerning the no-discrimination clause, the Swedish and the Dutch national competition authorities had prohibited its application for domestic transactions in their respective jurisdictions.<sup>34</sup> However, in a decision taken in 2001 by the EU Commission,<sup>35</sup> the no-discrimination clause was given negative clearance on the European level. As a result, the Swedish and Dutch competition authorities will probably have to reverse their previous decisions.

One of the Commission's arguments for accepting the clause was the observation that relatively few merchants (5–10 per cent) used the option of surcharging in Sweden and the Netherlands; another argument was the direct benefit cardholders get from knowing they will not be surcharged when paying with the card. The main argument for not accepting the

<sup>32</sup> See Wetter et al. (2002) p. 226.

<sup>33</sup> Rochet & Tirole (2002). Hence, the US rule is known as the "no-surcharge rule".

<sup>34</sup> Visa International only requires that the individual banks enforce the no-discrimination clause if the competent national authorities do not prohibit its application.

<sup>35</sup> Commission Decision of 9 August 2001, L 293/24, OJ 10.11.2001 (2001/782/EC).

clause was that it reduces the merchants' set of options, which, in turn, may weaken their bargaining position vis-à-vis the banks. With the clause in force, a merchant must either accept the card and not surcharge, or not accept the card at all. Without the clause, the merchants would have a third option: to accept the card, but surcharge the customer. This could exert pressure on the interchange fee.

Visa has also adopted the honour-all-cards clause, which obliges a merchant that accepts, e.g., a Visa direct-debit card to also accept Visa deferred-debit cards and Visa credit cards. This is so, despite of the fact that the merchant's fee can vary between cards of different types.<sup>36</sup> The Commission's argument for accepting this clause was that in its absence, customers could not be certain that a merchant would accept their card, even though the merchant purported to accept Visa cards. That, in turn, would endanger the universal acceptance of the system as a whole. In the US, merchants instigated a class action that focused on the honour-all-cards rule.<sup>37</sup> Note also that the honour-all-cards issue is related to the no-discrimination issue: if merchants were willing to accept direct-debit cards, but not credit cards, they could levy a large (or prohibitive) surcharge on customers that use the latter. Furthermore, it appears that much or all of the merchants' opposition would disappear if they could simply pass on the merchant fee to the final customers. They would then have no reason to opt for prohibitively high surcharges.<sup>38</sup>

In a follow-up decision in 2002, the EU Commission gave a five-year individual exemption to the multilateral interchange fee (the MIF).<sup>39</sup> EuroCommerce, a retail, wholesale and international trade organisation, had complained that the MIF in fact amounted to horizontal price fixing, i.e., a cartel between the member banks. The fee was set by Visa EU Region and was applicable as the default interchange fee for cross-border transactions and, when no domestic default fee had been set, as the default interchange fee also for national transactions. Despite the MIF, every pair of banks was free to set another fee bilaterally. Since the introduction of the MIF in 1974, it had gradually increased.

The EU Commission found that while the MIF restricted competition

**In a follow-up decision in 2002, the EU Commission gave a five-year individual exemption to the multilateral interchange fee.**

<sup>36</sup> The 2001 Decision, at 68.

<sup>37</sup> In the US, the merchants complained that the interchange fee was *the same* for off-line debit cards and for credit cards. In a settlement, reached 1 May 2003, Visa and Mastercard agreed to pay US \$2 billion and \$1 billion, respectively, to the merchants. Furthermore, they promised to lower their fees substantially and to eliminate the honour-all-cards clause. See Balto (2000, 2003).

<sup>38</sup> In addition, the 2001 Decision dealt with rules that restricted cross-border issuing and acquiring, with territorial licensing and with the no-acquiring-without-issuing rule. The Commission imposed a partial liberalisation of cross-border activities, but it accepted the no-acquiring-without-issuing rule.

<sup>39</sup> Commission Decision of 24 July 2002, L 318/17, OJ 22.11.2002 (2002/914/EC). This means that the multilateral interchange fee was permitted for five years. After that, the Visa would in theory have to re-apply for exemption. However, since the system of individual exemptions is being abolished, Visa will in practice not have to apply again. On the other hand, after the five-year-period, the Commission can itself initiate an investigation into the system and, at least in theory, prohibit the system or require modifications.

**The EU Commission found that the MIF restricted competition but that it potentially could be beneficial for consumers.**

it also contributed to the development of the Visa system and, therefore, potentially could be beneficial for consumers. The Commission recognised the network aspects of the Visa system: cardholders benefit from a high number of merchants accepting the card, while merchants benefit from a high number of cardholders. However, both cardholders and merchants prefer that the other party bear the cost. Cardholders prefer a high MIF, while merchants want a low MIF. The Commission identified maximum efficiency of the system with maximum size of the network, and argued that this would be achieved if each category of user paid a cost equal to that category's average marginal utility.<sup>40</sup>

**The Commission argued that some objective criteria must be used to prevent the MIF from being set at a super-optimal level.**

However, as marginal utilities are difficult to measure, an "objective benchmark" (see below) for the system's cost would be an acceptable proxy for the marginal utility. Such a principle would ensure that each category got a "fair share" of the benefits provided by the system. From the Commission's analysis, it is clear that it perceived a risk of the MIF being set too high. However, it recognised that in order to reap the full benefits of the Visa system a default MIF would have to be established at *some* level. In the absence of a MIF, there would be two possible outcomes. The first and more likely would be that the issuer had to charge the cardholder all of its costs. This would imply a drastic rebalancing of the fee structure, since currently the MIF constitutes approximately 80 per cent of acquirers' costs. If the acquirers were to stop contributing to the issuers, cardholders' fees would increase significantly. This increase would result in sub-optimal usage of the system, since, from a welfare perspective, cardholder fees would be too high and merchant fees would be too low. The second possible outcome would be bilaterally determined interchange fees. However, this would only be feasible in small national systems; Visa EU Region has 5,000 members. Hence, the Commission found that a higher-than-zero MIF was conducive to maximum efficiency. Although the Commission did not try to pin down the exact optimal level, it argued that some objective criteria must be used to prevent the MIF from being set at a super-optimal level.

In order to obtain exemption from the competition rules, Visa agreed to lower the volume-weighted MIF for debit-card transactions by more than 50 per cent, to EUR 0.28, and to keep the MIF at that level for at least five years. In addition, Visa agreed to lower the average MIF applicable to credit and deferred-debit card transactions from approximately 0.85 per cent in 2002 to 0.7 per cent in 2007. Finally, Visa undertook to set the MIF at a level that corresponded to the sum of three cost components, the "objective benchmark". The three components were the cost

<sup>40</sup> This is a sufficient, but not a necessary condition. See Rochet & Tirole (2002).

of processing transactions, the cost of the free funding period for cardholders and the cost of providing merchants with a “payment guarantee”. In order to verify that the MIF reflected these costs, Visa undertook to have accountancy firms make cost studies at regular intervals, and to present these to the Commission.

## CONCLUSIONS

The above examples show that competition law can be quite effective in promoting competition in payment systems, suggesting that there is less need for sector-specific regulations than there would be in the absence of such legislation. The strength of competition law is that, in principle, it can be applied to all types of anti-competitive behaviour. Hence, there is no need to foresee all possible means through which one or a few large firms can curb competition.

Compared to sector-specific (pro-competitive) regulation it has, however, certain disadvantages. In general, it imposes less strict behavioural limitations than can be achieved with sector-specific regulations. It will prevent dominant firms (or combinations of firms) from discriminating excessively against smaller rivals, but it may allow a certain degree of price differentiation. It can impose access to natural-monopoly infrastructure (“essential facilities”), but the access price will typically be above average costs.<sup>41</sup>

The third of the three proposed reasons for regulating payment infrastructure less than telecom infrastructure was that the general competition rules are sufficient to guarantee effective competition in the markets for payment services. While it is certainly true that the existence of competition law has played a part in promoting competition in payment services, it is also true that the competition laws apply to other network industries as well. However, the payment industry appears to differ systematically from many other network industries in one important respect, namely its greater dependence on infrastructural clubs. Since such clubs often rely on multilateral agreements, this means that the prohibition against anti-competitive agreements is applicable. In contrast, only the prohibition against abuse of dominance applies directly to single owners of infrastructure. Since the former prohibition is stricter than the latter, the competition authorities can impose more stringent pro-competitive conditions on infrastructural clubs than on infrastructure under single ownership.

**The strength of competition law is that, in principle, it can be applied to all types of anti-competitive behaviour.**

**Compared to sector-specific regulation it has, however, certain disadvantages.**

<sup>41</sup> As an illustrative example, the Swedish Competition Authority was able to reduce the incumbent telecom operator's (Telia's) interconnection fee from SEK 0.35 to SEK 0.215 for single-segment (local) access. Later on, regulatory decisions based on the telecom legislation reduced the fee to SEK 0.069.

## Discussion and conclusions

**In industries where access to infrastructure is essential, a useful distinction is the one between competition for the market and in the market.**

**In the payment-system industry, the regulatory initiatives have primarily been based on general competition rules.**

In industries where access to infrastructure is essential, a useful distinction is the one between competition *for* the market and competition *in* the market. Competition for the market means that firms compete by establishing rival infrastructures; competition in the market means that they compete within one or a few infrastructural systems, each of which is used by several competing firms. Clearly, maximising competition *for* the market by forcing individual banks to set up proprietary non-compatible payment systems is not likely to be conducive to efficiency. However, if privately owned and operated payment systems are left unregulated, maximal efficiency from competition *within* the market will not automatically follow.

The general competition rules constitute one important safeguard against anti-competitive behaviour. However, in some industries these rules have been deemed insufficient for the purpose of protecting the interest of consumers. In such industries, access regulation complements the competition rules. The telecom industry is perhaps the industry that has been subject to the strictest set of regulations. In this industry, the main regulatory tool has been the telecom analogy to the interchange fee – the interconnection fee. In the payment-system industry, the regulatory initiatives have primarily been based on general competition rules.<sup>42</sup>

As discussed above, a possible explanation for this divergence is that the two industries have different histories. The telecom industry has, in most countries, been a state-controlled monopoly, while banking has been a competitive or oligopolistic industry. Furthermore, while a national telecom network is often owned by a single firm, payment systems are typically infrastructural clubs. From a regulatory point-of-view, infrastructural clubs have two main advantages. Firstly, they are to some extent self-regulating, since none of the owners want to be discriminated against or want costs to be too high. Secondly, the general competition rules are more effective against infrastructural clubs than against single owners of bottleneck infrastructure, since in addition to the prohibition against abuse of dominance, the prohibition against anti-competitive agreements can be applied. On the other hand, the mobile-telephony industry has

<sup>42</sup> An interesting exception is the EU Commission's recent regulation of the *consumer* price of international payments. According to Regulation 2560/2001, the cost of international euro-denominated payment transactions within the Union should be the same as that of intranational payments. Perhaps even more interesting is the Australian legislation concerning payment systems. On its homepage, the Australian central bank writes that "[t]he Payments System Board (PSB) of the Reserve Bank oversees the payments system in Australia. The PSB is responsible for promoting the safety and efficiency of the payments system in Australia. Through the Payment Systems (Regulation) Act 1998 (PSRA), and the Payment Systems and Netting Act 1998 the Reserve Bank has one of the clearest and strongest mandates in the world to oversee the operation of the payments system."



evolved into an oligopolistic industry, where networks sometimes are jointly owned by several firms.

A possible regulatory strategy is to use competition law as a “first-line treatment”. Sector-specific regulation will then be considered only if competition law cannot resolve the problem. Probably more incidentally than by design, this strategy has been used in some of Sweden’s domestic deregulations.<sup>43</sup> However, at least when privatisation of a bottleneck is considered or when massive investments in new technologies are expected, it is likely to be less costly to implement a sector-specific regulation *before* the privatisation or the investments, respectively.

Theoretical analyses of the mobile telecom and the payment card industries suggest that there is a tendency for a multilateral interchange fee to be set too high. On the other hand, this tendency may not be very strong and it may not be relevant in all circumstances. From a more practical perspective, an important aspect appears to be the need for new entrants to be able to use existing payment systems on reasonable terms.

Despite the differences between telecom and payment services, the perceived success of the telecom regulation suggests that there may be a role for access regulation also in the payment-system industry. At this point, it would be premature to say that such a regulation *should* be introduced, but certainly there are good reasons for the authorities concerned to follow developments carefully and regularly evaluate whether the market is able to handle the bottleneck-access problem successfully. A possible regulatory initiative would be to introduce a “weak” access regulation, similar to the early Swedish telecom regulation. The 1993 Telecom Act required access prices to be cost based. The regulatory authority was given the role of mediator between conflicting parties, but in the end, it had no power to determine access prices. Such a solution would be in line with the tradition of moral suasion in the relations between banks, central banks and bank regulators. It would also be in line with one of the conclusions drawn in the Core Principles report referred to above: that although statute-based systems and non-statutory approaches can both be appropriate solutions, “[t]he potential benefits of a statute-based approach to oversight [...] deserve serious consideration in countries newly establishing or significantly revising the oversight role and related policies”.<sup>44</sup>

Finally, however, it seems appropriate to re-iterate Laffont & Tirole’s (2000) warning against drawing analogies between different network markets:

**A possible regulatory strategy is to use competition law as a “first-line treatment”.**

**The perceived success of the telecom regulation suggests that there may be a role for access regulation also in the payment-system industry.**

<sup>43</sup> Bergman (2002).

<sup>44</sup> BIS (2001) at 4.3.1.

**Before imposing regulations, a careful analysis of what incentives such a regulation would introduce is strongly recommended.**

'One should be careful, though, before importing lessons drawn from one industry into another, since networks can differ substantially.'<sup>45</sup>

Before imposing regulations, a careful analysis of what incentives such a regulation would introduce is strongly recommended. It is also worth recollecting the trade-off between, on the one hand, short-run competition and, on the other hand, the incentives for investments and long-run competition. Excessively strict regulations would clearly restrict the incentives to invest and to engage in facilities-based competition.

---

<sup>45</sup> Page 181. It appears that the same warning can be applied to a specific theoretical model of one network industry.

## ■ References

- Andersson, M., Guibourg, G. & Segendorff, B., (2001), The Riksbank's Oversight of the Financial Infrastructure, *Sveriges Riksbank Economic Review*, 3, Stockholm.
- Balto, D. A., (2000), Creating a Payment System Network: The Tie that Binds or an Honorable Peace?, *The Business Lawyer*, 55, 1391–1408.
- Balto, D. A., (2003), Life After the Wal-Mart Case, *Credit Card Management*, 16, 48–49.
- Bauer, P. & Ferrier, G., (1996), Scale Economies, Cost Efficiency, and Technological Change in Federal Reserve Payment Processing, *Journal of Money, Credit and Banking*, 28, 1004–39.
- Bergman, M., (2002), Competition Law, Competition Policy, and Deregulation, *Swedish Economic Policy Review*, 9, 95–128.
- Bergman, M., (2002), *Lärobok för regelnissar – en ESO-rapport om regelhantering vid avreglering*, Ds 2002:21, Stockholm. (A Textbook on Deregulation. Experiences from Six Network Industries.)
- Besen, S. M. & Farrell, J., (1994), Choosing How to Compete: Strategies and Tactics in Standardisation, *The Journal of Economic Perspectives*, 8, 117–131.
- BIS, Bank for International Settlement, (2001), *Core Principles for Systematically Important Payment Systems*, CPSS Publication No. 43, available at <http://www.bis.org/publ/cpss43.pdf>.
- Felgran, S. D., (1985), From ATM to POS Networks: Branching, Access, and Pricing, *New England Economic Review*, 44–61.
- Gonec, R. & Nicoletti, G., (2000), *Regulation, Market Structure and Performance in Air Passenger Transportation*, OECD, Economics Department Working Paper No. 254.
- Guibourg, G., (2001), *Interoperability and Network Externalities in Electronic Payments*, Sveriges Riksbank Working Paper Series, No. 126, Stockholm.
- Guibourg, G., (1998), Efficiency in the payment system – a network perspective, *Quarterly Review*, 3, Sveriges Riksbank, Stockholm.
- David, H., Willesson, M., Lindblom, T. & Bergendahl, G., (2003), What Does it Cost to Make a Payment?, *Review of Network Economics*, 2, 1–16.

- Laffont, J., Rey, P. & Laffont, J.-J., (1998a), Network Competition: I. Overview and Nondiscriminatory Pricing, *Rand Journal of Economics*, 29, 1–37.
- Laffont, J., Rey, P. & Laffont, J.-J., (1998b), Network Competition: II. Price Discrimination, *Rand Journal of Economics*, 29, 38–56.
- Laffont, J. & Laffont, J.-J., (2000), *Competition in Telecommunications*, MIT Press, Cambridge, MA.
- Lindquist, K.-G., (2003), *The Effect of New Technology in Payment Services on Banks' Intermediation*, working paper, Norges Bank, Oslo.
- Priest, G. L., (1977), Cartels and Patent License Agreements, *Journal of Law and Economics*, 20, 309–377.
- Rochet, J.-C. & Tirole, J., (2002), Cooperation Among Competitors: Some Economics of Payment Card Associations, *Rand Journal of Economics*, 33, 549–570.
- Rochet, J.-C. & Tirole, J., (2003), Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, 1, 990–1029.
- Scherer, F. M. & Ross, D., (1990), *Industrial Market Structure and Economic Performance*, Houghton Mifflin, Boston, MA.
- Schmalensee, R., (2002), Payment Systems and Interchange Fees, *Journal of Industrial Economics*, 50, 103–122.
- The Swedish Competition Authority, (1999), *Konkurrens på bankmarknaden – betalningsförmedling och villkor för nya aktörer* (Competition in the banking market – the mediation of payments and conditions for new players), Report 3, Stockholm.
- Tirole, J., (1988), *The Theory of Industrial Organization*, MIT Press, Cambridge, MA.
- Wetter, C., Karlsson, J., Rislund, O. & Östman, M., (2002), *Konkurrenslagen – en handbok* (The Competition Law – a Manual), Thomson Fakta, Stockholm.
- Wheelock, D. C. & Wilson, P. W., (2001), New Evidence on Returns to Scale and Product Mix among U.S. Commercial Banks, *Journal of Monetary Economics*, 47, 653–674.
- Winston, C., (1998), U.S. Industry Adjustment to Economic Deregulation, *Journal of Economic Perspectives*, 12, 89–110.